OXFORD

CANCER SYSTEMS BIOLOGY

Translational Mathematical Oncology

Edited by
Ravi **Salgia**
Mohit Kumar **Jolly**
Prakash **Kulkarni**
Govindan **Rangarajan**

Cancer Systems Biology

Cancer Systems Biology

Translational Mathematical Oncology

EDITED BY

Ravi Salgia

Mohit Kumar Jolly

Prakash Kulkarni

Govindan Rangarajan

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2025934642

ISBN 978–0–19–286763–6

DOI: 10.1093/med/9780192867636.001.0001

Printed in the UK by Bell & Bain Ltd., Glasgow

The manufacturer's authorised representative in the EU for product safety is
Oxford University Press España S.A., Parque Empresarial San Fernando de Henares,
Avenida de Castilla, 2 – 28830 Madrid (www.oup.es/en).

To our families for all their love and support.

Preface

Civilization over centuries has seen considerable advances in healthcare. Cancer is a challenging issue, and a number of discoveries have led to better care of some patients. Despite all the progress and the promise regarding early detection and precision medicine, we are still faced with the nettlesome problem—cancer is a moving target. Even within an individual tumour, deep sequencing analyses now indicate multiple, phenotypically distinct subpopulations whose representation seems to vary dramatically from one stage to the next as the tumour progresses.

The discovery of driver mutations and 'actionable' genes added to the genetic underpinning and fuelled the era of -omic approaches starting from whole tissue to the single-cell level. The reductionist perspective, together with these new developments, initially established cancer as a complex and heterogeneous disease driven by genetic mutations. According to this view, cancer per se, its progression, and the acquisition of drug resistance, whether innate or acquired, are highly deterministic. However, recent research has promoted the idea that phenotypic plasticity and non-genetic mechanisms are important in these events.

Increasingly, it is acknowledged that cancer is a complex adaptive and dynamic system and that, in addition to genetic mutations, intratumoral heterogeneity can also arise through non-genetic mechanisms. Furthermore, it is also becoming evident that the malignant phenotype results from complex interactions between genetic and non-genetic mechanisms. These interactions contribute to functional changes across multiple spatiotemporal scales, from the molecular to cellular and tissue level, creating a heterogeneous cancer cell population.

Technological advances to study single cells and single molecules, the availability of powerful computing platforms, and the application of evolutionary game theory to discern group behaviour have galvanized multiscale modelling enabling a quantitative view of the key processes in tumorigenesis. These efforts have served as a powerful means to investigate biological phenomena more comprehensively in experimentally relevant ways. Therefore, a deeper understanding of the disease can not only provide novel insight needed to personalize treatments, but this knowledge may also be used to address therapeutic resistance by administering 'adaptive' (intermittent) therapy rather than the standard (continuous) therapy that typically, albeit inadvertently, encourages the emergence of drug-resistant tumours. However, performing pre-clinical studies and clinical trials for multiple specific targets in varied dosing sequence and timing schedules is often too resource- and time-consuming, and hence challenging. Therefore, calibrated and validated mathematical models offer an attractive approach to evaluate untested protocols *in silico* to narrow the set of promising treatment schemas to be evaluated, to identify new treatment targets, and to reduce the risk of adverse clinical outcomes due to complex feedback mechanisms.

The main intent of this book, in addition to providing state-of-the-art reviews and thought-provoking ideas in a concise and succinct manner, is to encourage cross-pollination of ideas from multiple disciplines between clinicians and scientists interested in integrating both theoretical and experimental approaches to study cancer. The chapters provide new ideas and concepts outlining how a quantitative picture of cancer can provide a deeper understanding of the disease and how a systems-level perspective may hold the key to fully comprehend how cancer arises and progresses.

The book embodies 41 chapters that are organized in 9 sections. The first two introductory chapters provide an overview of the systems approach using multiple paradigms. The following section on single-cell -omics provides in-depth perspective on the analyses of big data derived from single cells at the systems level. The cardinal features of such single-cell multi-omic analysis include technologies for single-cell isolation, barcoding, and sequencing to measure multiple types of molecules from individual cells as well as the integrative analysis of molecules to characterize cell types and their functions regarding pathophysiological processes based on molecular signatures. Furthermore, new hypotheses about functionality and consequences of expression heterogeneity from time-resolved measurements of gene expression and how the principles from information theory can guide models of transcriptional regulation and gene network connectivity are presented.

Next, computational approaches to drug discovery are presented. Computational approaches form an important part of the tools employed in drug discovery and development. Their applications span almost all stages in the discovery and development pipeline, from target identification to lead discovery and from lead optimization to preclinical and even clinical trials. Thus, the chapters devoted to this section cover concepts of structure- and ligand-based drug designing, protein modelling and visualization, molecular docking, virtual screening, molecular dynamics simulation, pharmacophore modelling, and quantitative structure–activity relationship approaches that are typically used in conjunction with conventional biophysical techniques. Some also address the broad area of data analysis, including data mining algorithms, statistical approaches, and practical applications. Topics in this section include problems involving massive and complex datasets, solutions utilizing innovative data mining algorithms and/or novel statistical approaches, and the objective evaluation of analyses and solutions.

The volume and complexity of scientific and clinical data in oncology have shown a remarkable growth in the past decade that include electronic health records, radiographic and histologic data, and patient genomic/genetic information. New data-processing technologies have the potential to derive clinically meaningful insights from large-volume data revolutionizing medicine. Among those techniques is supervised machine learning, the study of computer algorithms that use self-improving models that learn from labelled data to solve problems. The chapters included here describe a framework to aid clinicians in understanding and critically evaluating studies applying supervised machine learning methods and techniques to the diagnosis, prognostication, and treatment of cancer.

Artificial intelligence (AI) refers to the ability of a machine to perform tasks commonly associated with intelligent human behaviour. AI includes disciplines from both computer science and mathematics and is a group of iterative, 'self-learning' techniques, which discover relationships within data that can evolve and often be performed faster over time. Deep learning, a subfield of AI that is highly flexible and supports automatic feature extraction, is increasingly being applied in various areas of both basic and clinical cancer research. This book includes chapters that describe numerous recent examples of the application of AI in oncology, including cases in which deep learning has efficiently solved problems that were previously thought to be unsolvable, and address obstacles that must be overcome before such application can become more widespread.

Today, medicine in general, and oncology in particular, has become a digital data-intensive endeavour, relying on secure and scalable computing, storage, and network infrastructure, which has traditionally been purchased, supported, and maintained locally. However, cloud computing has emerged as an alternative to locally maintained traditional computing approaches. It offers users pay-as-you-go access to services such as hardware infrastructure, platforms, and software for solving common biomedical computational problems. Cloud-computing services offer secure on-demand storage and analysis and are differentiated from traditional high-performance computing by their rapid availability and scalability of services. As such, cloud services are engineered to address big data problems and enhance the likelihood of data and analytics sharing, reproducibility, and reuse. The chapters included here provide a perspective on cloud computing and its evolving utility and demand in oncology.

Chapters on biomechanics are included in the next section. This is an up-and-coming field, especially in cancer. Biomechanics, a branch of biophysics, is the study of the structure, function, and motion of the mechanical aspects of biological systems. Furthermore, studies devoted to biomechanics concern multiple levels from whole organisms to organs, cells, and cell organelles, using the methods of mechanics. More recently, biomechanics also includes computational mechanics and goes beyond pure mechanics and involves other physical aspects such as heat and mass transfer, electric and magnetic stimuli, and many others. Together the chapters in this section provide an overview and outline how biomechanics can provide new insights into cancer biology.

The next section is devoted to translational mathematical oncology, i.e. the use of mathematics, modelling, and simulation to study cancer, and has a broad scope—ranging from theoretical studies to clinical trials designed with mathematical models. It has provided a framework in which these theoretical constructs are applied and biologic data are analysed to make predictions about cancer progression. Mathematical models can be directly applied to real scenarios and readily tested using large amounts of biologic and clinical data to analyse and understand big data. Together with evolutionary game theory (EGT), translational mathematical oncology has aided personalization of medicine through modelling and simulation. The chapters embodying this sub-section discuss how this is achieved using patient-specific clinical data to develop individualized screening strategies to detect cancer earlier; make predictions of response to therapy; design adaptive, patient-specific treatment plans to overcome therapy resistance; and establish domain-specific standards to share model predictions and to make models and simulations reproducible.

Cancer cells and stromal cells interact within a tumour to give both cooperative and competitive behaviours that have been attributed to various molecular signalling pathways. The application of EGT which studies the strategic interactions of biological agents based on frequency-dependent fitness functions can help to understand cancer–stroma interactions as well as to interpret counter-intuitive cooperative behaviours among cells in the tumour microenvironment. Further, EGT has been clinically employed in the form of adaptive therapy to take advantage of competition between various cancer clones. The power of EGT lies in its ability to understand current population-level behaviour through experimental fitting and simultaneously predicting future dynamics. Thus, the chapters included in this section provide in-depth reviews on various aspects of EGT and its application to cancer, especially in tackling drug resistance.

The last section is devoted to chaos theory and fractals in cancer. Here, tumour growth is considered as a dynamical system and hence is chaotic. The chapters provide an overview of the potential contribution of chaos theory and fractal mathematics to the study of cancer. Fractals are mathematical constructs that show self-similarity over a range of scales and non-integer (fractal) dimensions. Owing to these properties, fractal geometry can be used to efficiently estimate the geometrical complexity and the irregularity of shapes and patterns observed in lung tumour growth (over space or time), whereas the use of traditional Euclidean geometry in such calculations is more challenging. The application of fractal analysis in biomedical imaging and time series has shown considerable promise for measuring processes as varied as heart and respiratory rates, neuronal cell characterization, and vascular development. The chapters discuss how fractal analyses are used to quantify changes in nuclear and chromatin FD in primary and metastatic tumour cells, and clinical imaging studies that correlated changes in the FD of tumours on CT and/or PET images with tumour growth and treatment responses are reviewed. Moreover, the potential use of these techniques in the diagnosis and therapeutic management of lung cancer is discussed.

In summary, *Cancer Systems Biology and Translational Mathematical Oncology* is a comprehensive, up-to-date treatise contributed by multiple experts from across the globe, emphasizing quantitative biology integrating theory and experiments, to provide a systems perspective. We hope this compendium will encourage cross-pollination of ideas from experts in multiple disciplines, including basic and translational science researchers, systems biologists, physicists, and clinicians interested in decoding a systems-level

emergent view of cancer so that more effective treatment strategies may be designed. We trust the book will be a valuable companion to academic researchers working in the field of cancer biology, medical oncology, mathematical oncology, cancer systems biology, cancer evolutionary biology, and related topics. This book will also interest clinicians/scientists at pharmaceutical/biotechnology companies interested in a deeper understanding of phenotypic plasticity in cancer, its progression, and in the emergence of drug resistance. Finally, this book will serve as a reference for graduate students, postdoctoral scientists, and academic/practicing clinicians/scientists with an interest in cancer biology from an interdisciplinary perspective.

Acknowledgements

The idea to bring out this book was conceived following numerous discussions between us and our colleagues worldwide both formally at meetings and informally in various venues. Furthermore, the recent developments in analytical methods and the availability of large datasets from patient-derived samples, especially at the single-cell level as well as data obtained using live cell imaging, together with advances in our thinking about the mechanism underlying cancer and the progress we have seen on the theoretical and computational front, served as a strong motivation.

We would like to thank the members of our respective laboratories and clinics for their contributions that helped lay the foundation for the genesis of this book. In particular, we thank our patients and their families for putting their trust in us to advance cancer research and cancer care. It would be impossible to mention the long list of individuals to whom we are grateful, and we do sincerely thank them for their valuable input.

We are especially indebted to all the authors who so graciously contributed their thought-provoking chapters. We also thank the anonymous reviewers for their constructive criticism and thoughtful suggestions.

We would like to acknowledge Ms Caroline Smith, Acquisitions Editor, Medicine, Academic Division, Oxford University Press, for soliciting our book proposal and guiding us through the publication process which at times felt rather overwhelming, and Mr Adam Brevik, Project Editor, Oxford University Press, for his outstanding support and meticulous attention to detail. It was a real pleasure to work with them. Caroline and Adam were not only friendly, professional, and extremely helpful but also always went the extra mile to ensure we stayed the course.

And, last but not least, we express our gratitude to our families for their love, encouragement, and their incredible support.

Editors

Contents

SECTION 4
Computational approaches to drug development

SECTION 5
Statistical methods and data mining, machine learning, artificial intelligence, and cloud computing

SECTION 6
Biomechanics

SECTION 7
Translational mathematical oncology

SECTION 8
Ecology, evolution, and cancer

SECTION 9
Critical transitions and chaos in cancer

Abbreviations

ABM	Agent-based models
ACD	Asymmetric cell division
ACGME	Accreditation Council for Graduate Medical Education
AFM	Atomic force microscopy
AHIC	AMIA Health Informatics Certification [In Ref list]
AI	Artificial intelligence
AJCC	American Joint Committee on Cancer
ALCHEMIST	Adjuvant Lung Cancer Enrichment Marker Identification and Sequencing Trials
ALL	Acute lymphoblastic leukaemia
AMIA	American Medical Informatics Association
AML	Acute myeloid leukaemia
ANN	Artificial neural network
ANOVA	Analysis of variance
ANSIA	Accelerated Nanopatterned Stromal Invasion Assay
APC	Adenomatous polyposis coli
API	Application Programming Interface
AR	Androgen receptor
ATRA	All trans-retinoic acid
AUC	Area under the curve
AWSEM	Associative memory, Water-mediated, Structure and Energy Model
BERT	Bidirectional Encoder Representations from Transformer
BloodPAC	Blood Profiling Atlas in cancer [In Ref list]
BM	Basement membrane
BMJ	British Medical Journal
BSDE	Backward stochastic differential equations
CA	Cellular automata
CAF	Cancer-associated fibroblast
CAR	Chimeric antigen receptor
CART	Classification and regression tree
CDC	Centers for Disease Control and Prevention
CDS	Cancer Data Service
CHASM	Cancer-specific High-throughput Annotation of Somatic Mutations
CI	Composite index
CLSM	Confocal laser scanning microscopy
CNN	Convolutional neural network
CNV	Copy number variants
COD	Coefficient of determination
COG	Children's Oncology Group
CPM	Cellular Potts modelling
CPT	Current Procedural Terminology
CRC	Colorectal cancer
CRDC	Cancer Research Data Commons
CRS	Cytokine release syndrome
CRV	Carcinogenesis relevance value
CSC	Cancer stem cell
CSP	Chemical shift perturbations
CTC	Circulating tumour cells
CV	Collective variables
CWL	Common workflow language
CYM	Complex Young's modulus
DBR	Disordered binding regions
DCC	Data Coordinating Centers
DDD	Dynamic distribution decomposition
DDN	Differential dependency network
DEEPEST	Data-Enriched Efficient PrEcise STatistical
DGEA	Differential gene expression analysis
DIC	Differential interference contrast
DL	Deep learning
DLN	Deep-learning network
DNB	Dynamic network biomarkers
DPT	Diffusion pseudo-time
DR	Dimensionality reduction
DRS	Data Repository Service
DTEP	Drug-tolerant expanded persister
DTP	Drug-tolerant persister
DTW	Dynamics time warping
ECM	Extra-cellular matrix
EDAC	Epithelial defence against cancer
EGFR	Epidermal growth factor receptor
EGT	Evolutionary game theory
EHR	Electronic health record
ELI	Evolved Levels of Invasibility
ELT	Energy landscape theory
ELViM	Energy landscape visualization method
EM	Elastic modulus
EMT	Epithelial-to-mesenchymal transition
EpCAM	Epithelial cell adhesion molecule
ES	Essential space
EV	Extra-cellular vesicles
FAIR	Findable, Accessible, Interoperable, and Reusable

FAP	Fibroblast-activated protein
FDA	Federal Drug Administration
FES	Free energy landscapes/surfaces
FGF	Fibroblast growth factor
FMD	Fast mimicking diets
GAN	Generative adversarial network
GAP	GTPase-activating proteins
GCN	Graph convolutional network
GDC	Genomic Data Commons
GEF	Guanine exchange factors
GFP	Green fluorescent protein
GNN	Graph neural network
GRN	Gene regulatory network
GSC	Glioma stem cells
GSEA	Gene set enrichment analysis
HA	Hyaluronic acid
HGG	High-grade gliomas
HGNC	HUGO Gene Nomenclature Committee
HNSCC	Head and neck squamous cell carcinoma
HUGO	Human Genome Organization
ICANS	Immune cell-associated neurologic syndrome
ICD	International Classification of Diseases
ICD9	International Classification of Diseases, Ninth Revision
ICD10	International Classification of Diseases, Tenth Revision
ICDO3	International Classification of Diseases for Oncology, Third Edition
ICDC	Integrated Canine Data Commons
ICGC	International Cancer Genome Consortium
ICI	Immune checkpoint inhibitors
IDC	Imaging Data Commons
IDP	Intrinsically disordered proteins
IDR	Intrinsically disordered regions
IMANN	Interactive mathematical modelling-artificial neural network
IMC	Imaging mass cytometry
IT	Information technology
ITH	Intratumoral heterogeneity
LBC	Lung and bronchus cancer
LEUP	Least microenvironmental uncertainty principle
LIF	Leukaemia inhibitory factor
LIME	Local Interpretable Model-agnostic Explanations
LINC	Linker of nucleoskeleton and cytoskeleton
LLM	Large language models
LM	Language models
LNE	Local network entropy
LOINC	Logical Observation Identifiers Names and Codes
LR	Logistic regression
MAE	Mean absolute error
MAP	Minimum action path
MD	Molecular dynamics
MedNLI	Medical Natural Language Inference
MeHA	Methacrylated glycosaminoglycan hyaluronic acid
MERFISH	Multiplexed error-robust Fluorescent *in situ* Hybridization
MeSH	Medical subject headings
MET	Mesenchymal-to-epithelial transition
MHC	Major histocompatibility complex class
MI	Mutual information
ML	Machine learning
MoE	Ministry of Education [In Ack section]
MPS	Mononuclear phagocytic system
MRD	Minimal residual disease
MRI	Magnetic resonance imaging
MS	Mass spectrometry
MSE	Mean squared error
MSM	Markov state model
MTD	Maximum tolerated dose
MVD	Minimum viable disease
NCAM	Neural cell adhesion molecule [in Ref list]
NCI	National Cancer Institute
NCPI	NIH Cloud Platform Interoperability
NET	Neutrophil extra-cellular traps
NIH	National Institute of Health
NK	Natural killer
NLM	National Library of Medicine
NLP	Natural language processing
NMR	Nuclear magnetic resonance
NSCC	Non-stem cancer cells
NSCLC	Non-small-cell lung cancer
ODE	Ordinary differential equations
PC	Principal components
PCA	Principle component analysis
PCC	Pearson correlation coefficient
PDC	Proteomic Data Commons
PDE	Partial differential equations
PDMS	Poly-di-methoxy-silane
PDX	Patient-derived Xenograft
PGCC	Polyploidy giant cancer cells
PIN	Protein interaction networks
PINN	Physics-informed neural network
PMRF	Prime Minister's Research Fellowship
PPC	Percent positive cores
PRECINCT	PRE-medical Cancer Immunotherapy Network Canine Trials
PRO	Patient-reported outcomes
PSA	Prostate-specific antigen
PSMSR	Phenotype Switch Model with Stress Response
QPI	Quantitative phase imaging
RAVE	Reweighted autoencoded variational Bayes for enhanced
RC	Reaction coordinate
RCT	Randomized control trial
RMSD	Root-mean-square deviation
RNN	Recurrent neural network
ROS	Reactive oxygen species
RS	Recurrence score
RSV	Rous sarcoma virus
RTE	Real-time elastography
RWE	Real-World world Evidenceevidence

SAXS	Small-angle X-ray scattering
SBINN	Systems biology informed neural networks
SCC	Squamous cell carcinoma
SCD	Symmetric cell division
SCLC	Small cell lung cancer
SDE	Stochastic differential equations
SEER	Surveillance, Epidemiology, and End Results
SERB	Science and Engineering Research Board
SILCS	Site-Identification by Ligand Competitive Saturation
SLCC	Stem-like cancer cells
SM	Storage modulus
SMT	Somatic mutation theory
SNV	Single-nucleotide variants
SOC	Self-organized criticality
SP	Special Publication
SQL	Structured Query Language
ST	Spatial transcriptomics
SV	Structural variants
SVM	Support vector machine
SWE	Shear wave elastography
TACS	Tumour-associated collagen signatures
TAD	Topologically associating domains
TAM	Tumour-associated macrophage
TAN	Tumour-associated neutrophils
TBR	Tumour-to-background ratio
TCGA	The Cancer Genome Atlas
TCR	T-cell receptors
TF	Transcription factor
TFV	Trajectory feature vectors
TIL	Tumour infiltrating leukocytes
TNBC	Triple-negative breast cancer
TNF	Tumour necrosis factor
TOFT	Tissue organization field theory
TRR	Toll-related receptors
UMAP	Uniform manifold approximation and projection
VAE	Variational auto encoder
VAMP	Variational principle for Markov process
WBMT	Worldwide Network for Blood & and Marrow Transplantation
WDL	Workflow description language
WES	Whole-exome sequencing
WFO	Watson for Oncology
WGS	Whole genome sequencing
WT	Wild type
YAP	Yes-associated protein
YM	Young's modulus
ZIMS	Zoological Information Management System

Contributors

Srisairam Achuthan Vice President, Center for Informatics, Division of Research Informatics, City of Hope National Medical Center

J. Arvid Ågren Assistant Professor, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University

Syeda Subia Ahmed Department of Cell and Developmental Biology, Feinberg School of Medicine, Northwestern University; Center for Synthetic Biology, Northwestern University; Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine

Philipp M. Altrock Heisenberg Professor for Cancer Modeling and Evolution, Department of Internal Medicine II, University Hospital Schleswig-Holstein and Kiel University, Clinical Research Unit CATCH-ALL

Arya Amini Chief of Thoracic Radiotherapy, Associate Professor, Department of Radiation Oncology, City of Hope National Medical Center

Benedict Anchang Stadtman Tenure-Track Principal Investigator, Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences and Center for Cancer Research, National Cancer Institute, Bethesda, MD, United States

Alexander R.A. Anderson Richard O. Jacobson Distinguished Endowed Chair, Integrated Mathematical Oncology Department, H. Lee Moffitt Cancer Center and Research Institute

Asadullah Research Associate, Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay

Rowan Barker-Clarke Genomic Medicine Institute, Cleveland Clinic; Rose Ella Burkhardt Brain Tumor and Neuro-Oncology Center, Cleveland Clinic

Arnab Barua Technische Univesität Dresden, Center for Information Services and High Performance Computing (ZIH), Technical University of Dresden

Rowan Beck Velsera, Charlestown, MA, USA

Subhendu Bhandary PhD Student at Department of Mathematics, Indian Institute of Technology Ropar

Supriyo Bhattacharya Integrative Genomics Core, and Division of Research Informatics, Center for Informatics, City of Hope National Medical Center; Integrative Genomics and Bioinformatics Core, Beckman Research Institute

Nicholas Bodkin MD/PhD Trainee, Department of Cell and Developmental Biology, Northwestern University

Biplab Bose Associate Professor, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati

Jean-Pascal Capp Toulouse Biotechnology Institute, INSA / University of Toulouse, CNRS, INRAE, Toulouse, France

Rishov Chatterjee Center for Informatics, Division of Research Informatics, City of Hope National Medical Center

Guranda Chitadze Department of Internal Medicine II, University Hospital Schleswig-Holstein and Kiel University, Clinical Research Unit CATCH-ALL

Vittorio Cristini Mathematics in Medicine Program, Department of Medicine, Houston Methodist Research Institute

Amrapali Datta Department of Bioengineering, Indian Institute of Science

Brandi N. Davis-Dusenbery Velsera, Charlestown, MA, USA

Smita Deb Research Scholar, Department of Mathematics, Indian Institute of Technology Ropar

Prashant Dogra Mathematics in Medicine Program, Department of Medicine, Houston Methodist Research Institute; Department of Physiology and Biophysics, Weill Cornell Medical College

Christina R. Dollahon Graduate Research Assistant, Department of Biomedical Engineering, Texas A&M University

Wenqiang Du University of Connecticut Health Center, Connecticut, USA

William C. Eward Associate Professor, Department of Orthopaedics, Duke University Medical Center

Cera R. Fisher Velsera, Charlestown, MA, USA

Jasmine Foo Distinguished McKnight University Professor, University Northrop Professor, School of Mathematics, University of Minnesota

Jill Gallaher Integrated Mathematical Oncology Department, H. Lee Moffitt Cancer Center and Research Institute

Jason George Assistant Professor, Department of Biomedical Engineering, Texas A&M University

Jenniffer Ghith Omnichannel Strategy and Innovation Lead, Global Scientific Communications, Pfizer Biopharmaceuticals Group, Oncology

Subhashis Ghosh BRIC National Institute of Biomedical Genomics

Shreya Goel Department of Pharmaceutics and Pharmaceutical Chemistry, University of Utah; Department of Radiology and Imaging Sciences, University of Utah; Department of Biomedical Engineering, University of Utah

Vishaka Gopalan Cancer Data Science Laboratory, National Cancer Institute, National Institutes of Health

Abhishek Goswami Project Associate, Department of Bioengineering, Indian Institute of Science

Sidhartha Goyal Associate Professor, Department of Physics, St. George campus of the University of Toronto

Yogesh Goyal Department of Cell and Developmental Biology, Feinberg School of Medicine, Northwestern University; Center for Synthetic Biology, Northwestern University; Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine

Laurie Graves Medical Instructor of Pediatrics, Hematology & Oncology, Duke University Medical Center

Einar Bjarki Gunnarsson School of Mathematics, University of Minnesota; Division of Applied Mathematics, University of Iceland

Abhay Gupta Graduate Student, Department of Biological Sciences, Indian Institute of Science Education and Research (IISER) Mohali

Haralampos Hatzikirou Associate Professor, Mathematics Department, Khalifa University; Technische Univesität Dresden, Center for Information Services and High Performance Computing

Henry H. Heng Professor, Center for Molecular Medicine and Genetics, Pathology Department, Wayne State University School of Medicine

Sui Huang Institute for Systems Biology

Mohit Kumar Jolly Associate Professor, Department of Bioengineering, Indian Institute of Science

Luciane T. Kagohara Assistant Professor of Oncology, Sidney Kimmel Comprehensive Cancer Center, Department of Oncology Gastroaintestinal Cancer Division and Quantitative Science Division

Loukia G. Karacosta Assistant Professor, Cancer Systems Imaging, the University of Texas MD Anderson Cancer Center

Andrzej Kasperski Professor UZ, Institute of Biological Sciences, Department of Biotechnology, Laboratory of Bioinformatics and Control of Bioprocesses, University of Zielona Góra

Eshan S. King Medical Scientist Training Program, Case Western Reserve University School of Medicine

Mohammad Kohandel Associate Professor, Department of Applied Mathematics, University of Waterloo

Prakash Kulkarni Professor and Director of Translational Research, Department of Medical Oncology, City of Hope National Medical Center

Alka Kumari PhD Research Scholar, Department of Bioengineering, Indian Institute of Science

Kshitiz Department of Biomedical Engineering, University of Connecticut Health, Farmington; Center for Cell Analysis and Modeling, University of Connecticut Health

Colton Ladbury Assistant Professor, Department of Radiation Oncology, City of Hope National Medical Center

Jintong Lang PhD Candidate, Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University

Caterina A.M. La Porta Full Professor, Department Environmental Science and Policy, Center for Complexity and Biosystems, University of Milan and Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico di Milano

Jinzhi Lei Professor, School of Mathematical Science, Tiangong University

Vitor B.P. Leite Federal Institute of Education, Science and Technology of São Paulo (IFSP)

Tanmay P. Lele Department of Biomedical Engineering, Texas A&M University; Department of Chemical Engineering, Texas A&M University; Department of Translational Medical Sciences, Texas A&M University

Chunhe Li Associate Professor, Shanghai Center for Mathematical Sciences, Fudan University

Frederick L. Locke Chair, Department of Blood and Marrow Transplant and Cellular Immunotherapy, H. Lee Moffitt Cancer Center and Research Institute

Jeff Maltas Genomic Medicine Institute, Cleveland Clinic

Andriy Marusyk Associate Member, Department of Tumor Microenvironment and Metastasis, Moffitt Cancer Center

Cameron Meaney PhD Candidate, Department of Applied Mathematics, University of Waterloo

Paromita Mitra Senior Research Fellow, BRIC National Institute of Biomedical Genomics; Regional Centre for Biotechnology

Dorsa Mohammadrezaei PhD Canditate, Department of Applied Mathematics, University of Waterloo

Atish Mohanty Departmental of Medical Oncology and Therapeutics Research, City of Hope National Medical Center

Annice Najafi Assistant Professor, Department of Biomedical Engineering, Texas A&M University

Vera Pancaldi Cancer Research Center of Toulouse (CRCT), University of Toulouse, CNRS, INSERM

Maria Jose Peláez Mathematics in Medicine Program, Department of Medicine, Houston Methodist Research Institute; Neal Cancer Center, Houston Methodist Research Institute; Department of Imaging Physics, University of Texas M.D. Anderson Cancer Center; Physiology, Biophysics, and Systems Biology Program, Graduate School of Medical Sciences, Weill Cornell Medicine

Danielle Pi MD/PhD Trainee, Cell and Developmental Biology Department, Northwestern University

Priyanka Prakash Institute Research Investigator, Institute for Applied Cancer Science (IACS), The University of Texas MD Anderson Cancer Center, Houston, Texas

Govindan Rangarajan Professor of Mathematics and Director, Indian Institute of Science

Vito W. Rebecca Johns Hopkins University Bloomberg School of Public Health; Department of Biochemistry and Molecular Biology; Johns Hopkins School of Medicine; Sidney Kimmel Comprehensive Cancer Center and Department of Oncology

Sumaiyah Rehman Scientific Associate Princess Margaret Cancer Center, University Health Network

Mark Robertson-Tessi Integrated Mathematical Oncology Department, H. Lee Moffitt Cancer Center and Research Institute

Jay G. Ronquillo Center for Biomedical Informatics and Information Technology, National Cancer Institute

Uday Saha BRIC National Institute of Biomedical Genomics; Regional Centre for Biotechnology

Ravi Salgia Professor and Chair, Departmental of Medical Oncology, City of Hope national Medical Center

Murilo N. Sanches PhD Candidate, Physics Department, São Paulo State University

Madhurima Sarkar Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay

Suchitaa Sawhney Department of Biomedical Engineering, Texas A&M University and Texas A&M College of Medicine

Bob J.A. Schijvenaars Department of Urology, University of Tübingen

Jacob G. Scott Professor, Department of Radiation Oncology, Cleveland Clinic; Genomic Medicine Institute, Cleveland Clinic; Case Western Reserve University School of Medicine; Department of Physics, Case Western Reserve University; Case Comprehensive Cancer Center

Shamik Sen Professor, Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay

Manu Setty Assistant Professor, Basic Sciences Division and Translational Data Science IRC, Fred Hutchinson Cancer Research Center

Partha Sharathi Dutta Professor of Mathematics, Indian Institute of Technology Ropar

Divjoy Singh Centre for BioSystems Science and Engineering, Indian Institute of Science; Department of Physics, University of California Santa Barbara

Sandeep Singh Professor, BRIC National Institute of Biomedical Genomics

Jason A. Somarelli Assistant Professor, Department of Medicine, Duke University Medical Center, Durham

Arnulf Stenzl Professor and Chair, Department of Urology, University of Tübingen

Maximilian A.R. Strobl Cleveland Clinic, Translational Hematology and Oncology Research

Ayalur Raghu Subbalakshmi Centre for BioSystems Science and Engineering, Indian Institute of Science

Yasir Suhail Department of Biomedical Engineering, University of Connecticut Health; Center for Cell Analysis and Modeling, University of Connecticut Health

Dagim Shiferaw Tadele Oslo University Hospital, Department of Medical Genetics; Genomic Medicine Institute, Cleveland Clinic

Joseph Tandurella Research Technician, Johns Hopkins University, Sidney Kimmel Comprehensive Cancer Center

Ajay Tijore Department of Bioengineering, Indian Institute of Science

Arne Traulsen Director, Department Theoretical Biology, Max Planck Institute for Evolutionary Biology

Abicumaran Uthamacumaran Researcher, Oxford Immune Algorithmics, McGill University, Dept. of Surgical and Interventional Sciences. Department of Physics & Psychology (Alumni), Concordia University

Srinikhil S. Vemuri Undergraduate Researcher, Department of Biomedical Engineering, Texas A&M University

Rafael G. Viegas Federal Institute of Education, Science and Technology of São Paulo (IFSP); Department of Physics, São Paulo State University (UNESP), Institute of Biosciences, Humanities and Exact Sciences

Medhavi Vishwakarma Department of Bioengineering, Indian Institute of Science

Günter Wagner Department of Ecology and Evolutionary Biology, Yale University

Ting-Ching Wang Department of Chemical Engineering, Texas A&M University

Weikang Wang Institute of Theoretical Physics, Chinese Academy of Sciences

Zhihui Wang Mathematics in Medicine Program, Department of Medicine, Houston Methodist Research Institute; Neal Cancer Center, Houston Methodist Research Institute; Department of Physiology and Biophysics, Weill Cornell Medical College

Jeffrey West Assistant Professor, Integrated Mathematical Oncology Department, H. Lee Moffitt Cancer Center and Research Institute

Zelia F. Worman Velsera, Charlestown, MA, USA

Xiwei Wu Director, Integrative Genomics and Bioinformatics Core. Professor, Department of Computational and Quantitative Medicine, Beckman Research Institute City of Hope

Jianhua Xing Department of Computational and Systems Biology and UPMC-Hillman Cancer Center, University of Pittsburgh

Stefano Zapperi Full Professor, Department of Physics, Center for Complexity and Biosystems, University of Milan

SECTION 1

Cancer systems biology: An overview

Cancer Systems Biology and Translational Mathematical Oncology, Oxford University Press

1

The necessary existence of cancer and its progression from first principles of cell state dynamics

Sui Huang

1.1. Introduction: cancer progression as explainable shift in systems configuration

The traditional view of cancer has long been restricted by the tacit notion of *immutability* of objects that has pervaded natural philosophy since Aristotle. When in 1572 Tycho Brahe discovered a new star (now known as a supernova) in the constellation of Cassiopeia, it came as a shock [1]. The firmament was presumed to be an immutable divine creation. When in the 1830s Charles Lyell proposed that geological formations (mountains and valleys) had been shaped by slow changes of Earth due to known physical forces that follow uniform principles, and not caused by biblical catastrophic events, he met with great resistance [2]. A third challenge to the default assumption of immutability was of course the ideas emerging in the 19th century that biological species evolve, culminating in Charles Darwin's theory of the origin of species by natural selection [3]. Opposition to the 'mutability' of species defended the taxonomy of plants and animal as an eternal order created by God [4].

Today, the discovery of reprogramming of cell types in the adult mammalian body can be considered a modern version of the shock to the unquestioned assumption of an unchangeable order. The awareness rising in the past three decades of multi-potent stem cells in adult tissues introduced the notion that metazoan cell types do not represent permanent 'postmitotic' cells that obey a rigid anatomical system but exhibit 'plasticity' [5,6]. The concept of cell-type plasticity reached its peak with the demonstration that a set of transcription factors (TFs), when ectopically introduced into differentiated adult cells, resets the gene regulatory network (GRN) to produce a pluripotent state, the 'induced pluripotent stem cells' [7]. These discoveries eventually popularized the notion of *cellular plasticity*.

By now, explaining a change of phenotype by a genetic mutation is readily accepted. In organismal evolution, the generation of a new trait by genetic mutation followed by natural selection is integral to biological thinking. But interestingly, the very concept that phenotypic innovation starts with a random mutation that produces the new trait does not actually admit a process of 'phenotypic mutability' of organismal structures. Instead, it relegates any alteration of phenotype to the random, singular event of a mutation, a molecular *deus ex machina*. Such thinking obviates the need of explaining the physics of the actual process of change by some universal mechanism, much as geologists refuted Lyell's proposition to explain alterations of geological structures by uniform, physically plausible processes.

In cancer biology, the explanation of changes in traits of the cell during tumorigenesis or tumour progression has long borrowed from evolutionary biology: in the prevailing somatic evolution theory of cancer [8], the innovation of a malignant cell phenotype *P*, such as invasion, autonomous cell division, stemness, resistance to treatment or immune escape, or the acquisition of any other 'hallmark of cancer' [9], is thought to be the result of genetic mutations followed by selection in the tissue environment of the mutated cells for traits that confer a fitness advantage. A change of phenotype *P* is explained by an alteration of the cell's genotype *G*. This scheme tacitly assumes a bijective $P \rightarrow G$ mapping from genotype *G* (nowadays equivalent to genome) to phenotype *P* [10]. This simple scheme has been cemented by the facility of cancer genome sequencing and the promise and plausibility of pharmacological targeting of cancer-causing mutations. But one cannot ignore a similarity to the historical unwillingness to embrace a rule-governed change of a complex systems as an explainable process as such.

At its core, such thinking accepts a change of *P* only by reducing it to a discrete and irreducible causal event, the mutation that changes *G*. Thus, a mutation is akin to the catastrophic events in geology—from nowhere and in no need for an explanation. Invoking a random event helps to avoid the explanation of a change of a *system's configuration*, manifest as 'phenotype', by a rule-governed process, let alone a mathematical description.

The rigid habit to reduce phenotypic changes to genetic mutations ignores a central fact of metazoan development: that the vast diversity of distinct and stable cell phenotypes in the body is produced by the process of cell division and differentiation without altering the genomes of the cells. If cells as diverse as the liver cell, the neuron, or the granulocyte all share the same genome with the same set of genes, why would a cancer cell, which compared to its less malignant variant just divides faster, or is more resilient to toxic reagents, requires an alteration of the genome to generate these incremental features—differences dwarfed by that between cell types?

Because transitions from one particular cell phenotype P_1 to another phenotype P_2, be it the acquisition of a new cell-type trait during development, or new cancerous capabilities during tumour progression, are all recurrent, rule-governed processes requiring coordinated change in a vast number of subcomponents, we treat them here as a change of a *system configuration* **X**: the orchestrated shift in the material composition of a cell that embodies one coherent functional state or phenotype to another one. A question that we also address is then what is the role of genetic mutations if cancer progression is driven by changes in system configuration within the immense repertoire of phenotypes afforded by one same genome.

We do so by considering a more encompassing view in which phenotypic innovation in neoplasia is the result of entering the '*adjacent possible*', a powerful concept introduced by Stuart Kauffman [11]. By this we mean that all the complex and coherent functionalities of malignant cells, its 'hallmarks', are not in principle novel traits; in the healthy organism they are not just *actualized* (realized), but they are also in principle possible, that is, *available* to be actualized (realized). The cancer hallmarks are all functions built-in by evolution in the theoretical behavioural repertoire, similar to functions used in the physiological realm, but reconfigured and realized in a new context in cancer. This combinatorial principle of innovation diminishes but does not eliminate the explanatory burden for genetic mutations as source of phenotype innovation and entices us to embrace a profound concept that is a key message of this chapter: the *potentiality* of the cancerous states is somehow preordained by the repertoire of possible (actual and potential) phenotypic behaviours that emanate from genomic information.

The tumour cell is the realization of a potential that in the healthy state is unrealized: it is not in the *actual* but the *possible* just beyond it, '*adjacent*' to it. The adjacent possible is created as an inevitable, latent by-product during the evolution of the healthy state. In this sense, genetic mutations are not the *cause* of cancer but *catalyze* its development, and as explained from first principles of dynamical systems theory: mutations facilitate the occupation of a 'potential' state that is just adjacent to the 'actual'. Once a cell has entered the adjacent possible, freed from constraints of the actual that has been optimized for organismal function by evolution, it can undergo diversification. And having its own adjacent possible, it will move farther from the realm of the physiological actualized. A chain reaction ensures that malignant innovation begets malignant innovation.

This chapter presents the argument in a qualitative, simplified but logically coherent manner; we do not resort to equations to describe the underlying ideas (which are reviewed elsewhere and will be referenced) nor fill the space with examples of specific molecular pathways that so many review articles have already provided and only distract from the explanation of first principles. Our goal is to earn the interest of experimental and clinical cancer researchers who already possess immense knowledge of the specific facts but are in search of an integrating conceptual framework. To them, this chapter shall hopefully provide with generic tools for thought.

1.2. Challenging the paradigms: non-genetic dynamics and treatment-induced progression

Tumorigenesis and tumour progression are governed by a relentless arrow of progression from the normal to increasing malignancy of the cancerous tissue, of which one trait is the resistance to any virtual type of treatment, be it traditional chemo, targeted, radiation, or immune therapy—the ultimate cause of death from cancer [12–17]. Why does any current treatment, if they fail to eradicate the tumour upfront, consistently result in the recurrence of a more immature ('stem-like') and malignant, notably metastatic, and treatment-resistant tumour? This universal behaviour warrants the consideration of fundamental properties of cancer progression. Here, we discuss challenges to long-existing paradigms regarding cancer treatment and the biology of progression towards therapy resistance which have begun to exhibit cracks under the load of new data afforded by -omic technologies. We then introduce two principles that must be acknowledged in view of these findings: non-genetic phenotype dynamics (Section 1.2.1) and treatment-induced progression (Section 1.2.2).

The unquestioned notion that more efficient killing of cancer cells implies more effective treatment is plausible (as long as collateral damage can be avoided). This taken-for-granted concept has culminated in the development of target-selective drugs aimed at killing cancer cells. But the tumour that recurs is almost always refractory to the same treatment [18,19], and yet existing approaches are to double down on killing by identifying remaining vulnerabilities to kill even the relapsed tumours, typically with combination therapies, so far with scant lasting benefit [20]. Could the very act of attempting to kill cancer cells be the problem of recurrence itself?

Equally plausible and unquestioned has been the explanation of a link between the killing and development of resistance that invokes the neo-Darwinian evolution (see **Figure 1.1**A). In this second prevailing paradigm, any phenotypic innovation, such as the trait of treatment resistance, is the result of genetic mutations that afford mutant cells some (often imagined) selection advantage in the tumour cell population of the tumour tissue, relying on the aforementioned bijective $G \rightarrow P$ mapping. However, the rapid and consistent emergence of the complex stem-like phenotype of cancer cells that confer resistance to cytotoxic insults (stem cells are naturally more resilient to xenobiotic injury [13,21,22]), and the failure of cancer genome sequencing [23–30] to reveal a clonal architecture of the tumour cell population that is consistent with Darwinian selection, have begun to challenge this second paradigm [10,31–36].

1.2.1. Non-genetic phenotype dynamics

A first step of departure from neo-Darwin thinking still explains the acquisition of resistance by Darwinian *selection* but with the novelty that it selects cells that exist in a stable *non-genetic* ('epigenetic')

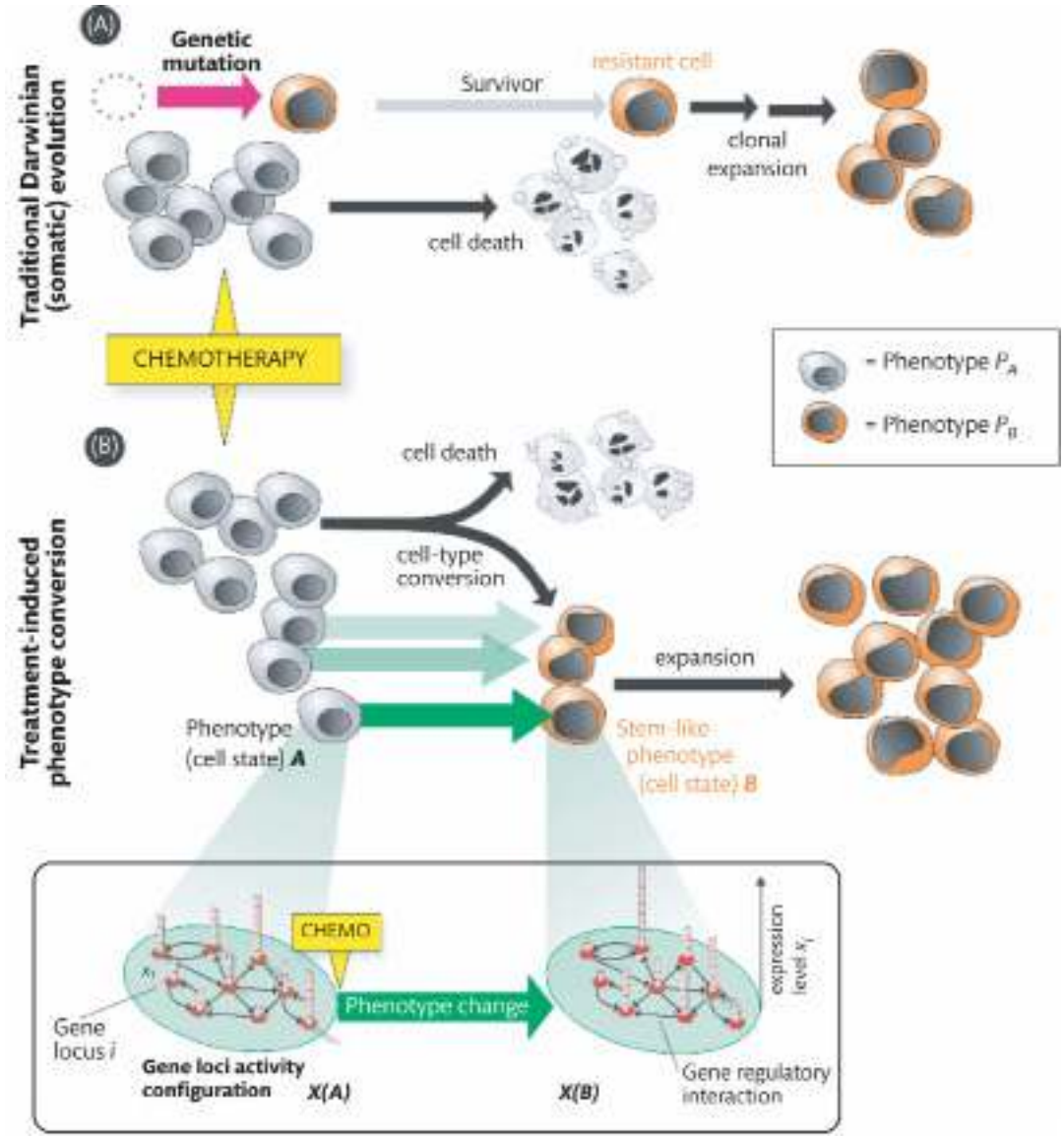


Figure 1.1. Selection versus induction of a new cell phenotype, such as stemness or treatment resistance. Following chemotherapy, cells either carrying (pre-existing) mutations or in an alternative 'epigenetic state' (orange cells with large nucleus) that confers that resistance will expand by natural *selection*. (**A**) Alternatively, chemotherapy *induces* the resistant phenotype, the Nietzsche effect. (**B**) Often the latter occurs via a state bifurcation in which cells either die or are stressed and undergo individually, without selection, a change in gene expression program that underlies a phenotypic conversion *PA* → *PB*. The shift in expression is governed by the gene regulatory network (box; see Figure 1.2 for more details)) that channels the change in the activity configuration of all the gene loci.

stem-like state within an isogenic (clonal) cell population [37,38]. Such cells exist because of non-genetic heterogeneity of cell phenotype in any clonal cell population that is now recognized, not least owing to now ubiquitous single-cell resolution analysis of gene expression profiles [30,39–44]. We will return to this non-genetic 'background heterogeneity' that is linked to phenotype plasticity. Because of such non-genetic heterogeneity of cell populations, some cells may by chance (likely at a higher rate than through random mutations [45]) occupy a stem-like state that is stable enough to be inherited across a sufficiently large number of generations [46] and that possess fitness advantage in the presence of anticancer agents. These cells would be expanded during treatment—this would constituting 'mutationless Darwinian evolution' [47]. But the idea still rests on selection. And by relying on random events to provide the new phenotypic properties, we still do not address the mechanism for the actual process of the phenotype shift.

1.2.2. The Nietzsche effect on cancer treatment

A closer examination of a series of *in vitro* treatment experiments, notably the aforementioned single-cell resolution analysis, as well as a direct observation of individual cells becoming resistant by video microscopy and quantitative models of the rate of acquisition of resistance [46,48], obliges us to see beyond the concept of *selection*, be it of genetic variant cells or of cells in an alternative non-genetic stable state, and to also consider an alternative: the *induction* by treatment itself in a *given* cell to the new resistant phenotype [49–54] (Figure 1.1B).

If the resistant cells represent a stem-like functional state not 'caused' by genetic mutations, we must entertain the idea that such alternative cell states are *regulated* cellular programs: a pathological form of 'stemness' (which may also underlie the old notion of 'cancer stem cell'). Regulation is more likely to achieve a coherent functional (biologically meaningful) state than the stepwise selection of random mutations to evolve the same trait. Regulation is a deterministic process and hence also likely to be faster than the stochastic occupation of a non-genetic alternative cell state followed by preferential expansion. A *regulated* state is the result of a *regulating* influence: a phenotypic switch of the cell, induced by an external signal that does not change the genome but the 'gene expression program' of the genome.

Thus, we have 'Lamarckian-like' dynamics, where a novel (inheritable) phenotype is directly induced by external influence on *individual* cells in a cell population. This process requires (1) a regulatory mechanism that converts a cell's phenotype by shifting $\boldsymbol{X}$ and (2) an external agent E that sends the signals that trigger this process. Then, we can formally talk about a state transition of a (plastic) phenotypic state P under invariance of the genome G caused by the external signal E.

Practically, it has become clear that cytocidal treatment, a perturbation to the system that seeks its destruction, is a double-edged sword: yes, killing tumour cells reduces the tumour burden. But if it fails to eliminate every single cell, it will still have stressed those non-killed cells with a sublethal blow. In doing so, treatment may elicit an active defence response that consists of switching to a stem-like, and thus, resistant state. In the last figure of this chapter, we present a formal explanation for the necessary double-edged nature of cell response to near-lethal stress. Most part of this chapter is dedicated to a stepwise build-up of the conceptual understanding of theories needed to comprehend this counter-intuitive phenomenon.

Cancer recurrence is thus inseparably linked to treatment, which *eo ipso* causes stemness in the non-killed cells. In this sense, treatment can 'backfire', manifesting, to use Nietzsche's aphorism, the principle that 'What does not kill me makes me stronger' [55,56].

But why will a near-lethal, non-specific stress so consistently cause a specific change of P (under invariant G) that produces stem-like traits? To address this broader question, we introduce some abstract concepts from which the natural origin of the cancerous state can be logically derived. We start with two premises and dynamical systems theory, followed by their application to cancer biology.

1.3. Premise I: Intrinsic cell state dynamics is governed by the gene regulatory network

We first consider how a shift of phenotype P of a cell is governed by the *internal* machinery of the cell that constrains the plasticity of P in specific ways [57,58]. This control is receptive to *external* influences that facilitate, trigger, or prevent the changes of P which is discussed in a second step (Section 1.4).

The picture of the internal regulation of the cell borrows the concept of an internal milieu as a homeostatic (self-stabilizing) entity, separated from its environment, yet it is open for a selective exchange of material, energy, and 'information'. The self-stabilization of the internal state affords a stable cell phenotype that we recognize as a 'cell type' or as a functional state (e.g. a secretory, proliferative, and migratory state). To further reduce a phenotype to its building blocks, we introduce the *system configuration* $\boldsymbol{X}$ as its material basis: $\boldsymbol{X}$ is the configuration of the activation status x_i in the cell of all the N gene loci i of genome G. The values of all x_i ($i = 1, 2, \ldots, N$) collectively define the gene expression profile that ultimately dictates the abundances of proteins that in turn exert their downstream effector functions, including the biosynthesis of all the structural and biochemical constituents of the cell that underlies the phenotype P. Thus, $\boldsymbol{X}$ is a high-dimensional vector state vector that contains the values of the (gene expression) activity level of all gene loci in genome G with N gene loci:

$$\boldsymbol{X} = \left[x_1, x_2, \ldots, x_i, \ldots, x_N\right]$$

The details are illustrated in **Figure 1.2** that uses pictures to guide us step by step through a series of concepts. The collective activity of all the N gene loci i (Figure 1.2A) that represents $\boldsymbol{X}$ (Figure 1.2A) maps into an apparent phenotype P (e.g. cell type). $\boldsymbol{X}$ can be measured at the granularity level of the components x_i using genome-wide technologies, most directly as the transcriptome or, with some caveats [59], as genome-wide pattern of 'epigenetic marks', such as chromatin openness (a prerequisite for transcription), DNA methylation, or post-translational histone modifications at the loci i, all of which are (loosely) associated with the activity of the respective genomic loci.

Now, the essential property of configuration $\boldsymbol{X}$ is that the changes of its components, the individual gene expression activities x_i, are not independent from each other. They influence each other in a preordained manner, namely following rules 'hardwired' in the genomic sequence. The latter determines all transcriptional regulation: which specific locus i influences the activity of which other specific gene loci, x_j and x_k, as described below. The central principle is that the interdependence of the loci constrains the ways in which the configuration $\boldsymbol{X}$ can change to predestined trajectories.

A change of cell phenotype (Figure 1.2C) can be imagined as a movement (green arrows in Figure 1.2C–E) of $\boldsymbol{X} = [x_1, x_2, \ldots, x_i, \ldots, x_N]$ in the N-dimensional *state space of* $\boldsymbol{X}$ (blue areas in Figure 1.2D) along trajectories. Each configuration $\boldsymbol{X}$ maps to a *position* (green balls in Figure 1.2) of the state space, each dimension of which represents the activity value x_i in $\boldsymbol{X}$ at a given time, such that the values x_1, $x_2, \ldots, x_i, \ldots, x_N$ are the coordinates in this N-dimensional *state space.*

As to the molecular implementation, locus-specific cross-regulation is mediated by 'locus-aware' TFs that are encoded by one locus but regulate other loci. Non-coding (nc) RNA [60] that prevents the synthesis of the protein expression of the target locus has emerged as another important mode of regulation between genomic loci. Such locus-aware regulation is supported by locus-agnostic 'epigenetic modification', chromatin confirmation, other structural constraints (looping), etc. [59]. In the case of TFs, locus awareness is epitomized by the specificity of their DNA-binding domains for binding to their cognate DNA sequence motifs in the regulatory regions of their target genes and their ability to recruit the locus-unaware chromatin-modifying enzymes that help to suppress or activate the transcription of these target loci [59]. In the case of regulatory non-coding RNA, such as miRNAs and many other ncRNAs, target specificity is directly readable from their DNA sequence. Thus, all the interactions between the gene loci are specified by the DNA sequence of the genome G.

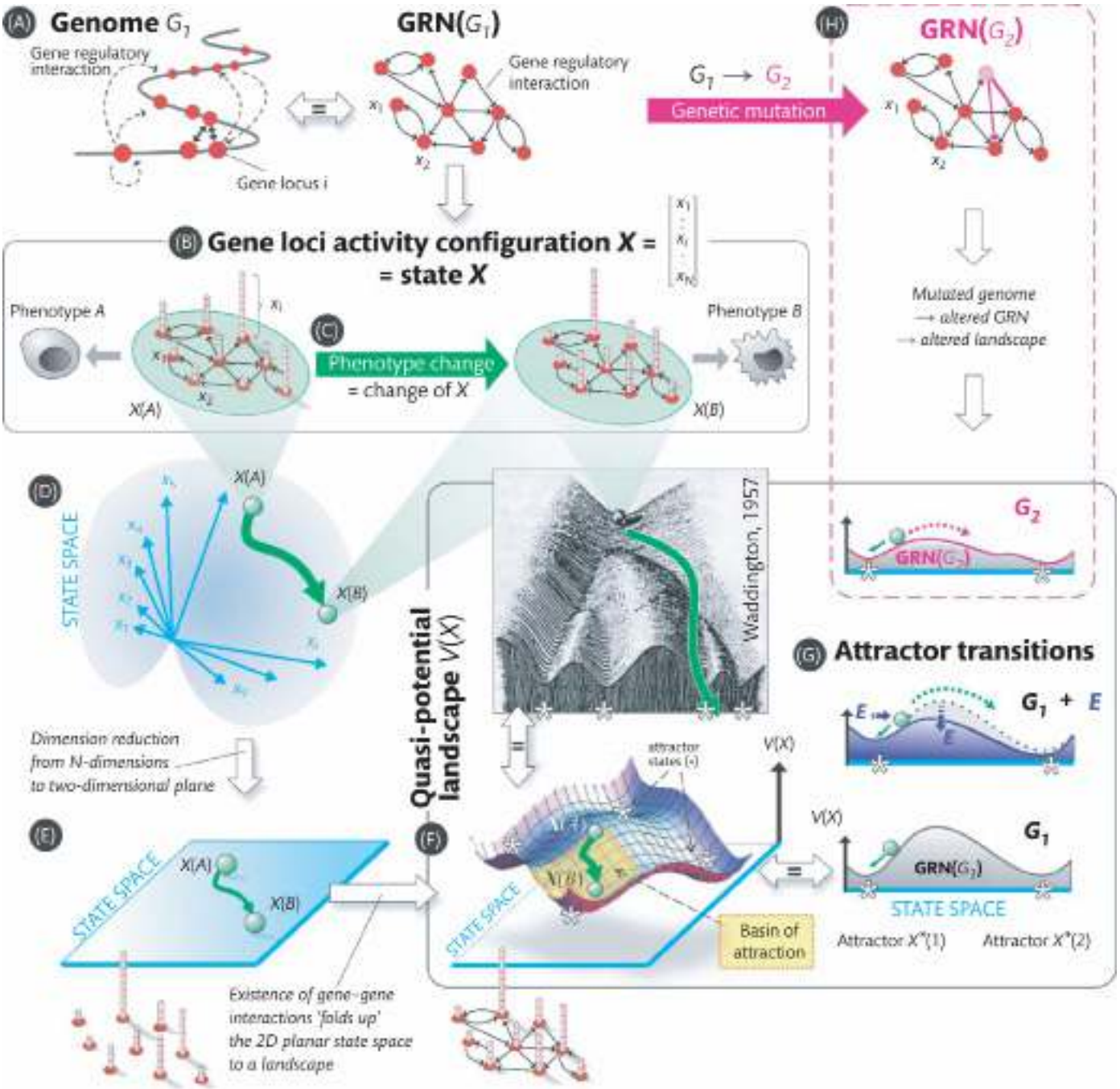


Figure 1.2. The key concepts, from dynamical systems theory in pictures, are explained step by step throughout this chapter. Red dots: gene loci *i*; pink towers: gene activation level *xi* of respective locus (vertical axis); green balls: systems configuration *X* (or cell state); light blue area or line: state space for *X* (projections); green (curved) arrows: state change (trajectory, chreods); *Gi*: genome (and associated GRN); *E*: external influence (panel (G)) promoting attractor transition either by shifting *X* (horizontal blue arrow) or by reducing the quasi-potential barrier (vertical arrow down); asterisks: attractor states; *V*(*X*): quasi-potential for each *X*; white shadowed arrows: conceptual relationships (equal sign indicates equivalence). Note the major dualism between non-genetic phenotype change via change of *X* (green thick arrow in panel (B) and via genetic mutation (pink thick arrow in panel (H)) that permanently changes the quasi-potential landscape (pink silhouette).

Collectively, these genome-encoded regulatory interactions between the entirety of all the *N* gene loci in a genome *G* form the GRN of genome *G* [61] and represent a *dynamical system* whose wiring diagram is hard-coded by the genome (Figure 1.2A and B). By defining all interactions between x_i, the GRN imposes the constrains on the movement of ***X*** in state space, as illustrated below. The wiring-diagram specification is the result of evolution and governs how the interdependence of the components x_i of the state vector ***X*** produces the repertoire of all biologically meaningful patterns of gene locus activation ***X*** and how ***X*** can change (Figure 1.2C): not freely but channelled by the interactions along preordained trajectories (green arrows in Figure 1.2D–F). But the very same GRN also

harbours the *potential* for states $\boldsymbol{X}$ that lie in the *adjacent possible* and are not realized by the normal tissue but play a role in cancer.

1.4. Premise II: Perturbation and stochasticity shift gene activation configurations X and hence cell states

In addition to the *intrinsic* regulation controlled by the GRN (Figure 1.2B), the activity x_i of a subset of gene loci is also subjected to direct *extrinsic* regulation by extracellular signals E, embodied by hormones, cytokines, and other mediators, as well as environmental signals (nutrients and their metabolites, drugs, and toxins). Their impact on gene expression x_i alters $\boldsymbol{X}$ and is referred to as *perturbation* of the GRN as a dynamical system. The external signals E bind and activate cell surface or cytoplasmic receptor molecules, triggering signal transduction cascades that impinge on response elements in the regulatory regions of genes. These cellular signalling pathways are also hardwired in the genome and determine what set of gene loci are directly receptive to such external signals. Thus, E can change $\boldsymbol{X}$ via coordinated changes in a set of x_i in a manner defined in G (see Figure 1.2G, bottom, blue horizontal arrow).

Perturbations of $\boldsymbol{X}$ can also be non-deterministic and come from within the cell, in the form of *molecular noise* [62,63]: the thermal fluctuations in abundances of biomolecules that do not 'average out' in the small reaction volumes of cellular compartments. Such stochastic fluctuations are intrinsic to the variable x_i, i.e. they represent a change that is 'unmotivated' by a deterministic cause E and affect the state $\boldsymbol{X}$ in a way that is best imagined as a 'wiggling motion' of $\boldsymbol{X}$ in the N-dimensional state space. Beyond molecular noise other sources of non-determinism have been proposed, including the random shapes of intrinsically disordered regions of proteins that introduce stochasticity in the choice of reaction partners [64,65] or structural arrangements of three-dimensional chromatin loops [66].

A given invariant genome G (of an organism) maps into a multitude of phenotypes P encoded by $\boldsymbol{X}$ that changes in time in processes controlled by the invariant GRN. Thus, the GRN, hardwired by G, establishes an entire *repertoire* of P. It is herein that the old notion of a 1:1 mapping between G and P collapses. The rule-constrained changes of P under invariant G are now referred to as (*non-genetic*) *plasticity*, a term increasingly used in cancer research that has barely been formally defined [9,10,33,67,68]. If $\boldsymbol{X}$ changes in response to influence of E, a large number of cells in a population may change in unison. By contrast, intrinsic fluctuations in $\boldsymbol{X}$ due to molecular noise would affect individual cells differently. Both regulated cell state changes and molecular noise are the source of the *non-genetic heterogeneity* (diversity of cells in isogenic cell populations) that has equally gained interest in cancer biology [33,37,63,64,69].

It is therefore of central importance to remember that the possibility of intrinsic 'wiggle room' of $\boldsymbol{X}$ not only makes the $G \rightarrow P$ mapping non-bijective but also non-deterministic: concretely, this means that two isogenic cells of the same cell type may, under identical conditions undergo, differentiate into distinct cell types as if they throw the dice for cell-fate decision. Thus, the choice of a particular phenotypic outcome (e.g. to divide or not and to differentiate into cell type A or B) cannot always be reduced to an upstream causal determinant, such as a regulatory signal E, but instead can appear to be stochastic, yet constrained to the repertoire (e.g. limited to a (random) binary choice between two specific fates for a given $\boldsymbol{X}$). The 'microscopic' stochasticity due to randomness of molecular fluctuations thus can be amplified by the GRN into apparent randomness of macroscopic cell phenotype. Such behaviours of individual cells defy genetic determinism—the mode of thought captured by the 'arrow–arrow' schemes of mechanistic pathways that underlie 'precision oncology'. In fact, such *macroscopic indeterminacy* manifests the instabilities [70] of tumour states and can only be comprehended using concepts of non-linear dynamics, as explained in the next section.

1.5. Elementary dynamical systems concepts—the quasi-potential landscape

To make the above premises more concrete for application to cells, the general idea of the (*quasi-*) *potential landscape* is considered useful (for more detailed introduction, see refs. [58,71]). This mathematical landscape captures the behaviour of $\boldsymbol{X}$ and is the formal basis for the 'epigenetic landscape' proposed by Waddington starting in the 1940s which led to the famous 1957 depiction (Figure 1.2F) [72] but has often been dismissed as metaphorical. The GRN as a dynamical system produces a 'quasi-potential landscape' on which every point (geographical position) is a gene activation configuration $\boldsymbol{X}(t)$, the state of the GRN or, for that matter, a cell, as a dynamical system at time t.

1.5.1. The GRN folds up the landscape

The characteristic topography of the quasi-potential landscape is determined by the specific wiring diagram of the GRN in the following way: first, for visualization by the human mind, we project every possible high-dimensional configuration $\boldsymbol{X} = [x_1, x_2, \ldots, x_N]$ (a point in the N-dimensional state space) into a two-dimensional (2D) plane (Figure 1.2E), preserving as much information as possible in that two neighbouring points in the 2D plane would represent configurations $\boldsymbol{X}$ that are closely similar with respect to their state $[x_1, x_2, \ldots, x_N]$ in the N-dimensional state space.

With this projection of all states $\boldsymbol{X}$ onto the 2D plane (Figure 1.2D and E), we free up the vertical axis for displaying a property $V(\boldsymbol{X})$ that smoothly varies between the configurations $\boldsymbol{X}$ in the plane. If gene loci would not interact, then all configurations of gene activities would be equally 'probable', that is, requiring equal 'effort' or 'action' to realize. The term 'action' is of central theoretical significance here [73]. Since the GRN introduces dependencies between the activity values x_i of the gene loci i, it makes each configuration $\boldsymbol{X}$ different in terms of how much regulatory effort is required for them to be realized—or how much one has to 'act' against the rules imposed by the GRN. For instance, if gene A is a repressor of gene B (as determined by the genome G), then an increase in x_A decreases x_B. A configuration in which both x_A and x_B are simultaneously high would thus be much harder to implement, requiring extra *action against regulatory constraints* imposed by the GRN. Equivalently, the GRN exerts a force that drives the cell in state $\boldsymbol{X}$ from a position with coordinates $[\ldots, x_A(\text{high}), x_B(\text{high}), \ldots]$ to the more stable state with coordinates $[\ldots, x_A(\text{high}), x_B(\text{low}), \ldots]$. The latter complies with the rules in the GRN's wiring diagram. Some configurations, such as $[\ldots, x_A(\text{high}), x_B(\text{high}), \ldots]$, are difficult or near impossible to realize, others would be very much favoured. Thus, roughly, due

to the inevitable presence of molecular noise, $\boldsymbol{X}$ randomly wiggles in the 2D plane, exploring neighbouring configurations $\boldsymbol{X}$ 'to probe which is more or less probable to be realized'. Driven by the stochastic fluctuations, the cells move along the path of *least action*. This becomes imaginable and intuitive with the landscape topography that displays the relationship between neighbouring points with respect to $V(\mathbf{X})$.

We therefore plot the relative regulatory effort (with some reference point) that needs to be spent for realizing a given configuration $\boldsymbol{X}$ as dictated by the GRN as elevation V on the vertical axis above each respective configuration $\boldsymbol{X}$ in the 2D plane (Figure 1.2F). In doing so, all the values $V(\boldsymbol{X})$ at positions $\boldsymbol{X}$ jointly establish a landscape of varying altitude over the 2D plane since different configurations require different efforts to be realized. Thus, the interactions in the GRN between components x_i of the state vector $\boldsymbol{X}$ collectively *fold up* the 2D plane into a landscape $V(\boldsymbol{X})$ with a characteristic *topography*: this is the quasi-potential landscape that is fully specified by the GRN. Configurations $\boldsymbol{X}$ with higher 'elevation' $V(\boldsymbol{X})$ are configurations that require relatively more action maintained by countering the regulatory rule imposed by the GRN or, crudely speaking, are less 'probable' and more 'unstable'.

Since similar configurations are placed next to each other in the 2D plane, and in the case of the class of networks to which the GRN belongs [74,75], they tend to also have similar values for V, the landscape is rather smooth and not overtly rugged. (There are deeper reasons being studied which relate the type of architecture of the GRN and the smoothness/ruggedness of the landscape, but this is beyond the scope of this chapter.) It suffices to state the hypothesis, based on observations, that GRNs seem to belong to a class of networks that create relatively smooth, yet multi-valley landscapes that are suited to govern metazoan development [74,76–78]). The slope between neighbouring points then indicates that a cell whose gene loci activation configuration $\boldsymbol{X}$ is 'less probable' would spontaneously 'move' (with respect to its position in the 2D plane) towards a more readily realizable state, following the direction of the steepest downhill slope indicated by the quasi-potential $V(\boldsymbol{X})$. The cells thus roll down towards the lowest points in the valleys as Waddington had anticipated in 1957 with his epigenetic landscape.

1.5.2. Theoretical caveats to the notion of 'energy landscape' and gradients

Now it must be said that however tempting it is, given the relative 'smoothness', to view this landscape of cell states as an 'energy landscape' with potential wells in which cells are balls that roll down a gradient because of gravity, here is where the equivalence to energy landscapes must stop [71,73,79]. The elevation is *not* an energy potential in the classical sense of a conservative system. For instance, the 'net efforts' required to go from $\boldsymbol{X}_A$ to $\boldsymbol{X}_B$ is path dependent [80]. We do not have 'true' gradients that generate the driving force, only approximately so, and in a formally distinct way, hence the prefix 'quasi'. Nevertheless, the notion of 'energy minimalization' and of slopes that act like gradients is to *some* extent warranted. The underlying formal theory is based on the 'least action principle' of 'large deviations' (due to molecular noise) that is concerned with the minimal action needed [73], driven by minimal noisy fluctuations, to 'climb' up the hill to exit a 'potential well'. In classical mechanics with energy conservation, the elevation represents a global free energy that is conserved: only energy differences (altitude difference) matter for the efforts required to climb to some higher spot on the hill, not the path. In such systems, the elevation represents the stability of a state of the system or the steady-state probability to find the system (cell) at that given state. The extent of admissibility of the gradient idea for landscapes produced by non-conservative systems depends on the specifics of the architecture of the underlying network. For GRNs, which are evolved networks, this requirement seems to often be satisfied and sufficient for the qualitative analysis of the dynamics of the GRN [74,75].

At the level of the concrete, the GRN is an open (non-linear) dynamical system that operates far from thermodynamical equilibrium and in which the maintenance of order and interesting stationary patterns hinges on the constant influx and consumption of free energy (in the form of adenosine and guanosine triphosphates (ATP and GTP) that power the cell, including the regulatory activities of the GRN). Such non-conservative systems also lack the inertia of classical system that would allow a ball rolling downhill to then also roll uphill on the other side of the valley, recovering the potential energy, hence the use of the term 'overdamped dynamical system' to describe the dynamics of biological regulatory networks. Yet, given these caveats, overall and with due caution, the landscape picture is useful as a mental aid for fathoming complex and robust systems dynamics as proposed by Waddington whose intuition of the 'epigenetic landscape' turned out to be largely correct.

As an interesting side note, if we focus on domains of the state space devoid of circular dynamics (in which $\boldsymbol{X}$ goes around a cycle, repeating itself, or has spiral-shaped trajectories), then the dynamics of some nonequilibrium systems, such as the GRN, can indeed be viewed as 'gradient-like' dynamics. For this reason, some theoreticians treat the landscape as a purely geometrical object to describe cell-fate behaviours, independent of the notion of an underlying GRN that creates the landscape [81]. In this gene-agnostic framework that assumes gradient-like dynamics, mathematicians use the tools of differential geometry to study the *global* changes of the landscape topography that must obey geometric constraints as a way to predict what behaviours, such as fate decisions, are possible. By contrast, here we maintain the conceptual link to the interacting genes in the GRN and examine *local* changes of the landscape that governs cell phenotype switching as described in the following sections.

1.5.3. Central concepts of dynamical systems: steady states and attractors

With the landscape image as a tool in mind, we can now return to explaining the GRN as a dynamical system that governs cell state dynamics. Since the landscape topography captures (with the above caveats) a 'driving force' on the cell state imposed by the GRN that seeks to 'satisfy' regulatory interactions, we can identify $\boldsymbol{X}$ positions in the 2D plane which are not on a slope but are 'flat': these points are configurations of $\boldsymbol{X}$ that do not experience a force from the GRN to change. They are the *steady states* in which all regulatory rules, such as 'gene *A* represses gene *B*', are 'satisfied'. A steady-state configuration is denoted by an asterisk, $\boldsymbol{X}^*$, which is defined by distinct values for the activation of the gene loci, x_i^* (Figure 1.2F). The surrounding of $\boldsymbol{X}^*$ determines (at least, for simplicity) two distinct types of steady states: a steady state in a dynamical system can be *stable* or *unstable*. The stable steady-state $\boldsymbol{X}^*$ is an *attractor* state if it is at the lowest point or a (round) valley or 'potential well', surrounded by 'uphill' slopes in every direction. These slops would retore the system to $\boldsymbol{X}^*$

when the system is perturbed (in the above sense) to deviate to $\boldsymbol{X}'$ from $\boldsymbol{X}^*$. Alternatively, a steady state is *unstable* when it is exactly balanced on a hilltop, but any slight departure to $\boldsymbol{X}'$ into any direction would cause it to fall along the fall line of a surrounding slop, increasing the distance away from the unstable steady-state $\boldsymbol{X}^*$. In the high-dimensional landscape, barely fathomable by human mind, steady states can be stable with respect to some dimensions of the state space while unstable with respect to other. Such geometry establishes, for instance, long valleys draining a 'river' at its bottom into deeper regions, with uphill slopes on both sides, or mountain ridges, with downhill slops on both sides.

1.5.4. Multi-stability: the presence of many attractors in the landscape of one GRN

A fundamental property of complex systems, such as the GRN, is that its landscape exhibits multi-stability, i.e. has *multiple* attractors (Figure 1.2F). *Multi-stability* is a characteristic of the GRN. Mathematically, multi-stability requires the presence of auto-regulatory loops (e.g. gene *A* activates gene *A*; or, indirectly, gene *A* inhibits gene *B* which inhibits gene *A*) with non-linear regulatory characteristics (e.g. sigmoidal input–output relationships at a locus *i*). Such regulatory circuitries abound in the genome-wide network of interactions of GRNs and countless small regulatory circuitries have been modelled to study multi-stability [82,83]. In fact, the GRN with thousands of regulatory genes and several times more regulatory interactions may have, according to simple simulations in toy model GRNs, thousands or more stable attractor states [74,76,78,84]. They correspond to the above-discussed repertoire of distinct, stable phenotypic behaviours that a genome can realize. By geometric necessity, stable attractors with their basin of attraction are all separated from each other by unstable states, as much as valleys are separated by ridges or hills. The curve containing adjacent unstable steady states that separates two basins of attraction, akin to a mountain ridge, is called a *separatrix*. The unstable steady states on it are poised to fall into either attractor $\boldsymbol{X}^*_1$ or $\boldsymbol{X}^*_2$ on either side of it. Thus, the *separatrix* enforces a binary decision on a cell placed on this unstable position to 'decide' to adopt either phenotype P_1 or P_2 commanded by the $\boldsymbol{X}^*$ of either one of these two attractor states.

1.5.5. Transient and permanent changes of the landscape topography

We can now consider changes of the landscape topography. Being a folded surface with gradient-like features, the topography can only change in some topologically permitted ways, some types of distortions. But since the topography directly maps into cell state dynamics, these surface distortions are at the core for explaining any rule-governed behaviours of cells. While the landscape topography is determined by the GRN, it still can change but at a timescale that is longer than that in which changes of $\boldsymbol{X}(t)$ operate. The two major ways of changing the landscape are as follows (Figure 1.2G vs Figure 1.2H):

(a) by *transient modulation of the strength/type of regulatory interactions* (Figure 1.2G) by external influences *E*, e.g. metabolites or hormones [85,86]; this alters the value of a *parameter* in the equations (differential equations) describing the regulatory function of locus *i* that maps the inputs of *j*, *k*, *l*, etc. to their target locus *i* to its output, the rate of change of x_i—the change of gene expression;

(b) by *genetic mutations* (Figure 1.2H); they permanently rewire the GRN since they directly or indirectly impact regulatory interactions, e.g. by deleting a DNA-binding site of a TF.

These localized changes in the GRN wiring diagram are relatively subtle relative to the complex intertwining interactions in a large network of thousands of regulator genes. Indeed, computational models and theory have shown that the mapping of the GRN to the landscape topography is relatively robust, in that a single change in an interaction of the GRN, even deletion of a gene, may only distort the landscape 'slightly', resulting in *quantitative* (gradual) changes of the height of hills between valleys or, equivalently, the size and depth of basins of attractions (see Figure 1.2G, bottom, and Figure 1.2H, bottom). Only rarely does an alteration of an interaction parameter in the GRN, or rewiring, cause *qualitative* changes, such as creating a new hill that separates valleys, which would create new attractors or, conversely, abolish entire attractors. These qualitative changes in which attractors are created or disappear are referred to as (local) *bifurcations* (see later, **Figure 1.3**C). In other words, most changes in the GRN wiring diagram will have no such qualitative, 'catastrophic' consequences. This property of the system to preserve the qualitative attractor structure is called *structural stability* [87] (e.g. Figure 1.2G) and must not be confounded with the stability of the attractors in that $\boldsymbol{X}^*$ resists perturbations (enforced alteration of the values of x_i) by staying in the basin of attraction. Structural stability plays a role in buffering against the effective genetic mutations; conversely, boundaries of structural stability can be crossed, for instance, by those rare homeotic mutations that drastically alter the body plan.

1.6. Application to cancer: cancer as entry into unused attractors

We now apply the above general principles to cancer biology. We consider a gene activation configuration $\boldsymbol{X}$ and its position on the quasi-potential landscape to represent a phenotype *P* and its 'relative stability' (tendency to change or lack thereof). Using these principles, we present the following three key biological corollaries.

1.6.1. Cell types are attractor states

A state $\boldsymbol{X}^*$ being an attractor states implies that a perturbation-induced state $\boldsymbol{X}'$ that is ΔX away from $\boldsymbol{X}^*$, and still in its basin, will spontaneously 'flow back' to the attractor state, $\boldsymbol{X}^*$, once the perturbation has subsided. The attractor configuration $\boldsymbol{X}^*$ other than being a self-stabilizing steady state also produces a characteristic gene activation profile that also includes the expression of non-regulatory effector genes and contributes to the characteristic of cell phenotype *P*. Since attractor states $\boldsymbol{X}^*_i$ are demarcated from each other by separatrices between their basins, the cell phenotypes P_i are discretely distinct from each other. Therefore, P_i corresponds to cell types that represent qualitatively distinct categories.

Waddington, of course, equated the valleys in his epigenetic landscape to cell types. More formally, the existence of multiple attractors in the very same molecular network and the concept that they correspond to the distinct cell types that one dynamical system can generate were first proposed by Max Delbruck in 1949 and then

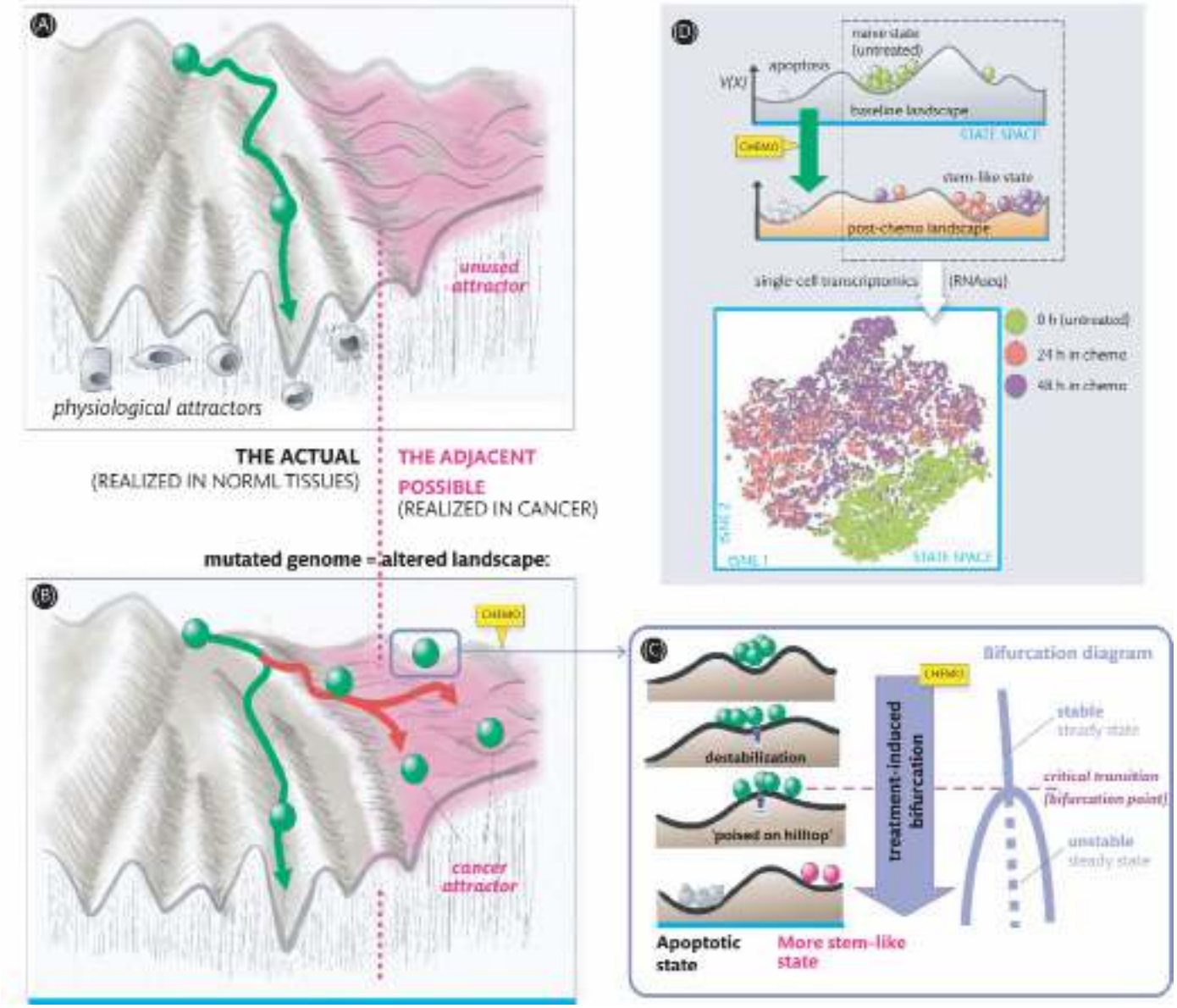


Figure 1.3. Cancer as a consequence of entry into the 'adjacent possible'. The state space, or the quasi-potential landscape, of the GRN that governs development and ensures that tissue homeostasis (green trajectories, leading to normal cell types, in grey) also contains regions that are normally not occupied by cells, the adjacent possible (pink regions). This non-physiological portion of state space contains unused attractors as by-products of GRN evolution (A). Genetic alterations distort the landscape, in such a way as to facilitate 'spilling' of cells into the adjacent possible (B) when they depart from the chreods (red arrows). These cells are trapped in the unused attractors that represent the gene expression configuration of cancerous cells. These pathological attractors lack access to the chreods and are not as deep, allowing easy transitions between attractors (producing phenotypic cell heterogeneity). These higher elevation states represent abnormal immature cells. Stress from cytocidal treatment can trigger bifurcation (critical transition) that destabilizes cells and induces attractor transitions in even more stem-like (resistant) cells (C). Panel (D) shows experimental data (underneath a schematic model) in which leukeamic cells (HL60) were treated with vincristine ('CHEMO') for 24 or 48 h; surviving cells were profiled for their transcriptomes individually using singe-cell RNAseq. The states X of each cell of each of the three conditions were pooled and projected into the same 2D plane (using the tSNE method). Each dot corresponds to a cell, and the colour label represents the experimental condition (untreated, treated for 24 and 48 h), not the cell state that is indicated by its position.

solidified by the discovery in 1963 that genes regulate each other (the very first notion of a GRN) by Monod and Jacob, who also observed the correspondence between attractors and stable differentiation states (see refs. in [58,88]). Stuart Kauffman then proposed in 1969 that high-dimensional attractors of the complex GRN of thousands of genes map to every cell type in the metazoan body [74,89]. If cell types are attractors, then it follows that differentiation from cell type A to cell type B is the transition from the attractor at gene activation configuration $\boldsymbol{X}^*(A)$ to that of $\boldsymbol{X}^*(B)$, as discussed in Section 1.6.3.

1.6.2. Cancer cells occupy 'unused' attractor states in the 'adjacent possible'

If the physiological cell types are attractors of the GRN dynamics, then cancer cells are also attractors—an idea first proposed by Stuart

Kauffman in the 1970s [90–93]. In other words, cancer cells represent abnormal cell types 'not meant to be'. Based on random Boolean network models of GRNs and without notion of the quasi-potential landscape, Kauffman suggested that cancer cells occupy attractors of large GRNs that normally remain unused by the set of 'normal' cell types (**Figure 1.3**). Developing this idea further, we can now ask: *why are there unused attractors, and how do they become occupied during tumorigenesis?*

By folding up a quasi-potential landscape, the GRN generates a set of attractors (valleys) that become occupied by cells during development, implementing the stable cell types P encoded by the genome G. But as the GRN evolves to fold a topography capable of *at least* producing these physiological attractors, due to mathematical constraints in how a complex, evolved GRN wiring-diagram maps into a quasi-potential landscape, it also contains many more valleys than needed to serve as cell-type attractors. These unused 'surplus' attractors are profound features of the phenotypic repertoire of the GRN and represent the *potential* behaviours in the adjacent possible (pink regions in **Figure 1.3**) that are not realized in the actual organism—unless in pathological situations (**Figure 1.3**B).

1.6.3. Cell phenotype conversions (differentiation) are transitions between attractor states

If attractors are cell types, then development involves a series of stepwise cell phenotype switching that create branching cell lineages to increase the diversity of cell types, starting from immature multipotent embryonic cells to multi-potent stem cells to mature, differentiated cells of the adult tissue. This process involves transitions from cell type to cell type, which is known as differentiation [58,88] and illustrated in Figure 1.2H.

What is the mechanism for physiological attractor transitions? Since attractor states are stable states at the bottom of a basin, a perturbation by E sufficient to alter $\boldsymbol{X}$ (Section 1.4; Figure 1.2H) and cause transition into another attractor will have to 'push' cells uphill out of the basin of attraction across the high potential barriers (hills between valleys), and thus, altering $\boldsymbol{X}$ against a steep gradient $\Delta V(\boldsymbol{X})$ (Figure 1.2G, horizontal blue E arrow). Realizing an uphill movement on the landscape requires major regulatory action against homeostatic regulation: it would necessitate a highly coordinated (unlikely) change of activity in a large number of gene loci to alter the set of x_i precisely so as to shift $\boldsymbol{X}$ uphill along the least action path to the hilltop and across the separatrix.

Instead, a mathematical principle of dynamical systems offers a shortcut. We have proposed that, instead, signals that cause physiological transitions to new attractors (differentiation) may have evolved to alter the *quantitative regulatory characteristics* of interactions [70,85,93]. As described in Section 1.5.5, (a) they alter parameters in the differential equation that describes a particular regulatory interaction. In doing this, signals temporarily and locally distort the landscape topography so as to lower the energy barrier surrounding the attractor, i.e. flatten the basin of attraction. This change of topography allows cells to exit the attractor basin with much lower action against homeostatic regulatory constraints of the slightly modified GRN wiring diagram, perhaps even purely driven by molecular noise (see Figure 1.2G, vertical blue E arrow). It may suffice that the gradient up the hills of the basin of attraction is only reduced with respect to one of the N state space dimensions—a distortion that would correspond to forming a 'saddle' in the separatrix, akin to a mountain pass that facilitates passage between two valleys.

In other words, attractor transitions are preceded by *destabilization* ('flattening') of the current attractor [70]. The lowering of a potential energy barrier that separates two energy wells bears superficial (but warranted) correspondence to chemical catalysis that lowers the activation energy necessary for a reaction to occur—with the now familiar caveat that the quasi-potential landscape of the GRN is not a classical energy potential landscape of a conservative system.

1.6.4. Destabilization underlying attractor switching

The *destabilization* that precedes attractor transitions warrants more elaboration (Figure 1.2E). If destabilization proceeds to a *qualitative* change of the landscape, e.g. disappearance of attractors or appearance of new ones, the bounds of structural stability of a system are breached, which is called a *bifurcation* event. In one canonical type of bifurcation (fold bifurcation), the topography gradually flattens the basin until it abolishes an attractor such that the position of $\boldsymbol{X}^\star$ is suddenly on a slope and the cell *has* no choice than to descend into a particular neighbouring attractor that was previously not accessible due to the separating hill. A second major type of bifurcation (pitchfork bifurcation and its stable, higher-order subforms) converts a valley into a hilltop or a stable steady state into an unstable one, thus introducing a separatrix. This imparts, in a previously stable state, maximal instability to the cell and forces it to make a choice to descend to either one of the two attractors on either side—akin to creating a watershed (see **Figure 1.3**C). This topography change results in a cell being poised to undergo a binary fate decision to implement one of the two phenotypes encoded by the two newly accessible attractors.

Phenomenologically, in the study of complex system, when the underlying network structure is unknown, a bifurcation appears as a *critical transition* (**Figure 1.3**C): *a sudden system flip* from one stable equilibrium into another, characterized by a preceding destabilization of the system. The latter can be measured as 'critical slowing down' that precedes the bifurcation: the return of perturbed system states $\boldsymbol{X}'$ to the attractor state $\boldsymbol{X}^\star$ takes longer and longer time in a flattening basin of attraction [94,95]. In cell populations, which represent the ensemble of thousands of replicated of the GRN as a system, cell phenotype diversity increases. But counter-intuitively, the activities of a subset of gene loci (x_i, x_j, ...) across these cells tend to become more similar to each other as the 'effective dimensionality' of the state space decreases, as cells, while increasingly diverse, align in state space along a trajectory that leads to the saddle for exiting the attractor state. Thus, a subset of genes x_i, x_j, ... exhibits increased correlation, $|\mathrm{corr}(x_i, x_j)|$, across the ensemble of cells in an attractor. This property can now be measured in single-cell transcriptomics [70,96] to identify impending critical transitions.

Taking together, if cancers are unused attractors, then tumorigenesis is the accidental, undue departure of cells from attractors in the physiological domain of the landscape and the entry into the adjacent possible. Cancer cells are trapped in excess, physiologically inaccessible attractors that are not used for making and maintaining the healthy organism but are by-products of the evolution of phenotype control by the GRN. Before asking how cells enter these unused

attractors, and why they encode the immature, stem-like traits of malignancy, we first need to discuss another property of the quasi-potential landscape that is a natural consequence of its evolution.

1.7. The fundamental inevitability of cancer—explained using the quasi-potential landscape

We now place the landscape-based thinking in the context of organismal *evolution* to articulate the (in principle) fundamental inevitability of cancer as undesired realization of the adjacent possible.

1.7.1. Evolution in the landscape: chreods ensure smooth descent

Natural selection in the evolution of organismal traits through genomic mutations acts on the GRN wiring diagram and thereby shapes the topography of the landscape since $V(\boldsymbol{X})$ is a function of the architecture of the GRN. New regulatory genes that arise during genome expansion are connected to the existing GRN, which is continuously rewired by mutations that affect regulatory regions in non-coding and binding sites of proteins in coding sequences [76,97]. Cis–trans regions shuffling in the course of larger scale genomic rearrangement also alter the wiring diagram [98]. Through such GRN rewiring, evolution has shaped the landscape topography to ensure that the existence of a set of stable configurations $\boldsymbol{X}$ encodes functional cell phenotypes. The scheme of ontogenic development is thus to a great extent hardwired [76,97], epitomized by the landscape such that, crudely speaking, *global* gradient-like forces drive a series of cellular differentiation events, channelling cells from embryonic stem cell gene activation configurations to that of increasingly mature cells. (For the sake of focusing on first principles of cell phenotype plasticity, we leave out the cell–cell interaction that adds a layer of control to developmental dynamics.)

Importantly, when new attractors, hence new cell types, are created during evolution (in those rare events of mutational wiring of the GRN with qualitative consequence), these new attractor states must have a lower 'quasi-potential energy' than existing ones so that they can be reached with little effort from within the latter. Evolution selects for reachable attractors that can be accessed with little efforts from the existing, used ones. Hence, during the evolution of new cell types, the landscape must have 'grown downwards' by adding attractors at lower and lower elevation so that they could be occupied with ease and also would further evolve to serve organismal fitness [99,100]. This process carves the downhill trajectories of development into the landscape that affords the 'arrow of development' towards implementing all the physiological cell types in the metazoan body while avoiding being trapped inside valleys at higher altitude that encode, by necessity, less mature phenotypes (Figure 1.2F). This principle of robust developmental trajectories corresponds to Waddington's insightful idea of '*chreods*' (Greek for 'necessary path') [101] (green lines/arrows in Figures 1.2F and 1.3) [100].

Only those gene activation configurations $\boldsymbol{X}$ that produce attractor states that are normally occupied by cells are seen by evolution. These attractors are subjected to fine-tuning by ensuing natural selection that optimizes the phenotype of these cells encoded by these attractor states for building cell societies and improving the fitness of the entire organism.

Evolution may also have optimized the chreods for resilience against perturbations and stochastic fluctuations that could accidentally steer cells into the side valleys in the adjacent possible that have no physiological function and may even confer neoplastic activities: the cancer attractors. Thus, tumour suppressor genes, such as p53 or Rb, may have evolved to guarantee smooth descent along robust or 'buffered' chreods into physiological attractors of mature cell types by increasing the height of hills that separate them from going astray into the unused domains of the landscape where cells could get 'stuck' and remain undifferentiated. In fact, these genes also suppress stemness and promote differentiation.

1.7.2. Entering unused attractors in the adjacent possible

If destabilization of a physiological attractor, e.g. by mutations, affects an undifferentiated stem cell (which is naturally at a higher elevation and poised to descend to a mature attractor), then the healthy stem cell may unduly enter the adjacent possible domain [100]. It will be trapped in one of the many unoccupied attractors nearby and thus remain in an undifferentiated state (**Figure 1.3**A and B). It is unable to descent (and thus differentiate) because the gene activation configuration encoded by such attractors are not tuned for normal development; they have never been shaped by evolution to 'drain' into the valleys where chreods of normal development run. In this sense, pathologists consider cancer cells to be 'maturation arrested'. We can now apply the above concepts for attractor transitions (points (a) and (b) in Section 1.5.5) to cancer: distortion of the landscape topography that lowers a barrier and allows cells to enter can result from non-genetic and genetic events:

(a) *Chronic non-genetic perturbations* (Figure 1.2E): Changes in tissue environment conditions in response to stress and injury, which cause chronic inflammation [102], transiently alter the regulatory interactions of the GRN through cytokine and other mediators as discussed in point (a) in Section 1.5.5. Among the many possible ensuing distortions of the landscape, the most consequential for cell state dynamics is the lowering of barrier heights that prevent 'spilling' of cells into the adjacent possible. While these changes of the landscape may be transient, the enduring nature of chronic injury and inflammation increases over time the probability that perturbations or any random fluctuations in $\boldsymbol{X}$ cause entry, via a lowered barrier, into the abnormal attractor states. This mechanism would explain why non-specific chronic irritation (e.g. bronchitis associated with smoking) would consistently produce similar types of tumours in a given tissue.

(b) *Permanent rewiring of the GRN by a genetic mutation* (Figure 1.2H): Mutations in coding regions of regulatory genes and regulatory regions of the genome in essence permanently rewire the GRN, as first discussed in point (b) in Section 1.5.5. Mathematically, the consequences of an altered wiring diagram for the landscape topography are rather subtle due to structural stability. But occasionally a distortion of the landscape may open up access to unused attractors in the adjacent possible. However, unlike the transient non-genetic changes of the landscape, those caused by genetic mutations are magnified because they are permanent and inheritable to all descendants of a cell, thus eventually spreading through an entire cell population

and thereby increasing the probability of accidental entry into cancer attractors and selection.

In both cases, the new access to nearby attractors is in line with the ubiquitous increase of cell phenotype plasticity ('identity confusion') in cancer, manifest as 'dedifferentiation', heterogeneity of cell phenotypes (pleomorphism) and transdifferentiation into neighbouring lineages (lineage infidelity) with large basin of attraction—best known in the appearance of neuroendocrine cell types in many carcinoma [103,104] or the myeloid differentiation in lymphoid-type leukaemia [105].

Distortions of the landscape caused by mutations with tumorigenic consequence are those that decrease quasi-potential hills ('energy barrier') that separate the physiological from unused attractors in the adjacent possible domain of the repertoire of the GRN. Thus, we can now see that *mutations do not 'cause' cancer but rather 'catalyze' cancer.* This concept is consistent with the increasing empirical recognition that mutations are necessary but not sufficient for tumorigenesis [106].

1.8. Wrapping up: biological and clinical implications

The conceptual framework presented in this chapter posits that non-specific, undirected, and random factors, non-genetic or genetic, can (jointly) facilitate the exit of cells from physiological attractors and enter into unused attractors of the adjacent possible phenotype repertoire where they become trapped in the rugged, unevolved terrain of the landscape that lacks chreods carved by evolution. These excess attractors of the GRN can produce quite stable but non-physiological gene activation configurations $\boldsymbol{X}$ that encode programs of the immature, stem-like malignant cell that cannot revert back to the normal phenotype.

A fundamental question then remains to be answered: why do excess attractor states in the adjacent possible accessed by cells in response to non-specific triggers so consistently encode *specifically* a coherent stemness program? This question is linked to the evolution of robustness of normal tissues and remains to be investigated. It suffices to propose here the idea that the inherent proclivity to activate stemness, a cell program central to tissue regeneration, is deeply 'baked-in' the system that has evolved to ensure such robustness of normal tissues by homeostatic negative feedback regulation, which is represented by the slope of the basin of attraction. Stemness and regeneration are just the biological embodiment of the slopes in the basin of attraction of the higher-level 'tissue attractors' that restore the physiological tissue phenotype upon injury.

Here, we only consider the robustness of the realized phenotype as a *generic* concept and the possibility of veering into the unrealized adjacent possible of the quasi-potential landscape. This framework allows us to reduce the very phenomenon of cancer to first principles, namely that of dynamical systems, and can now be tested for its ability to explain elementary properties of cancer not readily accounted for by the current characterization of oncogenic molecular pathways. As examples, we address below two elementary, rarely asked biological questions on the cause and treatment of cancer by revisiting the two major challenges of existing paradigms discussed in Section 1.2.

1.8.1. Why do mutagens require (non-mutagenic) tumour promoter agents to produce tumours?

An old two-stage concept of tumorigenesis that goes back to seminal animal experiments of the 1940s (reviewed in [106]) posits that after application of a mutagenic carcinogen (the *initiator*) the mutated cells may stay dormant for extended time and sometimes even regress. They only develop a tumour after repeated application of a chemical that is not (necessarily) mutagenic but an 'irritant' (the *promoter*), such as a phorbol ester. Many non-mutagenic promoters of a broad chemical diversity have since been identified. Environmental factors and signals from the tissue environment, such as inflammation, also serve as promoters. In the case of phorbol ester that turns out to activate the protein kinase C signal transduction pathway, the result is the activation of a large set of many gene loci [107]. Thus, it can substantially shift the GRN state $\boldsymbol{X}$—the cell's position on the landscape.

Stimulation by tumour promoters can be considered the rate-limiting factor of tumour development. With the discovery of the ubiquitous presence of oncogenic mutations in normal tissues in the era of systematic analysis by next-generation sequencing, the initiator–promoter concept has seen a revival [106,108]. Sequencing of 'initiated but not promoted' tumour-free tissue has indeed revealed a large number of oncogenic mutations in non-cancerous tissues. This finding defies the principle of Darwinian somatic evolution as the central driver behind tumorigenesis that would have selected among the mutated cells a clone that would have grown into a tumour (accumulating more fitness-increasing mutations)—independent of a tumour promoter.

The initiator–promoter principle is consistent with the concepts discussed in this chapter, notably the idea that mutations do not 'cause' but 'catalyze' cancer: the mutations inflicted by the initiator may only minimally shift $\boldsymbol{X}^*$ and thus, qualitatively, not change the cell phenotype because of structural stability. But they would lower the quasi-potential barrier that separates normal attractors from the potentially malignant attractors. Without promoters, cells might still reside in a shallower physiological attractor, manifest in increase of non-genetic cell phenotype plasticity as abundantly revealed by single-cell transcriptomics (see also Figure 1.3D), although with increased probability to 'spill over' into a cancer attractor. Conversely, initiated microtumours sometimes spontaneously differentiate and regress [109], suggesting that their cells can still access the chreods leading to physiological attractors—but perhaps less efficiently.

1.8.2. Why does treatment backfire?

We can now also comprehend why, as presented in the opening of this chapter, any cytocidal treatment itself promotes the non-killed fraction of tumour cells to progress to a more malignant phenotype. These massive ('near-lethal') perturbations further destabilize the cancerous state in the surviving cells, $\boldsymbol{X}^*\boldsymbol{c}$, and may drive cells through a *bifurcation* that forces upon the cell a binary branchpoint decision between two possible cell fates: cell death or, if not, entry into an unused attractor that could encode immature, stemness traits (Figure 1.3C and D) [55]. While the first outcome is therapeutically intended, some cells in the heterogenous tumour cell population almost inevitably survive the maximally tolerated dose and

enter the alternative fate that becomes accessible after the destabilizing bifurcation.

These two opposite outcomes are preceded by instability of the cell state, to be imagined as placing the cells onto a mountain ridge (separatrix) where it is easily tipped to fall down the slope on either side. This symmetry-breaking bifurcation explains why cytocidal therapy is a potential double-edged sword: treatment stress poises cells at the bifurcation point to either die or become stem-like. The decision in individual cells is governed by cell state $\boldsymbol{X}'$ within the attractor basin at the time of response to the perturbation, chance fluctuations, and deterministic bias of other signals that can tilt the bifurcation. This principle explains the near-inevitable *induction* of the stem-like resistant phenotype in non-killed cells (in addition to Darwinian *selection* of such resistant cells that operate at a slower timescale). Single-cell genetic barcode-based tracing of individual cells currently seeks to measure the relative contribution of these two mechanisms [110–112]. The bifurcation dynamics is thus at the core of Nietzsche's dualism of either dying or, if not, coming out stronger after an adversary intervention.

Taken together, we may even postulate that current treatment that almost invariably leaves surviving cells and also substantially irritates the tumour tissue may also actually act as a promoter on the mutated residual tumour and thus actively contribute to relapse. Thus, the Nietzsche effect would suggest that treatment acts as a tumour promoter in the two-stage model.

1.9. Concluding remarks

We have sought to present an integrated, qualitative but logically coherent narrative, without resorting to mathematical equations, that provides a new type of conceptual explanation for the very existence of cancer and its progression. The novelty was to depart from the prevailing epistemological habit that tacitly assumes immutability of objects and thus reduces cancer to random genetic mutations and ensuing proximal causes, such as oncogenic pathways, as a convenient way to explain phenotype innovation. This scheme of thought leaves many questions open regarding the generic properties of cancer, such as the insufficiency of mutations as cause of cancer and the near obligate backfiring of cytocidal treatment. Instead, we present established and more recent formal concepts to explain the immanent inevitability of tumorigenesis based on first principles of systems dynamics in the context of evolution, development, and homeostasis of metazoan tissues. In order to present these principles with clarity within the limited space, and to emphasize coherent reasoning, we had to simplify and forgo comprehensive enumeration of specific molecular pathways.

We also refrained from discussing all the higher layers of organization in which the tissue as such, not the cell, is the system that changes during tumorigenesis. The tumour tissue, which includes the tumour microenvironment and the immune system, too represents a dynamical system, composed of the network not of genes but of cells, defined not only by their gene activation configuration $\boldsymbol{X}$ but also by their physical location in the tissue, that interact with each other via communication signals. Instead of cell attractor states produced by gene–gene regulatory interactions, we have tissue attractors produced by the cell–cell communication network that represents tissue level programs, such as angiogenesis, inflammation, regeneration, and immune activation or suppression.

We believe that the principles of such higher-level tissue dynamics with homeostatic tissue attractors may be similar to the dynamics of cell states driven by the GRN: the occupation of non-physiological tissue attractor states in the 'adjacent possible' of the quasi-potential landscape of tissue configurations, trapping tissues in abnormal configurations of cell activities with the vicious cycles of structural disorganization, and futile attempts of aberrant regenerative processes as source for undue environmental signals that shift cell states towards stemness. With the arrival of single-cell resolution of multiomic profiling of tissues (an example is shown in Figure 1.3D), including *in situ* cell states, there will be a need for a theoretical framework in the years to come to make sense of the vast amount of data that is continuing to accumulate. Hopefully, the reader can use the concepts presented in this chapter to do so.

REFERENCES

1. Dreyer JLE. Tycho Brahe: A picture of scientific life and work in the sixteenth century. Cambridge, UK:Cambridge University Press; 2014.
2. Camardi G. Charles Lyell and the uniformity principle. Biol Philos. 1999;**14**(4):537–60.
3. Darwin C. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. 5th ed. London: John Murray; 1869. p. 91–2.
4. Dear P. The intelligibility of nature: How science makes sense of the world. Chicago: University of Chicago Press; 2006.
5. Joshi CV, Enver T. Plasticity revisited. Curr Opin Cell Biol. 2002;**14**(6):749–55.
6. Bonfanti P, Barrandon Y, Cossu G. 'Hearts and bones': the ups and downs of 'plasticity' in stem cell biology. EMBO Mol Med. 2012;**4**(5):353–61.
7. Yamanaka S. Induction of pluripotent stem cells from mouse fibroblasts by four transcription factors. Cell Prolif. 2008;**41** Suppl 1:51–6.
8. Attolini CS, Michor F. Evolutionary theory of cancer. Ann N Y Acad Sci. 2009;**1168**:23–51.
9. Hanahan D. Hallmarks of cancer: New dimensions. Cancer Discov. 2022;**12**(1):31–46.
10. Huang S. Reconciling non-genetic plasticity with somatic evolution in cancer. Trends Cancer. 2021;**7**(4):309–22.
11. Björneborn L. Adjacent possible. The Palgrave encyclopedia of the possible. Cham: Springer International Publishing; 2020. p. 1–12.
12. Barrueto L, Caminero F, Cash L, Makris C, Lamichhane P, Deshmukh RR. Resistance to checkpoint inhibition in cancer immunotherapy. Transl Oncol. 2020;**13**(3):100738.
13. Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. Nat Rev Cancer. 2005;**5**(4):275–84.
14. Elliot A, Adams J, Al-Hajj M. The ABCs of cancer stem cell drug resistance. IDrugs. 2010;**13**(9):632–5.
15. Mellor HR, Callaghan R. Resistance to chemotherapy in cancer: a complex and integrated cellular response. Pharmacology. 2008;**81**(4):275–300.
16. Rosenzweig SA. Acquired resistance to drugs targeting receptor tyrosine kinases. Biochem Pharmacol. 2012;**83**(8):1041–8.
17. Zhang K, Erkan EP, Jamalzadeh S, Dai J, Andersson N, Kaipio K, et al. Longitudinal single-cell RNA-seq analysis reveals

stress-promoted chemoresistance in metastatic ovarian cancer. Sci Adv. 2022;**8**(8):eabm1831.
18. Prasad V, Fojo T, Brada M. Precision oncology: origins, optimism, and potential. Lancet Oncol. 2016;**17**(2):e81–6.
19. Grossmann N, Del Paggio JC, Wolf S, Sullivan R, Booth CM, Rosian K, et al. Five years of EMA-approved systemic cancer therapies for solid tumours—a comparison of two thresholds for meaningful clinical benefit. Eur J Cancer. 2017;**82**:66–71.
20. Fares J, Kanojia D, Rashidi A, Ulasov I, Lesniak MS. Landscape of combination therapy trials in breast cancer brain metastasis. Int J Cancer. 2020;**147**(7):1939–52.
21. Donnenberg VS, Donnenberg AD. Multiple drug resistance in cancer revisited: the cancer stem cell hypothesis. J Clin Pharmacol. 2005;**45**(8):872–7.
22. Eyler CE, Rich JN. Survival of the fittest: cancer stem cells in therapeutic resistance and angiogenesis. J Clin Oncol. 2008;**26**(17):2839–45.
23. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;**499**(7457):214–8.
24. Ma CX, Ellis MJ. The Cancer Genome Atlas: clinical applications for breast cancer. Oncology. 2013;**27**(12):1263–9, 74–9.
25. Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, Marioni JC, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. Proc Natl Acad Sci U S A. 2013;**110**(10):4009–14.
26. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature. 2010;**467**(7319):1114–7.
27. Notta F, Mullighan CG, Wang JC, Poeppl A, Doulatov S, Phillips LA, et al. Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. Nature. 2011;**469**(7330):362–7.
28. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature. 2014;**506**(7488):328–33.
29. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;**481**(7382):506–10.
30. Morrissy AS, Garzia L, Shih DJ, Zuyderduyn S, Huang X, Skowron P, et al. Divergent clonal selection dominates medulloblastoma at recurrence. Nature. 2016;**529**(7586):351–7.
31. Baker SG, Kramer BS. Paradoxes in carcinogenesis: new opportunities for research directions. BMC Cancer. 2007;**7**:151.
32. Salgia R, Kulkarni P. The genetic/non-genetic duality of drug 'resistance' in cancer. Trends Cancer. 2018;**4**(2):110–8.
33. Bell CC, Gilan O. Principles and mechanisms of non-genetic resistance in cancer. Br J Cancer. 2020;**122**(4):465–72.
34. Noble D. Cellular Darwinism: Regulatory networks, stochasticity, and selection in cancer development. Prog Biophys Mol Biol. 2021;**165**(Oct):66–71.
35. Satgé D. Analysis of somatic mutations in cancer tissues challenges the somatic mutation theory of cancer. eLS. 2013 (September 2013).
36. Sonnenschein C, Soto AM. Over a century of cancer research: inconvenient truths and promising leads. PLoS Biol. 2020;**18**(4):e3000670.
37. Brock A, Chang H, Huang S. Non-genetic heterogeneity—a mutation-independent driving force for the somatic evolution of tumours. Nat Rev Genet. 2009;**10**(5):336–42.
38. Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, et al. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proc Natl Acad Sci U S A. 2011;**108**(19):7950–5.
39. Miles LA, Bowman RL, Merlinsky TR, Csete IS, Ooi AT, Durruthy-Durruthy R, et al. Single-cell mutation analysis of clonal evolution in myeloid malignancies. Nature. 2020;**587**(7834):477–82.
40. Nam AS, Kim KT, Chaligne R, Izzo F, Ang C, Taylor J, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. Nature. 2019;**571**(7765):355–60.
41. Velten L, Story BA, Hernandez-Malmierca P, Raffel S, Leonce DR, Milbank J, et al. Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics. Nat Commun. 2021;**12**(1):1366.
42. Williams N, Lee J, Mitchell E, Moore L, Baxter EJ, Hewinson J, et al. Life histories of myeloproliferative neoplasms inferred from phylogenies. Nature. 2022;**602**(7895):162–8.
43. Lee SE, Kang SY, Yoo HY, Kim SJ, Kim WS, Ko YH. Clonal relationships in recurrent B-cell lymphomas. Oncotarget. 2016;**7**(11):12359–71.
44. Oren Y, Tsabar M, Cuoco MS, Amir-Zilberstein L, Cabanos HF, Hutter JC, et al. Cycling cancer persister cells arise from lineages with distinct programs. Nature. 2021;**596**(7873):576–82.
45. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature. 2008;**453**(7194):544–7.
46. Pisco AO, Brock A, Zhou J, Moor A, Mojtahedi M, Jackson D, et al. Non-Darwinian dynamics in therapy induced cancer drug resistance. Nat Commun. 2013;**4**:2467.
47. Huang S. Tumor progression: chance and necessity in Darwinian and Lamarckian somatic (mutationless) evolution. Prog Biophys Mol Biol. 2012;**110**(1)::69–86.
48. Greene JM, Gevertz JL, Sontag ED. Mathematical approach to differentiate spontaneous and induced evolution to drug resistance during cancer treatment. JCO Clin Cancer Inform. 2019;**3**:1–20.
49. Chaudhary PM, Roninson IB. Induction of multidrug resistance in human cells by transient exposure to different chemotherapeutic drugs. J Natl Cancer Inst. 1993;**85**(8):632–9.
50. Boyd AL, Aslostovar L, Reid J, Ye W, Tanasijevic B, Porras DP, et al. Identification of chemotherapy-induced leukemic-regenerating cells reveals a transient vulnerability of human AML recurrence. Cancer Cell. 2018;**34**(3):483–98 e5.
51. Lagadec C, Vlashi E, Della Donna L, Dekmezian C, Pajonk F. Radiation-induced reprogramming of breast cancer cells. Stem Cells. 2012;**30**(5):833–44.
52. Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, et al. Rare cell variability and drug induced reprogramming as a mode of cancer drug resistance. Nature. 2017;546(7658):431–5.
53. Su Y, Wei W, Robert L, Xue M, Tsoi J, Garcia-Diaz A, et al. Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. Proc Natl Acad Sci U S A. 2017;**114**(52):13679–84.
54. Brock A, Huang S. Precision Oncology: between vaguely right and precisely wrong. Cancer Res. 2017;**77**(23):6473–9.
55. Huang S. The logic of cancer treatment: Why it is so hard to cure cancer? Treatment-induced progression, hyper-progression and the Nietzsche effect. In: Strauss B, Bertolaso M, Ernberg I, Bissell MJ, editors. Rethinking cancer: A new paradigm for the post-genomics era (Vienna Series in Theoretical Biology). Cambridge, MA: MIT Press; 2021. p. 456.
56. Nietzsche F. Twilight of the idols. Oxford: Oxford University Press; 1998. 124 p.

57. Huang S. Systems biology of stem cells: three useful perspectives to help overcome the paradigm of linear pathways. Philos Trans R Soc Lond B Biol Sci. 2011;**366**(1575):2247–59.
58. Huang S. The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology. Bioessays. 2012;**34**(2):149–55.
59. Huang S. Towards a unification of the 2 meanings of 'epigenetics'. PLoS Biol. 2022;**20**(12):e3001944.
60. Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. Nat Rev Mol Cell Biol. 2023.
61. Davidson EH. Regulatory genome: gene regulatory networks in development and evolution. San Diego: Academic Press; 2006.
62. Blake WJ, Balazsi G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, et al. Phenotypic consequences of promoter-mediated transcriptional noise. Mol Cell. 2006;**24**(6):853–65.
63. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. Development. 2009;**136**(23):3853–62.
64. Jolly MK, Kulkarni P, Weninger K, Orban J, Levine H. Phenotypic plasticity, bet-hedging, and androgen independence in prostate cancer: role of non-genetic heterogeneity. Front Oncol. 2018;**8**:50.
65. Kulkarni P, Bhattacharya S, Achuthan S, Behal A, Jolly MK, Kotnala S, et al. Intrinsically disordered proteins: critical components of the wetware. Chem Rev. 2022;**122**(6):6614–33.
66. Dekker J, Mirny L. The 3D genome as moderator of chromosomal communication. Cell. 2016;**164**(6):1110–21.
67. Vendramin R, Litchfield K, Swanton C. Cancer evolution: Darwin and beyond. EMBO J. 2021;**40**(18):e108389.
68. Shen S, Clairambault J. Cell plasticity in cancer cell populations. F1000Research. 2020;9.
69. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012;**12**(5):323–34.
70. Mojtahedi M, Skupin A, Zhou J, Castano IG, Leong-Quong RY, Chang H, et al. Cell fate decision as high-dimensional critical state transition. PLoS Biol. 2016;**14**(12):e2000640.
71. Zhou JX, Aliyu MD, Aurell E, Huang S. Quasi-potential landscape in complex multi-stable systems. J R Soc Interface. 2012;**9**(77):3539–53.
72. Waddington CH. The strategy of the genes. London: Allen and Unwin; 1957.
73. Freidlin M, Wentzell A. Random perturbations of dynamical system. New York: Springer-Verlag; 1984.
74. Kauffman SA. The origins of order. New York: Oxford University Press; 1993.
75. Rodriguez-Sanchez P, van Nes EH, Scheffer M. Climbing Escher's stairs: a way to approximate stability landscapes in multidimensional systems. PLoS Comput Biol. 2020;**16**(4):e1007788.
76. Aldana M, Balleza E, Kauffman S, Resendiz O. Robustness and evolvability in genetic regulatory networks. J Theor Biol. 2007;**245**(3):433–48.
77. Balleza E, Alvarez-Buylla ER, Chaos A, Kauffman A, Shmulevich I, Aldana M. Critical dynamics in genetic regulatory networks: examples from four kingdoms. PLoS ONE. 2008;**3**:e2456.
78. Shmulevich I, Kauffman SA, Aldana M. Eukaryotic cells are dynamically ordered or critical but not chaotic. Proc Natl Acad Sci U S A. 2005;**102**(38):13439–44.
79. Zhou P, Li T. Construction of the landscape for multi-stable systems: Potential landscape, quasipotential, A-type integral and beyond. J Chem Phys. 2016;**144**(9):094109.
80. Wang J, Xu L, Wang E, Huang S. The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation. Biophys J. 2010;**99**(1):29–39.
81. Saez M, Briscoe J, Rand DA. Dynamical landscapes of cell fate decisions. Interface Focus. 2022;**12**(4):20220002.
82. Zhou JX, Huang S. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. Trends Genet. 2011;**27**(2):55–62
83. Duddu AS, Sahoo S, Hati S, Jhunjhunwala S, Jolly MK. Multi-stability in cellular differentiation enabled by a network of three mutually repressing master regulators. J R Soc Interface. 2020;**17**(170):20200631.
84. Bagley RJ, Glass L. Counting and classifying attractors in high dimensional dynamical systems. J Theor Biol. 1996;**183**(3):269–84.
85. Huang S, Guo YP, May G, Enver T. Bifurcation dynamics of cell fate decision in bipotent progenitor cells. Dev Biol. 2007;**305**(2):695–713.
86. Ferrell JE, Jr. Bistability, bifurcations, and Waddington's epigenetic landscape. Curr Biol: CB. 2012;**22**(11):R458–66.
87. Thom R. Structural Stability And Morphogenesis: CRC Press; 2018.
88. Huang S. Reprogramming cell fates: reconciling rarity with robustness. Bioessays. 2009;**31**(5):546–60.
89. Kauffman S. Homeostasis and differentiation in random genetic control networks. Nature. 1969;**224**(215):177–8.
90. Kauffman S. Differentiation of malignant to benign cells. J Theor Biol. 1971;**31**(3):429–51.
91. Huang S, Ernberg I, Kauffman S. Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective. Semin Cell Dev Biol. 2009;**20**(7.):869–76.
92. Huang S. On the intrinsic inevitability of cancer: From foetal to fatal attraction. Semin Cancer Biol. 2011;**21**(3):183–99.
93. Huang S, Kauffman S. How to escape the cancer attractor: rationale and limitations of multi-target drugs. Semin Cancer Biol. 2013;**23**(4):270–8.
94. Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, et al. Anticipating critical transitions. Science. 2012;**338**(6105):344–8.
95. Trefois C, Antony PM, Goncalves J, Skupin A, Balling R. Critical transitions in chronic disease: transferring concepts from ecology to systems medicine. Curr Opin Biotechnol. 2015;**34**:48–55.
96. Bargaje R, Trachana K, Shelton MN, McGinnis CS, Zhou JX, Chadick C, et al. Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells. Proc Natl Acad Sci U S A. 2017;**114**(9):2271–6.
97. Teichmann SA, Babu MM. Gene regulatory network growth by duplication. Nat Genet. 2004;**36**(5):492–6.
98. Cordero OX, Hogeweg P. Feed-forward loop circuits as a side effect of genome evolution. Mol Biol Evol. 2006;**23**(10):1931–6.
99. Torres-Sosa C, Huang S, Aldana M. Criticality is an emergent property of genetic networks that exhibit evolvability. PLoS Comput Biol. 2012;**8**(9):e1002669.
100. Huang S. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. Cancer Metastasis Rev. 2013;**32**(3–4):423–48.
101. Gare A. Chreods, homeorhesis and biofields: finding the right path for science through Daoism. Prog Biophys Mol Biol. 2017;**131**:61–91.
102. Afify SM, Hassan G, Seno A, Seno M. Cancer-inducing niche: the force of chronic inflammation. Br J Cancer. 2022;**127**(2):193–201.

103. Stovold R, Meredith SL, Bryant JL, Babur M, Williams KJ, Dean EJ, et al. Neuroendocrine and epithelial phenotypes in small-cell lung cancer: implications for metastasis and survival in patients. Br J Cancer. 2013;**108**(8):1704–11.
104. Zou M, Toivanen R, Mitrofanova A, Floch N, Hayati S, Sun Y, et al. Transdifferentiation as a mechanism of treatment resistance in a mouse model of castration-resistant prostate cancer. Cancer Discov. 2017;**7**(7):736–49.
105. Bueno-Costa A, Pineyro D, Soler M, Javierre BM, Raurell-Vila H, Subirana-Granes M, et al. B-cell leukemia transdifferentiation to macrophage involves reconfiguration of DNA methylation for long-range regulation. Leukemia. 2020;**34**(4):1158–62.
106. Balmain A. The critical roles of somatic mutations and environmental tumor-promoting agents in cancer risk. Nat Genet. 2020;**52**(11):1139–43.
107. Angel P, Imagawa M, Chiu R, Stein B, Imbra RJ, Rahmsdorf HJ, et al. Phorbol ester-inducible genes contain a common cis element recognized by a TPA-modulated trans-acting factor. Cell. 1987;**49**(6):729–39.
108. Martincorena I. Somatic mutation and clonal expansions in human tissues. Genome Med. 2019;**11**(1):35.
109. Helfrich I, Chen M, Schmidt R, Furstenberger G, Kopp-Schneider A, Trick D, et al. Increased incidence of squamous cell carcinomas in Mastomys natalensis papillomavirus E6 transgenic mice during two-stage skin carcinogenesis. J Virol. 2004;**78**(9):4797–805.
110. Oren Y, Tsabar M, Cuoco MS, Amir-Zilberstein L, Cabanos HF, Hutter JC, et al. Cycling cancer persister cells arise from lineages with distinct programs. Nature. 2021; **596**(7873):576–82.
111. Simeonov KP, Byrns CN, Clark ML, Norgard RJ, Martin B, Stanger BZ, et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. Cancer Cell. 2021;**39**(8):1150–62 e9.
112. Gutierrez C, Al'Khafaji AM, Brenner E, Johnson KE, Gohil SH, Lin Z, et al. Multifunctional barcoding with ClonMapper enables high-resolution study of clonal dynamics during tumor evolution and treatment. Nat Cancer. 2021;**2**(7):758–72.

2

Non-genetic intratumoral heterogeneity and phenotypic plasticity as consequences of microenvironment-driven epigenomic dysregulation

Vera Pancaldi and Jean-Pascal Capp

2.1. Introduction

Cancer cells are now considered destabilized cells in which, besides genomic instability, a global increase in entropy is observed at different levels (e.g. in protein interaction networks [1,2] or in gene regulatory networks [3]). This increased cellular entropy is somehow related to the global increase in cellular stochasticity, especially at the epigenomic level (see, for instance, 'The stochastic epigenetic model of human cancer' in Feinberg [4]), and to increased cellular plasticity. In the past decade, the epigenome has been in the spotlight as a major determinant of malignant phenotypes that would be characterized by enhanced plasticity, conferring cancer cells stronger adaptability in changing environments and under therapy, ultimately leading to therapeutic resistance. Also, the less organized and less stable chromatin structure observed in cancer cells appears to increase stochasticity of gene expression at higher levels than in normal stem cells [5].

While the origins of the global cellular destabilization remain largely unknown, it seems to correspond to a de-differentiation or a non-differentiation process where cells return or remain in a state of entropy close to the ones of stem or progenitor cells (depending on their level of differentiation: cancer stem cell or non-stem cell) [6]. Interestingly, early studies identified that cancer cells revert to a 'pseudo-primitive' epigenetic status where features of embryonic stem cells and different developing lineages are observed [7]. Cells designated as 'cancer stem cells' have the lowest degree of differentiation, and from these cancer stem cells, cancer cells harbour a continuum of states with 'noisy' and skewed differentiation [8]. Moreover, stochastic transitions between alternate cellular phenotypic states driven by non-genetic mechanisms have been known for a decade [9] as a common and relevant process in tumours that shape intratumoral heterogeneity in conjunction with genetics and the tumour microenvironment. These phenotypic transitions could contribute to cancer cells' enhanced proliferation [10].

Dynamic heterogeneity and enhanced plasticity in cancer are commonly associated with the notion of cancer attractor state. In this framework, cancer cells would be able to explore the global regulatory network in an unusual way thanks to a reconfiguration of the epigenetic landscape [11,12]. This epigenetic landscape reconfiguration in cancer would allow cells to acquire gene expression patterns that are not accessible in normal tissues. Indeed, cancer cells lose the defined epigenomic hierarchy and lineage identity observed in normal tissues as exemplified in single chronic lymphocytic leukaemia cells, in which mutually exclusive activating and repressing histone modifications co-localize and normally exclusive phenotypic markers co-occur [13].

An interesting question is whether genetic alterations of chromatin modifiers are required for reshaping epigenomic plasticity or if fluctuations in their expression levels can already modulate phenotypes, especially in the light of recent findings that highlight a weak correlation between genetic and phenotypic heterogeneity levels in colorectal tumours [14,15]. While traditionally therapy resistance has been associated with enhanced plasticity of cellular phenotypes, recent experiments on patient-derived xenograft models in which cancer clones can be followed propose an alternative 'phenotypic inertia' principle as an almost diametrically opposed explanation for the selection of clones resistant to stress and therapies [16]. Starting from the observation that epigenomic disruption through mutations in histone genes and chromatin modifiers is ubiquitous across many cancers, the authors identify an effect of these mutations in disrupting the correct activation of stress responses, which normally limits proliferation, allowing these mutant clones to outcompete others. Instead of exploiting an improved capacity to adapt to their environment, these cells would be faulty in responding to external

cues, such as stress conditions, due to their inability to correctly rearrange the epigenome to achieve cell cycle arrest. This mechanism could explain the prevalence of chromatin modifier alterations in dominating cancer clones, but it might also be the key to understanding the 'phenotypic inertia' that might lead pre-cancerous cells to ignore their environmental context, by failing to rearrange their epigenome accordingly.

Classical oncogenic mutations, which have been traditionally implicated as causative mechanisms in oncogenesis, are unlikely sources for such epigenomic destabilization and enhanced plasticity. It is now accepted that healthy tissues can contain large percentages of cells carrying mutations in 'cancer-associated genes' in the absence of uncontrolled proliferation, to an extent that increases with the individual's age [17,18]. If genetic mutations are unlikely to be the sole determinants of phenotypic characteristics and plasticity, alternative hypotheses are needed to account for the global loss of chromatin and epigenome organization seen in cancer. We can envisage multiple factors that could maintain these cells under control, especially at the epigenetic level, and prevent cancer development, including interactions with cells in the surrounding tissues, such as immune cells, fibroblasts, and cytokines in the tumour microenvironment, which have recently been implicated in the cellular de-differentiation leading to pollution-induced lung cancer [19]. We propose that the interplay between microenvironmental cues and nuclear structuring during differentiation could constitute a relevant starting point for an alternative view on cancer cell destabilization.

2.2. Nuclear organization and cancer

During development, a nuclear structure specific to each cell type is established in differentiated cells [20]. Each chromosome is organized in the nucleus with a particular configuration defining 'territories' and determining the position of genes in the three-dimensional (3D) nuclear space. This core organization is therefore extremely structured, but it is also highly dynamic. Expressed genes tend to be co-localized, even if they are on different chromosomes or far apart on the same chromosome, in 'transcription factories' that concentrate the factors involved in gene expression [21]. But these structures are very flexible and can constantly be restructured and remade elsewhere with other genes, depending on the molecules that associate and dissociate [22]. Since these association–dissociation events are highly random, they would contribute to making changes in gene expression random. Recent findings on single-cell variability in genome organization confirm that the stochastic nature of chromatin organization is related to transcription dynamics [23], but the association between structure and gene regulation remains controversial, despite the availability of multiple orthogonal experimental techniques to assess it [24].

Some researchers, including Tom Misteli, have proposed that this structure self-organizes spontaneously during differentiation, without external influence, so as to find a state of maximum thermodynamic equilibrium through random fluctuations of local interactions between molecules of the chromatin [20]. This model would make it possible to reconcile the apparently contradictory facts that are the apparent structuring of the nucleus and its ability to reorganize itself in a very dynamic way. But it is also possible to link this nuclear structuring with external constraints since chromatin molecules and what might seem random gene expression are permanently affected by external signals from the cellular environment. A recent publication was indeed able to quantitatively describe how cells respond to signals in specific ways which are dependent on their internal state and prior states as well as on the signal itself [25].

Several types of microenvironmental factors are known to play on the 3D epigenome such as mechanical constraints that, through mechanotransduction, can affect chromatin and differentiation state. Advanced techniques for mechanical manipulation of cells accompanied by imaging of the resulting forces have unveiled sophisticated mechanisms of cellular response to physical stresses [26,27]. Cellular phenotypes, such as proliferation and invasiveness, have been related to the stiffness of the extra-cellular matrix via the coupling between mechanical force sensing and nuclear localization of transcription co-factors. One such factor is Yes1 associated transcriptional regulator (YAP), an oncogenic transcriptional co-activator responsible for the loss of contact inhibition typical of cancers [28].

It is likely that mechanical forces can also impact the global organization of chromatin in the nucleus. This could be via changes in the nuclear localization of other factors that have an impact on chromatin structure or through a much faster direct effect of the nuclear forces on chromatin compaction. Subjecting cells to pulsed forces produces temporary chromatin compaction changes that involve euchromatic regions preferentially, suggesting a possible coupling between mechanical stimuli and gene regulation [29]. Research in mesenchymal stem cells has shown that purely dynamical perturbations of cells in the absence of chemical factors can promote chromatin compaction changes that lead to differentiation. More specifically, it was shown that the dynamic loading of cells produces nuclear reorganization involving chromatin condensation mediated by actomyosin-based contractility. Repeated cycles of loading led to stronger chromatin condensation and the establishment of a mechanical memory in the cells, which seems to be encoded by altered activity of histone modifiers [30]. This mechanically induced chromatin compaction was found to be dependent on EZH2, a histone methyltransferase that is involved in gene silencing as a component of the Polycomb repressive complex [31]. Mechanical factors are now recognized as regulators of nuclear properties in stem cells, and properties of the extra-cellular matrix directly impact cellular fate [32]. Mechanical forces are now understood to also regulate shuttling of factors from the cytoplasm to the nucleus and back, with clear implications of a direct connection between mechanical processes and gene transcription regulation [33]. Interestingly, there does seem to be an impact of mechanical constraints on cell proliferation with relevance to the emergence of non-proliferating but more resistant clones in cancer [34].

The 3D nuclear organization and its stabilization would therefore crucially depend on the cell's external constraints. It would rather be a phenomenon of 'hetero-organization' because it stems from a permanent interaction between random fluctuations within the nucleus and external signals tending to stabilize it [35]. In that case, nuclear organization would therefore be the result of the canalization of cells towards a cell type by cell signalling that stabilizes certain chromatin conformations according to the genes that these conformations allow to be expressed. For instance, the influence of the interactions with the cellular microenvironment on nuclear organization during the differentiation of mammary cells was shown [36]. The position

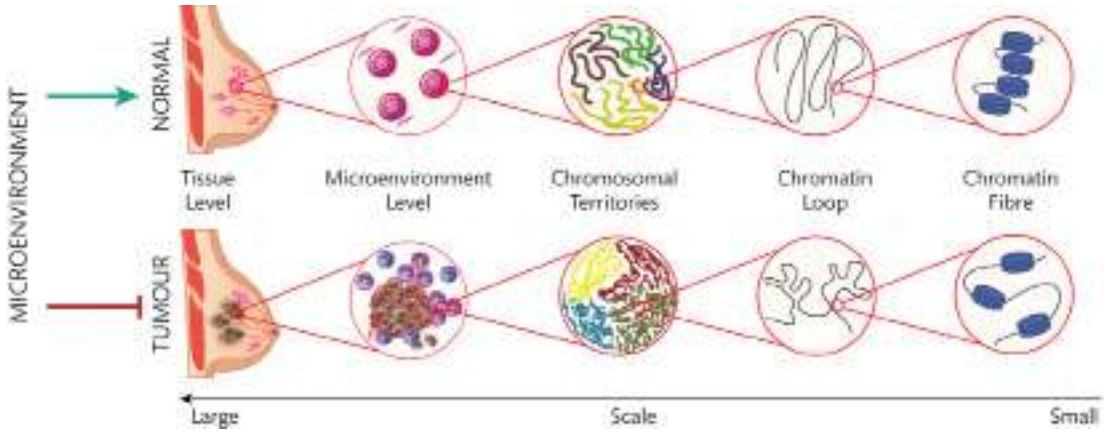


Figure 2.1. Microenvironmental disruption generates alterations of the 3D epigenome and chromatin structure. In a healthy context, normal microenvironmental cues allow a progressive canalization of cells towards differentiation that is characterized by well-defined chromosomal territories and chromatin loops (collectively called the 3D epigenome), as well as a more compacted chromatin at the genome scale. In a tumoral context, microenvironmental disruptions, such as a more destructured extra-cellular matrix and aberrant cell–cell interactions and communication, lead to a more diffused 3D epigenome at different scales and a more open chromatin. These phenomena collectively lead to increased cell stochasticity and entropy.

of certain genes involved in lactation is not the same in the cells of the intact tissue as the one observed in the cells cultured *in vitro* in monolayer [36]. The absence of tissue organization and signalling resulting from the cellular interactions in this 3D structure is probably the cause of this difference in nuclear organization. It is clear that these interactions strongly contribute to the establishment of the nuclear structure. But could their disruption be at the origin of the global 3D epigenome and nuclear organization destabilization in cancer?

Nuclear structure disruption in cancer has become a major theme in molecular oncology, either caused by somatic genomic rearrangements [37–39], with surprisingly little consequence on expression, or by chromatin alterations [40]. It has long been known that cancer cells have a very different nuclear organization from that of differentiated cells [41]. This is also exploited by pathologists when diagnosing samples [42]. Detailed studies have shown that many genes are 'repositioned' in cancer cells compared to normal cells, particularly in the early stages of carcinogenesis [43]. However, this repositioning does not affect the expression level of most of the genes tested [43], at least in these early stages and at the level of bulk population data.

This early nuclear destructuring in cancer cells is hardly conceivable in the self-organization model. How to explain that these early stages of cancer generate such destructuring when they are only supposed to be caused at most by a few genetic mutations? Are these mutations alone capable of disrupting this self-organization which is the result of a balanced and very complex process? On the contrary, if cellular interactions make it possible to gradually set up a nuclear organization characteristic of differentiated cell type, their disruption can certainly modify this organization. It would be therefore logical that this disorganization would be one of the characteristics of the early stages of cancer. Thus, we propose to consider the disrupted 3D epigenome as an important contributor in the production of cellular entropy and stochasticity and to envisage the origins of this disruption of the 3D epigenome at the extra-cellular microenvironmental level. Microenvironmental alterations would first affect the 3D epigenome and the nuclear organization that would become less constrained and more flexible and dynamic at the single-cell level, and more heterogeneous from cell to cell (Figure 2.1). This tissue disruption would consequently make cancer cells plastic and their differentiation state more reversible [44]. Those phenomena would thus not rely on broken regulatory links arising from mutations, even if the role of mutations and their 'promoting' effect must not be denied [44]. Indeed, evidence for the involvement of the cell's environment in establishing alternative epigenomic programs potentially leading to cancer was seen in a pancreatic cancer mouse model, in which the combination of tissue injury in addition to the common oncogenic mutation KRAS gave rise to a unique chromatin state that kick-starts the neoplastic transformation [45].

The early nuclear destructuring could have no particular 'functional significance': it might only be the result of the disruption of cellular interactions that normally maintain this structure in differentiated tissues. But it would directly generate a loss of epigenetic stability with an increased epigenomic stochasticity and, consequently, an increased cellular entropy driven by more variable and widespread gene expression patterns. The increased epigenetic stochasticity in cancer was previously proposed to be mediated by large-scale changes in DNA methylation and chromatin in domains associated with the nuclear lamina [4].

2.3. Clues from cellular reprogramming

Other indications on the interactions between the epigenome and the increased cellular stochasticity and their potential role in tumorigenesis come from the cell reprogramming process. This process corresponds to the reacquisition of pluripotency in fully differentiated cells achieved by a small subset of cells through stochastic transitions by ectopic expression of a set of transcription factors (TFs), known as OSKM [46] that are able to reconstitute a more open chromatin configuration. The cells produced are called induced pluripotent stem cells. These TFs through their binding of specific genome regions ensure the correct maintenance of chromatin interactions between enhancers and promoters [47]. Years after these first

studies, it was found that much more efficient and deterministic reprogramming of pluripotency could be achieved by expression of the OSKM factors combined with repression of Mbd3 [48], a central member in a nucleosome remodelling and deacetylation complex, suggesting that reprogramming could be controlled and is not purely a stochastic phenomenon. In reprogramming, OSKM factor binding can alter topologically associating domains (TADs), regions of chromosomal association which display high frequencies of physical interaction within a given domain but lower frequencies outside of these domains [49], which were identified as fundamental units of the 3D nuclear organization with compartments and chromatin loops [50,51]. The regions bordering TADs, called TAD boundaries, contribute to gene expression regulation by restricting interactions of cis-regulatory sequences to their target genes. OSKM factors can create new boundaries, strengthen them [52], and modify specific 3D enhancer–promoter contacts [47]. These chromatin alterations could affect enhancer–promoter interactions and impact gene expression.

Two mechanisms guard the genome against spurious regain of pluripotency, namely DNA methylation and the constitutive heterochromatin histone marks. The maintenance of these marks can depend on external signalling that can trigger a phenotypic shift known as epithelial to mesenchymal transition (EMT), which is often observed in initial phases of oncogenic transformation but was also found to precede reprogramming [53,54]. Thus, cytokine regulation in the microenvironment, likely due to immune and other stromal cells, has the potential to push the cells to lose their lineage identity.

Consistent with this idea, chromatin modifiers are very often mutated in cancer, and a recent pioneering study of single clones in colorectal cancer has detected positive selection for mutations in chromatin-modifying factors [15]. Moreover, chromatin accessibility alterations, which reflect changes in 3D organization and promoter–enhancer networks, were found to be stably inherited in the clonal populations and specific to malignant samples but not found in pre-neoplastic lesions [15]. Interestingly, these chromatin states were often associated with changes in gene expression, suggesting that they could more strongly impact phenotypes compared to mutations. In line with the concepts mentioned above, an integrated analysis of TF-binding motifs present in regions with altered DNA accessibility implicated the activation of EMT and developmental programs. The clonal and stable characteristics of the detected chromatin alterations suggests that they could be subject of Darwinian selection, contrary to gene expression alterations that could be just stochastic phenotypic fluctuations.

The study recorded a high number of somatic chromatin accessibility alterations also involving genes traditionally associated with colorectal cancer but not mutated in the specific samples. Interestingly, the authors were able to confirm that a majority of these alterations were clonal, suggesting their importance in the development of the tumour. Moreover, since the samples included pre-neoplastic adenomas as well as developed carcinoma, it was possible to note that contrary to the number of cancer-associated mutations, the number of chromatin alterations was increased after the onset of malignancy and many of them were associated with changes in gene expression. A specific TF motif accessibility analysis showed the reduction of interferon signalling and activation of EMT programs as well as re-activation of developmental programs via Hox genes, which was also confirmed by specific hypomethylation of developmentally regulated promoters that are stably methylated and repressed in normal samples.

During reprogramming chromatin that is closed in differentiated cells regains accessibility while regions specific for the original differentiated cell type are closed. It is not fully understood to which extent the malignant transformation resembles reprogramming, but it seems logical that what might be happening in oncogenesis is a very stochastic form of de-differentiation that will generate heterogeneous phenotypes in each cell, with increased stochasticity. Moreover, specific chromatin factors such as KDM5 have been implicated in therapy resistance via their impact on increasing phenotypic heterogeneity, essentially allowing a bet-hedging strategy to be developed in the tumour [55]. The up-regulation of lysine demethylases has also been associated with resistance in glioma cancer stem cells [56].

2.4. Towards an atavistic viewpoint on cancer

A complex interplay of external environmental cues modulating the activity of TFs can drive the establishment of specific transcriptional programs and reshaping of the 1D and 3D epigenome, without necessarily stemming from genetic alterations, which can ultimately increase the plasticity of the cells' phenotypes. In cancer, this could occur in cells that have lost coordination from interactions at the tissue level. They would then adopt a unicellular lifestyle, based on stochasticity, in which bet-hedging strategies aiming at maximizing the exploitation of nutritional resources dominate [57] and in which stress response elicits a diversification of phenotypes [58].

Davies and Lineweaver's atavistic idea from 2011 proposed to invoke that an ancestral program would be accidentally re-activated in malignant cells to explain cancer atavistic behaviour [59,60]. It is also possible to propose that cancer cells resemble unicellular cells in the way they generate specialized cells with optimal capacity to exploit nutritional resources for proliferation through a bet-hedging strategy based on cellular stochasticity, while keeping a subpopulation in a more plastic state [57,61,62]. Thus, selection would retain malignant cells that adopt a microbial cell lifestyle with highly stochastic phenotypic fluctuations (because of the lack of normal environmental constraints).

Of note, recent studies on colorectal cancers showed that the majority of intratumour variation in gene expression is not strongly heritable but rather plastic [14]. These data argue in favour of a mostly non-genetic origin for variation in tumoral phenotypes, contributing to intratumoral heterogeneity, confirmed by the observed lack of sub-clonal selection of putative genetic drivers. Thus, at least in these cases, most genetic variation seems to have no major phenotypic consequences, and transcriptional plasticity might, instead, be a major contributor in cancer progression [14]. Also, it was recently shown that disrupted epigenetic regulatory networks are frequently selected in evolving human cancers and do not promote the selection of genetically defined subclones [16]. This disruption rather provokes an inertia in the modulation of global transcriptional activity at the single-cell level that makes cells unable to respond rapidly to stress by correctly inducing cell cycle arrest. They then are more prone to acquire resistance to environmental stress through

long-term adaptation at the population level [16]. These results are all in accordance with a model of cancer progression in which loss of transcriptional coordination and altered transcriptional plasticity are at least as important as genetic alterations, if not more.

Some evidence of the pertinence of the atavistic theory comes from the investigation of genes that appeared at different stages of evolution. In cancer cells, genes that are disrupted tend to be those that appeared with the appearance of multicellularity. Whereas primitive genes normally ensure basic cellular processes such as cell cycle and metabolism, multicellularity and cancer involve genes related to cell–cell communication and growth control. Genes altered in cancer were found to often represent functional links between unicellular and multicellular processes [63], and analysing gene ages as projected on interactome networks has shown a strong connection between the topology of protein–protein interactions and evolutionary gene origins [64]. At the same time, the topology of protein–protein interaction networks can reflect the favourable or stressful external conditions of the cell [65].

Given the clear implication of 1D and 3D epigenome organization in both establishing and losing coordinated control in cell behaviour during development and cancer, one could expect similar associations between gene evolutionary history and position in the nucleus. Indeed, the analysis of patterns of 3D interaction of genomic regions representing chromatin as a network found relationships between gene functional categories and location of the genes inside the nucleus, identifying that highly connected nodes on these chromatin networks formed a rich club related to basic cellular processes [66]. We suggest that further studies of global principles of genome organization alterations across differentiation or oncogenesis, facilitated by network theory frameworks, could reveal that epigenomic network topology also carries the signs of successive evolutionary programs and identify strong structural–functional relationships in the epigenome [66,67]. Indeed genes of different evolutionary ages are organised in the 3D epigenome with patterns that can be characteristic of stem, differentiated and cancer cells. This view could suggest a correspondence between structural features of the chromatin network and phenotypic attractors reachable by the cells.

2.5. Conclusions

To go back to the tumour microenvironment context, if an external stress situation (mechanical, chemical, inflammatory, metabolic, etc.) disrupts the healthy microenvironment, different cell types would lose their perception of the context, compromising cooperativity between them and generating a sort of conflict in which acquiring stochastic, pluripotent, or unicellular-like behaviours would be advantageous. The increased stochasticity and entropy and the global epigenetic restructuring should be seen as a consequence of this tissue disruption. Finally, re-stabilizing the 3D epigenome and thus decreasing the cellular entropy and stochasticity would require playing on the microenvironmental context to restore cues that normally maintain the 3D epigenome in a fully differentiated state. As is often the case, the maximum stability of the differentiated state would also not be ideal in healthy tissues to achieve homeostasis, which requires some plasticity to allow rearranging the epigenome to respond to normal external cues (including stress, cytokines, and contact with other cells). This situation would be reminiscent of the need for an equilibrium in cancer cells between genetic stability and instability [68]. Indeed, genetic and epigenomic stability might even be related, in the light of recent findings that implicate the assembly of damaged DNA in repair factories [69].

Increasing experimental evidence indicates that the normal microenvironment is able to control genetically altered cells and stochastic fluctuations to reduce phenotypic plasticity [44,70–72]. This could also correspond to eliminating specific cancer cell attractors that are only reachable in the absence of external controls normally exerted by other cells and the physical environment on the cancer cell. One of us has already proposed, as a general concept [44,73] and applied to multiple myeloma [71,74], that acting adequately on the tumour microenvironment, i.e. restoring or mimicking healthy cellular interactions present in the initial tissue, would make it possible to 're-educate' it and to control cells by re-stabilizing their phenotypes, restoring differentiation and stopping cell proliferation. This microenvironment-centred therapy could be coupled with molecules that alter gene regulation (e.g. by 'epidrugs' acting on chromatin remodelers) towards the establishment of full differentiation [71,73,75]. This strategy would certainly help to decrease cancer cell entropy and stochasticity, while limiting overall cellular plasticity and therapy resistance.

REFERENCES

1 West J, Bianconi G, Severini S, Teschendorff AE. Differential network entropy reveals cancer system hallmarks. Sci Rep. 2012;**2**:802.
2 Ibáñez K, Guijarro M, Pajares G, Valencia A. A computational approach inspired by simulated annealing to study the stability of protein interaction networks in cancer and neurological disorders. Data Min Knowl Disc. 2016;**30**:226–42.
3 Nijman SMB. Perturbation-driven entropy as a source of cancer cell heterogeneity. Trends Cancer. 2020;**6**:454–61.
4 Feinberg AP. Epigenetic stochasticity, nuclear structure and cancer: the implications for medicine. J Intern Med. 2014;**276**:5–11.
5 Jenkinson G, Pujadas E, Goutsias J, Feinberg AP. Potential energy landscapes identify the information-theoretic nature of the epigenome. Nat Genet. 2017;**49**:719–29.
6 Banerji CR, Miranda-Saavedra D, Severini S, Widschwendter M, Enver T, Zhou JX, et al. Cellular network entropy as the energy potential in Waddington's differentiation landscape. Sci Rep. 2013;**3**:3039.
7 Stergachis AB, Neph S, Reynolds A, Humbert R, Miller B, Paige SL, et al. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell. 2013;**154**:888–903.
8 Lawson DA, Kessenbrock K, Davis RT, Pervolarakis N, Werb Z. Tumour heterogeneity and metastasis at single-cell resolution. Nat Cell Biol. 2018;**20**:1349–60.
9 Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell. 2011;**146**:633–44.
10 Ortega-Sabater C, Calvo GF, Pérez-García VM. Stochastic fluctuations drive non-genetic evolution of proliferation in clonal cancer cell populations. bioRxiv: 2021.06.08.447479.
11 Huang S, Ernberg I, Kauffman S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. Semin Cell Dev Biol. 2009;**20**:869–76.

12 Li Q, Wennborg A, Aurell E, Dekel E, Zou JZ, Xu Y, et al. Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape. Proc Natl Acad Sci U S A. 2016;**113**:2672–7.
13 Pastore A, Gaiti F, Lu SX, Brand RM, Kulm S, Chaligne R, et al. Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in CLL. Nat Commun. 2019;**10**:1874.
14 Househam J, Heide T, Cresswell GD, Spiteri I, Kimberley C, Zapata L, et al. Phenotypic plasticity and genetic control in colorectal cancer evolution. Nature. 2022;**611**:744–53.
15 Heide T, Househam J, Cresswell GD, Spiteri I, Lynn C, Mossner M, et al. The co-evolution of the genome and epigenome in colorectal cancer. Nature. 2022;**611**:733–43.
16 Loukas I, Simeoni F, Milan M, Inglese P, Patel H, Goldstone R, et al. Selective advantage of epigenetically disrupted cancer cells via phenotypic inertia. Cancer Cell. 2023;**41**:70–87.
17 Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science. 2015;**348**:880–6.
18 Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, et al. Somatic mutant clones colonize the human esophagus with age. Science. 2018;**362**:911–7.
19 Hill W, Lim EL, Weeden, CE, Lee C, Augustine M, Chen K, et al. Lung adenocarcinoma promotion by air pollutants. Nature. 2023; **616**: 159–67.
20 Misteli T. Beyond the sequence: cellular organization of genome function. Cell. 2007;**128**:787–800.
21 Cremer T, Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet. 2001;**2**:292–301.
22 Branco MR, Pombo A. Chromosome organization: new facts, new models. Trends Cell Biol. 2007;**17**:127–34.
23 Sood V, Misteli T. The stochastic nature of genome organization and function. Curr Opin Genet Dev. 2022;**72**:45–52.
24 Nollmann M, Bennabi I, Gotz M, Gregor T. The impact of space and time on the functional output of the genome. Cold Spring Harb Perspect Biol. 2022;**14**:a040378.
25 Kramer BA, Sarabia Del Castillo J, Pelkmans L. Multimodal perception links cellular state to decision-making in single cells. Science. 2022;**377**:642–8.
26 Roca-Cusachs P, Iskratsch T, Sheetz MP. Finding the weakest link: exploring integrin-mediated mechanical molecular pathways. J Cell Sci. 2012;**125**:3025–38.
27 Iskratsch T, Wolfenson H, Sheetz MP. Appreciating force and shape-the rise of mechanotransduction in cell biology. Nat Rev Mol Cell Biol. 2014;**15**: 825–33.
28 Dupont S, Morsut L, Aragona M, Enzo E, Giulitti S, Cordenonsi M, et al. Role of YAP/TAZ in mechanotransduction. Nature. 2011;**474**:179–83.
29 Iyer KV, Pulford S, Mogilner A, Shivashankar GV. Mechanical activation of cells induces chromatin remodeling preceding MKL nuclear transport. Biophys J. 2012;**103**:1416–28.
30 Heo SJ, Thorpe SD, Driscoll TP, Duncan RL, Lee DA, Mauck RL. Biophysical regulation of chromatin architecture instills a mechanical memory in mesenchymal stem cells. Sci Rep. 2015;**5**:16895.
31 Heo SJ, Cosgrove BD, Dai EN, Mauck RL. Mechano-adaptation of the stem cell nucleus. Nucleus. 2018;**9**:9–19.
32 Tang Y, Zhu L, Cho JS, Li XY, Weiss SJ. Matrix remodeling controls a nuclear lamin A/C-emerin network that directs Wnt-regulated stem cell fate. Dev Cell. 2022;**57**:480–95 e6.
33 Andreu I, Granero-Moya I, Chahare NR, Clein K, Molina-Jordan M, Beedle AEM, et al. Mechanical force application to the nucleus regulates nucleocytoplasmic transport. Nat Cell Biol. 2022;**24**:896–905.
34 Rizzuti IF, Mascheroni P, Arcucci S, Ben-Meriem Z, Prunet A, Barentin C, et al. Mechanical control of cell proliferation increases resistance to chemotherapeutic agents. Phys Rev Lett. 2020;**125**:128103.
35 Kupiec JJ. 2009. The origin of individuals. Singapore: World Scientific.
36 Kress C, Ballester M, Devinoy E, Rijnkels M. Epigenetic modifications in 3D: nuclear organization of the differentiating mammary epithelial cell. J Mammary Gland Biol Neoplasia. 2010;**15**:73–83.
37 Achinger-Kawecka J, Taberlay PC, Clark SJ. Alterations in three-dimensional organization of the cancer genome and epigenome. Cold Spring Harb Symp Quant Biol. 2016;**81**:41–51.
38 Akdemir KC, Le VT, Chandran S, Li Y, Verhaak RG, Beroukhim R, et al. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. Nat Genet. 2020;**52**:294–305.
39 Malod-Dognin N, Pancaldi V, Valencia A, Przulj N. Chromatin network markers of leukemia. Bioinformatics. 2020;**36**:i455–i63.
40 Morgan MA, Shilatifard A. Chromatin signatures of cancer. Genes Dev. 2015;**29**: 238–49.
41 Lever E, Sheer D. The role of nuclear organization in cancer. J Pathol. 2010;**220**: 114–25.
42 Zink D, Fischer AH, Nickerson JA. Nuclear structure in cancer cells. Nat Rev Cancer. 2004;**4**:677–87.
43 Meaburn KJ, Gudla PR, Khan S, Lockett SJ, Misteli T. Disease-specific gene repositioning in breast cancer. J Cell Biol. 2009;**187**:801–12.
44 Capp JP. Tissue disruption increases stochastic gene expression thus producing tumors: Cancer initiation without driver mutation. Int J Cancer. 2017;**140**:2408–13.
45 Alonso-Curbelo D, Ho YJ, Burdziak C, Maag JLV, Morris JPt, Chandwani R, et al. A gene-environment-induced epigenetic program initiates tumorigenesis. Nature. 2021;**590**:642–8.
46 Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;**126**:663–76.
47 Di Giammartino DC, Kloetgen A, Polyzos A, Liu Y, Kim D, Murphy D, et al. KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. Nat Cell Biol. 2019;**21**:1179–90.
48 Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, et al. Deterministic direct reprogramming of somatic cells to pluripotency. Nature. 2013;**502**:65–70.
49 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;**485**:376–80.
50 Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. 2016;**17**:661–78.
51 Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet. 2019;**20**:437–55.
52 Stadhouders R, Vidal E, Serra F, Di Stefano B, Le Dily F, Quilez J, et al. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat Genet. 2018;**50**:238–49.
53 Li R, Liang J, Ni S, Zhou T, Qing X, Li H, et al. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell. 2010;**7**:51–63.

54 Samavarchi-Tehrani P, Golipour A, David L, Sung HK, Beyer TA, Datti A, et al. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. Cell Stem Cell. 2010;**7**:64–77.

55 Hinohara K, Wu HJ, Vigneau S, McDonald TO, Igarashi KJ, Yamamoto KN, et al. KDM5 Histone demethylase activity links cellular transcriptomic heterogeneity to therapeutic resistance. Cancer Cell. 2018;**34**:939–53 e9.

56 Liau BB, Sievers C, Donohue LK, Gillespie SM, Flavahan WA, Miller TE, et al. Adaptive chromatin remodeling drives glioblastoma stem cell plasticity and drug tolerance. Cell Stem Cell. 2017;**20**:233–46 e7.

57 Capp JP, Thomas F. From developmental to atavistic bet-hedging: how cancer cells pervert the exploitation of random single-cell phenotypic fluctuations. Bioessays. 2022;**44**:e2200048.

58 Lopez-Maury L, Marguerat S, Bahler J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat Rev Genet. 2008;**9**:583–93.

59 Davies PC, Lineweaver CH. Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors. Phys Biol. 2011;**8**:015001.

60 Lineweaver CH, Bussey KJ, Blackburn AC, Davies PCW. Cancer progression as a sequence of atavistic reversions. Bioessays. 2021;**43**:e2000305.

61 Capp JP, Thomas F. A similar speciation process relying on cellular stochasticity in microbial and cancer cell populations. iScience. 2020;**23**:101531.

62 Csermely P, Hodsagi J, Korcsmaros T, Modos D, Perez-Lopez AR, Szalay K, et al. Cancer stem cells display extremely large evolvability: alternating plastic and rigid networks as a potential Mechanism: network models, novel therapeutic target strategies, and the contributions of hypoxia, inflammation and cellular senescence. Semin Cancer Biol. 2015;**30**:42–51.

63 Trigos AS, Pearson RB, Papenfuss AT, Goode DL. How the evolution of multicellularity set the stage for cancer. Br J Cancer. 2018;**118**:145–52.

64 Trigos AS, Pearson RB, Papenfuss AT, Goode DL. Somatic mutations in early metazoan genes disrupt regulatory links between unicellular and multicellular genes in cancer. Elife. 2019;**8**:e40947.

65 Lehtinen S, Marsellach FX, Codlin S, Schmidt A, Clement-Ziza M, Beyer A, et al. Stress induces remodelling of yeast interaction and co-expression networks. Mol Biosyst. 2013;**9**:1697–707.

66 Sandhu KS, Li G, Poh HM, Quek YL, Sia YY, Peh SQ, et al. Large-scale functional organization of long-range chromatin interaction networks. Cell Rep. 2012;**2**:1207–19.

67 Raynal F, Sengupta K, Plewczynski D, Aliaga B, Pancaldi V. Global chromatin reorganization and regulation of genes with specific evolutionary ages during differentiation and cancer. Nuc Acids Res. 2025; 53(4).

68 Andor N, Maley CC, Ji HP. Genomic instability in cancer: teetering on the limit of tolerance. Cancer Res. 2017;**77**:2179–85.

69 Arnould C, Rocher V, Finoux AL, Clouaire T, Li K, Zhou F, et al. Loop extrusion as a mechanism for formation of DNA damage repair foci. Nature. 2021;**590**:660–5.

70 Capp JP. Stochastic gene expression, disruption of tissue averaging effects and cancer as a disease of development. Bioessays. 2005;**27**:1277–85.

71 Capp JP, Bataille R. Multiple myeloma as a bone disease? The tissue disruption-induced cell stochasticity (TiDiS) theory. Cancers (Basel). 2020;**12**:2158.

72 Solary E, Laplane L. The role of host environment in cancer evolution. Evol Appl. 2020;**13**:1756–70.

73 Capp JP. Stochastic gene expression stabilization as a new therapeutic strategy for cancer. Bioessays. 2012;**34**:170–3.

74 Capp JP, Bataille R. A bone paradigm challenging the standard model of myeloma oncogenesis. Crit Rev Oncol Hematol. 2022;**172**:103640.

75 Yan M, Liu Q. Differentiation therapy: a promising strategy for cancer treatment. Chin J Cancer. 2016;**35**:3.

3

Dimensions of cellular plasticity: Epithelial–mesenchymal transition, cancer stem cells, and collective cell migration

Caterina A.M. La Porta and Stefano Zapperi

3.1. Cellular plasticity in physiological processes and injuries

The concept of phenotypic plasticity in biology can be used to describe multiple processes associated with the effect of the environment on the phenotype of the cells. In recent years, biologists have started to show that this process is able to drive phenotypic divergence within individual species that is as great as that between different species [1]. The possible role of phenotypic switching for the large scale of evolutionary processes such as in macroevolution is also debated currently [2, 3]. In this regard, there is ample evidence that phenotypic plasticity is related to the natural evolution of plants and animals [4, 5, 6, 7], the latter supporting the role of plasticity in evolutionary innovation.

Phenotypic plasticity appears to work at the level of cellular evolution as well: it is well documented that the entire intestinal epithelial lining is replaced approximately every 3–5 days, and a constant renewal required to maintain intestinal homeostasis and tissue integrity [8]. It is, however, quite remarkable that after an event damaging the stem cell compartment the intestinal epithelium displays an incredible ability to regenerate. De-differentiation of intestinal cells into multipotent stem cells, a process cumulatively termed intestinal cell plasticity, provides a mechanism to maintain barrier integrity and homeostatic stability against persistent injury. Multiple factors could damage the intestinal epithelium, including infection, acute or chronic inflammatory disease, and genotoxic stress associated with chemotherapy/radiation treatment (for more biological details, see [9]). Cellular heterogeneity and plasticity was also shown in skin epithelial cells under certain circumstances [10], and similar evidence has been reported for liver regeneration. The liver shows a remarkable regenerative capacity performing multiple physiological functions. Since this organ plays a central role in metabolism and detoxification, it often suffers extraneous injury and can gradually lose its regenerative ability. Cell reprogramming of hepatocytes and biliary epithelial cells has been recently identified as a major pathway to generate new hepatocytes in response to liver injury (for more biological details, see [11]). The concept of plasticity and the role of the environment were better understood analysing macrophages. These cells are usually considered plastic, although the environment could limit their plasticity. Macrophages are responsible for a mechanism by which tissues can avoid excessive immune reactions and protect themselves from collateral damage [12]. Another interesting example of cellular plasticity is represented by T cells, which are able to repolarize changing their phenotype in response to cytokine milieu, microbial products, and products of metabolism [13,14].

Cellular plasticity is also shown by Schwann cells that are glial cells present in the peripheral nervous system, playing a crucial role in the development, maintenance, function, and regeneration of peripheral nerves. These cells were shown to have important reprogramming and morphogenic changes promoting nerve regeneration and functional recovery during traumatic injuries and peripheral neuropathies [15]. Therefore, considering all this evidence, it emerges that plasticity is an important feature of all the cells whose role is mainly related to repair damage and restore tissue homeostasis.

3.2. Cancer stem cells: evolution and plasticity of the tumour

In the past 20 years, the literature clearly showed that the evolution of cancer cells depends on the capability of tumour cells to be plastic and to be able to change their phenotype depending on the environment, a feature known as tumour plasticity. An analysis of the

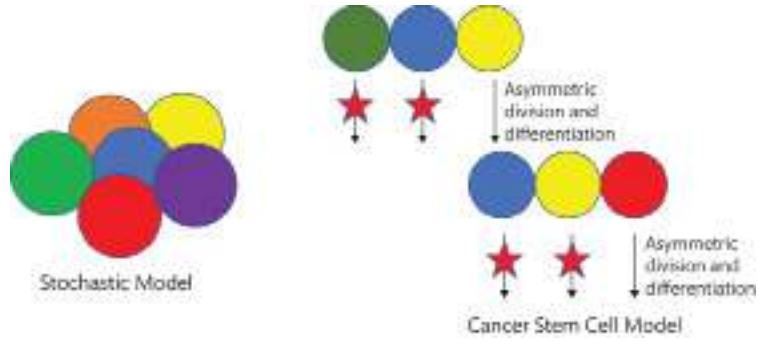


Figure 3.1. The competing models of cancer initiation and progression. Stochastic model of cancer initiation posits where most cells in a given tumour would be able to initiate a new tumour, and the cancer stem cell model points where only CSCs can initiate tumours. CSCs are maintained through symmetric cell division, while asymmetric cell division leads the cells to die.

scientific literature shows that the evolution of tumours was first described as a stochastic process in which the accumulation of genetic alterations in cancer cells was considered to be responsible for the heterogeneity of the tumours [16]. According to this view, genetic or epigenetic alterations may arise during each cellular replication, providing a fitness advantage to mutated cells in terms of increased proliferation, migration, or invasion capabilities [17]. Once cells have acquired a sufficiently aggressive phenotype, they can escape from homeostatic control mechanisms typical of healthy tissues, where homeostatic mechanisms induce compensatory regulatory responses via cellular signalling, apoptosis, and other processes [18], and initiate the tumorigenetic process (**Figure 3.1**) [19-23]. This view of tumour evolution changed over time by accumulating evidence of the presence of cell subpopulation, called cancer stem cells (CSCs), which was described to be resistant to drugs and apoptosis, mainly composed by quiescent cells, overall resembling many biological characteristics of stem cells [24] (**Figure 3.1**). CSCs were identified in several tumours, including brain, melanoma, and breast [25-34]. The consequence in terms of therapeutic intervention is that instead of using drugs targeting common biological markers for all the heterogeneous cells within the tumour, CSCs were investigated with the aim of finding out their specific molecular characteristics [24,35,36]. According to the CSC hypothesis, the eradication of the tumour is only possible after eliminating this subpopulation since even a single residual CSC might be able to induce tumour relapse under appropriate environmental conditions [35-38].

The picture of cancer evolution illustrated in **Figure 3.1**, contrasting stochastic evolution with the presence of a special subpopulation that maintains itself thanks to a symmetric division or CSCs, is however too simplified. Several studies have provided evidence that both CSCs and non-CSCs are plastic and capable of undergoing phenotypic transitions in response to appropriate stimuli, in analogy with stem cell plasticity [39,40] (see **Figure 3.2**). The impact of the occurrence of tumour plasticity within tumour populations is crucial from the perspective of therapeutic strategy. While targeting CSCs or inhibiting generically the cellular proliferation of the tumours by chemotherapy, the tumour can shrink even to a few cells; if these cells have the capability to change their phenotype into a more aggressive one in response to environmental stimuli, then the tumour can restart growing. The complexity of tumour evolution therefore depends on the interaction with the environment that under specific conditions might help tumour cells switch to a more aggressive phenotype [39,41]. Understanding the molecular mechanisms, underlining phenotypic switching can help in the development of new therapeutic strategies targeting specific markers involved in crucial steps of this process [39,41].

Our group extensively investigated the phenotypic switching of human melanoma cells, discovering key factors involved in this complex process. We have shown that a complex network of micro RNAs (miRNAs) is able to modulate the phenotypic switching of human melanoma cells [42]. Within this intricate network of miRNAs, we found that miRNa222 is released by cancer cells into the environment to control the plasticity of neighbouring cells [43]. We suggested that this mechanism is able to control the numbers of CSCs in the tumour and is modulated by the same cells in response to specific signals received by the environment [42,43]. We also showed that circRNAs can crosstalk with miRNAs and modulate critical circuits controlling the phenotypic switching in human melanoma cells [43-44]. In particular, we showed a correlation between the level of ZEB1 and SNAI1 and the fraction of CSCs in the population [14]. Furthermore, to predict the switching between the two biological states of the cells, we developed a mathematical model describing the regulatory circuit ZEB1/SNAIL showing that a back-splicing factor was needed to limit circRNA production [44].

Cancer plasticity is not a feature of melanoma only, but it is possible to find it in all types of tumours. Accordingly, a recent paper described the plasticity of breast cancer in xenografts and its dependence on the environment, reflecting their adaptation to particular environmental cues [45]. Other studies on colorectal cancer showed how an inflammatory environment activates nuclear factor kappa-light-chain-enhancer of activated B cells

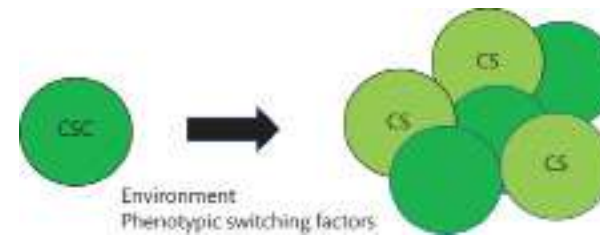


Figure 3.2. Figure shows the plasticity of tumour cells. CSCs expand symmetrically and asymmetrically, and under specific environmental conditions, CS can revert into CSC.

(NF-κB) signalling, promoting the tumour-initiating potential of non-stem cells by triggering their de-differentiation [46]. Recent experiments using cell ablation with a CRISPR–Cas9 approach to insert an inducible version of the suicide-gene caspase 9 (iCasp9) into the LGR5 locus in human colorectal cancer organoids confirmed tumour plasticity [47]. This study showed that in xenografts produced by these organoids the induction of apoptosis reduced the tumour size and, upon removal of the inducer, the tumours regrew [47]. Moreover, lineage-tracing experiments from differentiated tumour cells demonstrated that these cells regained their proliferative potential and restored the LGR5 + CSC pool [47]. In another study, using mouse colorectal cancer organoids engineered to express the diphtheria–toxin receptor under the control of the LGR5 locus56, it was shown that the ablation of Lgr5 + CSC cells halted tumour growth, but tumours resumed growth upon the cessation of diphtheria–toxin treatment, thus illustrating their phenotypic plasticity [48].

3.3. Epithelial–mesenchymal transition, metastasis, and predictive strategies for cancer aggressiveness

Nowadays people do not die for primary tumours but for metastasis [49]. Metastasis or secondary tumours are due to a sequence step in which tumour cells leave their primary site, circulate in the bloodstream, endure pressure in blood vessels, acclimate to new cellular surroundings in a secondary site, and escape deadly combat with immune cells [39]. Dissemination and invasion start from chromosomal instability inside the cells which is caused by continuous errors in chromosome segregation during mitosis. *In vivo* and *in vitro* studies show that metastatic cancer cells migrate individually [50] but in humans, and it is believed that seeding requires the joint action of a cluster of tumour cells moving together [51]. Metastatic cells are supposed to undergo an epithelial–mesenchymal transition (EMT), which implies crucial changes in the cells. Epithelial cells displaying apical–basal polarity are held together by tight junctions, adherens junctions and desmosomes, and are tethered to the underlying basement membrane by hemidesmosomes. These cells express molecules that are associated with the epithelial state and help maintain cell polarity. Transition to the mesenchymal state leads to the expression of the EMT-inducing transcription factors ZEB, SNAIL, and TWIST, which inhibit the expression of genes associated with the epithelial state and concomitantly activate the expression of genes associated with the mesenchymal state. These changes in gene expression result in cellular changes that include the disassembly of epithelial cell–cell junctions and the dissolution of apical–basal cell polarity. This progressive loss of epithelial features is accompanied by the acquisition of a partial set of mesenchymal features with retention of certain epithelial features; under certain circumstances, a complete set of mesenchymal features may be acquired. Mesenchymal cells display front-to-back polarity and an extensively reorganized cytoskeleton and express a distinct set of molecules that promote and maintain the mesenchymal state. During EMT, cells become motile and acquire invasive capacities. Note that the EMT is a reversible process, and mesenchymal cells can also revert to the epithelial state by undergoing mesenchymal–epithelial transition (MET).

Recent evidence has clearly showed that when epithelial cancer cells acquire a mesenchymal gene program they increase their capability for migration and invasion [52,53]. We have recently investigated the EMT by using a combination of numerical simulations of a Boolean network model of the EMT pathways and the analysis of bulk and single-cell gene expression data [54]. We showed that the EMT involves the transit through a multitude of meta-stable states, corresponding to highly aggressive hybrid cells that can easily switch under external and internal perturbations [54]. Our study, therefore, allowed to reconstruct the topography of the phenotypic landscape, as originally envisaged in general terms by Waddington many decades ago [55], describing possible attractors of the relevant gene regulatory network (for more details, see [39]). Moreover, phenotypic switching can take multiple paths and produce a variety of outcomes corresponding to the astonishing complexity of a cancer cell population. Furthermore, the large number of states expressed by the network confirms that we should abandon the rigid distinction between epithelial or mesenchymal cells [39]. Instead, a continuum of possible cell phenotypes with varying degrees of plasticity appears to be more reasonable. In the light of these findings, metastasis could be due to one of these cancer cells with a hybrid phenotype that can easily migrate and switch in the right environment [39]. Dormant cells can also be explained in the light of these findings as a state of the cells related to the environment [39].

Given the difficulties in curing cancer metastasis, the possibility to predict the metastasis risk in patients with a primary tumour appears to be crucial. The prediction of this risk can help avoid the overtreatment of patients suffering side effects with high economical and social costs. Considering breast cancer, about 20% of these tumours are due to triple-negative breast cancer (TNBC), and since there is no expression of markers such as hormone and HER2, these tumours are lacking specific treatment, growing, and spreading rapidly with limited treatment options and typically worse prognosis [56]. Using an innovative strategy targeting the complexity of the intracellular gene network, we developed ARIADNE, a general algorithmic strategy to assess the risk of metastasis from transcriptomic data of patients with TNBC [56]. Using this mapping, we were able to stratify patients with TNBC according to their prognosis, as we showed by validating the strategy with three independent cohorts of TNBC patients [56]. More recently, we have compared ARIADNE with immunological strategy to predict the aggressiveness of TNBC, stratifying a high-risk population with high immune markers that is, however, not properly classified by the tumour immune microenvironment-based strategy [57].

3.4. Collective cell migration and tumour plasticity

Tumour plasticity plays a critical role in the evolution of cancer, as discussed in the previous section. EMT plays an important role in this process, thereby identifying crucial targets involved in this complex pathway could be critical for identifying new therapeutic strategy for metastasis. An extensive overview of these aspect was discussed in depth in a recent book [41]. In particular, it is interesting that the immune privilege status of CSCs [41] seems to be linked to its quiescent states and not due to an intrinsic property of

CSCs. Therefore, the environment is able to modulate the status of the cells and helps switching from a quiescent state to a migratory phenotype or vice versa. Moreover, many papers showed that collective cell migration occurs at the invasive front of tumours as well as observed that CTC clusters in the bloodstream are more effective than single CTCs for seeding metastasis [41].

Collective cell migration has been described in terms of the so-called jamming–unjamming transition (JUT) and its reverse unjamming–jamming transition (UJT) [58] that is believed to relevant both in tissues and cancer cells (for a review, see [59]). JUT/UJT are observed in the rheology of soft materials, such as colloidal suspensions, gels, or foams, where flow can be hindered by kinematic constrains [60]: jamming corresponds to a dynamic arrest into a solid-line state, while unjamming leads to a fluid-like state. Several recent studies explored in depth the connection between EMT/MET and JUT/UJT, but how the two transitions are related is still debated. Can the two transition act together or are they mutually exclusive? As discussed in [61], an answer to these questions can be obtained by framing the discourse in the broad context of the physics of nonequilibrium phase transitions. From this point of view, EMT and JUT are fundamentally different because they occur in a different phase space although both lead to the mobilization of the cells. JUT/UJT occur within the physical space of the tumour and can be described by the positions and velocities of the cells themselves. The cell velocity field [62] or the effective diffusion constant D [63] can then be used to distinguish a jammed from an unjammed state. For instance, recent experiments on cancer cells, both *in vitro* and *in vivo*, showed phase transitions between jammed solid, active fluid, and active nematic [64] controlled by cell density and adhesion and described velocity and vorticities [65,66]. Contrary to the JUT, in the case of EMT, the relevant degrees of freedoms are not located in a physical space but are instead the nodes of a complex gene regulatory network [54]. Simulations of Boolean network models show that EMT/MET are controlled by a large set of biochemical and physical parameters. While cells undergoing EMT can induce a JUT in the tumour by reducing cell–cell adhesion, the reverse is not necessarily true.

3.5. Conclusions

In this chapter, we have reviewed recent results on phenotypic plasticity of cancer cells. We discussed the relevance of this concept for CSCs and highlighted the role of the EMT in driving metastasis. Finally, we discussed the interplay between unjamming and EMT in driving collective cancer cell migration.

Competing interests

The authors declare the following competing interests: Complexdata S.R.L. has filed an Italian patent application related to the present work. Inventors: F. Font-Clos, S. Zapperi, C.A.M. La Porta. Patent status: granted. Date of application: 13 December 2019. Application number: 102019000023946. The patent concerns a method to screen breast cancer patients using transcriptomic data and Boolean networks. S.Z. and C.A.M.P.L. hold 7.36 and 14.72% shares of Complexdata S.R.L., respectively.

Author contributions statement

C.A.M.L.P. and S.Z. wrote and reviewed the manuscript.

REFERENCES

1. Liem KF, Kaufman LS. Intraspecific macroevolution: functional biology of the polymorphic cichlid species Cichlasoma minckleyi. Orono, ME:University of Maine Press; 1984. p. 203–15.
2. Pigliucci M, Murren CJ. Perspective: Genetic assimilation and a possible evolutionary paradox: can macroevolution sometimes be so fast as to pass us by? Evolution. 2003 Jul;**57**(7):1455–64.
3. Jablonski D. Approaches to macroevolution: 1. general concepts and origin of variation. Evol Biol. 2017;**44**(4):427–50.
4. Schlichting CD, Wund MA. Phenotypic plasticity and epigenetic marking: an assessment of evidence for genetic accommodation. Evolution. 2014 Mar;**68**(3):656–72.
5. Schwander T, Leimar O. Genes as leaders and followers in evolution. Trends Ecol Evol. 2011 Mar;**26**(3):143–51.
6. Levis NA, Pfennig DW. Evaluating 'plasticity-first' evolution in nature: key criteria and empirical approaches. Trends Ecol Evol. 2016 July;**31**(7):563–74.
7. Gibert J-M. The flexible stem hypothesis: evidence from genetic data. Dev Genes Evol. 2017 Sep;**227**(5):297–307.
8. Barker N. Adult intestinal stem cells: critical drivers of epithelial homeostasis and regeneration. Nat Rev Mol Cell Biol. 2014 Jan;**15**(1):19–33.
9. Meyer AR, Brown ME, McGrath PS, Dempsey PJ. Injury-induced cellular plasticity drives intestinal regeneration. Cell Mol Gastroenterol Hepatol. 2022;**13**(3):843–56.
10. Wang J, He J, Zhu M, Han Y, Yang R, Liu H, et al. Cellular heterogeneity and plasticity of skin epithelial cells in wound healing and tumorigenesis. Stem Cell Rev Rep, 2022 Aug;**18**(6):1912–25.
11. Li W, Li L, Hui L. Cell plasticity in liver regeneration. Trends Cell Biol. 2020 April;**30**(4):329–38.
12. Guilliams M, Svedberg FR. Does tissue imprinting restrict macrophage plasticity? Nat Immunol. 2021 Feb;**22**(2):118–27.
13. Ueno A, Jeffery L, Kobayashi T, Hibi T, Ghosh S, Jijon H. Th17 plasticity and its relevance to inflammatory bowel disease. J Autoimmun. 2018 Feb;**87**:38–49.
14. Bal SM, Golebski K, Spits H. Plasticity of innate lymphoid cell subsets. Nat Rev Immunol. 2020 Sep;**20**(9):552–65.
15. Nocera G, Jacob C. Mechanisms of Schwann cell plasticity involved in peripheral nerve repair after injury. Cell Mol Life Sci. 2020 Oct;**77**(20):3977–89.
16. Vogelstein B, Kinzler KW. The genetic basis of human cancer. New York: McGraw-Hill Professional; 2002.
17. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000 Jan;**100**(1):57–70.
18. Jacobson MD. Programmed cell death: a missing link is found. Trends Cell Biol. 1997 Dec; **7**(12):467–9.
19. Levine AJ. The tumor suppressor genes. Annu Rev Biochem. 1993;**62**:623–51.
20. Kinzler KW, Vogelstein B. Cancer-susceptibility genes. gatekeepers and caretakers. Nature. 1997 Apr;**386**(6627):761, 763.
21. Kinzler KW, Vogelstein B. Landscaping the cancer terrain. Science. 1998 May; **280**(5366):1036–7.

22. Knudson AG. Two genetic hits (more or less) to cancer. Nat Rev Cancer. 2001 Nov; **1**(2):157–62.
23. Hahn WC, Weinberg RA. Rules for making human tumor cells. New Engl J Med. 2002 Nov; **347**(20):1593–603.
24. Galmozzi E, Facchetti F, La Porta CAM. Cancer stem cells and therapeutic perspectives. Curr Med Chem. 2006;**13**(6):603–7.
25. Uckun FM, Sather H, Reaman G, Shuster J, Land V, Trigg M, et al. Leukemic cell growth in SCID mice as a predictor of relapse in high-risk b-lineage acute lymphoblastic leukemia. Blood. 1995 Feb;**85**(4):873–8.
26. Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature. 1994 Feb;**367**(6464):645–8.
27. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med. 1997 July;**3**(7):730–7.
28. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A. 2003 Apr; **100**(7):3983–8.
29. O'Brien CA, Pollett A, Gallinger S, Dick JE. A human colon cancer cell capable of initiating tumour growth in immunodeficient mice. Nature. 2007 Jan;**445**(7123):106–10.
30. Dalerba P, Dylla SJ, Park I-K, Liu R, Wang X, Cho RW, et al. Phenotypic characterization of human colorectal cancer stem cells. Proc Natl Acad Sci U S A. 2007 Jun;**104**(24):10158–63.
31. Singh SK, Hawkins C, Clarke ID, Squire JA, Bayani J, Hide T, et al. Identification of human brain tumour initiating cells. Nature. 2004 Nov;**432**(7015):396–401.
32. Monzani E, Facchetti F, Galmozzi E, Corsini E, Benetti A, Cavazzin C, et al. Melanoma contains CD133 and ABCG2 positive cells with enhanced tumourigenic potential. Eur J Cancer. 2007 Mar;**43**(5):935–46.
33. Salmaggi A, Boiardi A, Gelati M, Russo A, Calatozzolo C, Ciusani E, et al. Glioblastoma-derived tumorospheres identify a population of tumor stem-like cells with angiogenic potential and enhanced multidrug resistance phenotype. Glia. Dec 2006;**54**(8):850–60.
34. Taghizadeh R, Noh M, Huh YH, Ciusani E, Sigalotti L, Maio M, et al. Cxcr6, a newly defined biomarker of tissue-specific stem cell asymmetric self-renewal, identifies more aggressive human melanoma cancer stem cells. PLoS ONE. 2010 Dec;**5**(12):e15183,
35. La Porta CAM. Drug resistance in melanoma: new perspectives. Curr Med Chem. 2007;**14**(4):387–91.
36. La Porta C. Cancer stem cells: lessons from melanoma. Stem Cell Rev Rep. 2009 Mar; **5**(1):61–5.
37. La Porta CAM, Zapperi S. Human breast and melanoma cancer stem cells biomarkers. Cancer Lett. 2013 Sep;**338**(1):69–73.
38. La Porta CAM. CXCR6: the role of environment in tumor progression. challenges for therapy. Stem Cell Rev Rep. 2012 Dec;**8**(4):1282–5.
39. La Porta CAM, Zapperi S. Complexity in cancer stem cells and tumor evolution: toward precision medicine. Semin Cancer Biol. 2017 June;**44**:3–9.
40. Lionetti MC, Fumagalli MR, La Porta CAM. Cancer stem cells, plasticity, and drug resistance. Cancer Drug Resist. 2020;**3**(2):140–48.
41. La Porta CAM. Phenotypic plasticity: the emergence of cancer stem cells, collective cell migration, and the impact on immune surveillance. In: Successes and challenges of NK immunotherapy. Elsevier; 2021. p. 183–90.
42. Sellerio AL, Ciusani E, Ben-Moshe NB, Coco S, Piccinini A, Myers CR, et al. Overshoot during phenotypic switching of cancer cell populations. Sci Rep. 2015 Oct;**5**:15464.
43. Lionetti MC, Cola F, Chepizhko O, Fumagalli MR, Font-Clos F, Ravasio R, et al. MicroRNA-222 regulates melanoma plasticity. J Clin Med. 2020 Aug;**9**(8):2573.
44. Fumagalli MR, Lionetti MC, Zapperi S, La Porta CAM. Cross-talk between circRNAs and mRNAs modulates miRNA-mediated circuits and affects melanoma plasticity. Cancer Microenviron. 2019 Dec;**12**(2–3):95–104.
45. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell. 2011 Aug;**146**(4):633–44.
46. Schwitalla S, Fingerle AA, Cammareri P, Nebelsiek T, Göktuna SI, Ziegler K, et al. Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties. Cell. 2013 Jan;**152**(1–2):25–38.
47. Shimokawa M, Ohta Y, Nishikori S, Matano M, Takano A, Fujii M, et al. Visualization and targeting of LGR5+ human colon cancer stem cells. Nature. 2017 May;**545**(7653):187–92.
48. de Sousa e Melo F, Kurtova AV, Harnoss JM, Kljavin N, Hoeck JD, Hung J, et al. A distinct role for LGR5+ stem cells in primary and metastatic colon cancer. Nature. 2017 Mar;**543**(7647):676–80.
49. Steeg PS. Tumor metastasis: mechanistic insights and clinical challenges. Nat Med. 2006 Aug;**12**(8):895–904.
50. Clark AG, Vignjevic DM. Modes of cancer cell invasion and the role of the microenvironment. Curr Opin Cell Biol. 2015 Oct;**36**:13–22.
51. Cheung KJ, Ewald AJ. A collective route to metastasis: seeding by tumor cell clusters. Science. 2016 Apr;**352**(6282):167–9.
52. Cano A, Pérez-Moreno MA, Rodrigo I, Locascio A, Blanco MJ, del Barrio MG, et al. The transcription factor SNAIL controls epithelial-mesenchymal transitions by repressing e-cadherin expression. Nat Cell Biol. 2000 Feb;**2**(2):76–83.
53. La Porta CAM, Zapperi S. Explaining the dynamics of tumor aggressiveness: at the crossroads between biology, artificial intelligence and complex systems. Semin Cancer Biol. 2018 Dec;**53**:42–7. .
54. Font-Clos F, Zapperi S, La Porta CAM. Topography of epithelial-mesenchymal plasticity. Proc Natl Acad Sci U S A. 2018 June;**115**(23):5902–907.
55. Waddington CH. The strategy of the genes. Abingdon: Routledge; 2014.
56. Font-Clos F, Zapperi S, Caterina A M La Porta. Classification of triple-negative breast cancers through a Boolean network model of the epithelial-mesenchymal transition. Cell Syst. 2021 May;**12**(5):457–62.e4.
57. Font-Clos F, Zapperi S, La Porta CAM. Classification of triple negative breast cancer by epithelial mesenchymal transition and the tumor immune microenvironment. Sci Rep. 2022 June;**12**(1):9651.
58. Park J-A, Kim JH, Bi D, Mitchel JA, Qazvini NT, Tantisira K, et al. Unjamming and cell shape in the asthmatic airway epithelium. Nat Mater. 2015;**14**(10):1040–48.
59. La Porta CAM, Zapperi S. Cell migrations: causes and functions. Vol. **1146**. Cham: Springer; 2019.
60. Liu AJ, Nagel SR. Jamming is not just cool any more. Nature. 1998;**396**(6706):21–2.
61. La Porta CAM, Zapperi S. Phase transitions in cell migration. Nat Rev Phys. 2020;**2**(10):516–17.

62. Garcia S, Hannezo E, Elgeti J, Joanny J-F, Silberzan P, Gov NS. Physics of active jamming during collective cellular motion in a monolayer. Proc Natl Acad Sci. 2015;**112**(50):15314–19.
63. Bi D, Yang X, Marchetti MC, Manning ML. Motility-driven glass and jamming transitions in biological tissues. Phys Rev X. 2016;**6**(2):021011.
64. Saw TB, Doostmohammadi A, Nier V, Kocgozlu L, Thampi S, Toyama Y, et al. Topological defects in epithelia govern cell death and extrusion. Nature. 2017;**544**(7649):212–16.
65. Ilina O, Gritsenko PG, Syga S, Lippoldt J, La Porta CAM, Chepizhko O, et al. Cell-cell adhesion and 3d matrix confinement determine jamming transitions in breast cancer invasion. Nat Cell Biol. 2020 Sep;**22**(9):1103–15.
66. Chepizhko O, Armengol-Collad JM, Alexander S, Wagena E, Weigelin B, Giomi L, et al. Confined cell migration along extracellular matrix space in vivo. Proc Natl Acad Sci. 2025 Jan 7;**122**(1):e2414009121.

4

Phenotypic switching in cancer: A systems-level perspective

Divjoy Singh, Abhay Gupta, Mohit Kumar Jolly, and Prakash Kulkarni

4.1. Introduction

Waddington's epigenetic landscape is a commonly used metaphor in developmental biology, often invoked to explain the trajectory of differentiation of a stem or progenitor cell into a differentiated one. In this metaphor, a cell is represented as a ball rolling downhill through a rugged landscape of bifurcating valleys which further lead to sub-valleys at the foot of the hill. Each of these sub-valleys denotes a distinct differentiated cell state or phenotype that is 'locked' unless perturbed significantly [1] (**Figure 4.1**).

Initially, cell-fate decisions during embryonic development were thought of as irreversible, but remarkable investigations into cellular reprogramming have shown that differentiated cells can be induced to gain pluripotency, often through ectopic overexpression of specific transcription factors. These induced pluripotent stem cells can typically proliferate and self-renew indefinitely *in vitro* and can generate the three primary germ layers (endoderm, mesoderm, and ectoderm) [2]. Similarly, in oncology, a long-standing conceptual framework has been that of a 'hierarchical (cancer) stem cell' model where cancers have been thought of as organized in hierarchical structures based on their differentiation capacity. At the apex are cancer stem cells (CSCs) that are the most stem-like and capable of regenerating the lower tiers of cell types, often non-CSCs, thus displaying their potential to repopulate a heterogeneous tumour and facilitating tumour recurrence and drug resistance [3]. However, research in the past decade across many cancer types has highlighted that non-CSCs can give rise to CSCs, under varying environmental conditions [4–8], establishing the 'plasticity' model [9]. Together, these observations suggest that in both non-cancerous and cancerous contexts, 'stemness' is a dynamic trait that can be gained or lost in individual cells.

What factors control the acquisition of 'stemness'? Both intracellular and extra-cellular stimuli have been shown to mediate the underlying complex dynamics. For instance, cells undergoing epithelial–mesenchymal transition (EMT)—a dynamical process through which epithelial cells weaken their cell–cell adhesion and gain migratory and invasive features—have been shown to enable 'stemness' through crosstalk among the gene regulatory networks (GRNs) regulating EMT and stemness [10–12]. Recent *in vitro* and *in vivo* observations have revealed that the hybrid epithelial/mesenchymal (E/M) cancer cells, instead of the 'fully epithelial' or 'fully mesenchymal' ones, are the most stem-like ones and possess the maximal tumour-initiating and metastatic abilities [13–15]. Computational models of many EMT-related GRNs have shown them to be 'multistable', i.e. enabling the coexistence of multiple cell states, including one or more hybrid E/M ones, that can reversibly switch among one another [16]. The rates of transitions among these cell states depend on various parameters, such as chromatin state, metabolic state, drug treatment being given, and the topology of underlying GRNs [17–19]. Not just stemness, other cellular traits often enabling metastasis—immune evasion, resistance to targeted therapies, resistance to anoikis (cell death due to matrix detachment), and metabolic reprogramming—are dynamically connected to EMT through interconnected GRNs [20–24]. Thus, metastasizing cells often traverse in a high-dimensional cell-state space dynamically acquiring or losing many phenotypes with varying 'fitness', as witnessed in phenotypic heterogeneity in circulating tumour cells in cancer patients [25]. This plasticity, i.e. the ability to transition to different phenotypes, is thus regarded as a hallmark of cancer metastasis that cells employ to navigate different time-varying bottlenecks during the invasion-metastatic cascade. Therefore, it becomes imperative to investigate the systems-level dynamics of phenotypic plasticity. Here, we discuss some mechanisms implicated in phenotypic plasticity in cancer.

4.2. Mechanisms underlying phenotypic switching

Phenotypic switching or plasticity refers to the reversible transition of cells from one state (phenotype) to another, often in response to changing environment [26]. It has been implicated in cancer metastasis and in enabling cells to evade therapeutic attacks. Here, we discuss some commonly observed mechanisms that govern the dynamics of phenotypic switching:

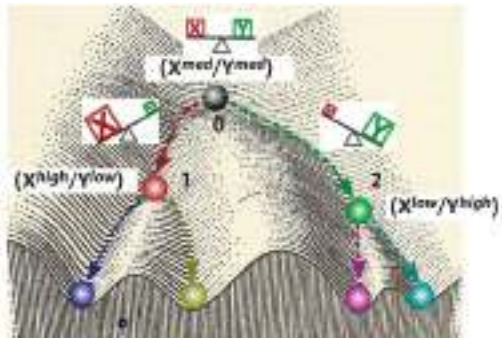


Figure 4.1. Schematic illustration of Waddington's epigenetic landscape. *Source:* Adapted from Jia et al. [1]. This rugged topography is a metaphorical representation of cell-fate trajectories in a high-dimensional molecular space. Different coloured balls shown on this landscape represent functionally different cell phenotypes. Starting from the top, each ball (or cell) settles steadily in one of the sub-valleys at the foot of the hill that depicts terminal cell states. X and Y are master regulators driving a cell to attain the phenotypes '1' and '2', respectively. The phenotype '0' co-expresses both X and Y at a medium level X^{med}/Y^{med}, reminiscent of the progenitor state. The two differentiated states are depicted as '1' and '2'—X^{high}/Y^{low} and X^{low}/Y^{high}, respectively. Due to inherent stochasticity in progenitor cell '0', the level of one of the molecules (say X) becomes higher than that of the other (say Y). This initial asymmetry can further trigger an amplifying cascade of events where the levels of X keeps further increasing while that of its repressor Y keeps decreasing, eventually rendering its own inhibition by Y as ineffective. Consequently, the cell attains one of the two differentiated states X^{high}/Y^{low}. Similarly, the progenitor cell-state can also give rise to X^{low}/Y^{high}.

4.2.1. Multi-stability in regulatory networks

Mathematical modelling of GRNs has been an important endeavour to understand the emergent dynamics of complex interactions and feedback loops constituting these networks. In the context of cell-fate decisions, a GRN commonly noted is a toggle switch, i.e. two master regulators (say, X and Y) transcriptionally inhibiting each other. This GRN allows for the existence of two mutually exclusive states—(X^{low} and Y^{high}) and (X^{high} and Y^{low})—both of which correspond to a specific phenotype/cell state [27] (Figure 4.1). These two states can switch among themselves stochastically. When X and/or Y can self-activate, it allows for the existence of another stable state—(medium X and medium Y)—which corresponds to the progenitor or undifferentiated state that is primed to differentiate to one of the two above-mentioned mutually exclusive phenotypes [28]. Such self-activating toggle switches have been reported at various bifurcation points in terms of developmental decision-making, such as PTF1A and NKX6 inhibiting each other transcriptionally during differentiation of pancreatic progenitor cells (medium PTF1A and medium NKX6) into exocrine (high PTF1A and low NKX6) or endocrine (low PTF1A and high NKX6) states [29].

The two master regulators engaged in a toggle switch need not be both transcription factors; such feedback loops can consist of microRNAs [30] as well. For instance, EMT and its reverse mesenchymal–epithelial transition (MET) decisions involve a chimeric toggle switch between the microRNA-200 family and ZEB transcription factor family, driving epithelial and mesenchymal cell states, respectively [31,32]. While ZEB1 and ZEB2 transcriptionally repress miR-200 family members, they are inhibited at post-transcriptional level. Similarly, RNA-binding proteins can engage in toggle switch topology together with microRNAs, for instance, LIN28 and let-7 inhibiting each other and controlling 'stemness' [33,34]. Moreover, GTPases Rac and Rho also are involved in a mutually inhibitory feedback loop such that Rac1-mediated cell polarization and lamellipodia formation facilitate a mesenchymal-type migration, while Rho-driven actomyosin contractility allows for protease-independent amoeboid-type migration [35,36]. Computational modelling of all these networks has revealed multi-stability as a common trait in their dynamics, thus explaining interconvertibility of phenotypes seen experimentally, driven by various environmental factors such as hypoxia, confinement, and matrix stiffness [37–39].

Among the above-mentioned instances of cancer cell plasticity, the dynamics of EMT/MET has been quite well studied. Many mathematical models for large and complex GRNs for EMT/MET have been developed in the past decade, all of which indicate multi-stability as a hallmark of these networks, irrespective of the modelling strategy used (continuous or discrete; deterministic or stochastic), cancer subtype being studied, or the scope of interactions considered in the model, i.e. considering only intra-cellular interactions or also cell–matrix and/or cell–cell interactions [40–46]. A recent analysis of topology of many such EMT/MET GRNs highlighted that they consist of 'teams' of nodes such that members within a team activate each other effectively and those across teams inhibit each other, thus leading to a 'toggle switch' between two teams, one driving EMT and other pushing MET [17]. Together, these observations indicate that multi-stability, as enabled by mutually inhibitory feedback loops, is a cornerstone of cancer cell plasticity.

Another instance where such multi-stability can have important clinical implications is the therapy-induced adaptive response in terms of cell-state switching. For instance, in oestrogen–receptor positive breast cancer, tamoxifen resistance can drive EMT and, in turn, EMT can drive tamoxifen resistance because the dynamics of underlying GRN allows for phenotypic switch among (epithelial, tamoxifen-sensitive) and (mesenchymal, tamoxifen-resistant) cell states [47]. Similar analysis of coupling of EMT circuits with those of stemness and metabolic switching helps unravel a mechanistic basis for their interconnected multi-stable dynamics [10,22].

4.2.2. Asymmetric cell division

Asymmetric cell division (ACD) refers to the event of two daughter cells having different fates after mitosis. Often reported in somatic stem cells, it allows for one daughter cell to retain 'stemness' while the other attains a non-stem differentiated state. A balance between ACD and symmetric cell division (SCD)—where both daughter cells attain the fate of the parent post-mitosis—is essential for homeostasis and development [48]. In mammalian models, a switch from ACD to SCD can trigger disruption of tissue homeostasis and drive tumour formation [49]. It is also considered to be more closely associate with CSCs in early-stage tumours, while late-stage tumours prioritize SCD and higher proliferation [50]. Various signalling molecules, such as p53, CD133, Numb, and Notch, have been shown to be implicated in ACD [48]. For instance, reduced p53 levels are associated with reduced ACD in aged human epidermis, possibly contributing to hypoplasia with age. Treatment of ALDH + CD44+ keratinocyte stem cells with nultin-3 (p53-activator) restored p53 levels and ACD frequency to adult levels [51].

ACD is reported in multiple cancers such as glioblastoma (GBM) and breast cancers. A pre-clinical model of GBM revealed the role

of ACD in generating daughter cells with enhanced therapy resistance driven by EGFR, thus amplifying tumour heterogeneity [52]. ACD can also help generate a subpopulation of 'G0-like' progeny cells through division of rapidly proliferating cells, as seen to be enriched upon chemotherapy in breast cancer [53]. Similarly, in triple-negative breast cancer, where three major subpopulations were seen (K14 + K18+, K18 +, and K18 + VIM–), the progenitor-like K14 + K18+ and luminal-like K14– K18 + could convert to one another through ACD, thus facilitating heterogeneity [54]. Further, in PC-3-derived prostate CSCs, the frequency of ACD depended on direct cell–cell interaction between CSC and non-CSC subpopulations [55], suggesting a mode akin to 'quorum sensing' in cellular decision-making at a population level [56].

The concept of ACD was recently invoked in two computational modelling efforts to explain the population dynamics of EMT noted experimentally [57,58]. PMC42-LA cells showed a bimodal distribution of levels of EpCAM (epithelial cell adhesion molecule), and when segregated, the EpCAM-high and EpCAM-low subpopulations could recapitulate the parental distribution of 80% cells in EpCAM-high state and 20% in EpCAM-low state [59]. A computational model that incorporated ACD via considering the asymmetric distribution in levels of SNAIL, an EMT-inducing transcription factor, could recapitulate these experimental observations [57]. While a causal role of ACD in driving this EMT dynamics remains to be established experimentally, the role of ACD in CSC dynamics and the association between EMT and CSCs [10,60] together argue for it being an important regulator of phenotypic switching in cancers through cell division.

4.2.3. Conformational noise

Many proteins lack rigid three-dimensional structures and instead exist as an ensemble of interconvertible conformations. Such proteins are known as intrinsically disordered proteins (IDPs). They can interact with multiple binding partners and often undergo disorder-to-order transitions upon binding to their specific targets, leading to promiscuous interactions among proteins involved in a signal transduction pathway [61]. Such stochastic fluctuations in protein conformations of IDPs or proteins containing intrinsically disordered regions (IDRs) are defined as 'conformational noise' [62] with possible long-reaching impact through its ability to amplify pre-existing noise in signalling pathways and in gene transcription [63,64]. IDPs or IDRs can lead to phenotypic switching through rewiring of multiple protein interaction networks. Moreover, feedback loops involving one or more IDPs can also enable multi-stability, thereby increasing stochastic cell-state switching [65] (**Figure 4.2**).

Prostate-associated gene 4 (PAGE4) is an IDP that acts as cancer/testis antigen and is highly expressed in human foetal prostate but undetectable in normal adult prostate. Its expression is up-regulated in prostate cancer (PCa) in response to various stress factors [66]. PAGE4 can be phosphorylated by two kinases: homeodomain-interacting protein kinase 1 (HIPK1) and CDC-like kinase 2 (CLK2). Phosphorylation of PAGE4 by these two kinases results in opposing functions: HIPK1-phosphorylated PAGE4 (HIPK1-PAGE4) potentiates c-Jun while CLK2-phosphorylated PAGE4 (CLK2-PAGE4) attenuates c-Jun activity. HIPK1-PAGE4 also exhibits a more compact conformation that can bind AP-1 compared to CLK2-PAGE4, which has a loose random coiling conformation making it hard for

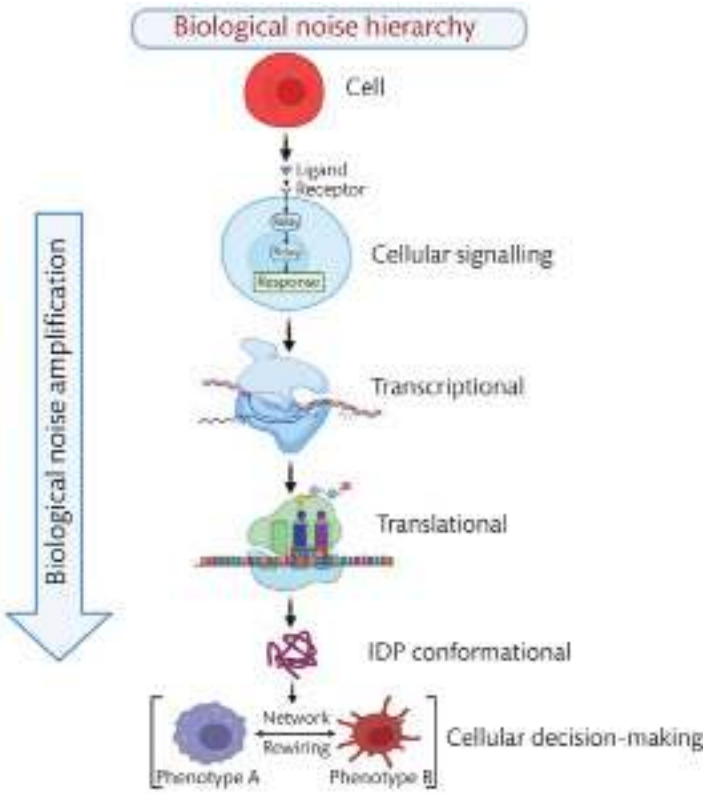


Figure 4.2. Different forms of cellular noise impacting phenotypic switching. *Source:* Adapted from Kulkarni et al. [62]. Noise or stochasticity can operate at each of the regulatory levels shown here, and the implications of noise shown at one of the levels here can amplify and/or propagate to the next level, thus governing phenotypic switching.

AP-1 to bind [67]. These different aspects can impact the dependence of PCa cells on androgen receptor (AR). Dynamical modelling of emergent dynamics of feedback loops involving PAGE4/AP-1/AR regulatory circuit reveals that cells can oscillate between the androgen-dependent and androgen-independent phenotypes [67]. Thus, in an isogenic PCa cell population, individual cells at a given timepoint may have varying levels of androgen dependence or independence, thus enhancing phenotypic heterogeneity.

A recent computational model investigated the emergent dynamics of a coupled circuit between SNAIL/miR-200/ZEB (core EMT circuit) and PAGE4/AP-1/AR circuits [65]. This coupled circuit enabled both oscillatory dynamics (the standalone behaviour of PAGE4/AP-1/AR circuit) and non-oscillatory one, where various phenotypes—epithelial, mesenchymal, and hybrid E/M—could coexist, showcasing multi-stability, depending on the strength of inhibition of ZEB1 by AR and that of AR by ZEB1. While epithelial phenotype usually co-occurred with oscillatory dynamics of PAGE4/AP-1/AR circuit, a transition to hybrid E/M or mesenchymal phenotype led to quenched oscillations and relatively low levels of AR. Thus, EMT induction may promote therapy resistance by enabling an androgen-independent low-AR PCa phenotype. Overall, this example highlighted how different mechanisms such as multi-stability in GRNs and conformational noise can crosstalk and enable phenotypic plasticity and heterogeneity in a cancer cell population.

4.2.4. Epigenetic remodelling

Cancer has been typically viewed as a clonal disease driven by inheriting or acquiring genetic mutations. However, many chromatin-based epigenetic mechanisms such as DNA methylation and histone modification have been identified in cancer progression. Given the dynamic and tissue-specific traits of epigenetic changes, their numbers can be potentially much more than that of genetic mutations [68]. Recent efforts in genome-wide sequencing have revealed that genetic and epigenetic changes are often intricately connected. For instance, consider the DNA repair enzyme MGMT that is often hypermethylated and silenced in many cancers. It normally protects from mutations happening at guanine base; thus, its inactivation can accelerate genome mutability. In GBM, MGMT promoter methylation sequence can influence the efficacy of treatments. Further, epigenetic changes can respond quite quickly to their microenvironmental cues compared to genetic changes that would need cell division(s) to accrue. With evidence pointing towards the possible absence of metastasis-driving mutations, analysing epigenetic alterations has gained a centre stage in metastasis [68].

Chromatin-based cell-state changes have been implicated in drug-tolerant cells as well. Upon exposure of PC9 lung cancer cells to gefitinib at a concentration 100-fold greater than the corresponding IC50 value for 9 days, a small (~0.3%) subpopulation of largely quiescent 'drug-tolerant persisters' (DTPs) was obtained. Approximately 20% of DTPs resumed proliferation even in the presence of the drug, yielding 'drug-tolerant expanded persisters' (DTEPs). When cultured in drug-free medium, DTPs resumed growth and reacquired sensitivity to drug. Chromatin-modifying enzyme KDM5A was found to be up-regulated in DTPs and DTEPs, and its down-regulation significantly reduced DTPs and DTEPs upon treatment of gefitinib or cisplatin. Thus, histone demethylase KDM5A was essential to allow phenotypic switching for reversible drug tolerance [69]. A follow-up study demonstrated that DTPs serve as reservoirs for the emergence of genetically diverse drug-resistant clones [70], thus showcasing another instance of interplay between genetic and epigenetic layers. Similar observations are reported in other cancer types such as melanoma and colorectal cancer [71–73].

The epigenetic cell state also underlies varying susceptibility of cells to undergo EMT/MET. When epithelial and mesenchymal populations from human breast cancer metastatic biopsies were isolated and functionally evaluated *in vivo*, the ones showing loss of EpCAM were 'locked' in a mesenchymal state and consequently had compromised metastatic potential. ZEB1 and GRHL2—key transcription factors driving EMT and MET, respectively—impacted global epigenetic programs that governed cellular plasticity and metastatic spread [19]. Such 'epigenetic memory' is likely to be lost after specific cell divisions, as predicted *in silico*, and shown conceptually *in vitro* [74], but corresponding *in vivo* analysis remains to be carried out to examine for plasticity. Further, cells undergoing EMT also undergo epigenetic remodelling, thus making an intuitive understanding hard to state, owing to feedback interplay between the dynamic genetic and epigenetic regulatory programs.

4.3. Conclusions

The dynamics of phenotypic switching and consequent non-genetic heterogeneity in cancer cells is beginning to be elucidated, owing to technological advancements such as single-cell analysis, lineage tracing, and integration with computational modelling [75]. Acquiring a predictive understanding of the multi-dimensional phenomenon of phenotypic switching can accelerate design of better therapeutic strategies to overcome adaptive drug resistance and restrict metastasis. Here, we highlight some key processes enabling phenotypic switching in multiple cancer types—multi-stability in GRNs, conformational noise, ACD, and epigenetic reprogramming—and present an argument to strengthen our conceptual understanding of these dynamical hallmarks of cancer too.

Competing interests

The authors declare no conflict of interest.

Author contributions statement

All authors contributed to writing and reviewing the manuscript.

REFERENCES

1. Jia D, Jolly MK, Kulkarni P, Levine H. Phenotypic plasticity and cell fate decisions in cancer: insights from dynamical systems theory. Cancers (Basel). 2017;**9**(7):70.
2. Liang G, Zhang Y. Embryonic stem cell and induced pluripotent stem cell: an epigenetic perspective. Cell Res. 2013;**23**(1):49–69.
3. Cole AJ, Fayomi AP, Anyaeche VI, Bai S, Buckanovich RJ. An evolving paradigm of cancer stem cell hierarchies: therapeutic implications. Theranostics. 2020;**10**(7):3083–98.
4. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to

phenotypic equilibrium in populations of cancer cells. Cell. 2011;**146**(4):633–44.
5. Enderling H. Cancer stem cells: small subpopulation or evolving fraction? Integr Biol (Camb). 2015;**7**(1):14–23.
6. Auffinger B, Tobias AL, Han Y, Lee G, Guo D, Dey M, et al. Conversion of differentiated cancer cells into cancer stem-like cells in a glioblastoma model after primary chemotherapy. Cell Death Dis. 2014;**21**:1119–31.
7. Andriani F, Bertolini G, Facchinetti F, Baldoli E, Moro M, Casalini P, et al. Conversion to stem-cell state in response to microenvironmental cues is regulated by balance between epithelial and mesenchymal features in lung cancer cells. Mol Oncol. 2016;**10**(2):253–71.
8. Vipparthi K, Hari K, Chakraborty P, Ghosh S, Patel AK, Ghosh A, et al. Emergence of hybrid states of stem-like cancer cells correlates with poor prognosis in oral cancer. iScience. 2022;**25**(5):104317.
9. Thankamony AP, Saxena K, Murali R, Jolly MK, Nair R. Cancer stem cell plasticity—a deadly deal. Front Mol Biosci. 2020;**7**:79.
10. Pasani S, Sahoo S, Jolly MK. Hybrid E/M phenotype(s) and stemness: a mechanistic connection embedded in network topology. J Clin Med. 2021;**10**(1):60.
11. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial–mesenchymal transition generates cells with properties of stem cells. Cell. 2008;**133**:704–15.
12. Deshmukh AP, Vasaikar S V., Tomczak K, Tripathi S, Den Hollander P, Arslan E, et al. Identification of EMT signaling cross-talk and gene regulatory networks by single-cell RNA sequencing. Proc Natl Acad Sci U S A. 2021;**118**(19):e2102050118.
13. Kröger C, Afeyan A, Mraz J, Eaton EN, Reinhardt F, Khodor YL, et al. Acquisition of a hybrid E / M state is essential for tumorigenicity of basal breast cancer cells. Proc Natl Acad Sci U S A. 2019;**116**(15):7353–62.
14. Sahoo S, Duddu AS, Biddle A, Jolly MK. Interconnected high-dimensional landscapes of epithelial–mesenchymal plasticity and stemness. Clin Exp Metastasis. 2022;**39**(2):279–90.
15. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. Nature. 2018 Apr;**556**(7702):463–8.
16. Tripathi S, Levine H, Jolly MK. The physics of cellular decision-making during epithelial–mesenchymal transition. Annu Rev Biophys. 2020;**49**:1–18.
17. Hari K, Ullanat V, Balasubramanian A, Gopalan A, Jolly MK. Landscape of epithelial mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. Elife. 2022;**11**:e76535.
18. Jia W, Tripathi S, Chakraborty P, Chedere A, Rangarajan A, Levine H, et al. Epigenetic feedback and stochastic partitioning during cell division can drive resistance to EMT. Oncotarget. 2020;**11**(27):2611–24.
19. Saini M, Schmidleitner L, Moreno HD, Donato E, Falcone M, Bartsch JM, et al. Resistance to mesenchymal reprogramming sustains clonal propagation in metastatic breast cancer. Cell Rep. 2023;**42**(6):112533.
20. Jolly MK, Ware KE, Xu S, Gilja S, Shetler S, Yang Y, et al. E-cadherin represses anchorage-independent growth in sarcomas through both signaling and mechanical mechanisms. Mol Cancer Res. 2019;**17**(6):1391–402.
21. Burger GA, Nesenberend DN, Lems CM, Hille SC, Beltman JB. Bidirectional crosstalk between epithelial–mesenchymal plasticity and IFNγ-induced PD-L1 expression promotes tumour progression. R Soc Open Sci. 2022;**9**(11):220186.
22. Galbraith M, Levine H, Onuchic JN, Jia D. Decoding the coupled decision-making of the epithelial–mesenchymal transition and metabolic reprogramming in cancer. iScience. 2023 Jan 20;**26**(1):105719.
23. Huang RYJ, Wong MK, Tan TZ, Kuay KT, Ng a HC, Chung VY, et al. An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (AZD0530). Cell Death Dis. 2013;**4**:e915.
24. Sahoo S, Mishra A, Kaur H, Hari K, Muralidharan S, Mandal S, et al. A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. NAR Cancer. 2021 Mar;**3**(3):zcab027.
25. Yu M, Bardia A, Wittner BS, Stott SL, Smas ME, Ting DT, et al. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science. 2013;**339**(6119):580–4.
26. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic plasticity: driver of cancer initiation, progression, and therapy resistance. Cell Stem Cell. 2019;**24**(1):65–78.
27. Gardner TS, Cantor CR, Collins JJ. Construction of a genetic toggle switch in Escherichia coli. Nature. 2000;**403**(6767):339–42.
28. Jia D, Jolly MK, Harrison W, Boareto M, Ben-Jacob E, Levine H. Operating principles of tristable circuits regulating cellular differentiation. Phys Biol. 2017;**14**(3):035007.
29. Zhou JX, Huang S. Understanding gene circuits at cell-fate branch points for rational cell reprogramming. Trends Genet. 2011;**27**(2):55–62.
30. Lu M, Jolly MK, Gomoto R, Huang B, Onuchic J, Ben-Jacob E. Tristability in cancer-associated microRNA-TF chimera toggle switch. J Phys Chem B. 2013;**117**(42):13164–74.
31. Brabletz S, Brabletz T. The ZEB/miR-200 feedback loop--a motor of cellular plasticity in development and cancer? EMBO Rep. 2010;**11**(9):670–7.
32. Lu M, Jolly MK, Levine H, Onuchic JN, Ben-Jacob E. MicroRNA-based regulation of epithelial-hybrid-mesenchymal fate determination. Proc Natl Acad Sci U S A. 2013 Nov;**110**(45):18144–9.
33. Yang X, Lin X, Zhong X, Kaur S, Li N, Liang S, et al. Double-negative feedback loop between reprogramming factor LIN28 and microRNA let-7 regulates aldehyde dehydrogenase 1-positive cancer stem cells. Cancer Res. 2010 Nov;**70**(22):9463–72.
34. Lu M, Jolly MK, Onuchic J, Ben-Jacob E. Toward decoding the principles of cancer metastasis circuits. Cancer Res. 2014;**74**(17):4574–87.
35. Parri M, Chiarugi P. Rac and Rho GTPases in cancer cell motility control. Cell Commun Signal. 2010;**8**:23.
36. Huang B, Lu M, Jolly MK, Tsarfaty I, Onuchic J, Ben-Jacob E. The three-way switch operation of Rac1/RhoA GTPase-based circuit controlling amoeboid-hybrid-mesenchymal transition. Sci Rep. 2014;**4**:6449.
37. Wang M, Cheng B, Yang Y, Liu H, Huang G, Han L, et al. Microchannel stiffness and confinement jointly induce the mesenchymal-amoeboid transition of cancer cell migration. Nano Lett. 2019;**19**(9):5949–58.
38. Aggarwal V, Sahoo S, Donnenberg VS, Chakraborty P, Jolly MK, Sant S. P4HA2: a link between tumor-intrinsic hypoxia, partial EMT and collective migration. Adv Cancer Biol– Metastasis. 2022 Oct 1;**5**:100057.
39. Matte BF, Kumar A, Placone JK, Zanella VG, Martins MD, Engler AJ, et al. Matrix stiffness mechanically conditions EMT and

migratory behavior of oral squamous cell carcinoma. J Cell Sci. 2019;**132**(1):jcs224360.
40. Sullivan E, Harris M, Bhatnagar A, Guberman E, Zonfa I, Ravasz Regan E. Boolean modeling of mechanosensitive epithelial to mesenchymal transition and its reversal. iScience. 2023 Apr 21;**26**(4):106321.
41. Li C, Hong T, Nie Q. Quantifying the landscape and kinetic paths for epithelial–mesenchymal transition from a core circuit. Phys Chem Chem Phys. 2016;**18**(27):17949–56.
42. Hong T, Watanabe K, Ta CH, Villarreal-Ponce A, Nie Q, Dai X. An Ovol2-Zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. PLoS Comput Biol. 2015;**11**(11):e1004569.
43. Tian XJ, Zhang H, Xing J. Coupled reversible and irreversible bistable switches underlying TGFβ-induced epithelial to mesenchymal transition. Biophys J. 2013;**105**:1079–89.
44. Font-Clos F, Zapperi S, Porta CAM La. Topography of epithelial–mesenchymal plasticity. Proc Natl Acad Sci U S A. 2018;**115**(23):5902–7.
45. Steinway SN, Zañudo JGT, Michel PJ, Feith DJ, Loughran TP, Albert R. Combinatorial interventions inhibit TGFβ-driven epithelial-to-mesenchymal transition and support hybrid cellular phenotypes. NPJ Syst Biol Appl. 2015;**1**:15014.
46. Subbalakshmi AR, Sahoo S, Manjunatha P, Goyal S, Kasiviswanathan VA, Mahesh Y, et al. The ELF3 transcription factor is associated with an epithelial phenotype and represses epithelial–mesenchymal transition. J Biol Eng. 2023;**17**:17.
47. Sahoo S, Mishra A, Kaur H, Hari K, Muralidharan S, Mandal S, et al. A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. NAR Cancer. 2021;**3**(3):zcab027.
48. Chhabra SN, Booth BW. Asymmetric cell division of mammary stem cells. Cell Div. 2021;**16**(1):5.
49. Li Z, Zhang YY, Zhang H, Yang J, Chen Y, Lu H. Asymmetric cell division and tumor heterogeneity. Front Cell Dev Biol. 2022;**10**:938685.
50. Bu P, Chen KY, Lipkin SM, Shen X. Asymmetric division: a marker for cancer stem cells? Oncotarget. 2013;**4**(7):950.
51. Charruyer A, Weisenberger T, Li H, Khalifa A, Schroeder AW, Belzer A, et al. Decreased p53 is associated with a decline in asymmetric stem cell self-renewal in aged human epidermis. Aging Cell. 2021;**20**(2):e13310.
52. Hitomi M, Chumakova AP, Silver DJ, Knudsen AM, Pontius WD, Murphy S, et al. Asymmetric cell division promotes therapeutic resistance in glioblastoma stem cells. JCI Insight. 2021;**6**(3):e130510.
53. Dey-Guha I, Wolfer A, Yeh AC, Albeck JG, Darp R, Leon E, et al. Asymmetric cancer cell division regulated by AKT. Proc Natl Acad Sci U S A [Internet]. 2011 Aug 2[cited 2023 Sep 1];108(31):12845–50. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.1109632108
54. Granit RZ, Masury H, Condiotti R, Fixler Y, Gabai Y, Glikman T, et al. Regulation of cellular heterogeneity and rates of symmetric and asymmetric divisions in triple-negative breast cancer. Cell Rep. 2018;**24**(12):3237–50.
55. Suzuki E, Masaka N, Urabe T, Sasaki M, Hasumi K. Direct cell–cell interaction regulates division of stem cells from PC-3 human prostate cancer cell line. Biochem Biophys Res Commun. 2022 Nov 26;**631**:25–31.
56. Ben-Jacob E, Coffey DS, Levine H. Bacterial survival strategies suggest rethinking cancer cooperativity. Trends Microbiol. 2012;**20**(9):403–10.
57. Jain P, Bhatia S, Thompson EW, Jolly MK. Population dynamics of epithelial-mesenchymal heterogeneity in cancer cells. Biomolecules. 2022;**12**(3):348.
58. Tripathi S, Chakraborty P, Levine H, Jolly MK. A mechanism for epithelial–mesenchymal heterogeneity in a population of cancer cells. PLoS Comput Biol. 2020;**16**(2):e1007619.
59. Bhatia S, Monkman J, Blick T, Pinto C, Waltham A, Nagaraj SH, et al. Interrogation of phenotypic plasticity between epithelial and mesenchymal states in breast cancer. J Clin Med. 2019 Jun;**8**(6):893.
60. Shibue T, Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. Nat Rev Clin Oncol. 2017;**14**(10):611–29.
61. Mahmoudabadi G, Rajagopalan K, Getzenberg RH, Hannenhalli S, Rangarajan G, Kulkarni P. Intrinsically disordered proteins and conformational noise Implications in cancer. Cell Cycle. 2013;**12**(1):26–31.
62. Kulkarni P, Salgia R, Rangarajan G. Intrinsically disordered proteins and conformational noise: The hypothesis a decade later. iScience. 2023;**26**(7):107109.
63. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008;**135**(2):216–26.
64. Baptista ISC, Ribeiro AS. Stochastic models coupling gene expression and partitioning in cell division in Escherichia coli. BioSystems. 2020;**193–194**:104154.
65. Singh D, Bocci F, Kulkarni P, Jolly MK. Coupled feedback loops involving PAGE4, EMT and notch signaling can give rise to non-genetic heterogeneity in prostate cancer cells. Entropy. 2021;**23**(3):288.
66. Kulkarni P, Dunker AK, Weninger K, Orban J. Prostate-associated gene 4 (PAGE4), an intrinsically disordered cancer/testis antigen, is a novel therapeutic target for prostate cancer. Asian J Androl. 2016;**18**(5):695.
67. Kulkarni P, Jolly MK, Jia D, Mooney SM, Bhargava A, Kagohara LT, et al. Phosphorylation-induced conformational dynamics in an intrinsically disordered protein and potential role in phenotypic heterogeneity. Proc Natl Acad Sci U S A. 2017;**114**(13):E2644–53.
68. Chatterjee A, Rodger EJ, Eccles MR. Epigenetic drivers of tumourigenesis and cancer metastasis. Semin Cancer Biol. 2018 Aug 1;**51**:149–59.
69. Sharma S V., Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell. 2010;**141**(1):69–80.
70. Ramirez M, Rajaram S, Steininger RJ, Osipchuk D, Roth MA, Morinishi LS, et al. Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. Nat Commun. 2016;**7**:10690.
71. Goyal Y, Busch GT, Pillai M, Li J, Boe RH, Grody EI, et al. Diverse clonal fates emerge upon drug treatment of homogeneous cancer cells. Nature. 2023;**620**(7974):651–9.
72. Rehman SK, Haynes J, Collignon E, Brown KR, Wang Y, Nixon AML, et al. Colorectal cancer cells enter a diapause-like DTP state to survive chemotherapy. Cell. 2021;**184**(1):226–242.e21.
73. Househam J, Heide T, Cresswell GD, Spiteri I, Kimberley C, Zapata L, et al. Phenotypic plasticity and genetic control in colorectal cancer evolution. Nature. 2022;**611**(7937):744–53.
74. Jain P, Corbo S, Mohammad K, Sahoo S, Ranganathan S, George JT, et al. Epigenetic memory acquired during long-term EMT induction governs the recovery to the epithelial state. J R Soc Interface. 2023;**20**(198):20220627.
75. Pillai M, Hojel E, Jolly MK, Goyal Y. Unraveling non-genetic heterogeneity in cancer with dynamical models and computational tools. Nat Comput Sci. 2023;**3**(4):301–13.

5

Morphological state transition during epithelial–mesenchymal transition

Biplab Bose

5.1. Introduction

Biologists for ages have collected information on the shape, size, and behaviour of living organisms. With the invention of the microscope, a new world opened before us. Using microscopes, we catalogued life at the cellular level. It allowed us to characterize human cells using size and shape. Together, size and shape define the morphology of a cell. We can also include information on subcellular structures such as the shape and size of the nucleus or cytoskeletal structures to define the morphology of a cell.

Morphology is controlled by the underlying molecular processes and connected with cellular functions and tissue organization. Therefore, cell morphology provides clues for many diseases. Since the early 20th century, cytological examinations of tumour samples have been used to characterize cancers [1]. Malignant cells are irregular in shape and size, with an unusually large nucleus [2]. These cells also display aberrations in the cellular ultrastructures. As morphology provides vital clues, pathologists, even in this age of multi-omics, look for morphological aberrations of cells to detect and grade tumours [3].

Microscopy is now a well-developed field of quantitative investigation. Developments in digital imaging, image processing, and statistical learning have opened several opportunities for the quantitative analysis of cell morphology. Imaging provides high-throughput, higher dimensional data that complements other cell profiling experiments like single-cell RNAseq. Morphological information obtained through imaging is now used to screen drugs and detect drug resistance [4–7]. Using high-throughput imaging, Wu et al. [8] observed that the morphology of cancer cells is highly heritable and predictive of their metastatic potential. Even attempts have been made to connect the gene expression profile of cells with their morphology [9,10].

This chapter focuses on the quantitative analysis of morphological changes during the epithelial–mesenchymal transition (EMT). In EMT, epithelial cells lose anchorage dependence, cell–cell adhesion, apico-basal polarity, and become more motile and mesenchymal-like [11]. Cells gain the ability to migrate and invade surrounding tissues, which is crucial for embryonic development and tissue repair [12]. However, EMT can also promote cancer cell metastasis and the emergence of drug resistance in cancer cells [13].

During EMT, cells undergo characteristic changes in gene expressions with a decrease in the expression of molecules associated with the epithelial phenotype (e.g. E-cadherin and β-catenin) and a concurrent increase in the expression of mesenchymal molecules (e.g. N-cadherin and vimentin) [14]. Coupled feedback loops involving multiple transcription factors of the Snail, Twist, and Zeb families, and several microRNAs, regulate the transcriptional reprogramming during EMT [15].

Expression of EMT markers and their temporal changes are widely used to investigate EMT *in vivo* and *in vitro*. Molecular techniques can be scaled up for many samples and a large panel of EMT markers. One can perform marker-based studies using patient samples. High-throughput techniques such as microarray and single-cell RNAseq further popularized the marker-based studies of EMT.

However, the expression of marker molecules is a proxy, not a direct measure of the phenotype of a cell. The structure and function of a cell define its phenotype. In every experimental system, cells undergo two observable changes during EMT—change in cell morphology and cell motility. The underlying molecular processes for these two changes are broadly conserved. Even then, molecular-level observations vary across different experimental systems, leading to confusion in defining EMT [11,16,17]. Further, most studies follow the change in the expression of a handful of markers and do not investigate other associated molecular processes, such as post-transcriptional or post-translational modifications. Therefore, it has been recommended [11] that the EMT status should be defined in terms of cellular properties, such as morphology and motility, rather than relying only on molecular markers.

Prior to the induction of EMT, epithelial cells remain tightly packed with strong cell–cell and cell–basement adhesion. Epithelial cell form tightly packed monolayer *in vitro* cell cultures. Under the microscope, these cells look like cobblestones [18]. With the progression of EMT, these cells lose contact and change shape to elongated or spindle types [18]. These cells disperse individually or in tinny clusters of cells. In certain cell lines, cell shape may change to circular [19].

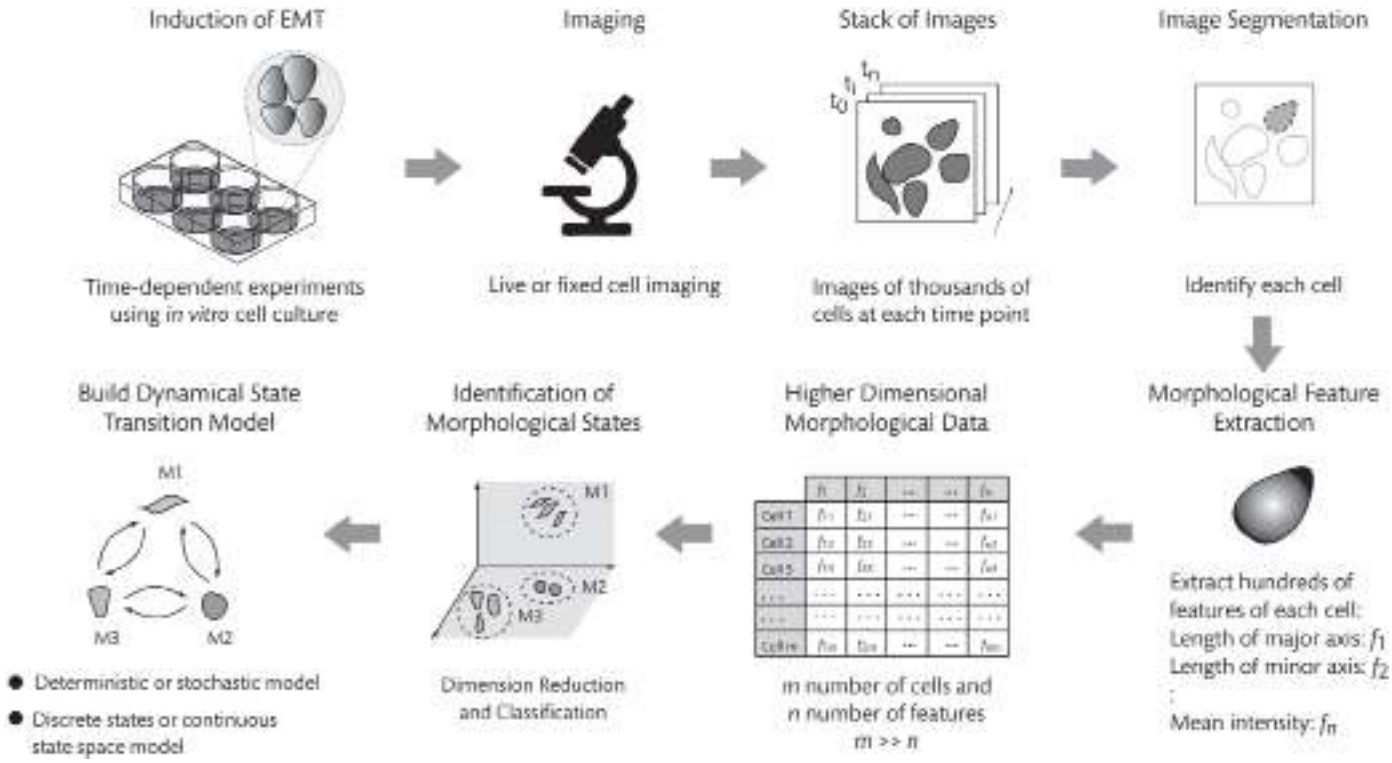


Figure 5.1. Quantitative investigation of the morphological state transition during EMT—different components.

This chapter defines a cell's phenotype or phenotypic state in terms of its morphology. For example, the cobblestone shape is one phenotype, whereas the spindle-type cell represents another phenotype. Therefore, the change in the phenotype of a cell during EMT is its transition through different morphological states. Through quantitative image analysis, we can identify cells of different morphology and study the dynamics of morphological state transition during EMT. The quantitative imaging-based approach can be extended to include information on cell motility and expression of the molecular markers. The experimental techniques discussed in this chapter are on *in vitro* cell culture-based studies. However, the concepts and mathematics presented in this chapter can be used in suitable *in vivo* experiments.

Figure 5.1 shows different components in the quantitative morphodynamical study of EMT. These components are discussed at length in this chapter. First, we discuss different imaging techniques. Subsequently, we discuss data generation through image processing and the identification of morphological states. Following these, we explore different mathematical modelling approaches for EMT and discuss the use of morphological data in those models. We also discuss the lessons learned through such morphodynamical studies of EMT.

5.2. Imaging the morphological dynamics in EMT

Phase contrast microscopy is the simplest way to observe the morphological changes during EMT. With a suitable digital camera, one can generate digital images that can be used for subsequent quantitative image analysis. However, these images are usually unsuitable for image analysis. The first step in image analysis is cell segmentation—identifying individual cells in the image. Images from phase contrast microscopy lack the adequate contrast required for automated cell segmentation.

Staining the cells with a dye resolves the contrast problem. Fluorescent dyes, like HCS cell mask, are used to stain the cytoplasm, and images are captured using an epi-fluorescence microscope [19]. These images are suitable for common cell segmentation protocols. One can additionally use DAPI and FITC-Phalloidin conjugate to counterstain the cell nucleus and the cytoskeleton, generating additional contrast and morphological information.

A popular multi-dye approach is the Cell Painting assay, where several cellular components are stained using six dyes [20]. One should use a multi-channel high-content imaging system to make the most of this assay.

Traditionally, fluorophore-tagged antibodies against EMT markers, like E-cadherin, and vimentin, are used in EMT experiments [18]. Fixed cells are treated with fluorophore-tagged antibodies and imaged using an epi-fluorescence microscope. Images of fluorescently labelled cells are suitable for image analysis. They provide additional molecular information apart from the morphology. However, cells usually have a heterogeneous and time-varying expression of these markers. Therefore, labelling with antibodies may not always help segment cells. We recommend that readers consult the article by Moreno-Bueno et al. [18] prior to designing any imaging experiment.

One limitation of dye/antibody-based techniques is that we must fix the cells before staining and imaging. Therefore, the cells imaged at different time points are different. For these experiments, multiple

flasks/plates are identically treated and fixed at different times for imaging. The data generated through these experiments are aggregate snapshot data and add limitations in data analysis.

Live-cell imaging is an alternative to fixed-cell imaging. Through live-cell imaging, one can track the changes in the morphology of individual cells over time and make direct quantitative estimates of the morphological dynamics. However, live-cell imaging is technically challenging as it requires a special imaging set-up and additional computational tools for tracking individual cells. We recommend that readers consult the articles by Ettinger and Wittmann [21] and Nketia et al. [22] to obtain an overview of live-cell imaging.

We can use cells expressing fluorescent proteins to facilitate the quantitative analysis of live-cell images [23,24]. Fluorescent markers have been developed targeting specific cellular components that work like the dyes used in the Cell Painting assay [23,24].

Wang et al. [25] used a cell line that expresses fluorescently tagged vimentin. They performed live-cell imaging simultaneously using differential interference contrast (DIC) and fluorescent microscopy to chase the TGF-β-induced EMT in these cells. DIC images generally have better contrast than phase contrast microscopy. They extracted the morphological information from the DIC images. The fluorescently tagged vimentin complimented the morphological data to decipher the cell's trajectory during the EMT.

Digital holographic microscopy and quantitative phase imaging (QPI) are gaining popularity as they allow label-free, three-dimensional imaging of live cells [26,27]. Lam et al. [28] used QPI to classify cells as epithelial or mesenchymal. Kamlund et al. [29] used holographic microscopy to monitor the progression of EMT in live cells.

Algorithms and computational pipelines have also been developed to process label-free images [30–32]. These algorithms expand the scope of conventional transmitted light imaging techniques (such as brightfield, phase contrast, and DIC).

5.3. Image processing and feature extraction

Good-quality images are the raw data for morphological studies. Post imaging, we process those images to identify each cell and extract quantitative morphological information.

For a few images with a handful of cells, we can manually identify each cell. However, quantitative image analysis needs many images with thousands of cells. Therefore, we have to use algorithms to automate the process.

Identifying individual cells in an image using an algorithm is known as segmentation. Cell segmentation is not a trivial job. Cells are usually irregular in shape and are not standard geometric objects. Further, in monolayers, multiple cells remain adhered to each other. That makes it difficult to differentiate two cells, particularly for unlabelled cells.

There are several types of segmentation algorithms [22]. The traditional methods are threshold-based algorithms and identify the cells using the difference in pixel intensities/colour values of a cell and the background [33,34]. Other algorithms, like the watershed, also use the difference in pixel intensities to differentiate a cell from the background [35].

Of late, machine learning algorithms have been gaining popularity in image segmentation. Several general-purpose machine learning algorithms have been implemented to segment and identify cells [36,37].

Extraction of quantitative morphological data goes hand in glove with image segmentation. As mentioned earlier, cells are not of defined geometric shapes. A cell identified by us as circular or elliptical need not be precisely a circle or an ellipse. Further, two cells that we perceive as similar in morphology may have subtle but crucial differences. Therefore, we need to define and measure different quantitative attributes related to the morphology of a cell. These attributes are called 'morphological features' and collectively used to define the morphology of a cell.

Popular image analysis tools can extract hundreds of morphological features from the image of a cell. Morphological features can be of two types. One set of features is related to the cell's size and geometry, and the other quantifies non-geometric aspects such as textures and brightness. Features such as area and perimeter capture the size of a cell. Some common geometric features are form factor, eccentricity, circularity, compactness, aspect ratio, solidity, Hu moment, and lengths of the major and minor axes.

The texture and brightness of a cell image provide additional morphological information. Texture-related features, like Haralick features [38], provide information on the roughness and smoothness of a cell. Different pixel intensity-related features, such as integrated intensity, mean intensity, and distribution of pixel intensities, help differentiate different cells.

When cells are counterstained for the nucleus or EMT markers, we can perform additional image segmentation for those counterstains. Subsequently, relevant feature information can be extracted for these secondary objects [39]. For example, if DAPI is used to stain the nucleus, additional segmentation is used to identify the nucleus of each cell and then extract morphological features of each nucleus.

All the morphological features that we have discussed are predefined. The purpose of feature extraction is to represent a cell in terms of the quantitative values of these morphological features. When we use predefined features, every cell is a point in a multidimensional hyperspace defined by those predefined features. However, one can argue that a cell image may have some unknown features that are more effective in capturing the morphological heterogeneity in a population of cells.

Machine-learning-based techniques are now developed as an alternative to predefined feature-based analysis. Deep learning algorithms, such as variational auto encoder (VAE) and convolution neural network, are used to embed each cell image in a low-dimensional hyperspace [40–43]. The quantitative information from this embedding space is used as features in subsequent analysis. The dimensions of the embedding hyperspace do not have any direct one-to-one connection with any particular morphological aspect. Even then, these methods effectively identify cellular heterogeneity, classify cells, and capture morphodynamics [40,41,44,45].

Popular programming languages, such as MATLAB, Python, and R, have functions/packages for image analysis that can be used to create a complete image analysis pipeline—from image segmentation to feature extraction. There are standalone software for performing large-scale analysis for microscopy images. ImageJ [46] and CellProfiler [47] are two popular open-source platforms widely

used to segment images, identify cells, and extract morphological features.

The choice of image processing and analysis tools depends upon several factors—the imaging technique, the number of images, the type of quantitative data to be generated, and the user's skills in computation. Therefore, one should decide on the image analysis tools early in the study. Algorithms used for image segmentation and feature extraction requires user-defined hyperparameters. Therefore, one must keep a good image analysis record and use the pipeline consistently with all batches of images. Further, we must be careful about image processing and feature extraction quality. The reader may consult the article by Caicedo et al. [48] to understand the critical issues in single-cell image analysis.

5.4. Identifying the morphological states

In EMT, epithelial cells acquire mesenchymal features. These cell types, epithelial and mesenchymal-like, are two phenotypic states. However, recent studies indicate that EMT is not a two-state process and may involve one or more intermediate states [49–51]. These intermediate states are known by different names—hybrid states, partial EMT, and incomplete EMT [51].

Whatever the number of states, we need to identify those in a population of cells undergoing EMT and investigate the state transition dynamics. Therefore, the dynamical study in EMT involves two steps—classifying cells in different phenotypic states and developing a quantitative model for the dynamics of cells over these states.

Most of the existing works in EMT define the state of a cell using molecular markers. However, in this chapter, we define the phenotypic states based on the morphology of cells. As discussed in the preceding sections, we can generate quantitative data on hundreds of morphological features of cells using image analysis. Subsequently, we have to identify the phenotypic state of a cell using those morphological features.

So, we have a classification problem where we assign a cell to a particular state, such as epithelial, hybrid, or mesenchymal, based on morphological features. However, the morphology of a cell cannot be defined using a handful of features. One extracts hundreds of features to characterize the morphology of a cell. Therefore, the identification of the morphological state of a cell is a multivariate multi-class classification problem.

Multivariate classification is not new in cell biology. Single-cell gene expression data are often used to identify cell types using multivariate classification algorithms [52,53]. We can use similar algorithms but use morphological feature information instead of gene expression data.

One approach could be where we have already decided on the phenotypic states in EMT and use a classifier [23,54,55] to identify cell states. Devaraj and Bose [19] used this approach in their work on morphological state transition dynamics in EMT. Through microscopy, they observed that cells in their experiments could be divided into three morphological groups—cobble, spindle, and circular. Upon induction of EMT, the population distribution of cells in these three morphological states changed with time. Further experiments showed that spindle and circular cells were migratory, but cobbles were non-migratory.

They used CellProfiler Analyst [56] to classify cells using the morphological features extracted from microscopic images of fixed cells stained with a dye. CellProfiler Analyst provides several classifiers, such as Random Forest, Support Vector Machines, and AdaBoost [57]. The user trains any of these classifiers using the training dataset and subsequently uses it to classify the cells in the test data in predefined morphological states.

To train the classifier, Devaraj and Bose [19] used a training dataset (~100 images per cell type) and manually labelled those cells as cobble or spindle or circular based on their visual perception. The training data must have a balanced representation of all the cell types. Therefore, one may have to pool individual cell images from samples imaged at different time points.

Whatever the method used, the trained classifier must be evaluated for its performance [58]. Confusion matrix and ROC are widely used to evaluate the performance of a classifier. CellProfiler Analyst also provides tools for the evaluation of a classifier. The number of images used for training affects the performance of a classifier. Therefore, the size of the training data must be optimized by comparing the performance of classifiers trained with datasets of different sizes.

Devaraj and Bose [19] assumed three morphological states. However, while investigating EMT in MCF-10A cells, Leggett et al. [59] divided cells into two discrete states, epithelial and mesenchymal. They trained a Gaussian mixture model for these two types of cells and used that classifier to estimate changes in the distribution of these two subpopulations during TGF-β-induced EMT.

We must decide the number of classes or labels *a priori* for any classification problem. EMT must have at least two classes, epithelial and mesenchymal. However, deciding the number of intermediate phenotypic states is not trivial. To circumvent this issue, we can use clustering algorithms to decide the number of phenotypic states during EMT.

Clustering is unsupervised. We partition the given samples into distinct subsets through clustering. An algorithmic approach is used to decide the suitable number of subsets or clusters. Clustering algorithms are widely used for phenotypic identification using flow cytometry [60] and single-cell gene expression data [52,61].

The *k*-means algorithm and its variants are popular clustering algorithms [62–64]. Hierarchical clustering [65], density-based clustering [66], and graph-based clustering [67] algorithms are also used to identify cell types. We can re-purpose these algorithms to identify distinct clusters in morphological feature space. Each cluster would represent a unique phenotypic state.

Single-cell gene expression data is higher dimensional, involving thousands of genes. Dimension reduction techniques, like principle component analysis (PCA) [68], are used to reduce the dimension of the data prior to clustering. Morphological feature space is also of higher dimension with hundreds of features. Therefore, the data dimension must be reduced before clustering. Dimension reduction helps in the clustering and visualization of clusters.

PCA is a linear method that captures the global variations within data. Therefore, it may not perform well for imaged-based data. Non-linear dimension reduction techniques, such as *t*-distributed stochastic neighbour embedding [69] and uniform manifold approximation and projection [70], can be used in place of PCA.

Deep learning has also been used to identify cells of different phenotypes from microscopic images [71]. Often, these algorithms

perform feature extraction and phenotypic classification directly from the segmented images. UPSIDE is a deep learning pipeline that uses brightfield images and performs unsupervised discovery of morphological states of cells [40]. It uses a VAE to learn the lower dimensional latent features that can be used to regenerate the images. Upon training, each cell is represented by a vector of latent features. Subsequently, UPSIDE clusters the cells in the latent space using the Louvain method for community detection. Each cluster is a morphological state.

Wu et al. [31] used another VAE, vector quantized variational autoencoders, to represent cells in a lower dimensional latent feature space. They performed PCA of the data in the latent space to find that the first few principal components well correlate with cellular properties such as geometry and optical density. Subsequently, they used the top principle components and displacement of cells during live-cell imaging to create trajectory feature vectors (TFVs) for each cell. A Gaussian mixture model was used to identify clusters of cell types in the TFV space.

Longden et al. [4] used standard feature extraction tools to extract hundreds of morphological features from millions of cells in an assay for drug resistance. Subsequently, they reduced the dimension of the feature space using an autoencoder. The lower dimensional latent feature space allowed them to discriminate between different types of cells.

An autoencoder encodes the information in such a way that images can be recreated by decoding the information in the latent space. That makes the latent space embedding interpretable in terms of the morphology of a cell. Further, one can use the autoencoder to predict a cell shape, making it a predictive model.

5.5. The dynamics of morphological state transition

Once we have classified cells in different morphological states, we can count the number of cells in a particular state at a particular time. So, the quantitative image analysis of EMT provides the population distribution of cells in different morphological states. During EMT, this population distribution changes with time. How can we use this information to understand the underlying dynamics and the governing principles of EMT?

The search for the governing principle behind the phenotypic changes of cells is an old problem in biology. In 1957, Waddington [72] proposed the metaphor of the epigenetic landscape for the emergence of diverse cell types through the differentiation of cells during embryogenesis. He visualized the unidirectional differentiation process as rolling a pebble over a landscape of hills and valleys. A pluripotent cell rolls down from a high hill to valleys through a series of branching points. Each valley corresponds to a particular phenotypic state. In this metaphor, differentiation is a journey over a potential landscape, from higher to lower potential.

We can use a similar potential landscape metaphor to understand EMT. In the landscape view of EMT, each phenotypic state is a well (or local minima) in the potential landscape, and cells move from one well to another. However, how would we connect the metaphor of potential landscape with the quantifiable cellular processes?

Works of Jacob and Monod [73], Kauffman [74], Thomas [75], and others led to the idea that gene regulatory networks can have multiple stable steady states (or attractors) and that the phenotypes of a cell are those stable steady states. This idea connected Waddington's landscape to the dynamical systems theory. Each low-potential well or valley of the landscape is a stable steady state of the underlying molecular network.

This dynamical systems view is the prime driver in the mathematical modelling of EMT. We can represent the gene regulatory network of EMT by a system of ordinary differential equations. Concentrations of molecules are the dependent variables in this model. Therefore, the molecules involved in EMT define the state space, and the trajectories in this state space represent the dynamics of cells during EMT.

From an initial position in this state space, a cell follows a trajectory and eventually reaches a stable steady state. The system can have multiple stable steady states, each representing a unique phenotypic state. Interestingly, the number of stable steady states in an EMT network can vary with the value of one or more control parameters. This property is known as the bifurcation in a dynamical system. The EMT-inducing signals control bifurcation parameters. Depending upon the architecture, EMT networks show different types of bifurcation. Readers may consult the articles by Tripathi et al. [15] for a comprehensive outline of the dynamical systems view of EMT.

In molecular network-based studies, one measures the expression of different EMT-related molecules and represents their relationships using a system of ODEs. In the morphology-based approach, we measure changes in different morphological features during EMT. However, we do not know how these features are related. Therefore, we cannot construct an ODE-based model using these morphological features.

Alternatively, one may relate the morphological state of a cell with its molecular state and then use the ODE-based model of molecular networks to study the morphological state transition dynamics. Attempts have been made to connect the morphology of a cell to its gene expression state [9,10]. However, the mapping between gene expression and morphology may not be injective. Therefore, such an approach should be used with adequate caution.

Discrete state transition models are better suited for morphology-based studies. Such models assume that there are several distinct morphological states, and during EMT, cells jump from one state to another. If state transitions are stochastic, we can consider them Markov processes [76]. In a Markov process, the system stochastically jumps from one state to another with a probability that depends only upon the system's current state.

There are different state transition models. The simplest model is a sequential, irreversible transition through intermediates from the epithelial to the mesenchymal state. The sequential model can also be reversible. The most elaborate model assumes reversible transitions between any two states.

Suppose, during EMT, a cell can be in m possible morphological states and jump from one state to another. Let the probability that a cell is at the kth state at time t be π_k^t, and the probability of transition from the ith to the jth state be p_{ij}. The probability transition matrix is $\mathbf{P} = \left[p_{ij}\right]_{i,j=1}^{m}$. $\mathbf{P}$ is an $m \times m$ stochastic matrix such that $0 \le p_{ij} \le 1$ and $\sum_{j=1}^{m} p_{ij} = 1$.

If the state of the system at time t is $\prod_t = \left[\pi_1^t, \pi_2^t, \ldots, \pi_m^t\right]$, then the state at $t+1$ is

$$\prod_{t+1} = \prod_t \mathbf{P} \tag{5.1}$$

Equation (5.1) is the governing equation of the system and provides the lineage trajectory during EMT.

Here, we have assumed that the transition probabilities do not change with time. Though this assumption of time homogeneity is reasonable for many systems, one must be careful before making it.

Through the quantitative image analysis, we calculate the frequency of cells in each state. These frequencies are used to construct the state vectors. In live-cell imaging, the same plate/flask is imaged at sequential time points. There are algorithms to track a cell in those sequential image frames. Some tools perform both cell tracking and cell classification [23]. Following individual cells, one can calculate the transition frequency from one cell type to another. These frequencies are used to construct the transition probability matrix **P**.

Gordonov et al. [24] developed a live-cell image analysis pipeline (SAPHIRE) that performs cell tracking, phenotypic identification, and estimating state transition probabilities. SAPHIRE uses PCA for dimension reduction. Subsequently, it uses a hidden Markov model with Bayesian model selection to generate the state transition model with estimated states and state transition probabilities.

Estimating **P** from fixed-cell imaging data is not trivial. Here, we are not following the same cells over time. Instead, different samples are fixed and imaged at different time points. Therefore, we obtain aggregate data by estimating the number of cells in different states for each time point.

There are methods to estimate the state transition probabilities from aggregate data [76]. Different regression models are used to fit the data to Equation (5.1) and estimate **P** [77–79]. Buder et al. [80] developed a root-finding method to estimate **P** for experiments where one starts with a pure population of cells. Farahat and Asada [81] proposed a Bayesian estimation of transition probabilities. Karacosta et al. [82] used convex optimization with sparsity constraint and bootstrapping to estimate transition probabilities for an EMT model from aggregate data.

One can use a differential-equation-based state transition model. Let there be m discrete morphological states. The probability of a cell in the kth state is q_k. The rate of change of this probability is given by

$$\frac{dq_k}{dt} = \sum_{i \neq k} k_{ik} q_i - \sum_{j \neq k} k_{kj} q_k \tag{5.2}$$

Here, k_{ik} is the rate constant for the transition from the ith to the kth state.

Considering all the states, we obtain a system of ODEs:

$$\frac{d}{dt}\mathbf{q} = \mathbf{K}\mathbf{q} \tag{5.3}$$

Here, **q** is the state vector with the probabilities for different states, and **K** is the coefficient matrix with the rate constants for state transitions. **K** is estimated by fitting this model to data.

Goetz et al. [83] used an ODE-based model to study EMT dynamics. They assumed a sequential irreversible three-state model—epithelial, intermediate, and mesenchymal. These three states were defined in terms of the expression of E-cadherin and vimentin.

However, they observed that the three-state model did not fit well with the data. Therefore, they assumed that these three experimentally observed states are macrostates, each having multiple hidden microstates. So, they increased the number of intermediate states in their model. They used an iterative model-fitting approach to decide the number of intermediate states.

Another approach for modelling state transition is to use discrete-time difference equations. For a system with m states, we can write

$$N_k\left(t + \Delta t\right) = \sum_{i=1}^{m} r_{ik} N_i\left(t\right) \tag{5.4}$$

Here, $N_k\left(t\right)$ and $N_k\left(t + \Delta t\right)$ are the number of cells in the kth state at time t and $t + \Delta t$, respectively. r_{ik} is the rate of transition from the ith to the kth state.

Devaraj and Bose [19] used difference equations to model EMT. They assumed three morphological states in EMT and included cell death and birth in their model. Further, they assumed time-varying rate parameters. They fitted the model using a constrained optimization algorithm and determined the dominant state transition paths. Their model showed that the morphological state transition in their experimental system was reversible and had the signature of hysteresis or memory.

The morphological state space is continuous. Nevertheless, till now, we have considered discrete morphological states. We can eliminate this assumption and investigate EMT as a continuous movement of cells in the multidimensional morphological space. In this regard, one can use the approach taken by Chang and Marshall [84].

They used live-cell imaging to study the morphodynamics of mouse embryonic cells. First, they reduced the multidimensional feature data to two dimensions by PCA. Subsequently, transition vectors for all the cells were calculated in this reduced state space. Interestingly, their analysis showed that the detailed balance is valid in this system. Therefore, they used the equilibrium formalism and calculated a potential landscape using the probability density of occupancy of different positions in the two-dimensional state space. The assumption of detailed balance may not be valid for EMT. Even then, one can use the transition vectors to investigate the dynamics in reduced state space and identify novel dynamical features.

Wang et al. [25] used live-cell imaging to study TGF-β-induced EMT in cells expressing fluorescent vimentin. They used PCA to separately reduce the dimension of morphological and Haralick feature spaces and selected four dominant principal components. Time-varying data of each cell was projected in this four-dimensional space, and single-cell trajectories were identified. Trajectories analysis showed that this experimental system had two types of paths for EMT.

They further extended this work [85] by reconstructing reaction coordinates (RCs) and pseudo-potentials to understand the transition paths. As per the reaction rate theory, an RC is a scalar geometric parameter that describes the progression along a reaction path in a multidimensional state space. They concluded that TGF-β treatment destabilizes the epithelial state attractor that eventually disappears, and cells move to the mesenchymal attractor. However, this process involves two sequential collisions between an attractor and saddle, generating two paths for state transition.

5.6. Conclusions and future directions

High-throughput imaging platforms are getting more sophisticated, cheaper, and easy to use. Quantitative analysis of images captured through these platforms generates vast higher dimensional morphological data. Like single-cell gene expression data, single-cell morphological data comes in the bracket of 'big data'. Therefore, big data analysis tools should be fully utilized to study cellular morphodynamics.

Several algorithms have been developed in the past decade for lineage tracing and inferring cellular trajectory from single-cell transcriptomic and flow cytometry data [86]. These algorithms consider a cell or a cluster of similar cells as a dot in the multidimensional molecular expression space and connect those dots to create trajectories for the emergence of one cell type from another. These algorithms can be suitably appropriated for morphological data to identify different cell types during EMT and to infer the trajectories between these cell types.

The trajectory inference algorithm SCUBA [87] uses time-course single-cell gene expression data. It assumes that every branching point in the trajectory is binary. It uses an iterative clustering algorithm to estimate the cellular hierarchy and build the trajectory. Several other trajectory inference algorithms, such as Waddington-OT [88], CSHMM [89], and Cstreet [90], use time series data and possibly are suited for inferring the morphological state transition trajectories during EMT.

We often have vast data in many natural and engineering problems whose underlying governing equations are unknown. In some instances, we may know the governing equations, but the high dimensionality of state space limits computational analysis. Data-driven algorithms are very successful in identifying dominant spatial and temporal modes in the data and creating reduced-order models [91,92]. We believe biologists can leverage those algorithms to find the governing principles for morphological state transition in EMT.

In the same vein, Taylor-King et al. [93] have developed the algorithm 'dynamic distribution decomposition (DDD)' to capture the dynamics of cells in a higher dimensional space using a linear combination of a set of basis functions. They used this algorithm to analyse single-cell mass cytometry data of iPSC reprogramming. DDD may be used to analyse higher dimensional morphological feature data and capture the key states and trajectories during EMT.

According to the dynamical systems theory, bifurcation in the regulatory network gives rise to EMT [15]. Investigation of bifurcation requires knowledge of the regulatory network and an ODE-based model. However, morphological data is not suitable for an ODE-based study of bifurcation. Even then, one may capture signatures of bifurcation directly from morphological data.

The transition through the bifurcation point is equivalent to the phase transition across a critical point. Near the critical point, systems show critical slowing down and high fluctuation. Correlations and fluctuation in the state variables can be used as signatures of critical transition [94,95]. Mojtahedi et al. [96] used this approach to study critical transition during the differentiation of murine multipotent cells. They showed that increased gene–gene correlation and reduced cell–cell correlation are signatures of critical transition during differentiation.

Similar ideas were used to study gene expression in cancer metastasis [97] and mathematical investigation of the EMT network [98]. We propose that one may use single-cell morphological feature data in place of gene expression data to detect signatures of critical transition during EMT.

Till now, only a handful of studies have used morphological data to study the dynamics of EMT [19,25,29,39,85]. These investigations have shown that the data from quantitative image analysis are robust and suitable for mathematical modelling. Further, these morphodynamical studies successfully detected critical dynamical features of EMT—bifurcation and hysteresis. We believe that the quantitative morphological study will soon become an integral part of EMT research and complement the molecular data.

REFERENCES

1. Hajdu SI, Ehya H. A note from history: foundation of diagnostic cytology. Ann Clin Lab Sci. 2008;**38**(3):296–9.
2. Baba AI, Catoi C. Comparative oncology. Bucharest: Publishing House of the Romanian Academy; 2007.
3. Sherman ME, Troester MA, Hoadley KA, Anderson WF. Morphological and molecular classification of human cancer. In: Thun M, Linet MS, Cerhan JR, Haiman CA, Schottenfeld D, editors. Cancer epidemiology and prevention. New York, NY: Oxford University Press; 2017.
4. Longden J, Robin X, Engel M, Ferkinghoff-Borg J, Kjær I, Horak ID, et al. Deep neural networks identify signaling mechanisms of ErbB-family drug resistance from a continuous cell morphology space. Cell Rep. 2021;**34**(3):108657.
5. Xin L, Xiao W, Che L, Liu J, Miccio L, Bianco V, et al. Label-free assessment of the drug resistance of epithelial ovarian cancer cells in a microfluidic holographic flow cytometer boosted through machine learning. ACS Omega. 2021;**6**(46):31046–31057.
6. Tanaka M, Bateman R, Rauh D, Vaisberg E, Ramachandani S, Zhang C, et al. An unbiased cell morphology-based screen for new, biologically active small molecules. PLoS Biol. 2005;**3**(5):0764–76.
7. Ziegler S, Sievers S, Waldmann H. Morphological profiling of small molecules. Cell Chem Biol. 2021;**28**(3):300–19.
8. Wu PH, Gilkes DM, Phillip JM, Narkar A, Cheng TWT, Marchand J, et al. Single-cell morphology encodes metastatic potential. Sci Adv. 2020;**6**(4):1–8.
9. Chlis NK, Rausch L, Brocker T, Kranich J, Theis FJ. Predicting single-cell gene expression profiles of imaging flow cytometry data with machine learning. Nucleic Acids Res. 2020;**48**(20):11335–46.
10. Wakui T, Negishi M, Murakami Y, Tominaga S, Shiraishi Y, Carpenter AE. Predicting gene expression from cell morphology in human induced pluripotent stem cells. bioRxiv. 2022.04.19.488786. https://doi.org/10.1101/2022.04.19.488786
11. Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M, et al. Guidelines and definitions for research on epithelial–mesenchymal transition. Nat Rev Mol Cell Biol. 2020;**21**(6):341–52.
12. Thiery JP, Acloque H, Huang RYJ, Nieto MA. Epithelial–mesenchymal transitions in development and disease. Cell. 2009;**139**(5):871–90.
13. Mittal V. Epithelial mesenchymal transition in tumor metastasis. Annu Rev Pathol Mech Dis. 2018;**13**:395–412.

14. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial–mesenchymal transition. Nat Rev Mol Cell Biol. 2014;**15**(3):178–96.
15. Tripathi S, Levine H, Jolly MK. The physics of cellular decision making during epithelial–mesenchymal transition. Annu Rev Biophys. 2020;**49**:1–18.
16. Nilsson GM a, Akhtar N, Kannius-Janson M, Baeckström D. Loss of E-cadherin expression is not a prerequisite for c-erbB2-induced epithelial–mesenchymal transition. Int J Oncol. 2014;**45**(1):82–94.
17. Hollestelle A, Peeters JK, Smid M, Timmermans M, Verhoog LC, Westenend PJ, et al. Loss of E-cadherin is not a necessity for epithelial to mesenchymal transition in human breast cancer. Breast Cancer Res Treat. 2013;**138**(1):47–57.
18. Moreno-Bueno G, Peinado H, Molina P, Olmeda D, Cubillo E, Santos V, et al. The morphological and molecular features of the epithelial-to-mesenchymal transition. Nat Protoc. 2009;**4**(11):1591–613.
19. Devaraj V, Bose B. Morphological state transition dynamics in EGF-induced epithelial to mesenchymal transition. J Clin Med. 2019;**8**(7):911.
20. States C, Gustafsdottir SM, Ljosa V, Sokolnicki KL, Wilson JA, Walpita D, et al. Multiplex cytological profiling assay to measure diverse. PLoS ONE. 2013;**8**(12):1–7.
21. Ettinger A, Wittmann T. Fluorescence live cell imaging. Methods Cell Biol. 2014;**123**:77–94.
22. Nketia TA, Sailem H, Rohde G, Machiraju R, Rittscher J. Analysis of live cell images: Methods, tools and opportunities. Methods. 2017;**115**(2017):65–79.
23. Held M, Schmitz MHA, Fischer B, Walter T, Neumann B, Olma MH, et al. CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. Nat Methods. 2010;**7**(9):747–54.
24. Gordonov S, Hwang MK, Wells A, Gertler FB, Lauffenburger DA, Bathe M. Time series modeling of live-cell shape dynamics for image-based phenotypic profiling. Integr Biol (United Kingdom). 2016;**8**(1):73–90.
25. Wang W, Douglas D, Zhang J, Kumari S, Enuameh MS, Dai Y, et al. Live-cell imaging and analysis reveal cell phenotypic transition dynamics inherently missing in snapshot data. Sci Adv. 2020;**6**(36):eaba9319.
26. Balasubramani V, Kujawińska M, Allier C, Anand V, Cheng CJ, Depeursinge C, et al. Roadmap on digital holography-based quantitative phase imaging. J Imaging. 2021;**7**(12):252.
27. Park YK, Depeursinge C, Popescu G. Quantitative phase imaging in biomedicine. Nat Photonics. 2018;**12**(10):578–89.
28. Lam VK, Nguyen TC, Bui V, Chung BM, Chang L-C, Nehmetallah G, et al. Quantitative scoring of epithelial and mesenchymal qualities of cancer cells using machine learning and quantitative phase imaging. J Biomed Opt. 2020;**25**(02):026002.
29. Kamlund S, Janicke B, Alm K, Judson-Torres RL, Oredsson S. Quantifying the rate, degree, and heterogeneity of morphological change during an epithelial to mesenchymal transition using digital holographic cytometry. Appl Sci. 2020;**10**(14):4726.
30. Li Y, Nowak CM, Pham U, Nguyen K, Bleris L. Cell morphology-based machine learning models for human cell state classification. NPJ Syst Biol Appl. 2021;**7**(1):1–9.
31. Wu Z, Chhun BB, Popova G, Guo SM, Kim CN, Yeh LH, et al. DynaMorph: self-supervised learning of morphodynamic states of live cells. Mol Biol Cell. 2022;**33**(6):ar59.
32. Vicar T, Balvan J, Jaros J, Jug F, Kolar R, Masarik M, et al. Cell segmentation methods for label-free contrast microscopy: Review and comprehensive comparison. BMC Bioinform. 2019;**20**(1):1–25.
33. Dima AA, Elliott JT, Filliben JJ, Halter M, Peskin A, Bernal J, et al. Comparison of segmentation algorithms for fluorescence microscopy images of cells. Cytom Part A. 2011;**79**A(7):545–59.
34. Sezgin M, Sankur B. Survey over image thresholding techniques and quantitative performance evaluation. J Electron Imaging. 2004;**13**(1):220.
35. Gamarra M, Zurek E, Escalante HJ, Hurtado L, San-Juan-Vergara H. Split and merge watershed: A two-step method for cell segmentation in fluorescence microscopy images. Biomed Signal Process Control. 2019;**53**:101575.
36. Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. Deep learning for cellular image analysis. Nat Methods. 2019;**16**(12):1233–46.
37. Lucas AM, Ryder P V., Li B, Cimini BA, Eliceiri KW, Carpenter AE. Open-source deep-learning software for bioimage segmentation. Mol Biol Cell. 2021;**32**(9):823–9.
38. Haralick RM, Dinstein I, Shanmugam K. Textural features for image classification. IEEE Trans Syst Man Cybern. 1973;SMC-3(6):610–21.
39. Leggett SE, Sim JY, Rubins JE, Neronha ZJ, Williams EK, Wong IY. Morphological single cell profiling of the epithelial–mesenchymal transition. Integr Biol. 2016;**8**(11):1133–44.
40. Nguyen P, Chien S, Dai J, Monnat RJ, Becker PS, Kueh HY. Unsupervised discovery of dynamic cell phenotypic states from transmitted light movies. PLOS Comput Biol. 2021;**17**(12):e1009626.
41. Yao K, Rochman ND, Sun SX. Cell type classification and unsupervised morphological phenotyping from low-resolution images using deep learning. Sci Rep . 2019;**9**(1):1–13.
42. Chow YL, Singh S, Carpenter AE, Way GP. Predicting drug polypharmacology from cell morphology readouts using variational autoencoder latent space arithmetic. PLoS Comput Biol. 2022;**18**(2):1–21.
43. Chandrasekaran SN, Ceulemans H, Boyd JD, Carpenter AE. Image-based profiling for drug discovery: due for a machine-learning upgrade? Nat Rev Drug Discov. 2021;**20**(2):145–59.
44. Imoto D, Saito N, Nakajima A, Honda G, Ishida M, Sugita T, et al. Comparative mapping of crawling-cell morphodynamics in deep learning-based feature space. PLoS Comput Biol. 2021;**17**(8): e1009237.
45. Buggenthin F, Buettner F, Hoppe PS, Endele M, Kroiss M, Strasser M, et al. Prospective identification of hematopoietic lineage choice by deep learning. Nat Methods. 2017;**14**(4):403–6.
46. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012;**9**(7):671–5.
47. Stirling DR, Swain-Bowden MJ, Lucas AM, Carpenter AE, Cimini BA, Goodman A. CellProfiler 4: improvements in speed, utility and usability. BMC Bioinform. 2021;**22**(1):1–11.
48. Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, et al. Data-analysis strategies for image-based cell profiling. Nat Methods. 2017;**14**(9):849–63.
49. Zhang J, Tian XJ, Zhang H, Teng Y, Li R, Bai F, et al. TGF-β-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. Sci Signal. 2014;**7**(345):ra91.
50. Jolly MK, Murphy RJ, Bhatia S, Whitfield HJ, Redfern A, Davis MJ, et al. measuring and modelling the epithelial- mesenchymal hybrid state in cancer: clinical implications. Cells Tissues Organs. 2022;**211**(2):110–33.

51. Pastushenko I, Blanpain C. EMT Transition states during tumor progression and metastasis. Trends Cell Biol. 2019;**29**(3):212–26.
52. Oller-Moreno S, Kloiber K, Machart P, Bonn S. Algorithmic advances in machine learning for single-cell expression analysis. Curr Opin Syst Biol. 2021;**25**:27–33.
53. Cao X, Xing L, Majd E, He H, Gu J, Zhang X. A systematic evaluation of supervised machine learning algorithms for cell phenotype classification using single-cell RNA sequencing data. Front Genet. 2022;**13**:1–14.
54. Rämö P, Sacher R, Snijder B, Begemann B, Pelkmans L. CellClassifier: Supervised learning of cellular phenotypes. Bioinformatics. 2009;**25**(22):3028–30.
55. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, et al. Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. Proc Natl Acad Sci U S A. 2009;**106**(6):1826–31.
56. Stirling DR, Carpenter AE, Cimini BA. CellProfiler Analyst 3.0: accessible data exploration and machine learning for image analysis. Bioinformatics. 2021;**37**(21):3992–4.
57. Dao D, Fraser AN, Hung J, Ljosa V, Singh S, Carpenter AE. CellProfiler analyst: interactive data exploration, analysis and classification of large biological image sets. Bioinformatics. 2016;**32**(20):3210–2.
58. Lever J, Krzywinski M, Altman N. Classification evaluation. Nat Methods. 2016;**13**(8):603–4.
59. Leggett SE, Sim JY, Rubins JE, Neronha ZJ, Williams EK, Wong IY. Morphological single cell profiling of the epithelial–mesenchymal transition. Integr Biol. 2016;**8**(11):1133–44.
60. Liu P, Liu S, Fang Y, Xue X, Zou J, Tseng G, et al. Recent advances in computer-assisted algorithms for cell subtype identification of cytometry data. Front Cell Dev Biol. 2020;**8**:234.
61. Peng L, Tian X, Tian G, Xu J, Huang X, Weng Y, et al. Single-cell RNA-seq clustering: datasets, models, and algorithms. RNA Biol. 2020;**17**(6):765–83.
62. Chen L, Wang W, Zhai Y, Deng M. Single-cell transcriptome data clustering via multinomial modeling and adaptive fuzzy K-means algorithm. Front Genet. 2020;**11**:1–16.
63. Jothi R, Mohanty SK, Ojha A. DK-means: a deterministic K-means clustering algorithm for gene expression analysis. Pattern Anal Appl. 2019;**22**(2):649–67.
64. Peyvandipour A, Shafi A, Saberian N, Draghici S. Identification of cell types from single cell data using stable clustering. Sci Rep. 2020;**10**(1):1–12.
65. Guo M, Wang H, Potter SS, Whitsett JA, Xu Y. SINCERA: a pipeline for single-cell RNA-seq profiling analysis. PLoS Comput Biol. 2015;**11**(11):1–28.
66. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell. 2015;**161**(5):1202–14.
67. Stassen S V., Siu DMD, Lee KCM, Ho JWK, So HKH, Tsia KK. PARC: Ultrafast and accurate clustering of phenotypic data of millions of single cells. Bioinformatics. 2020;**36**(9):2778–86.
68. Lever J, Krzywinski M, Altman N. Points of Significance: principal component analysis. Nat Methods. 2017;**14**(7):641–2.
69. Kobak D, Berens P. The art of using t-SNE for single-cell transcriptomics. Nat Commun. 2019;**10**:5416. https://doi.org/10.1038/s41467-019-13056-x
70. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2019;**37**(1):38–44.
71. Choi HJ, Wang C, Pan X, Jang J, Cao M, Brazzo JA, et al. Emerging machine learning approaches to phenotyping cellular motility and morphodynamics. Phys Biol. 2021;**18**(4):041001.
72. Waddington CH. The strategy of the genes: a discussion of some aspects of theoretical biology. London: George Allen & Unwin Ltd; 1957.
73. Jacob F, Monod J. Genetic repression, allosteric inhibition, and cellular differentiation. Cytodifferentiation and macromolecular synthesis. New York: Academic Press;1963. Vol. **21**. p. 30–64.
74. Kauffman SA. Homeostasis and differentiation in random genetic control networks. Nature. 1969;**224**:177–8.
75. Thomas R. Logical analysis of systems comprising feedback loops. J Theor Biol . 1978;**73**(4):631–56.
76. Devaraj V, Bose B. The mathematics of phenotypic state transition: paths and potential. J Indian Inst Sci. 2020;**100**(3):451–64.
77. Kalbfleisch JD, Lawless JF. Least-squares estimation of transition probabilities from aggregate data. Can J Stat. 1984;**12**(3):169–82.
78. Dent W, Ballintine R. A review of the estimation of transition probabilities in Markov chains. Aust J Agric Econ. 1971;**15**:69–81.
79. Kaur I, Rajarshi MB. Ridge regression for estimation of transition probabilities from aggregate data. Commun Stat Simul Comput. 2012;**41**(4):524–30.
80. Buder T, Deutsch A, Seifert M, Voss-Boehme A. CellTrans: An R package to quantify stochastic cell state transitions. Bioinform Biol Insights. 2017;**11**. doi:10.1177/1177932217712241
81. Farahat WA, Asada HH. Estimation of state transition probabilities in asynchronous vector Markov processes. J Dyn Syst Meas Control. 2012;**134**(6):061003.
82. Karacosta LG, Anchang B, Ignatiadis N, Kimmey SC, Benson JA, Shrager JB, et al. Mapping lung cancer epithelial–mesenchymal transition states and trajectories with single-cell resolution. Nat Commun. 2019;**10**(1):1–15.
83. Goetz H, Melendez-Alvarez JR, Chen L, Tian XJ. A plausible accelerating function of intermediate states in cancer metastasis. PLoS Comput Biol. 2020;**16**(3):1–22.
84. Chang AY, Marshall WF. Dynamics of living cells in a cytomorphological state space. Proc Natl Acad Sci U S A. 2019;**116**(43):21556–62.
85. Wang W, Poe D, Yang Y, Hyatt T, Xing J. Epithelial-to-mesenchymal transition proceeds through directional destabilization of multidimensional attractor. Elife. 2022;**11**:1–20.
86. Deconinck L, Cannoodt R, Saelens W, Deplancke B, Saeys Y. Recent advances in trajectory inference from single-cell omics data. Curr Opin Syst Biol. 2021;**27**:100344.
87. Eugenio M, Karp RL, Guo G, Robson P, Hart AH, Trippa L, et al. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. Proc Natl Acad Sci U S A. 2014;**111**(52):E5643–50.
88. Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. Cell. 2019;**176**(4):928–43.e22.
89. Lin C, Bar-Joseph Z. Continuous-state HMMs for modeling time-series single-cell RNA-Seq data. Bioinformatics. 2019;**35**(22):4707–15.
90. Zhao C, Xiu W, Hua Y, Zhang N, Zhang Y. CStreet: a computed Cell State trajectory inference method for time-series single-cell RNA sequencing data. Bioinformatics. 2021;**37**(21):3774–80.
91. Brunton SL, Kutz JN. Data-driven science and engineering: machine learning, dynamical systems, and control. Cambridge: Cambridge University Press; 2019.

92. Kutz JN, Brunton SL, Brunton BW, Proctor JL. Dynamic mode decomposition. Philadelphia, PA: Society for Industrial and Applied Mathematics; 2016.
93. Taylor-King JP, Riseth AN, Macnair W, Claassen M. Dynamic distribution decomposition for single-cell snapshot time series identifies subpopulations and trajectories during iPSC reprogramming. PLoS Comput Biol. 2020;**16**(1):e1007491.
94. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, et al. Early-warning signals for critical transitions. Nature. 2009;**461**(7260):53–9.
95. Scheffer M, Carpenter SR, Lenton TM, Bascompte J, Brock W, Dakos V, et al. Anticipating critical transitions. Science. 2012;**338**(6105):344–8.
96. Mojtahedi M, Skupin A, Zhou J, Castaño IG, Leong-Quong RYY, Chang H, et al. Cell fate decision as high-dimensional critical state transition. PLoS Biol. 2016;**14**(12): e2000640.
97. Yang B, Li M, Tang W, Liu W, Zhang S, Chen L, et al. Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. Nat Commun. 2018;**9**:678.
98. Sarkar S, Sinha SK, Levine H, Jolly MK, Dutta PS. Anticipating critical transitions in epithelial-hybrid-mesenchymal cell-fate determination. Proc Natl Acad Sci U S A. 2019;**116**(52):26343–52.

SECTION 2

Cancer systems biology: New paradigms

6

Evolution-informed multilayer networks: Overlaying comparative evolutionary genomics with systems-level analyses for cancer drug discovery

Laurie Graves, Ayalur Raghu Subbalakshmi, William C. Eward, Mohit Kumar Jolly, and Jason A. Somarelli

6.1. Cancer is a speciation event that exists within a dynamic ecological system

We tend to think of the evolution of life as a never-ending continuum that is unidirectional. While this unidirectionality is true on a timescale, it is also true that, on a phenotypic scale, evolutionary selection produces reversions in the phenotypes that are observed depending on environmental conditions. Indeed, the evolution of life on Earth has included countless examples in which populations of organisms have reverted to previously observed phenotypes. For example, Earth's first eukaryotic cells were likely to have been unicellular. From early unicellular organisms, there arose all of the multicellular eukaryotic life we observe today; however, multicellular organisms exhibit features reminiscent of speciation events in which single cells revert from phenotypes consistent with multicellular behaviour to those akin to unicellular behaviour. These reversions enable single cells within a multicellular organism to compete for space and resources to survive within the dynamic ecology of the body (**Figure 6.1**). Reversion events from multicellular- to unicellular-like phenotypes are observed across almost all multicellular life. When these reversion events possess specific features (e.g. survival and proliferation, and ability to spread), they are collectively referred to as cancer. The reversion events known as cancer are strongly associated with age, suggesting a relative loss in the evolutionary pressure to restrain cells in multicellular states beyond reproductive age [1].

The cellular phenotype of cancer is mediated by underlying alterations in genotype, rendering the cancer and normal tissue of the host genetically distinct. In this way, the appearance of cancer within individual organisms can be considered a speciation event according to the biological definition of a species—a population of individuals able to successfully interbreed and give rise to fertile offspring. In the case of cancer, the offspring are the daughter cells produced by cell division. The appearance of new cancer species can be found across all multicellular life, suggesting these events are a common feature of multicellularity [2]. Cancer speciation events are constrained by the host environment such that if cancer species sufficiently disrupt the ecosystem of the body, they can directly and indirectly induce death in the host. Put simply, the speciation events that lead to the pathology, we call cancer, are—in almost all cases—evolutionary dead ends. Fascinating exceptions to this are the transmissible cancers observed in dogs, Tasmanian devils, and molluscs [3,4]. In the latter case, cancer has even been transmitted between different species of clams [3]. Another example that illustrates the potential for cancer to live on outside the confines of the body is in the establishment of cell cultures from cancer cells, some of which have been propagated continuously for decades. These examples lend support to the notion that cancer speciation events can propagate if given an appropriate environment.

6.2. The complexity of cancer necessitates systems-level approaches

Cancer speciation events are subject to evolutionary constraints in the same way as any other population of species in the natural world. Despite being genetically and epigenetically diverse both within and between individuals, cancer cells converge onto key 'hallmark' phenotypes known as the hallmarks of cancer. The cancer hallmarks—observed across all cancers—are a striking example of convergent evolution and exemplify the need for cancer cells to possess certain key phenotypic features in order to survive and reproduce in the body [5]. The body's ever-changing environment

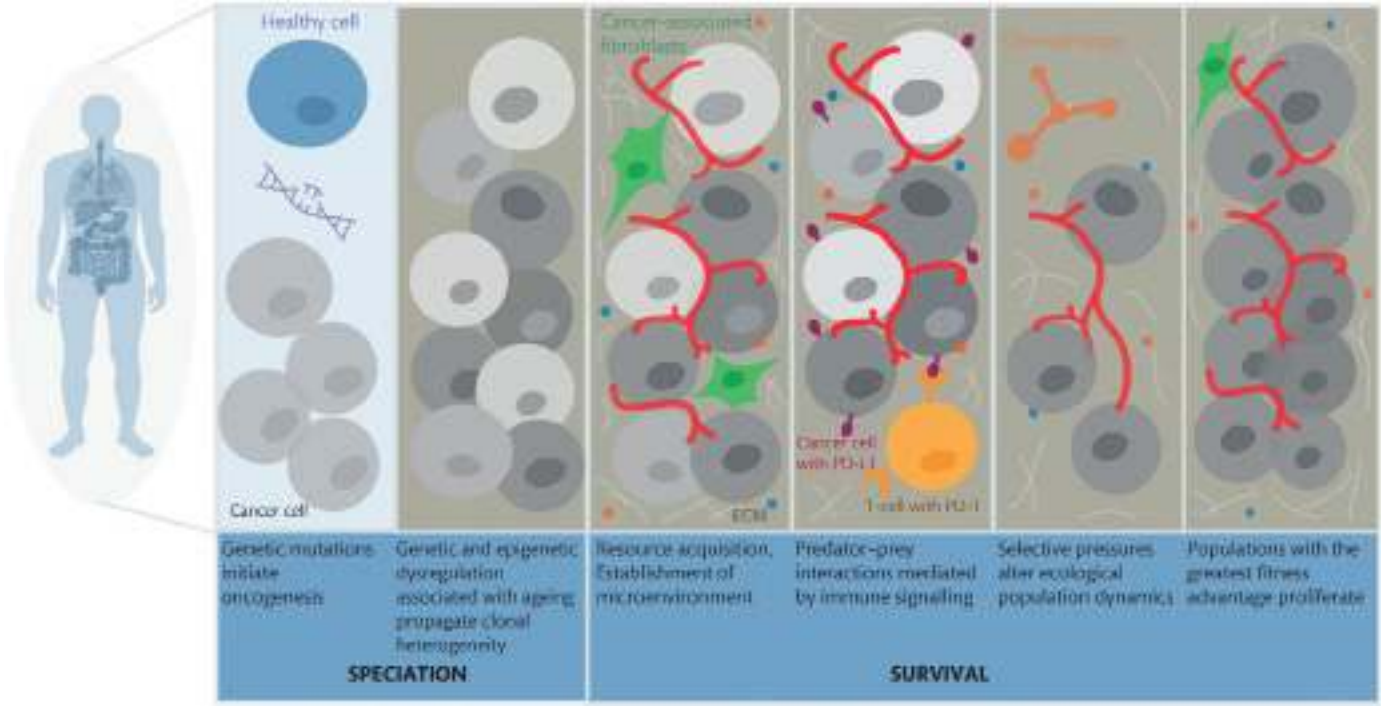


Figure 6.1. Cancer as a speciation event within a heterogeneous ecological system. Genetic dysregulation initiates oncogenesis, and subsequent unchecked cell replication within a cancer-permissive microenvironment of the tissue due to age, environmental insult, etc. induces clinically observable cancer. Cell and population survival is impacted by the cell's ability to acquire resources, including oxygen, growth factors, and glucose, to establish a pro-tumour microenvironment and to withstand selective pressures, such as nutrient and oxygen deprivation, immune predation, or chemotherapy. Cell populations with the greatest fitness advantage are able to survive and proliferate.

induces alterations to the selective pressures and subsequently shifts the genetic and molecular make-up of the cancer cell population as ecological fitness dynamics change. One well-studied example of this phenomenon is age. As we age, the ecological landscape of the body shifts to a more cancer-permissive state, with alterations in immune function, fibroblasts and other stromal components, senescence-associated signalling, and extra-cellular matrix composition [1,6]. Combined with increased mutational burden, the cancer-permissive, aged environment allows for cancer to take hold, survive, grow, and spread.

Another way in which the dynamics of the body change with time is with cancer treatment. Cancer therapy can often introduce rapid and substantial changes to selective pressures within the body through the removal of key resources necessary to sustain survival and reproduction (i.e. cell division). Resource removal can lead to shifts in population dynamics by creating new ecological niches through competitive release and, in some instances, by fuelling additional adaptive processes [7,8].

The complex and ever-changing genetic and molecular diversity of cancer cell populations requires approaches that can enable integration of multiple layers of genetic, genomic, and phenotypic information with the variables of time, treatment, and spatial context in an interpretable framework. Systems biology provides such a framework. This systems-level approach has been made more accessible through improvements in genomics technologies, particularly with respect to innovations in single cell and spatial '-omic' techniques [9,10]. However, although these technologies have provided an unprecedented understanding of the genetic and non-genetic complexity and dynamics of cancer cell populations, the high costs and demand for computationally intensive analysis pipelines have also thus far delayed the clinical applicability and broad research utility of these tools. Even with unlimited financial and technological resources, the vast quantities of data generated by these studies requires a systems-level perspective to distill key insights. These insights must then be translated into clinical action and decision-making, and most physicians outside of major academic medical centres are largely untrained in these techniques and how or whether to incorporate this information into their clinical practice. Continued education and efforts at integrating transdisciplinary approaches are needed to bring these new tools and concepts to bear on clinical treatment. One promising approach may be the use of multidisciplinary tumour boards—which not only include clinicians from different specialties but also engage genomics experts, systems biologists, statisticians, and evolutionary biologists—to integrate these new layers of knowledge into clinical decision-making [11].

6.3. A comparative evolutionary paradigm expands opportunities to understand cancer and rapidly evaluate new treatment options

Despite tremendous advances in our understanding and clinical management of cancer, there remains a substantial need to unravel the complexity of cancer ecology, particularly with respect to how cancer cell population genetic structure is shaped by the surrounding microenvironment and how the microenvironment is shaped by cancer cells. These aspects of cancer biology are challenging to model

using traditional platforms, which typically seek a reductionist approach of studying features in isolation (e.g. cells and single genes/proteins). While these approaches have been—and continue to be—successful in identifying fundamental features of cancer and have led to the discovery of nearly all of our currently approved treatments for cancer, it is also clear that widely used preclinical models often do not accurately recapitulate the complexity of human cancer, leading to many potential cancer 'cures' in preclinical models that ultimately fail in human clinical trials. To overcome some of the limitations of current systems, a subset of researchers have turned to comparative oncology, in which naturally occurring cancers in other animals serve as models for human tumours. This comparative evolutionary approach frames cancer systems and gene regulatory networks analyses within the fundamental paradigm of evolution, which enables a more complete understanding of all other layers of biological organization and interaction. This comparative evolutionary perspective can take several forms, including the study of cancer rates across the tree of life, leveraging cross-species comparisons to understand fundamental processes of somatic mutation and cancer signalling within and between species, and capitalizing on comparative oncology studies in companion animals or animals in human care as naturally occurring 'models' of human disease [2,12–14]. To date, comparative oncology has largely focused within cancer in companion animals, including that of dogs, cats, and horses [15]. However, cancer is prevalent in many mammalian species, with cancer-related mortality approaching 20–40% in some, and there is much to learn about similarities and differences in cancer risk and cancer biology across all species [12]. Utilization of a systems-based comparative oncology approach in which cancer is studied broadly across species holds great potential in the exploration of cancer biology and in the identification of novel therapeutic strategies.

One unique way in which comparative oncology can inform cancer biology is to study the conserved and divergent patterns of cancer risk. For example, early observations in the cross-species study of cancer biology led to the generation of Peto's paradox, based on the observation that cancer risk does not scale proportionally with size or with lifespan [16]. There are a number of hypothesized mechanisms to explain Peto's paradox, including cross-species differences in rates of stem cell division and acquisition of somatic mutations, efficiency of DNA repair mechanisms, and tumour suppressor gene expression. Cross-species analyses of mammals have demonstrated that cetaceans, with lifespans approaching 100–200 years, are among those with the lowest number of DNA substitutions per site per million years, suggestive of slower somatic mutation rates [17]. An additional study demonstrated that, despite wide variability in yearly somatic mutation rates, end-of-life mutational burdens across species tend to be relatively similar and that the somatic mutation rate per year inversely correlates with lifespan rather than with mammal size [13]. For example, despite a 23,000-fold difference in adult body mass, the giraffe and naked mole rat (*Heterocephalus glaber*) have similar lifespans and somatic mutation rates, whereas even with similar body mass, the lifespan and mutation rates between a mouse and a naked mole rat differ significantly [13]. The naked mole rat's lifespan may exceed 30 years, yet they are highly resistant to cancer with no reported cases in the literature [18–21]. In addition to low rates of somatic mutations, cancer resistance in the naked mole rat is mediated by sensitivity to early contact inhibition by fibroblasts that secrete high molecular weight hyaluronic acid and have low hyaluronidase activity [18,19]. Furthermore, the genome of the naked mole rat is under positive selection for genes associated with protection of telomere length, which likely contributes to their long lifespan and cancer resistance [18].

In addition to differences in somatic mutation rates across lifespans, comparative genomic studies have also identified that cancer resistance in other mammals has evolved due to conservation or expansion of tumour suppressor genes. The *TP53* gene in the blind mole rat (*Spalax*) is highly conserved, which contributes to its cancer resistance through cell cycle regulation and IFN-β-induced cell necrosis [18,22]. In some larger mammals, *TP53* gene expansion is believed to be an evolutionary protection against carcinogenesis. Among whales and other cetaceans, whole genome analysis has revealed that the evolution of cetacean giantism correlates with large segmental duplications of tumour suppressor genes and genes related to apoptosis, cell cycle regulation, B-cell immunity, complement activation, cell adhesion, and cell signalling [17]. Studies of the genome of African (*Loxodonta africana*) and Asian (*Elephas maximus*) elephants found 20 copies of the *TP53* retrogene, and that in response to DNA damage, elephant cells demonstrated a significant apoptotic response [16,23]. The study of *TP53* retrogene evolution within the Proboscidean lineage discovered expansion in *TP53* retrogene copy number over a 25 million year evolutionary scale, ranging from 3–8 copies in the extinct prehistoric American mastodon (*Mammut americanum*), to 14 copies in the Woolly (*Mammuthus primigenius*) and Columbian (*Mammuthus columbi*) mammoths, and to 12–17 copies in the modern Asian elephant genome [24]. Furthermore, increase in *TP53* copy number over time correlated with increasing body size in the Proboscidean lineage, providing additional support for the hypothesis that *TP53*'s tumour suppressive effects are an important evolutionary adaptation against cancer in larger mammals [24]. In addition to *TP53*, the genome of the African elephant, manatee, and rock hyrax have evolved through ancestral and lineage-specific duplication events to contain 7–11 copies of the leukaemia inhibitory factor (LIF) gene [25]. LIF, an interleukin 6 class cytokine, is known in different circumstances to function either as a tumour suppressor or an oncogene. Studies in African elephants reveal that LIF6, one of the duplicated genes with low expression under normal physiology, is up-regulated by *TP53* in response to DNA damage, inducing enhanced apoptosis. While the other duplicate LIF genes are not expressed and remain pseudogenes, LIF6 evolved to become a functional gene early in the Proboscidean lineage. These findings suggest that through its enhanced apoptotic signalling, LIF6 expression is another potential mechanism through which evolution supported the large body size of African elephants and other Proboscideans in concordance with Peto's paradox [25].

The study of the cancer genomes of animals has also pinpointed conserved oncogenic drivers across species. Chromosomal aberrations associated with several human haematologic malignancies are conserved in canine disease, including the 'Raleigh chromosome', a *t*(9;26) translocation leading to the BCR–ABL fusion gene, which is homologous to the *t*(9;22) translocation known as the Philadelphia chromosome in humans [26]. A comparison of canine oral squamous cell carcinoma (SCC) to human head and neck squamous cell carcinoma (HNSCC) revealed high homology in pathways involving cell cycle regulation, immune function, and transcriptional reprogramming and also identified shared therapeutic vulnerabilities, including increased expression of PD-L1, CTLA4, and CDK4/6 [27].

Across canine mammary tumours and human breast cancers, mutations in PIK3CA have been identified as a conserved oncogenic driver [28].

Studies of cancer across species have also enabled researchers unprecedented access to tissues and patient samples, particularly in cases of rare human cancers that are more prevalent in other species. For example, osteosarcoma is a rare but highly aggressive bone tumour that occurs in children and adolescents, with an incidence of just 4 per million cases yearly in the United States [29]. By contrast, it is one of the most common tumours in canines. Human and canine osteosarcoma share significant genomic and clinical overlap, yet canine osteosarcoma occurs with substantially higher frequency (~25–30,000 cases/year) and follows an accelerated course when compared to that in humans [15,30]. Comparative genomic studies of osteosarcoma have demonstrated conserved genomic features between the two species [30,31]. Leveraging the expanded access to genomic data from canines, these studies identified novel genes that correlated with patient outcomes in both canines and humans, including an association between CD86 expression and metastasis-free survival and an association between IL-8 and SLC1A3 and poor overall survival [30,31].

Similarities in tumour biology between human and canine tumours as well as greater accessibility to canine osteosarcoma and other sarcoma tumour samples through collaboration with veterinarians contribute to the application of a comparative oncology model to study novel therapeutic strategies in osteosarcoma and other sarcomas [32]. These comparative oncology pipelines have identified (1) heterogeneity in drug response across osteosarcoma cell lines from different species, (2) species-specific clustering by drug response, and (3) key similarities in drug response among canine and human preclinical models, such as the combined sensitivity to CRM1 nuclear export and proteasome inhibitors, which was validated in PDX models [32]. A similar drug discovery pipeline was applied as a precision medicine tool in the case of canine leiomyosarcoma, in which the patient-derived cell line underwent drug screening and *in vivo* PDX validation prior to use of the most effective agent, bortezomib, back in the canine patient [33]. The patient experienced a mixed response to bortezomib, reflecting extensive tumour heterogeneity that was also observed on whole exome sequencing of the patient's primary and recurrent tumours [33]. Among numerous additional canine comparative oncology clinical trials, the study of recombinant listeria vaccines that express a chimeric HER2/neu fusion protein has been particularly impactful, as the clinical benefit seen in canines has been translated into a Children's Oncology Group (COG) Phase 2 trial for children and young adults with recurrent osteosarcoma (NCT04974008) [34].

Comparative studies across a range of species, from humans, to canines, to naked mole rats and whales, provide tremendous insight into diverse mechanisms of carcinogenesis and natural cancer resistance. Through the study of both overlapping and divergent biological features within oncology using a comparative oncology model, scientists, veterinarians, and physicians are able to identify novel targets and vulnerabilities for the ultimate benefit of patients across species. Use of a comparative oncology platform can also permit more efficient study of novel therapies for prioritization into human clinical trials. In particular, research in rare cancers that affect non-human patients would benefit from implementation of comparative oncology models to expand knowledge in tumour biology and to identify novel treatment strategies.

6.4. Combining systems-level analyses across evolutionary scales has potential to highlight fundamental actionable targets for investigation across species

In addition to the multiple layers of biological information within human cancers, a deeper understanding of cancer systems across species has the potential to provide key insights (**Figure 6.2**).

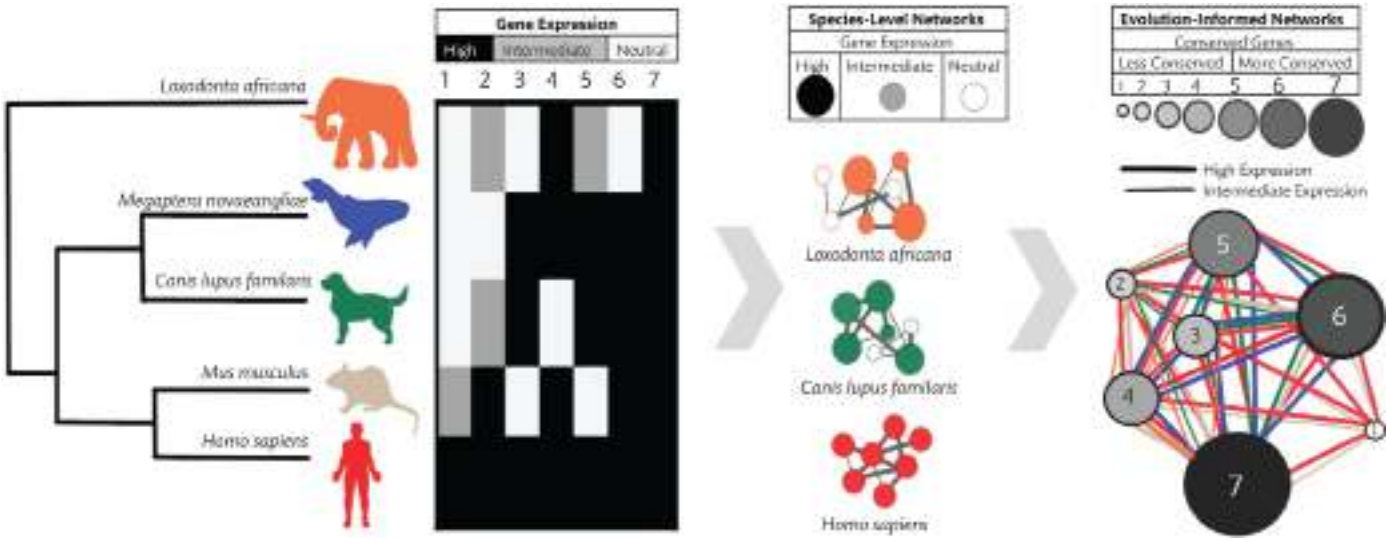


Figure 6.2. Comparative systems biology frameworks to reveal conserved and unique network features in cancers across species. A proposed schematic is shown for the construction of evolutionary-informed gene regulatory networks. Data from a range of relevant species would be used as input for networks. Species choice could be scaled by evolutionary time depending on the question. Input matrices could include gene gain/loss, mRNA or protein expression, protein–protein interactions, phospho-protein data, etc. Inter- and cross-species gene regulatory network analyses would be used to pinpoint common and unique interactions between species.

This comparative systems biology framework aims to identify similarities and differences between species based on the structure of gene regulatory networks rather than differences in expression or activity levels of specific mRNAs or proteins [35]. For instance, mouse models are often used to study human diseases under the assumption that orthologous genes give rise to similar phenotypes [36]. However, if the gene regulatory networks containing orthologous genes become rewired across species, due to gene gain/loss events, changes in cis-acting sequences, or epigenetic differences, this can result in the emergence of different phenotypic consequences [37]. This concept is exemplified in the analysis of liver samples from 15 mammalian species for their promoters, enhancers, and transcriptional activity levels, which demonstrated that the robustness of gene regulatory networks depended on the level of conservation of promoters and enhancers [38]. The conserved sequences were found to function as buffers for gene expression changes, whereas newly evolved regulatory networks can bring about complexity to the underlying regulatory landscape [38]. This difference in gene regulatory networks between species (e.g. humans and mice) has also been observed in the p53 regulatory network [39]. While expression levels of genes regulated by p53 that are involved in DNA damage response, metabolism, apoptosis, and cell cycle regulation varied extensively between humans and mice, a comparison of genome-wide enhancer sequence datasets studied across six mammalian species also revealed substantial conservation [40]. In this case, even when variation existed in the gene regulatory sequences of the mammalian species, the sequences responsible for the enhancer regulatory activity were found to be conserved for over 180 million years of evolution [40]. Similarly, a common mechanism was implicated in both primary tumours and organ development [41]. This observation was based on comparing gene expression profiles of mouse cerebral development between postnatal day 1 through day 60 with that of medulloblastoma cells. Consistent results were reported upon comparing a developing mouse lung with lung cancer cells. Such conserved gene regulatory network components can also be found among different cancer types [42]. For instance, network analysis performed on breast cancer, ovarian cancer, and renal cancer identified a core gene regulatory network consisting of genes responsible for key hallmarks of cancer, including chromatin remodelling, cell cycle regulation, and immune response modulation. Incorporating species-level information into network-based models can provide important new insights into both common features and species-specific features of cancer. As a conceptual example, one could envision a scenario in which gene regulatory network models were inferred using, e.g. data on gene gain/loss, protein–protein interaction, and co-expression across species to identify how network properties change over evolutionary time and infer the fundamental, core properties of these networks. Analogous to the use of evolutionary sequence conservation as a means to pinpoint functional domains within a nucleotide or protein sequence, this cross-species network inference may identify core features of cancer-relevant networks (**Figure 6.2**).

6.5. Conclusions

With rapid growth in the fields of multi-omics, deep learning, advanced imaging, and spatial 'omics', we have begun to further elucidate the complexity of cancer systems; however, studies using these novel technologies remain predominantly focused on mice and humans. Toolkits that can interrogate multi-'omics' and spatial 'omics' across species are needed to expand our knowledge of cancer's fundamental properties. Similarly, functional genomics platforms must be developed for rapid and low-cost deployment across species ([15], and references therein) to functionally perturb systems and understand gene regulatory network structure across species. Platforms such as DepMap and Expression Atlas serve as exemplary benchmarks for quality and ease of use. Repositories that include species-level genomics data, such as The Integrated Canine Data Commons (https://caninecommons.cancer.gov/#/), and databases of medical information across species, such as The Species360 Zoological Information Management System (ZIMS) (https://www.species360.org/), will greatly facilitate cross-species study of gene regulatory networks in cancers and their association with clinical outcomes. These datasets can be integrated into network-level analyses to pinpoint actionable nodes within and between species.

Beyond the development of tools, technologies, and data repositories, however, a fundamental change in mindset is needed from practitioners across fields. Researchers, clinicians, and the public should collectively and urgently embrace the interconnected fates of the environment, animal health, and human health through a One Health perspective. Through this One Health lens that includes evolutionary-informed analyses of gene regulatory networks, cancer research can be made more robust, spurring further insights into the fundamental properties of cancer systems and their vulnerabilities. At the clinical/translational level, comparative oncology clinical trials can be integrated within existing preclinical/translational pipelines to speed both discovery and clinical evaluation of novel agents and treatment approaches for cancer. These efforts will serve to improve the health and well-being of cancer patients from any species and can underscore the need to sustainably steward the health and well-being of all life on Earth.

REFERENCES

1. Aunan JR, Cho WC, Søreide K. The biology of aging and cancer: a brief overview of shared and divergent molecular hallmarks. Aging Dis. 2017;**8**(5):628–42. Published 2017 Oct 1. doi:10.14336/AD.2017.0103
2. Aktipis CA, Boddy AM, Jansen G, Hibner U, Hochberg ME, Maley CC, et al. Cancer across the tree of life: cooperation and cheating in multicellularity. Philos Trans R Soc Lond B Biol Sci. 2015;**370**(1673):20140219. doi:10.1098/rstb.2014.0219
3. Kattner P, Zeiler K, Herbener VJ, La Ferla-Bruhl K, Kassubek R, Grunert M, et al. What animal cancers teach us about human biology. Theranostics. 2021 May 3;**11**(14):6682–702. doi:10.7150/thno.56623
4. Kwon YM, Gori K, Park N, Potts N, Swift K, Wang J, et al. Evolution and lineage dynamics of a transmissible cancer in Tasmanian devils. PLoS Biol. 2020 Nov 24;**18**(11):e3000926. doi:10.1371/journal.pbio.3000926
5. Somarelli JA. The hallmarks of cancer as ecologically driven phenotypes. Front Ecol Evol. 2021;**9**:661583. doi:10.3389/fevo.2021.661583
6. Fane M, Weeraratna AT. How the ageing microenvironment influences tumour progression. Nat Rev Cancer. 2020;**20**(2):89–106. doi:10.1038/s41568-019-0222-9

7. West J, Ma Y, Newton PK. Capitalizing on competition: an evolutionary model of competitive release in metastatic castration resistant prostate cancer treatment. J Theor Biol. 2018;**455**:249–60. doi:10.1016/j.jtbi.2018.07.028
8. Russo M, Crisafulli G, Sogari A, Reilly NM, Arena S, Lamba S, et al. Adaptive mutability of colorectal cancers in response to targeted therapies. Science. 2019;**366**(6472):1473–80. doi:10.1126/science.aav4474
9. Nam AS, Chaligne R, Landau DA. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. Nat Rev Genet. 2021;**22**(1):3–18. doi:10.1038/s41576-020-0265-5
10. Rozenblatt-Rosen O, Regev A, Oberdoerffer P, Nawy T, Hupalowska A, Rood JE, et al. The Human Tumor Atlas Network: charting tumor transitions across space and time at single-cell resolution. Cell. 2020;**181**(2):236–49. doi:10.1016/j.cell.2020.03.053
11. Robertson-Tessi M, Brown J, Poole M, Luddy K, Marusyk A, Gallaher J, et al. Evolutionary Tumor Board: implementing dynamic personalized therapy using evolutionary theory and mathematical modeling for clinical decision support [abstract]. In: Proceedings of the AACR Special Conference on the Evolutionary Dynamics in Carcinogenesis and Response to Therapy; 2022 Mar 14–17. Philadelphia (PA): AACR; Cancer Res 2022;82(10 Suppl):Abstract nr PR010.
12. Vincze O, Colchero F, Lemaître JF, Conde DA, Pavard S, Bieuville M, et al. Cancer risk across mammals. Nature. 2022;**601**(7892):263–67. doi:10.1038/s41586-021-04224-5
13. Cagan A, Baez-Ortega A, Brzozowska N, Abascal F, Coorens THH, Sanders MA, et al. Somatic mutation rates scale with lifespan across mammals. Nature. 2022;**604**(7906):517–24. doi:10.1038/s41586-022-04618-z
14. Nair NU, Cheng K, Naddaf L, Sharon E, Pal LR, Rajagopal PS, et al. Cross-species identification of cancer resistance-associated genes that may mediate human cancer risk. Sci Adv. 2022;**8**(31):eabj7176. doi:10.1126/sciadv.abj7176
15. Somarelli JA, Boddy AM, Gardner HL, DeWitt SB, Tuohy J, Megquier K, et al. Improving cancer drug discovery by studying cancer across the tree of life. Mol Biol Evol. 2020;**37**(1):11–7. doi:10.1093/molbev/msz254
16. Caulin AF, Graham TA, Wang LS, Maley CC. Solutions to Peto's paradox revealed by mathematical modelling and cross-species cancer gene analysis. Philos Trans R Soc Lond B Biol Sci. 2015;**370**:1850–60.
17. Tollis M, Robbins J, Webb AE, Kuderna LFK, Caulin AF, Garcia JD, et al. Return to the sea, get huge, beat cancer: an analysis of cetacean genomes including an assembly for the humpback whale (Megaptera novaeangliae). Mol Biol Evol. 2019;**36**(8):1746–63.
18. Tollis M, Schiffman JD, Boddy AM. Evolution of cancer suppression as revealed by mammalian comparative genomics. Curr Opin Genet Dev. 2017;**42**:40–47.
19. Tian X, Azpurua J, Hine C, Vaidya A, Myakishev-Rempel M, Ablaeva J, et al. High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. Nature. 2013;**499**(7458):346–49. doi:10.1038/nature12234
20. Buffenstein R. Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. J Comp Physiol [B]. 2008;**178**:439–45.
21. Kim EB, Fang X, Fushan AA, Huang Z, Lobanov AV, Han L, et al. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature. 2011 Oct 12;**479**(7372):223–7. doi: 10.1038/nature10533.
22. Gorbunova V, Hine C, Tian X, Ablaeva J, Gudkov AV, Nevo E, et al. Cancer resistance in the blind mole rat is mediated by concerted necrotic cell death mechanism. Proc Natl Acad Sci U S A. 2012;**109**(47):19392–396. doi:10.1073/pnas.1217211109
23. Abegglen LM, Caulin AF, Chan A, Lee K, Robinson R, Campbell MS, et al. Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. JAMA. 2015;**314**(17):1850–60. doi:10.1001/jama.2015.13134
24. Sulak M, Fong L, Mika K, Chigurupati S, Yon L, Mongan NP, et al. TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. Elife. 2016;**5**:e24307. doi: 10.7554/eLife.24307.
25. Vazquez JM, Sulak M, Chigurupati S, Lynch VJ. A zombie LIF gene in elephants is upregulated by TP53 to induce apoptosis in response to DNA damage. Cell Rep. 2018;**24**(7):1765–76. doi:10.1016/j.celrep.2018.07.042
26. Breen M, Modiano JF. Evolutionarily conserved cytogenetic changes in hematological malignancies of dogs and humans—man and his best friend share more than companionship. Chromosome Res. 2008;**16**(1):145–54. doi:10.1007/s10577-007-1212-4
27. Guscetti F, Nassiri S, Beebe E, Rito Brandao I, Graf R, Markkanen E. Molecular homology between canine spontaneous oral squamous cell carcinomas and human head-and-neck squamous cell carcinomas reveals disease drivers and therapeutic vulnerabilities. Neoplasia. 2020;**22**(12):778–88. doi:10.1016/j.neo.2020.10.003
28. Lee KH, Hwang HJ, Noh HJ, Shin TJ, Cho JY. Somatic mutation of *PIK3CA* (H1047R) is a common driver mutation hotspot in canine mammary tumors as well as human breast cancers. Cancers (Basel). 2019 Dec 12;**11**(12):2006. doi:10.3390/cancers11122006
29. Ottaviani G, Jaffe N. The epidemiology of osteosarcoma. Cancer Treat Res. 2009;**152**:3–13. doi:10.1007/978-1-4419-0284-9_1
30. Paoloni M, Davis S, Lana S, Withrow S, Sangiorgi L, Picci P, et al. Canine tumor cross-species genomics uncovers targets linked to osteosarcoma progression. BMC Genomics. 2009 Dec 23;**10**:625. doi:10.1186/1471-2164-10-625
31. Jin Z, Liu S, Zhu P, Tang M, Wang Y, Tian Y, et al. Cross-species gene expression analysis reveals gene modules implicated in human osteosarcoma. Front Genet. 2019 Aug 7;**10**:697. doi:10.3389/fgene.2019.00697
32. Somarelli JA, Rupprecht G, Altunel E, Flamant EM, Rao S, Sivaraj D, et al. a comparative oncology drug discovery pipeline to identify and validate new treatments for osteosarcoma. Cancers (Basel). 2020 Nov 11;**12**(11):3335. doi:10.3390/cancers12113335
33. Rao SR, Somarelli JA, Altunel E, Selmic LE, Byrum M, Sheth MU, et al. From the clinic to the bench and back again in one dog year: how a cross-species pipeline to identify new treatments for sarcoma illuminates the path forward in precision medicine. Front Oncol. 2020 Feb 11;**10**:117. doi: 10.3389/fonc.2020.00117.
34. Mason NJ, Gnanandarajah JS, Engiles JB, Gray F, Laughlin D, Gaurnier-Hausser A, et al. Immunotherapy with a HER2-targeting listeria induces HER2-specific immunity and demonstrates potential therapeutic effects in a phase I trial in canine osteosarcoma. Clin Cancer Res. 2016;**22**(17):4380–90. doi:10.1158/1078-0432.CCR-16-0088
35. Teusink B, Westerhoff HV, Bruggeman FJ. Comparative systems biology: from bacteria to man. Wiley Interdiscip Rev Syst Biol Med. 2010;**2**(5):518–32. doi:10.1002/wsbm.74
36. Aitman TJ, Boone C, Churchill GA, Hengartner MO, Mackay TF, Stemple DL. The future of model organisms in human disease

research. Nat Rev Genet. 2011;**12**(8):575–82. Published 2011 Jul 18. doi:10.1038/nrg3047
37. Ha D, Kim D, Kim I, Oh Y, Kong J, Han SK, et al. Evolutionary rewiring of regulatory networks contributes to phenotypic differences between human and mouse orthologous genes. Nucleic Acids Res. 2022;**50**(4):1849–63. doi:10.1093/nar/gkac050
38. Berthelot C, Villar D, Horvath JE, Odom DT, Flicek P. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. Nat Ecol Evol. 2018;**2**(1):152–63. doi:10.1038/s41559-017-0377-2
39. Fischer M. Conservation and divergence of the p53 gene regulatory network between mice and humans. Oncogene. 2019;**38**(21):4095–109. doi:10.1038/s41388-019-0706-9.
40. Chen L, Fish AE, Capra JA. Prediction of gene regulatory enhancers across species reveals evolutionarily conserved sequence properties. PLoS Comput Biol. 2018 Oct 4;**14**(10):e1006484. doi:10.1371/journal.pcbi.1006484.
41. Kho AT, Zhao Q, Cai Z, Butte AJ, Kim JYH, Pomeroy SL, et al. Conserved mechanisms across development and tumorigenesis revealed by a mouse development perspective of human cancers. Genes Dev. 2004;**18**(6):629–40. doi:10.1101/gad.1182504
42. Knaack SA, Siahpirani AF, Roy S. A pan-cancer modular regulatory network analysis to identify common and cancer-specific network components. Cancer Inform. 2014 Oct 28;**13**(Suppl 5):69–84. doi:10.4137/CIN.S14058

7

Landscape of cell-fate decisions in cancer cell plasticity

Jintong Lang, Chunhe Li, and Jinzhi Lei

7.1. Introduction

Cancer is a disease involving unregulated cell growth, and its rates are rising as more people live to older ages and as lifestyle changes in the developing world [1]. The chances of surviving cancer vary greatly by the type, the location, and the time of starting treatment. For example, the tumour progression of breast cancer can be significantly delayed by current therapies. However, the recurrence is often inevitable, which results in high mortality rates [2].

Cancer can be defined as a disease in which some abnormal cells ignore the normal rules of cell division and grow uncontrollably. Unlike normal cells, cancer cells will grow and proliferate uncontrollably, which can be fatal. It is assumed that mammalian cells have similar molecular networks that control cell proliferation, differentiation, and cell death, which suggests that the changes in these networks at the molecular, biochemical, and cellular levels are all important factors leading to the transformation from normal cells to cancer cells [3–9]. Moreover, the dynamics of cancer evolution is essentially a process of cell regeneration, during which cancer cells show plasticity with random changes in their epigenetic state and result in heterogeneity of cells with diverse abilities of cellular regeneration. Hence, the interactions between cell stochastic plasticity and tumour growth form complex multiscale dynamics in tumour evolution [10–15].

Great efforts have been devoted to understanding the mechanisms of cancerization, e.g. using gene regulatory network (GRN) models, either using deterministic or stochastic models. However, there are still challenges in many respects. In this chapter, we review some approaches using mathematical models to study cancer evolution. We introduce cancer gene network models, stochastic analysis approaches (energy landscape and transition paths) of cancer networks, as well as multiscale modelling for tumour evolution.

7.2. Method

7.2.1. Cancer gene network model

Based on the regulatory relationship between genes, the GRN can be constructed by literature mining or network inference approaches from gene expression data. Hanahan and Weinberg proposed 10 hallmarks of cancer [4,5]. These hallmarks are characterized by certain key cancer marker genes, such as EGFR for proliferative signal, VEGF for angiogenesis, HGF for metastasis, hTERT for unlimited replication, HIF1 for glycolysis, CDK2 and CDK4 for evading growth suppressors, and so forth. Starting from these cancer marker genes and some critical tumour suppressor genes, such as P53, RB, P21, etc., a cancer GRN can be constructed by searching for the interactions among these key genes as well as the interactions among other cancer-associated genes (**Figure 7.1**A). In Figure 7.1A, arrows represent activation and short bars represent repression. The network mainly includes three kinds of marker genes: apoptosis marker genes (green nodes, including BAX, BAD, BCL2, and Caspase), cancer marker genes (magenta nodes, including AKT, MDM2, CDK2, CDK4, CDK1, NFKB, hTERT, VEGF, HIF1, HGF, and EGFR), and tumour repressor genes (light blue nodes, including P53, RB, P21, PTEN, ARF, and CDH1).

Based on the GRN structure, one can write down the ordinary differential equations (ODEs) to describe the time evolution of the expression levels for each component. The Hill functions are often used to describe the activation and inhibition regulations among different genes [17–19]. The ODEs take the form

$$F_i = \dot{x}_i = \frac{dx_i}{dt} = \sum_{j\in A_i} \frac{a_{ji}x_j^n}{s_{ji}^n + x_j^n} + \sum_{j\in B_i} \frac{b_{ji}s_{ji}^n}{s_{ji}^n + x_j^n} - k_i x_i \qquad (7.1)$$

Here, F_i represents the driving force for the time evolution of the activity of the ith gene ($i = 1,2,\ldots,16$). k_i is the degradation rate. A_i denotes the aggregate of genes that activate the ith gene in the network, while B_i denotes the aggregate of genes that inhibit the ith gene in the network. a_{ji} is the activation constant from the jth to the ith gene, while b_{ji} is for the inhibition. s_{ji} is the threshold level for the regulation from x_j to x_i, and n is the Hill coefficient of the regulation.

7.2.2. Energy landscape for stochastic analysis

To provide a global picture and study the stochastic dynamics of the cancer system, the energy landscape approach provides a powerful

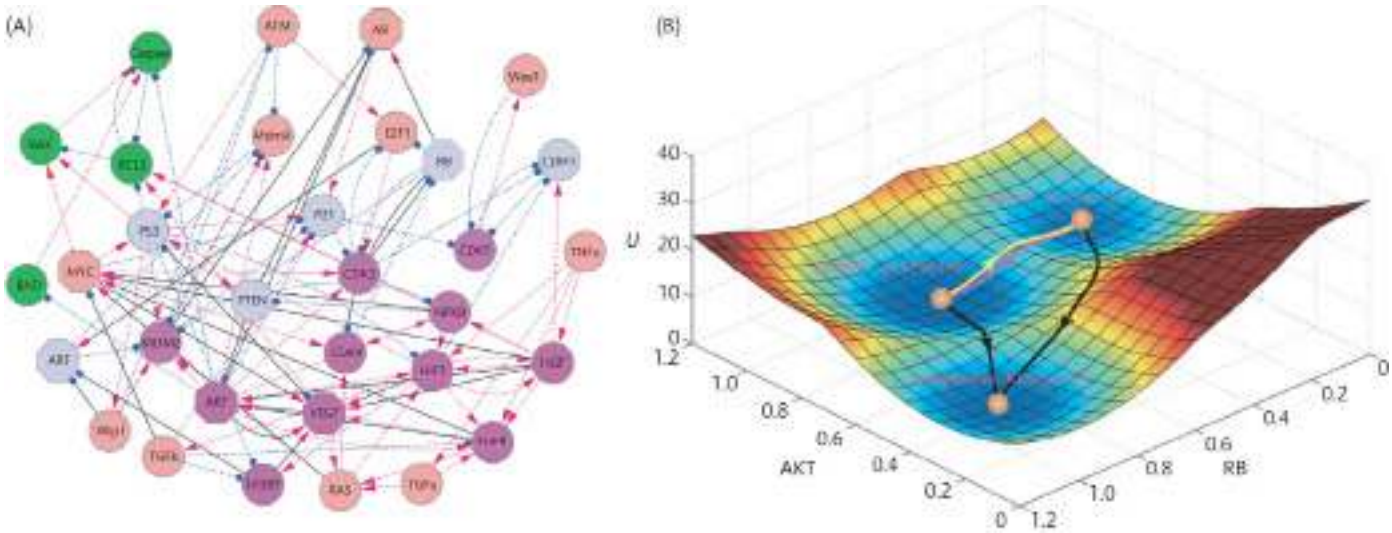


Figure 7.1. Energy landscape of cancer. (A) The diagram for the cancer network. Red arrows represent activation and blue filled circles represent repression. The network includes 32 nodes (genes) and 111 edges (66 activation interactions and 45 repression interactions). (B) The tristable landscape for the cancer network. The yellow path represents the path from normal state attractor to cancer state attractor, and the magenta path represents the path from cancer state attractor to normal state attractor. Black paths represent the apoptosis paths for normal and cancer states.
Source: Taken from [16].

tool. The time evolution of a dynamical system is determined by a probabilistic diffusion equation (Fokker–Planck equation). Given the system state $P(x_1, x_2, \ldots, x_N, t)$, with $x_1, x_2, \ldots, x_N$ representing the levels of components (e.g. the gene expression level), we have an N-dimensional partial differential equation. In the above cancer GRN, there are 32 genes, so $N = 32$. Usually, it is hard to solve a high-dimensional diffusion equation due to the huge state space of the system. Following a self-consistent mean-field approach [20–23], we can split the probability into the products of probabilities of individual ones: $P(x_1, x_2, \ldots, x_N, t) \sim \prod_i^n P_i(x_i, t)$ and solve the probability self-consistently. In this way, we effectively reduce the dimensionality of the system from M^N to MN (M is the dimension for each gene expression level), and the computation of the above problem becomes feasible.

However, for a multidimensional system (as a complex gene network system), it is still challenging to solve the diffusion equations directly. One possible way is to start from the moment equations and assume specific probability distribution based on physical constraints. For example, using Gaussian distribution as an approximation, one needs to calculate two moments, the mean and the variance. When the diffusion coefficient (quantifying the noise level) is small, the moment equations can be approximated to [24,25]

$$\dot{\bar{\mathbf{x}}}(t) = \mathbf{F}\left[\bar{\mathbf{x}}(\mathbf{t})\right] \tag{7.2}$$

$$\dot{\sigma}(t) = \sigma(t)\mathbf{A}^{\mathrm{T}}(t) + \mathbf{A}(t)\sigma(t) + \mathbf{2D}\left[\bar{\mathbf{x}}(\mathbf{t})\right] \tag{7.3}$$

Here, $\bar{\mathbf{x}}$, $\boldsymbol{\sigma}(t)$, and $\mathbf{A}(t)$ are vectors and tensors, and $\mathbf{A}^{\mathrm{T}}(t)$ is the transpose of $\mathbf{A}(t)$. The elements of matrix $\mathbf{A}$ are specified as $A_{ij} = \dfrac{\partial F_i\left[\mathbf{X}(t)\right]}{\partial x_j(t)}$. $\mathbf{D}\left[\bar{\mathbf{x}}(\mathbf{t})\right]$ is the diffusion matrix. For the external noise only cases, $\mathbf{D}[\bar{\mathbf{x}}(\mathbf{t})]$ is not dependent on $\bar{\mathbf{x}}(\mathbf{t})$, i.e. $\mathbf{D} = \mathrm{diag}(d, d, \ldots, d)$. For the intrinsic noise cases, $\mathbf{D}[\bar{\mathbf{x}}(\mathbf{t})]$ is dependent on $\bar{\mathbf{x}}(\mathbf{t})$.

Based on these equations, we can solve $\bar{\mathbf{x}}(t)$ and $\boldsymbol{\sigma}(t)$. Here, we only consider the diagonal elements of $\boldsymbol{\sigma}(t)$ from the mean-field approximation and define $\boldsymbol{\sigma}(t) = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_N)$. Therefore, the evolution of probability distribution for each variable can be acquired from the Gaussian approximation:

$$P(x_i, t) = \frac{1}{\sqrt{2\pi\sigma_i(t)}} e^{-\frac{\left[x_i - \bar{x}_i(t)\right]^2}{2\sigma_i(t)}} \tag{7.4}$$

Here, $\bar{\mathbf{x}}(t)$ and $\boldsymbol{\sigma}(t)$ are the solutions of Equations (7.2) and (7.3). From the mean-field approximation, we can extend this formulation to the multidimensional case by assuming that the total probability is the product of each individual probability for each variable: $P(x_1, x_2, \ldots, x_N, t) \sim \prod_{i=1}^{N} P(x_i, t)$, namely

$$\boldsymbol{P}(x_1, x_2, \ldots, x_N, t) = \frac{1}{(\boldsymbol{2\pi})^{(N/2)} \prod_{i=1}^{N} \sqrt{\sigma_i(\boldsymbol{t})}} e^{-\sum_{i=1}^{N} \frac{\left[x_i - \bar{x}_i(t)\right]^2}{2\sigma_i(t)}} \tag{7.5}$$

When $t \to \infty$, $\bar{x}_i(t)$ and $\sigma_i(t)$ approach the constants $\bar{x}_i$ and σ_i, the probability distribution obtained above corresponds to one steady state or basin of attraction. If the system has multiple stable states, there are several probability distributions localized at each basin with different variances. Therefore, the total probability is the weighted sum of all these probability distributions:

$$P(x_1, x_2, \ldots, x_N, t) = \sum_{j=1}^{N_{ss}} P_j(x_1, x_2, \ldots, x_N, t).w_j \tag{7.6}$$

where N_{ss} denotes the total number of the sable states, and j denotes the index of the corresponding stable states.

The weighting factors (w_j) characterize the relative sizes of a different basin of attraction. We determine the weights w_j by giving a

large number of random initial conditions to the ODEs to be solved and collecting the statistics from these different solutions. Finally, we can construct the potential landscape by $U(\mathbf{x}) = -\ln P_{ss}(\mathbf{x})$ [17,22], with P_{ss} representing the steady-state probability distribution (in practice we let t to be very large to obtain the steady state). To visualize the landscape, we can choose two critical marker genes to display a two-dimensional landscape (Figure 7.1B).

7.2.3. Transition path quantification

The energy landscape provides a global description of the stability of cell types. To study the stochastic dynamics in the cell-fate transition process, one can calculate the most probable transition path between stable states based on the large deviation theory [26,27]. From the Freidlin–Wentzell theorem, the most probable transition path from the stable state i at time 0 to the stable state j at time T, $\phi_{ij}^{*}(t), t \in [0,T]$ can be acquired by minimizing the action functional $S_T[\phi_{ij}]$:

$$S_T\left[\phi_{ij}\right] = \frac{1}{2}\int_0^T \left|\dot{\phi}_{ij} - \mathbf{F}(\phi_{ij})\right|^2 dt. \tag{7.7}$$

This is called the minimum action path (MAP).

However, one limitation for the transition path from the large deviation theory is that it assumes zero noise limitation [28]. To overcome this limitation, a path integral approach has also been developed to calculate the transition paths between stable states so that we can study the effects of noise on these transition paths. Based on the path integral approach [17], we have

$$\begin{aligned} &P_t(\mathbf{x}_{final}, t, \mathbf{x}_{initial}, 0) \\ &= \int \mathbf{x} \exp\left[-\int dt\left(\frac{1}{2}\nabla\cdot\mathbf{F}(\mathbf{x}) + \frac{1}{4}\left(d\mathbf{x}/dt - \mathbf{F}(\mathbf{x})\right)\cdot\frac{1}{\mathbf{D}(\mathbf{x})}\cdot\left(d\mathbf{x}/dt - \mathbf{F}(\mathbf{x})\right)\right)\right] \\ &= \int \mathrm{D}\mathbf{x} \exp\left[-S(\mathbf{x}(t))\right] \\ &= \int \mathrm{D}\mathbf{x} \exp\left[-\int L(\mathbf{x}(t))dt\right] \end{aligned} \tag{7.8}$$

where P_t represents the transition probability, and $S(\mathbf{x}(t))$ is the action and $L(\mathbf{x}(t))$ is the Lagrangian.

To calculate the most probable transition path from one stable state to another one, we need to minimize the action S to maximize the transition probability. Here, the Lagrangian is written as [17,29]

$$L(\mathbf{x}) = \frac{1}{4D}\dot{\mathbf{x}}^2 + V(\mathbf{x}) - \frac{1}{2D}\mathbf{F}(\mathbf{x})\cdot\dot{\mathbf{x}} \tag{7.9}$$

where $V(\mathbf{x}) = \frac{1}{4D}\mathbf{F}^2 + \frac{1}{2}\nabla\cdot\mathbf{F}(\mathbf{x})$

So, we can write the generalized momentum and Hamiltonian:

$$\mathrm{P}_{\mathbf{m}}(\mathbf{x}) = \frac{\partial L}{\partial \dot{\mathbf{x}}} = \frac{1}{2D}(\dot{\mathbf{x}} - \mathbf{F}(\mathbf{x})) \tag{7.10}$$

$$H(\mathbf{x}) = -L(\mathbf{x}) + \mathbf{P}_{\mathbf{m}}(\mathbf{x})\cdot\dot{\mathbf{x}} = \mathrm{E}_{eff} \tag{7.11}$$

We consider E_{eff} as a hyper-parameter and choose $E_{\text{eff}} = -V_{\min}$ (here $V_{\min}$ is the minimum of effective potential). Each path connects two stable states in this case, so V will reach its minimum when $\mathbf{x}$ is the most stable state among multiple stable states.

Then, we substitute Equation (7.11) into the action, we obtain $S(\mathbf{x}) = \int(\mathbf{P}_{\mathbf{m}}(\mathbf{x})\cdot\dot{\mathbf{x}} - H(\mathbf{x}))dt$. To calculate the action of the path, we need to transform the formulations into a different representation in $\mathbf{x}$ space and discretize the integral. The target function can be written as

$$S = \sum_{n=1}^{N_{tp}-1}\left(\sqrt{\left(E_{eff} + V(n)\right)/D} - \frac{1}{2D}F_l(n)\right)\Delta l_{n,n+1} + \lambda P_{\lambda} \tag{7.12}$$

where N_{tp} is the total number of points on the transition path, and P_λ is a penalty function keeping all the length elements close to their average:

$$\begin{aligned} P &= \sum_{n=1}^{N_{tp}-1}\left(\Delta l_{n,n+1} - \langle\Delta l\rangle\right)^2 \\ \Delta l_{n,n+1} &= \sqrt{\sum_{i=1}^{N}\left(\mathbf{x}_i(n+1) - \mathbf{x}_i(n)\right)^2} \\ F_l(n) &= \sum_{i=1}^{N}\mathbf{F}_i(\mathbf{x}(n))\left(\mathbf{x}_i(n+1) - \mathbf{x}_i(n)\right)/\Delta l_{n,n+1} \\ V(n) &= \sum_{i=1}^{N}\left(\frac{1}{4D}\mathbf{F}^2(\mathbf{x}_i) + \frac{1}{2}\sum_{j=1}^{N}\frac{\partial \mathbf{F}_j(\mathbf{x}_i)}{\partial x_j}\right) \end{aligned} \tag{7.13}$$

Here, N represents the number of components, and i is the index for different components in the system. In this way, we can calculate the transition action of any path given one starting point and one ending point. Finally, we can obtain the most probable transition path by minimizing the transition action S.

7.2.4. Multiscale modelling of heterogeneous stem cell regeneration

In addition to the gene network and transition of cell states, the process of cancer cell regeneration is essential for the cancer evolution dynamics, during which cell plasticity and heterogeneity are crucial. Here, we review a general mathematical framework for modelling the evolutionary dynamics of cancer with plasticity and heterogeneity in cells. This model integrates the classical G0 cell cycle model with epigenetic state transitions during cell divisions and results in a differential–integral equation that involves different scale interactions. The model framework enables us to formulate the dynamics of tumour cells landscape during tumour evolution and numerically investigate the dynamics through the technique of individual cell-based modelling.

7.2.4.1. Homogeneous stem cell regeneration

A simple model of stem cell regeneration, the G0 cell cycle model, was introduced by Burns and Tannock in 1970 (Figure 7.2A) [30]. This model assumes a resting phase (G0) between two cell cycles. Stem cells in cell cycles are classified into a resting or proliferating phases. The resting-phase cells can either re-enter the proliferative phase at a rate β that involves negative feedback or are removed from the resting pool with a rate κ due to differentiation, aging, or death. The proliferating cells are assumed to undergo mitosis at a fixed time τ after entry into the proliferative compartment and to be lost randomly at a rate μ during the proliferating phase.

Let $s(t, a)$ be the number of stem cells at time t with age a in the proliferating phase and $Q(t)$ be the number of cells in the resting

phase. The above biological process can be described by the following age-structured equation [30,31]:

$$\begin{aligned}&\frac{\partial s(t,a)}{\partial t}+\frac{\partial s(t,a)}{\partial a}=-\mu s(t,a),\ \ t>0, 0<a<\tau\\&\frac{dQ}{dt}=2s(t,\tau)-\big(\beta(Q)+\kappa\big)Q,\ \ t>0.\end{aligned}\tag{7.14}$$

The boundary condition at age $a = 0$ is given by

$$s(t,0)=\beta\big(Q(t)\big)Q(t).\tag{7.15}$$

Here, the proliferation rate of resting-phase cells is represented by the function $\beta(Q)$, which is dependent on the number of cells in the resting phase.

We can integrate the first equation in (7.14) through the method of characteristic line and obtain a delay differential equation model

$$\frac{dQ}{dt}=-\big(\beta(Q)+\kappa\big)Q+2e^{-\mu\tau}\beta(Q_\tau)Q_\tau,\tag{7.16}$$

where $Q_\tau(t)=Q_\tau(t-\tau)$. This equation describes the general population dynamics of stem cell regeneration.

Biologically, the self-renewal ability of a cell is associated with both microenvironmental conditions, e.g. growth factors and various types of cytokines, and intracellular signalling pathways [32,33]. Despite the complex signalling pathways, the phenomenological formation of Hill function dependence can be derived from simple assumptions regarding the interactions between signalling molecules and receptors [34,35], and is given by

$$\beta(Q)=\beta_0\frac{1}{1+\big(Q/\theta\big)^n}+\beta_1,\tag{7.17}$$

where β_0 represents the maximum proliferation rate of normal cells, and θ is a constant for the half-effective cell number. Here, the positive parameter β_1 is introduced to represent the possible mutations in cancer cells that enable a cell to achieve sustained proliferative signalling or to evade growth suppressors, which represent hallmarks of cancer [36].

7.2.4.2. Heterogeneous stem cell regeneration

The G0 cell cycle model describes the population dynamics of homogeneous stem cell regeneration. To model the heterogeneity in stem cells, we introduce a variable $\mathbf{x}$ (often a high-dimensional vector) for the epigenetic state of cells and Ω for the space of all possible epigenetic states in resting-phase stem cells [12,35,37]. The epigenetic state $\mathbf{x}$ represents intrinsic cellular states that may change during cell division. Biologically, the epigenetic state of a cell can be any molecular levels changes that are independent of the DNA sequences, including the patterns of DNA methylation, nucleosome histone modifications, and transcriptomics [38–43].

Through the epigenetic state $\mathbf{x}\in\Omega$, let $Q(t,\mathbf{x})$ represent the number of cells at time t in the resting-phase and with epigenetic state $\mathbf{x}$. Now, the total cell number is given by

$$Q(t)=\int_\Omega Q(t,\mathbf{x})d\mathbf{x}\tag{7.18}$$

The proliferation of each cell is regulated by the signalling pathways that are dependent on extracellular cytokines released by all cells in the niche and the epigenetic state $\mathbf{x}$ of the cell [34,44,45]. Let $\zeta(\mathbf{x})$ be the rate of cytokine secretion by a cell with state $\mathbf{x}$, and

$$c(t)=\int_\Omega Q(t,\mathbf{x})\ \zeta(\mathbf{x})d\mathbf{x}\tag{7.19}$$

represents the effective concentration of cytokines to regulate cell proliferation. The proliferation rate β can be written as a function of cytokine concentration c and the epigenetic state $\mathbf{x}$, i.e.

$$\beta(c,\mathbf{x})=\beta_0(\mathbf{x})\frac{1}{1+\big(c/\theta(\mathbf{x})\big)^n}+\beta_1(\mathbf{x})\tag{7.20}$$

Moreover, the apoptosis rate μ, the cell cycle duration τ, and the differentiation rate κ are dependent on the epigenetic state $\mathbf{x}$ and are denoted by $\mu(\mathbf{x})$, $\tau(\mathbf{x})$, and $\kappa(\mathbf{x})$, respectively. Here, we assume that these rates depend only on the state of each cell. The cell-to-cell interactions are not included in the current model framework.

To consider cell plasticity in each cell cycle, we introduce a transition function $p(\mathbf{x}, \mathbf{y})$ for the inheritance probability, which represents the conditional probability that a daughter cell of state $\mathbf{x}$ comes from a mother cell of state $\mathbf{y}$ after cell division, i.e.

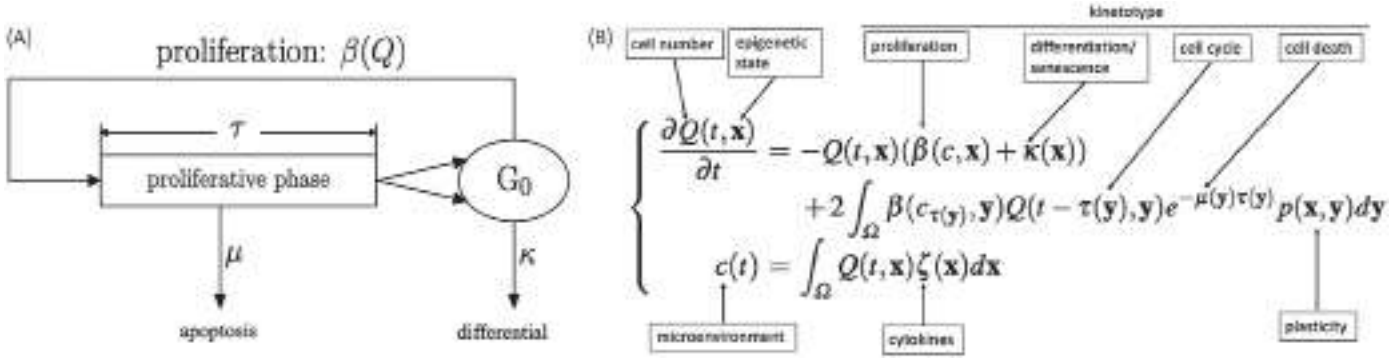


Figure 7.2. Dynamical model of stem cell regeneration. (A) Illustration of the G0 cell cycle model. (B) Multiscale model framework of heterogeneous stem cell regeneration.

$$p(\mathbf{x},\mathbf{y}) = \mathrm{P}(\text{state of daughter cell} = \mathbf{x} \mid \text{state of mother cell} = \mathbf{y}).$$

It is obvious that

$$\int_\Omega p(\mathbf{x},\mathbf{y})d\mathbf{x} = 1$$

for any $\mathbf{y} \in \Omega$.

Now, similar to (7.14), when stem cell heterogeneity is included, we obtain the corresponding age-structured model equation

$$\begin{aligned} &\nabla' s(t,a,\mathbf{x}) = -\mu(\mathbf{x})s(t,a,\mathbf{x}), (t>0,\ \ 0<a<\tau(\mathbf{x})) \\ &\frac{\partial Q(t,\mathbf{x})}{\partial t} = 2\int_\Omega s(t,\tau(\mathbf{y}),\mathbf{y})p(\mathbf{x},\mathbf{y})d\mathbf{y} \\ &\qquad\qquad -(\beta(c(t),\mathbf{x}) + \kappa(\mathbf{x}))Q(t,\mathbf{x}),\ (t>0). \end{aligned} \tag{7.21}$$

and

$$s(t,0,\mathbf{x}) = \beta(c(t),\mathbf{x})Q(t,\mathbf{x}),\ c(t) = \int_\Omega Q(t,\mathbf{x})\zeta(\mathbf{x})d\mathbf{x}.$$

Here, $\nabla' = \partial/\partial t + \partial/\partial a$, and the epigenetic state $\mathbf{x}$ can be considered as a parameter for the first equation. Hence, we can apply the characteristic line method and obtain

$$s(t,\tau(\mathbf{x}),\mathbf{x}) = \beta(c(t-\tau(\mathbf{x})),\mathbf{x})Q(t-\tau(\mathbf{x}),\mathbf{x})e^{-\mu(\mathbf{x})\tau(\mathbf{x})}.$$

Thus, substituting $s(t,\tau(\mathbf{x}),\mathbf{x})$ into the second equation in (7.21), we obtain the following delay differential–integral equation (here, we only show the equation for $t \geq \tau$ that is important for the long-term behaviour):

$$\begin{cases} \dfrac{\partial Q(t,\mathbf{x})}{\partial t} = -Q(t,\mathbf{x})\big(\beta(c,\mathbf{x}) + \kappa(\mathbf{x})\big) \\ \qquad\qquad + 2\displaystyle\int_\Omega \beta(c_{\tau(y)},\mathbf{y})Q(t-\tau(\mathbf{y}),\mathbf{y})e^{-\mu(\mathbf{y})\tau(\mathbf{y})}p(\mathbf{x},\mathbf{y})d\mathbf{y}, \\ c(t) = \displaystyle\int_\Omega Q(t,\mathbf{x})\ \zeta(\mathbf{x})d\mathbf{x}. \end{cases} \tag{7.22}$$

Here, $c_\tau = c(t-\tau)$.

Equation (7.22) provides a general mathematical framework for modelling the dynamics of heterogeneous stem cell regeneration with the epigenetic transition. Biologically, equation (7.22) connects different scale components (Figure 7.2B): the gene expression values at the single-cell level ($\mathbf{x}$), the population dynamic properties ($\beta(c, \mathbf{x})$, $\kappa(\mathbf{x})$, and $\mu(\mathbf{x})$), cell cycle ($\tau(\mathbf{x})$), cytokine secretion ($\zeta(\mathbf{x})$), and the transition of epigenetic states ($p(\mathbf{x}, \mathbf{y})$). In this equation, the functions $\beta(c, \mathbf{x})$, $\kappa(\mathbf{x})$, $\mu(\mathbf{x})$, $\tau(\mathbf{x})$ describe the kinetic properties of cell cycling and are termed as the kinetotype of a cell [35]. This framework can be applied to different problems related to cell regeneration, such as development, aging, and tumour evolution.

7.2.4.3. The transition function $p(\mathbf{x}, \mathbf{y})$

The transition function $p(\mathbf{x}, \mathbf{y})$ is key to describe the plasticity of cells. However, the exact formula of the transition function is difficult to determine biologically, which is dependent on the complex biochemical reactions during the biological process of cell division. Nevertheless, while we consider $p(\mathbf{x}, \mathbf{y})$ as a conditional probability density, we can focus on the epigenetic state before and after cell division and omit the intermediate complex process.

Usually, we are interested in the expressions of genes that are involved in the gene network. In this case, dynamics of the expression of genes, $\mathbf{x}(t)$, can often be described by a stochastic process of a deterministic chemical rate equation

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}). \tag{7.23}$$

The function $\mathbf{F}$ describes the regulatory relationship that is determined by the structure of gene regulation networks, e.g. the epithelial–mesenchymal transition (EMT) network. Equation (7.23) often depends on the parameters $\mathbf{p}$ that are stochastically changed over time and hence can be written in the form of nonautonomous equations:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x};\ \mathbf{p}(t)). \tag{7.24}$$

The above equations are often valid only within one cell cycle. During cell division, epigenetic information (histone modification, DNA methylation, etc.) redistributes to two daughter cells followed by the re-establishment of epigenetic marks. Meanwhile, the proteins and mRNA are also redistributed to the two daughter cells. Mathematically, cell division brings discontinuous boundary conditions in variables $\mathbf{x}$ and parameter $\mathbf{p}$. Hence, the above equation is extended below over divisions

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x},\mathbf{p}(t)), & \text{Between cell divisions} \\ (\mathbf{x},\mathbf{p}) \mapsto (\mathbf{x}',\mathbf{p}') \sim (\wp(\psi(\mathbf{x},\mathbf{p}),\wp(\mathbf{x},\mathbf{p}))) & \text{cell division} \end{cases} \tag{7.25}$$

Here, $\wp(f(\mathbf{x},\mathbf{p}))(f = \psi,\phi)$ represents the random number with a probability density function given by $f(\mathbf{x}, \mathbf{p})$. The equations of form (7.25) provide a general framework for describing the dynamics of single-cell cross-cell divisions. In this way, we can obtain phenomenological formulations for the transition function through numerical simulation based on GRNs and the laws of epigenetic state inheritance during cell division [46-48]. For more discussions, refer to [49].

Alternatively, we can assume the phenomenological formulations directly based on experimental observations. Let the epigenetic state $\mathbf{x} = (x_1, \ldots, x_n)$ represent n independent state variables, and assume that these states vary independently during cell division, then

$$p(\mathbf{x},\mathbf{y}) = \prod_{i=1}^{n} p_i(x_i,\mathbf{y}),$$

where $p_i(x_i, \mathbf{y})$ means the transition function of x_i given the state $\mathbf{y}$ of the mother cells.

The transition function $p_i(x_i, \mathbf{y})$ can be represented as the density function of x_i, which is often phenomenological assumed based on the biological implications. For example, we can assume the beta distribution for the normalized nucleosome modifications [46] or gamma distribution for transcription levels [50].

Here, we assume that $0 < x_i < 1$, and $p_i(x_i, \mathbf{y})$ is given by the density function of a beta distribution, i.e.

$$p_i(x_i,\mathbf{y}) = \frac{x_i^{a_i(\mathbf{y})-1}(1-x_i)^{b_i(\mathbf{y})-1}}{B(a_i(\mathbf{y}),b_i(\mathbf{y}))},\ B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad (7.26)$$

where $\Gamma(\cdot)$ represents the gamma function. Here, the transition function depends on two share parameters a and b, which are functions of the epigenetic state $\mathbf{y}$ of the mother cell. To determine the functions $a_i(\mathbf{y})$ and $b_i(\mathbf{y})$ from experimental data, if we write the mean and variance of x_i, given the state $\mathbf{y}$, as

$$\mathrm{E}(x_i \mid \mathbf{y}) = \phi_i(\mathbf{y}), \quad \mathrm{Var}(x_i \mid \mathbf{y}) = \frac{1}{1+\eta_i(\mathbf{y})}\phi_i(\mathbf{y})(1-\phi_i(\mathbf{y})), \quad (7.27)$$

through predefined functions $\phi_i(\mathbf{y})$ and $\eta_i(\mathbf{y})$, the shape parameters are given by

$$a_i(\mathbf{y}) = \eta_i(\mathbf{y})\phi_i(\mathbf{y}), \quad b_i(\mathbf{y}) = \eta_i(\mathbf{y})(1-\phi_i(\mathbf{y})). \quad (7.28)$$

Here, the functions $\phi_i(\mathbf{y})$ and $\eta_i(\mathbf{y})$ always satisfy

$$0 < \phi_i(\mathbf{y}) < 1,\ \eta_i(\mathbf{y}) > 0.$$

Since the predefined function $\phi_i(\mathbf{y})$ and $\eta_i(\mathbf{y})$ are associated with conditional expectation and variance through (7.27), they provide a bridge to connect the above model formulation with experimental data at single-cell level.

7.2.4.4. Hybrid computational model of multicellular tissues

Equation (7.22) provides general mathematical frameworks to model stem cell regeneration when heterogeneity and plasticity of epigenetic or genetic states are included. This framework can be used to describe many biological processes associated with stem cell regeneration, including development, aging, and cancer evolution [35]. Nevertheless, it is too expensive to numerically solve Equation (7.22) when high-dimensional epigenetic states are considered. In applications, we often develop hybrid computational models for multicellular tissues based on the above frameworks.

Based on the above mathematical framework, a hybrid numerical scheme can be developed that combines a discrete stochastic process for the epigenetic/genetic states of individual cells with a continuous model of cell population growth. In numerical simulation, a multicellular system is represented by a collection of epigenetic states for each cell as $\Omega_t = \left\{\left[C_i(\mathbf{x}_i)\right]_{i=1}^{Q(t)}\right\}$, where $Q(t)$ represents the number of resting-phase stem cells at time t. During a time interval $(t, t + dt)$, each cell $(C_i(\mathbf{x}_i))$ undergoes proliferation, apoptosis, or terminal differentiation with a probability given by the kinetic rates. The probabilities of proliferation, apoptosis, or differentiation are given by $\beta(c, \mathbf{x}_i)dt$, $\mu(\mathbf{x}_i)dt$, or $\kappa(\mathbf{x}_i)dt$, respectively, and hence are dependent on the epigenetic state of each cell. The total cell number $Q(t)$ changes after a time step dt in accordance with the behaviours of all cells. When a cell undergoes proliferation, the epigenetic states of daughter cells randomly change according to the transition function $p(\mathbf{x}, \mathbf{y})$. In this hybrid model, all detailed molecular interactions are hidden within the kinetic rates and the transition function. Moreover, we can also include a stochastic process for differential equations for the signalling dynamics within one cell cycle as well as microenvironmental variables that may depend on the cell behaviour of all cells. The proposed hybrid model can be implemented by single-cell-based models through GPU architecture [51].

The above hybrid numerical scheme can also be integrated with the stochastic modelling of gene networks through which the gene expression dynamics in individual cells are described with mathematical models of form (7.25). The kinetotype (β, μ, κ, and τ) of each cell is dependent on the gene expression state $\mathbf{x}$. The above procedure gives a multiscale dynamical model.

7.3. Applications of stochastic analysis to cancer and EMT network

Cancer metastasis is the most fatal stage of cancer which accounts for over 90% of cancer deaths [52]. The EMT plays a critical role in embryonic development and might contribute to cancer metastasis [53–56]. Many classical EMT marker genes are related to cancer metastasis [57]. However, it remains elusive how to elucidate the mechanistic connections between EMT and cancer metastasis quantitatively. In this section, we introduce the EMT model and the relationship between EMT and cancer metastasis based on the method in the previous section.

7.3.1. Landscape and transition path of EMT network

EMT, a basic developmental process that might promote cancer metastasis, has long been shown to be related to the acquisition of malignant cell traits, such as motility, invasiveness, and tumour-initiating potential, and is therefore associated with the progression of cancer metastasis [54,58–60]. In this subsection, we review some results of EMT gene network models from stochastic dynamics perspective.

As mentioned above, a GRN should be constructed first. By combining the EMT network [61], an important transcriptional factor Ovol2 [62], and an important microRNA miR-145 [27,63], and merging three microRNAs with similar functions (miR-200a, miR-200b, and miR-200c), we obtained a GRN for EMT including 16 components. So, the EMT model includes 16 representative gene/microRNA components [64]. Here x_i ($i = 1, 2, \ldots, 16$) stands for the expression levels of TGF-β, ZEB1, ZEB2, SNAI1, SNAI2, TWIST1, FOXC2, GSC, TCF3, VIM, miR-145, miR-141, miR-200, miR-34a, Ovol2, and CDH1, respectively. The circuit consists of 12 transcriptional factors and 4 microRNAs. The network diagram is shown in Figure 7.3A.

Then, the landscape can be quantified to study the stochastic dynamics of a GRN (Figure 7.3B). Here, we choose the key variables ZEB1 and CDH1 as the coordinates and projected the landscape to two-dimensional space since ZEB1 is a major M marker gene and CDH1 is a major E marker gene.

On the landscape, the basins (namely attractors, blue region) represent stable states or phenotypes (Figure 7.3B). It follows that the closer to the stable states, the lower the potential energy, and the cell is more likely to stay there. The six basins of attraction on the landscape represent six different cell states characterized by different gene expression patterns in the 16-dimensional state space, namely the six stable-state solutions of the ODEs. These states correspond to E state (epithelial state, high CDH1, and low VIM/ZEB1 expression), M state (mesenchymal state, high VIM/ZEB1, and low CDH1 expression), three intermediate states close to E state (IE1, IE2, and IE3, intermediate expression close to E state), and one intermediate

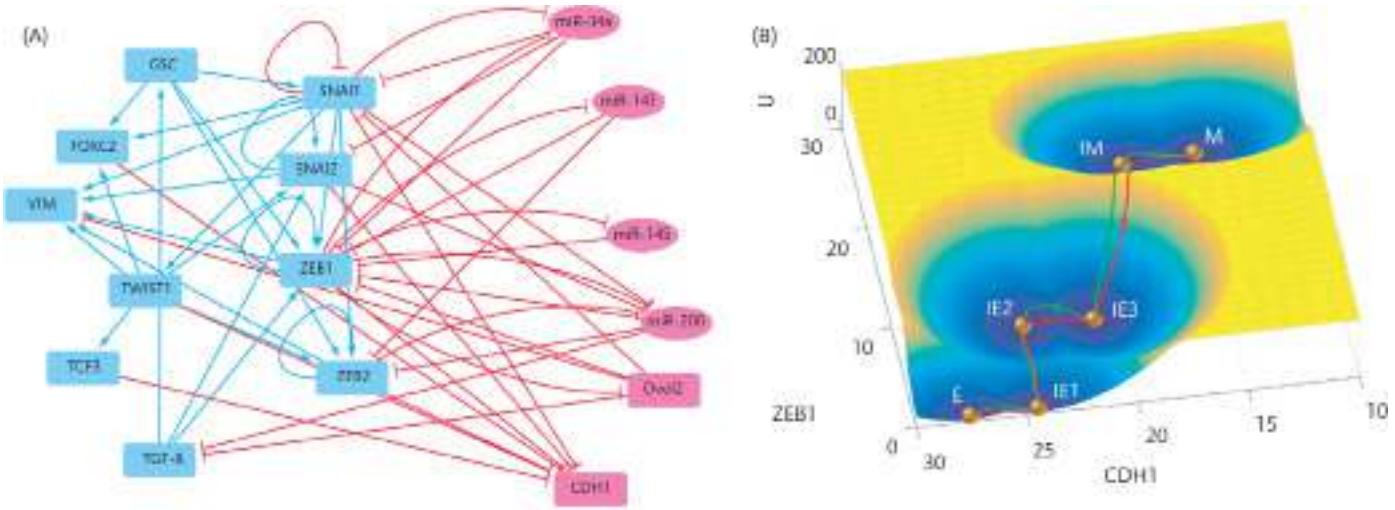


Figure 7.3. Landscape and transition path for EMT. (A) The diagram of EMT gene regulatory network. The network consists of 12 transcriptional factors (rectangle nodes), 4 microRNAs (ellipse nodes), and 53 regulatory links. Blue arrows represent activation and red bars represent repression. The six genes represented by the pink nodes (miR-145, miR-141, miR-200, miR-34a, Ovol2, and CDH1) are E markers, and the other genes represented by the blue nodes are M markers. (B) The landscape and transition paths between different attractor states in the CDH1/ZEB plane. The landscape surface is characterized by different colours, where the blue region represents a lower potential or higher probability and the yellow region represents a higher potential or lower probability. The red curve represents the transition path of EMT and the green one represents the transition path of MET.

Source: Taken from [64].

state close to M state (IM, intermediate expression close to M state), respectively.

Further, the 16-dimensional transition path or MAPs between neighbouring stable states can be calculated by the path integral approach. The MAPs are shown in Figure 7.3B on the landscape. The red MAP from the E state to the M state corresponds to the EMT process, while the green MAP from the M state to the E state corresponds to the MET process. Here, we minimize the transition action by the method of simulated annealing. The results of landscape and transition paths indicate that the cell-fate transition process for EMT needs to go through the intermediate states.

7.3.2. Landscape and transition path of EMT–metastasis network

In this subsection, we aim to introduce the relationship between EMT and metastasis, and discuss the mechanism of EMT-promoting metastasis from a GRN perspective. To do so, we first construct an EMT–metastasis regulatory network by merging an EMT gene network and a metastasis network obtained by literature mining (Figure 7.4A). The network includes 10 nodes and 26 links (8 activations and 18 repressions).

Similarly, following the self-consistent approximation approach, we determined the steady-state probability distribution and then mapped out the potential landscape for the EMT–metastasis system. Because it is difficult to visualize the landscape in a 10-dimensional space, we selected two variables as the coordinates and projected the 10-dimensional landscape into this two-dimensional space, by integrating out the other eight gene variables. We chose the two key variables ZEB and BACH1 as the two coordinates for the landscape since ZEB is a major EMT marker gene and BACH1 is a major metastasis marker and regulator gene. We found four stable cell states emerging on the landscape for the EMT–metastasis system (Figure 7.4B). The landscape surface is characterized by different colours, where the blue region represents a lower potential or higher probability, and the red region represents a higher potential or lower probability. The four basins of attraction on the landscape represent four different cell states characterized by different gene expression patterns in the 10-dimensional state space. These states separately correspond to M state (metastatic state, high ZEB/high BACH1 expression), A state (anti-metastatic state, low ZEB/low BACH1 expression), and two intermediate states (I1 and I2, intermediate ZEB and BACH1 expression).

To study the transitions among individual cell types, we calculated kinetic transition paths by minimizing the transition actions between attractors, obtaining MAPs (namely the above method). The MAPs for different transitions are shown on the landscape in Figure 7.4B. magenta solid lines represent the MAP from A state (antimetastatic cell state) to I states, and to M state (metastatic cell state), and the white solid lines represent the MAP from M to I, and to A state. The dashed lines represent the direct MAP from A to M and from M to A states, respectively. The lines represent the MAPs, with the arrows denoting the directions of the transitions. The MAP for pro-metastatic process and the MAP for anti-metastatic process are irreversible since the forward and reverse kinetic paths are not identical. This irreversibility of kinetic transition paths is caused by the non-gradient force or curl flux [22,66].

7.3.3. Landscape analysis of EMT–metastasis–metabolism network

Abnormal metabolism is an important hallmark of cancer, which has been explored in previous studies [5,67]. A computational model for cancer metabolism has been constructed using ODEs, which includes two genes (AMPK and HIF-1) and two metabolites (noxROS and mtROS) [68]. As mentioned above, the EMT has been suggested

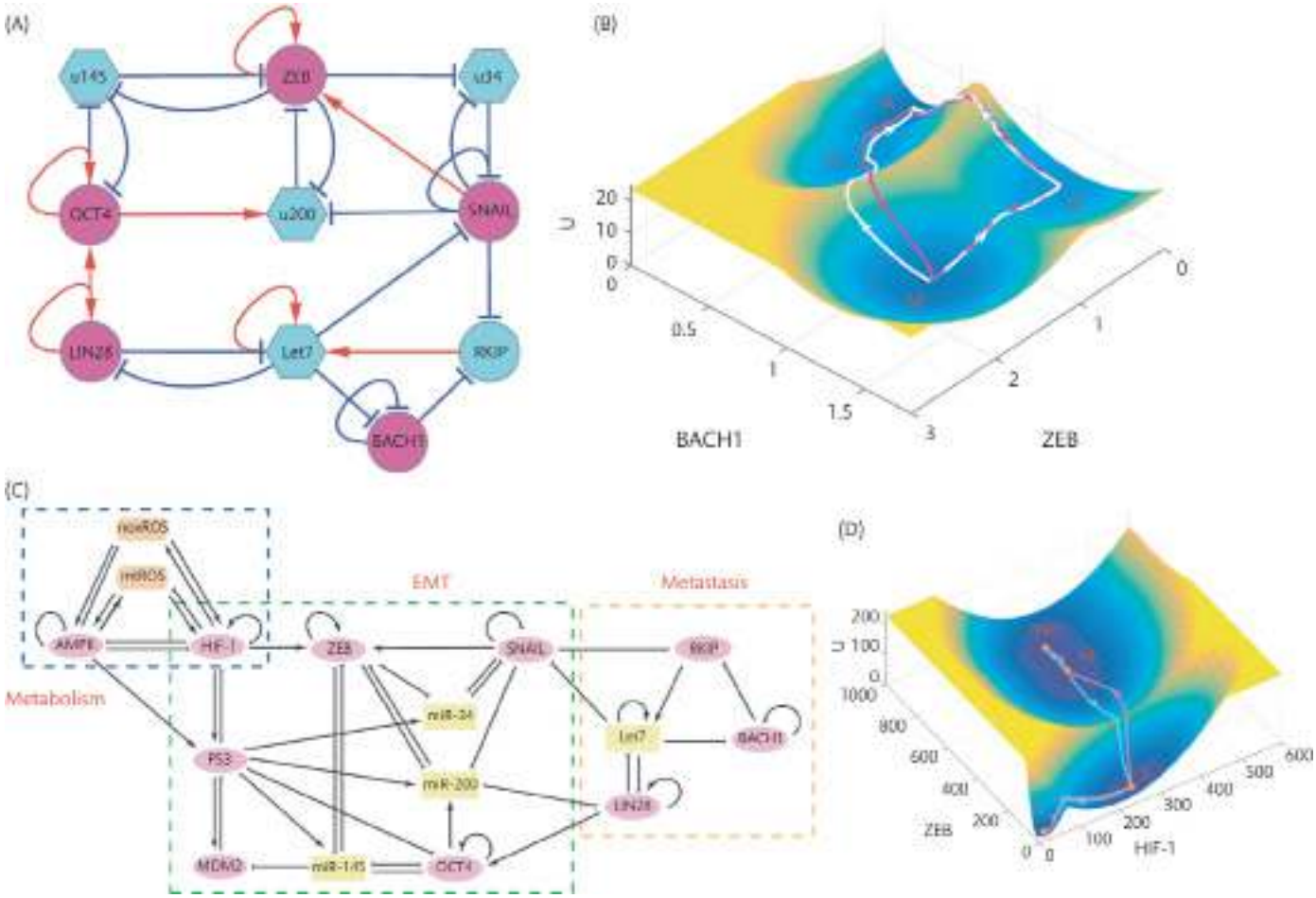


Figure 7.4. Landscape of cancer by coupling EMT, metastasis and metabolism circuit. (A) The diagram for the core circuit of EMT–metastasis network. Red arrows represent activation and blue bars represent repression. Magenta nodes represent pro-metastatic genes and cyan nodes represent anti-metastatic genes. Circle nodes represent proteins and hexagonal nodes represent microRNAs—u34: miR34; u200: miR200; u145: miR145. (B) The landscape and corresponding MAPs for the EMT–metastasis network. A: anti-metastatic state; M: metastatic state; I1, I2: intermediate state. Here, ZEB and BACH1 are selected as the two coordinates. (C) The regulatory network for the interplay among EMT, metabolism, and cancer metastasis. (D) The landscape for the interplay among EMT, metabolism, and cancer metastasis. *Source:* Panels (A) and (B) are taken from [27] and Panels (C) and (D) are taken from [65].

to be related to metastatic progression. Therefore, a key question is whether there is any relationship between EMT, metabolism, and cancer metastasis.

To uncover the mechanisms underlying the interplay between EMT, metabolism, and metastasis, we established a metabolism–EMT–metastasis GRN by incorporating the core components for each process through mining the experimental literature (Figure 7.4C). The metabolism–EMT–metastasis network involves 16 components (10 genes, 2 metabolites, and 4 microRNA) and 51 regulation links (22 activations and 29 inhibitions). AMPK, HIF-1, mtROS, and noxROS are the core components controlling cellular metabolism (Figure 7.4C, blue box). SNAIL, ZEB, OCT4, MDM2, miR-145, miR-200, miR-34, and P53 are the core components governing the EMT (Figure 7.4C, green box). RKIP, BACH1, LIN28, and Let7 are the core components governing the metastasis (Figure 7.4C, yellow box) [69].

Based on the network structure, we can write down the ODEs describing the time evolution of relative expression levels for each of the 16 genes or metabolites. Then, we calculated the steady-state probability distribution of the system and acquired the potential landscape (Figure 7.4D). Similarly, it is hard for visualization because we are dealing with a 16-dimensional potential landscape. Here, we pick two representative marker genes, HIF-1 and ZEB, as two coordinates and project the 16-dimensional landscape into the two-dimensional space.

On the landscape, the blue region represents a high probability or low potential, and the yellow region represents a low probability or high potential. We identified four stable states on the landscape, which characterize epithelial (E), abnormal metabolism (A), mesenchymal (M), and metastatic (Met) cell states, respectively (Figure 7.4D). Importantly, we identified a new intermediate state, which we defined as the abnormal metabolism (A) state since it has an increased expression of the glycolysis marker gene HIF-1. Here, the E state has a low HIF-1 and low ZEB expression. So, the marker genes for the abnormal metabolism, the EMT, and the metastasis are all off. The A state has a high HIF-1 and low ZEB expression and therefore corresponds to an abnormal aggressive metabolic phenotype (i.e. aerobic glycolysis state) whereby cancer cells change their

metabolism to produce energy more quickly. Besides, we have a mesenchymal state with a high HIF-1 and high ZEB expression and a metastatic state with a high HIF-1 and high ZEB expression.

To quantify the kinetic transitions among these states, we calculate the kinetic transition paths among different cell states by minimizing the transition actions S, namely MAPs. Figure 7.4D shows the landscape and transition paths for the quadrastable system. The magenta paths denote the transitions from the epithelial state to the metastatic state (metastatic progression process), whereas the cyan paths denote the transitions from the metastatic state to the epithelial state (de-metastasis process). The transition paths for the E to metastasis state transition and the backward transition paths from metastasis to the E state are not identical, reflected by the disparity between the forward and backward kinetic transition paths. This irreversibility of MAPs is a consequence of non-gradient force, i.e. curl flux [22,66].

7.4. Cancer cell plasticity and drug resistance

The landscape of EMT network is important to explore the transition of a cancer cell between different epigenetic states. In tumour progression, EMT is a process in which epithelial cells lose their junctions and polarity to gain a motile mesenchymal phenotype. Epigenetic plasticity of cancer cells is essential for EMT and drug resistance of cancer cells [70,71]. Thus, modelling of cancer cell plasticity is essential for the understanding of tumour evolution [12].

To apply the model framework (7.22) to cancer evolution with cell plasticity, we need to select proper epigenetic variable $\mathbf{x}$ and the dependence of the kinetotype of a cell on the epigenetic state. We usually do not include the transcriptome of all genes in the epigenetic variable. Instead, specific genes are often selected into the epigenetic state, such as the genes in the EMT network when concerning cancer cell plasticity and drug resistance. Alternatively, we can also define the epigenetic variable as macroscopic variables that can describe the state of a cell, such as the stemness that is associated with the potential of self-renewal and differentiation [72] or the single-cell entropy that is associated with the malignance of a cell [73]. Here, two examples are introduced to show how the framework (7.22) can be applied to model cancer cell plasticity and drug resistance.

7.4.1. Modelling tumour evolution with cancer cell plasticity

Now, we consider the situation with epigenetic state $\mathbf{x} = (x_1, x_2)$, where x_1 represents the stemness of a cell and x_2 stands for the malignance of a cell. Here, x_1 and x_2 are normalized to the interval $[0, 1]$ so that the epigenetic state $\mathbf{x} \in \Omega = [0, 1] \times [0, 1]$. Moreover, to consider the effect of the microenvironment on cancer evolution, we introduce a microenvironment index u, which represents the effects of the microenvironment on malignance and cell survival.

Mathematical formulations for the dependence of kinetic rates on the epigenetic state are given below. First, both proliferation rate β and differential rate κ are dependent on the stemness x_1. The differentiation rate usually decreases with the increase of stemness. The proliferation rate, however, depends on the stemness in a more complex way. A larger value x_1 means a higher level stemness, which means the potential of a quiescent state with a low proliferation rate. The proliferation rate increases with the decrease of x_1, which means the state of progenitor cells. Furthermore, the proliferation rate decrease to zero when the stemness further decreases to zero, which means the state of terminal differentiation. From these assumptions, referring to (7.20), we define the proliferation rate and differentiation rate as

$$\beta(c,\mathbf{x}) = \beta_0(\mathbf{x})\frac{1}{1+(c/\theta)^n}, \quad \beta_0(\mathbf{x}) = \frac{a_1x_1 + (a_2x_1)^{s_1}}{1+(a_3x_1)^{s_1}} \tag{7.29}$$

and

$$\kappa(\mathbf{x}) = \frac{\kappa_0}{1+(b_1x_1)^{s_1}}. \tag{7.30}$$

Here, we simply take c as the cell number Q. The coefficient $\theta(\mathbf{x})$ represents the repression of cell proliferation through cell responses to microenvironmental cytokines and is dependent on the malignancy. Hence, we assume that θ increases with the malignancy and hence

$$\theta(\mathbf{x}) = \theta_0 + \theta_1\frac{x_2^{s_2}}{\theta_1^{s_2} + x_2^{s_2}}, \tag{7.31}$$

here θ_0, θ_1, and θ_2 are parameters.

To model the dependence of cell survival on the microenvironment and malignancy, we introduce a fitness function

$$g(u,x_2) = 3.0x_2^u(1-x_2)^{1-u} \tag{7.32}$$

to represent how a cell with malignancy index x_2 may fit the microenvironmental condition u. Better fitness implies a lower apoptosis rate of the cell, and hence the apoptosis rate can be defined as

$$\mu(\mathbf{x},u) = \frac{\mu_0}{1+\rho e^{g(u,x_2)}}. \tag{7.33}$$

Here, μ_0 and ρ are constants, so there is a maximum apoptosis rate $\mu_0 = (1 + \rho)$ when the fitness $g = 0$.

Similar to the previous argument, the transition function

$$p(\mathbf{x},\mathbf{y}) = p_1(x_1,\mathbf{y}) \times p_2(x_2,\mathbf{y}),$$

where $p_i(x_i, \mathbf{y})$ are density functions of beta distribution

$$p_i(x_i,\mathbf{y}) = \frac{x_i^{a_i(\mathbf{y})-1}(1-x)_i^{b_i(\mathbf{y})-1}}{B(a_i(\mathbf{y}),(b_i(\mathbf{y}))}, \quad B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

The shape parameters $a_i(\mathbf{y})$ and $b_i(\mathbf{y})$ are defined by the predefined functions $\phi_i(\mathbf{y})$ and $\eta_i(\mathbf{y})$ according to (7.27) and (7.28).

We can usually take $\eta_1(\mathbf{y}) = \eta_2(\mathbf{y}) = \eta$ as constants. The stemness of the daughter cell usually depends on the stemness of the mother cells, and hence $\phi_1(\mathbf{y}) = \phi_1(y_1)$, and the malignancy of the daughter cell depends on the malignancy of the mother cell so that $\phi_2(\mathbf{y}) = \phi_2(y_2)$. The functions ϕ_i are often increased functions because of the inertial effects. Here, we take the Hill type functions

$$\phi_1(\mathbf{y}_1) = c_1 + d_1 \times \frac{(\alpha_1y_1)^{1.5}}{1+(\alpha_1y_1)^{1.5}}, \tag{7.34}$$

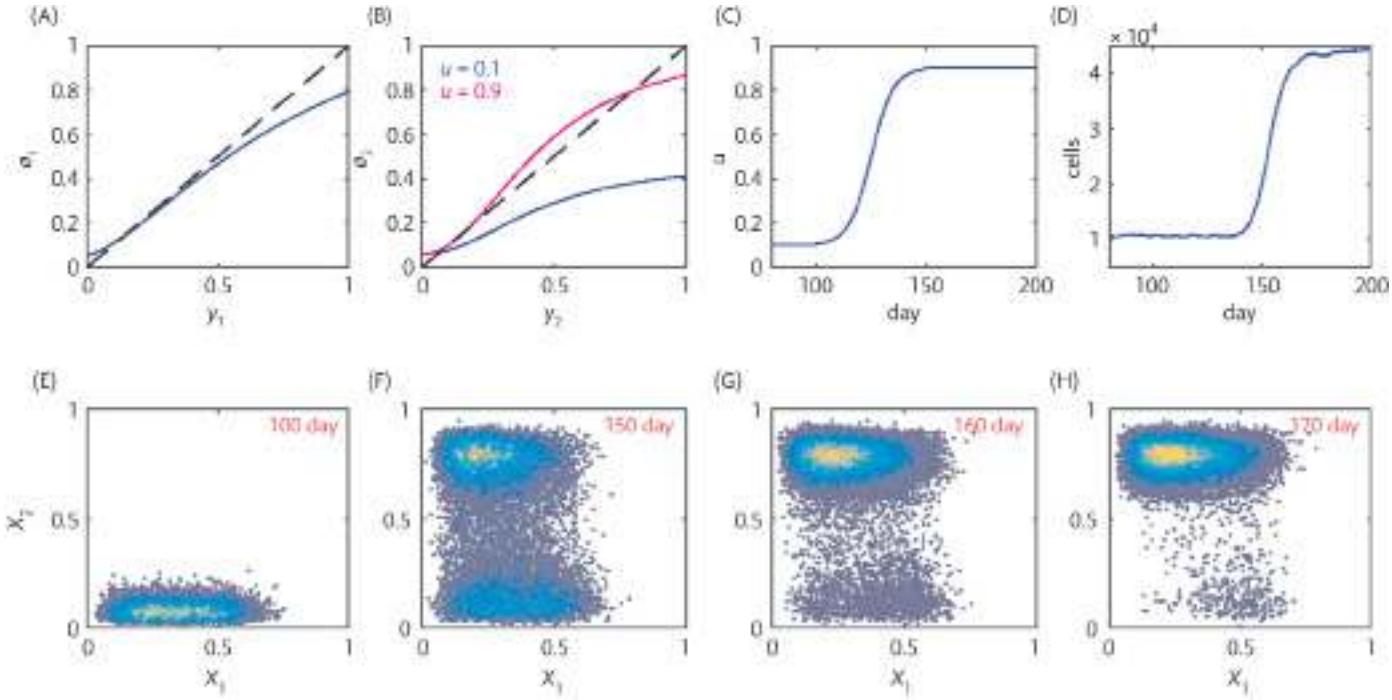


Figure 7.5. Simulation of microenvironmental-change-induced abnormal cell growth. (A) The function $\phi_1(y_1)$. (B) The function $\phi_2(y_2)$ with $u = 0.1$ and $u = 0.9$, respectively. (C) The function $u(t)$ in simulation. (D) Cell number $Q(t)$ following the changes in the microenvironmental index. (E–H) Epigenetic state of cells at different time points. Parameter values are $s_0 = 1$, $a_1 = 6.8$, $a_2 = 2.2$, $a_3 = 3.25$, $s_1 = 3.5$, $\theta_1 = 1{,}000$, $\theta_2 = 0.4$, $s_2 = 8.5$, $b_1 = 2.1$, $s_3 = 6.0$, $\mu_0 = 0.3$, $c = 1.0$, $c_1 = 0.055$, $c_2 = 0.06$, $d_1 = 1.3$, $\alpha_1 = 1.2$, $\alpha_2 = 2.2$, $\beta_0 = 10.0$, $\kappa_0 = 0.5$, $\eta = 70$, and $d_2 = 0.68u + 0.35$. Source: Replotted from [74].

$$\phi_2(\mathbf{y}_2) = c_2 + d_2 \times \frac{(\alpha_2 y_2)^{2.1}}{1+(\alpha_2 y_2)^{2.1}}. \tag{7.35}$$

Here, c_1, c_2, d_1 are constants, and d_2 may depend on the microenvironmental index u. When the microenvironment becomes abnormal (increases of u), the cells tend to be more malignancy so that d_2 increases with u. Figure 7.5A and B shows the plots of functions $\phi_1(y_1)$ and $\phi_2(y_2)$.

Based on the above formulations, Figure 7.5 shows an example simulation of abnormal cell growth dynamics when the microenvironmental index changes from normal value ($u = 0.1$) to abnormal value ($u = 0.9$). From the simulation results, obvious changes in the epigenetic state of cells are seen on day 150 before the obvious increase in the cell number. These results show how cell plasticity drives abnormal cell growth in response to changes in the microenvironment.

7.4.2. Cell plasticity and immune escape after CAR-T therapy

Cancer immunotherapy has been a breakthrough in recent years. However, cancer immune escape often occurs after the administration of immunotherapy [75–77]. Here, we introduce an example of how cancer cell plasticity induces immune escape after CAR-T therapy.

Chimeric antigen receptor (CAR) T-cell therapy targeting CD19 has been proven to be an effective therapy for B-cell acute lymphoblastic leukaemia (B-ALL). The majority of patients achieve a complete response following a single infusion of CD19-targeted CAR-T cells; however, many patients suffer a relapse after therapy, and the with leukaemic cells, 60% of the mice relapsed within 3 months, and the relapsed tumours underlying mechanism remains unclear.

When second-generation CAR-T cells are injected into mice retained CD19 expression but exhibited a profound increase in CD34 transcription [78]. Based on these observations, an individual cell-based computational model was developed to show the mechanism of cancer cell-plasticity-induced immune escape after CAR-T therapy. Here, we introduce the main assumption and formulations of the model. For detailed discussions, refer to [78].

Experimental results show that relapsed tumours in mice after infusion with CD19-28z-T cells retained CD19 expression but exhibit a subpopulation of $CD19^+$ $CD34^+$ and $CD123^+$ $CD34^+$ tumour cells, which are not shown in control NGFT-28z-treated mice. Based on this observation, key assumptions were proposed that CAR-T-induced tumour cells transition into haematopoietic stem-like cells (by promoting CD34 expression) and myeloid-like cells (by promoting CD123 expression), and hence escape CAR-T cell targeting. In the model, each cell was represented by the epigenetic state of marker genes CD19, CD22, CD34, and CD123, which play important roles in the CD19 CAR-T cell response and cell lineage dynamics. The proliferation rate β and differentiation rate κ depend on CD34 expression level through

$$\beta = \beta_0 \frac{\theta}{\theta+N} \times \frac{5.8[\mathrm{CD34}] + (2.2[\mathrm{CD34}])^6}{1+(3.75[\mathrm{CD34}])^6},$$

$$\kappa = \kappa_0 \frac{1}{1+(4.0[\mathrm{CD34}])^6}.$$

Here, N represents the total cell number. Here, the expression level of CD34 can be considered as the stemness of a cell. Hence, similar to the above argument, the functions β and κ were utilized so that a cell has a low proliferation rate and differentiation rate in the quiescent state with high level [CD34], a high proliferation rate and low differentiation rate in the progenitive state with intermediate level [CD34], and a low proliferation rate and high differentiation rate when the cell is mature and ready for differentiation. The apoptosis rate μ includes a basal rate μ_0 and a rate associated with the CAR-T signal

$$\mu = \mu_0 + \mu_1 \times \text{Signal},$$

$$\text{Signal} = f([\text{CD34}],[\text{CD123}])\frac{\gamma_{19}[\text{CD19}]}{1+\gamma_{19}[\text{CD19}]+\gamma_{22}[\text{CD22}]}R(t),$$

$$f([\text{CD34}],[\text{CD123}]) = \frac{1}{1+([\text{CD34}]/X_0)^{n_0})(1+([\text{CD123}]/X_1)^{n_1})}$$

Here, $R(t)$ is the predefined CAR-T activity. CD34 and CD123 are marker genes of stem-like cells and myeloid-like cells, respectively, which were assumed to inhibit CAR-T signalling.

Similar to the previous discussions, the expression levels of marker genes changed randomly following the transition probability of beta distributions, and the shape parameters were dependent on the state of the mother cells and the CAR-T signal. For example, given the expression level of CD34 in cycle k as u_k, the expression level for cycle $k+1$ (denoted by u_{k+1}) is a random number of beta distribution with probability density function

$$P(u_{k+1} = u \mid u_k) = \frac{u^{a_{34}-1}(1-u)^{b_{34}-1}}{B(a,b)}, \quad B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

where the shape parameters a and b are dependent on the conditional expectation and the conditional variance of u_{k+1}. When

$$E(u_{k+1} \mid u_k) = \phi_{34}(u_k), \ \text{Var}(u_{k+1} \mid u_k) = \frac{1}{1+\eta_{34}}\phi_{34}(u_{34})(1-\phi_{34}(u_{34})),$$

then

$$a = \eta_{34}\phi_{34}(u_k), \ b = \eta_{34}(1-\phi_{34}(u_k)).$$

In the model, it was assumed that $\eta_{34} = 60$ and

$$\phi_{34}(u_k) = 0.08 + 1.06\frac{(\alpha_{34}u_k)^{2.2}}{1+(\alpha_{34}u_k)^{2.2}},$$

and let

$$\alpha_{34} = 1.45 + 0.16 \times [\text{CD19}] + A_{34} \times \text{Signal}$$

to represent the promotion of CD34 expression by CD19 and the CAR-T signal. For details of the model, refer to [78].

Simulations shown in [78] effectively reproduced experimental results and predicted that CAR-T cell-induced cell plasticity can lead to tumour relapse in B-ALL after CD19 CAR-T treatment. Specifically, simulations show the increase of the fractions of $CD34^+$ and $CD123^+$ cells from day 5 to day 23 after treatment, during which the cell number begins to increase. The fractions of $CD34^+$ and $CD123^+$ cells decrease to low levels at the latter stage due to the exhaustion of CAR-T cells. These results suggest the occurrence of cell plasticity in response to therapy stress that is essential for tumour relapse. Cancer cell plasticity is a major obstacle to the effective treatment of many cancers. The modelling technique introduced in this chapter provides a general framework to study this process.

7.5. Evolution of the Waddington landscape

We have seen that Waddington's epigenetic landscape is a general and qualitative concept of understanding cell-fate decisions and plasticity. Waddington landscape provides a great picture of how a cell selects its type during development as well as the possibility of cell-type switches due to trans-differentiation. The concept of landscape is also important when we try to describe the dynamic change of the epigenetic state of cells in a multicellular tissue. In the situation of tumour evolution, many cells undergo regeneration and cell-type transitions to form a landscape of the epigenetic state of all tumour cells. Here, we introduce a mathematical model to describe the evolution of the Waddington landscape.

From Equation (7.22), let

$$Q(t) = \int_\Omega Q(t,\mathbf{x})\,d\mathbf{x} \tag{7.36}$$

to represent the total cell number. The relative cell number with epigenetic state $\mathbf{x}$ is given by

$$f(t,\mathbf{x}) = \frac{Q(t,\mathbf{x})}{Q(t)}. \tag{7.37}$$

The function $f(t, \mathbf{x})$ shows the evolution of the probability density of epigenetic states, and the evolution of the landscape is given by

$$W(t,\mathbf{x}) = -\log f(t,\mathbf{x}). \tag{7.38}$$

From (7.22), and integrating both sides of the equation, we obtain

$$\begin{aligned}\frac{dQ}{dt} &= -\int_\Omega Q(t,\mathbf{x})(\beta(c,\mathbf{x})+\kappa(\mathbf{x}))d\mathbf{x} \\ &\quad + 2\int_\Omega \beta(c_{\tau(\mathbf{x})},\mathbf{x})Q(t-\tau(\mathbf{x}),\mathbf{x})e^{-\mu(\mathbf{x})\tau(\mathbf{x})}d\mathbf{x}.\end{aligned} \tag{7.39}$$

If we omit the heterogeneity, all rate functions are independent of $\mathbf{x}$ and $c(t) = Q(t)$, we re-obtain the delay differential equation (7.16) for homogeneous stem cell regeneration.

From (7.37), (7.39), and (7.22), we obtain the equation for $f(t, \mathbf{x})$ as

$$\begin{aligned}&\frac{\partial f(t,\mathbf{x})}{\partial t} \\ &= \frac{2}{Q(t)}\int_\Omega \beta(c_{\tau(\mathbf{y})},\mathbf{y})Q(t-\tau(\mathbf{y}),\mathbf{y})e^{-\mu(\mathbf{y})\tau(\mathbf{y})}(p(\mathbf{x},\mathbf{y}) \\ &\quad - f(t,\mathbf{x}))d\mathbf{y} - f(t,\mathbf{x})\int_\Omega f(t,\mathbf{y})((\beta(c,\mathbf{x})+\kappa(\mathbf{x})) \\ &\quad - ((\beta(c,\mathbf{y})+\kappa(\mathbf{y})))d\mathbf{y}.\end{aligned} \tag{7.40}$$

Since

$$\frac{\partial W(t,\mathbf{x})}{\partial t} = -\frac{1}{f(t,\mathbf{x})}\frac{\partial f(t,\mathbf{x})}{\partial t},$$

we have

$$\frac{\partial W(t,\mathbf{x})}{\partial t} = \frac{2}{Q(t,\mathbf{x})}\int_\Omega \beta(c_{\tau(\mathbf{y})},\mathbf{y})Q(t-\tau(\mathbf{y}),\mathbf{y})e^{-\mu(\mathbf{y})\tau(\mathbf{y})}(p(\mathbf{x},\mathbf{y}))-f(t,\mathbf{x})d\mathbf{y} + \int_\Omega f(t,\mathbf{y})((\beta(c,\mathbf{x})+\kappa(\mathbf{x}))-(\beta(c,\mathbf{y})+\kappa(\mathbf{y})))d\mathbf{y}. \quad (7.41)$$

Equation (7.41) gives the evolution of the Waddington landscape of a system of stem cell regeneration with cell heterogeneity and plasticity.

Particularly, when the system reaches the equilibrium state that $Q(t)$ and $f(t, \mathbf{x})$ are independent of the time t, we can write

$$Q(t)=Q^*, f(t,\mathbf{x})=f^*(\mathbf{x}),$$
$$W(t,\mathbf{x})=W^*(\mathbf{x}), c(t)=\int Q(t,\mathbf{x})\zeta(\mathbf{x})d\mathbf{x}=c^*,$$

then equation (7.40) becomes

$$2\int_\Omega \beta(c^*,\mathbf{y})e^{-\mu(\mathbf{y})\tau(\mathbf{y})}f^*(\mathbf{y})(p(\mathbf{x},\mathbf{y})-f^*(\mathbf{x}))d\mathbf{y} - f^*(\mathbf{x})\int_\Omega f^*(\mathbf{y})((\beta(c^*,\mathbf{x})+\kappa(\mathbf{x}))-(\beta(c^*,\mathbf{y})+\kappa(\mathbf{y})))d\mathbf{y}=0.$$

Thus, define a nonlinear operator F_c as

$$F_c[f]=2\int_\Omega \beta(c,\mathbf{y})e^{-\mu(\mathbf{y})\tau(\mathbf{y})}f(\mathbf{y})(p(\mathbf{x},\mathbf{y})-f(\mathbf{x}))d\mathbf{y} - f(\mathbf{x})\int_\Omega f(\mathbf{y})((\beta(c,\mathbf{x})+\kappa(\mathbf{x}))-(\beta(c,\mathbf{y})+\kappa(\mathbf{y})))d\mathbf{y}.$$

Equation (7.42) becomes a nonlinear eigenvalue problem

$$F_c[f]=0. \quad (7.42)$$

The equilibrium state density function $f^*(\mathbf{x})$ corresponds to the eigenfunction of the operator F_c with a positive eigenvalue c^*. The Waddington landscape is given by

$$W^*(\mathbf{x})=-\log f^*(\mathbf{x}). \quad (7.43)$$

Thus, the mathematical formulation provides a general method of calculating the evolution of Waddington's landscape during tumour evolution.

7.6. Conclusions

In this chapter, we reviewed mathematical modelling approaches of cancer gene networks. From these approaches, we can quantify the underlying potential landscape of the cancer network and identify the attractor states in gene network state space as distinct biological functional states (normal, cancer, and apoptosis states). Through the landscape topography and kinetic transition rates between attractors, the global stability and the capability of the transitions between normal and cancer states can be quantified. Based on the path integral method, one can uncover the underlying mechanism of the state transition and quantify the transition process among different cell states. Further, using the landscape and transition path framework, one can interrogate the important relationship between cancer metastasis, EMT, and metabolism. Several examples are introduced to illustrate the procedure for stochastic analysis of these dynamical processes. The results for the landscape and kinetic transition paths for the cancer network, as well as the EMT–metastasis–metabolism network, can help to develop strategies for cancer prevention and treatment.

Tumour evolution is a multiscale dynamical process in which stochastic gene network regulations interconnect with cell cycling to form a complex interaction relationship. It is not enough to focus on cell plasticity in a single cell while we try to understand the dynamics of tumour evolution. Here, we introduce a mathematical framework to describe the dynamics of cell regeneration with highlights on cell heterogeneity and plasticity. The model provides an integration of multiple scales of interactions, including single-cell epigenetic state, cell behaviour, cytokine secretion, the transition of epigenetic state during cell cycling, microenvironmental conditions, and population dynamics. This model framework can be applied to different problems related to cell regeneration, such as development, ageing, and tumour evolution. Through the model, we are able to develop a computational model to simulate the process of microenvironmental-change-induced abnormal growth and cell-plasticity-induced immune escape after CAR-T therapy. Moreover, the evolution of Waddington's epigenetic landscape can be obtained from the model. Hence, the model introduced in the chapter provides a logical connection between cell-fate decisions at the single-cell level and the tissue growth level.

REFERENCES

1. Jemal A, Bray F, Center M, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J. Clin*. **61**:69–90, 2011.
2. Cowin P, Rowlands T. Cadherins and catenins in breast cancer. *Curr Opin Cell Biol*. 2005;**17**:499–508.
3. Kauffman S. Differentiation of malignant to benign cells. *J Theor Biol*. 1971;**31**:429–51.
4. Hanahan D, Weinberg R. The hallmarks of cancer. *Cell*. 2000;**100**:57–70.
5. Hanahan D, Weinberg R. Hallmarks of cancer: the next generation. *Cell*. 2011;**144**:646–74.
6. Weinberg R. *The biology of cancer*. Garland Science; New York, 2007.
7. Hejmadi M. *Introduction to cancer biology*. Holstebro, Denmark: Ventus Publishing; 2010.
8. Vogelstein B, Kinzler K. Cancer genes and the pathways they control. *Nat Med*. 2004;**10**:789–99.
9. Ao P, Galas D, Hood L, Zhu X. Cancer as robust intrinsic state of endogenous molecular–cellular network shaped by evolution. *Med. Hypotheses*. 2008;**70**:678–84.
10. Frank SA. *Dynamics of cancer: incidence, inheritance, and evolution*. Princeton (NJ): Princeton University Press; 2007.
11. Beerenwinkel N, Greenman CD, Lagergren J. Computational cancer biology: an evolutionary perspective. *PLoS Comput Biol*. 2016 Feb;**12**(2):e1004717.
12. Lei J. Evolutionary dynamics of cancer: from epigenetic regulation to cell population dynamics—mathematical model framework, applications, and open problems. *Sci China Math*. 2020 Jan;**63**(3):411–24.
13. Rozhok AI, DeGregori J. The evolution of lifespan and age-dependent cancer risk. *Trends Cancer*. 2016 Sep;**2**(10):552–60.
14. Wu H, Zhang X-Y, Hu Z, Hou Q, Zhang H, Li Y, et al. Evolution and heterogeneity of nonhereditary colorectal cancer

revealed by single-cell exome sequencing. *Oncogene*. 2017 May;**36**(20):2857–67.
15. Horne SD, Pollick SA, Heng HHQ. Evolutionary mechanism unifies the hallmarks of cancer. *Int J Cancer*. 2015 May;**136**(9):2012––21.
16. Li C, Wang J. Quantifying the underlying landscape and paths of cancer. *J R Soc Interface*. 2014;**11**:20140774.
17. Wang J, Zhang K, Xu L, Wang E. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc Natl Acad Sci U S A*. 2011;**108**(20):8257–62.
18. Lu M, Onuchic J, Ben-Jacob E. Construction of an effective landscape for multistate genetic switches. *Phys Rev Lett*. 2014;**113**(7):078102.
19. Li C, Wang J. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: Landscape and biological paths. *PLoS Comput Biol*. 2013;**9**(8):e1003165.
20. Sasai M, Wolynes PG. Stochastic gene expression as a many-body problem. *Proc Natl Acad Sci U S A*. 2003;**100**:2374–79.
21. Li C, Wang J. Quantifying cell fate decisions for differentiation and reprogramming of a human stem cell network: Landscape and biological paths. *PLoS Comput Biol*. 2013;**9**(8):e1003165.
22. Li C, Wang J. Landscape and flux reveal a new global view and physical quantification of mammalian cell cycle. *Proc Natl Acad Sci U S A*. 2014;**111**(39):14130–5.
23. Zhang B, Wolynes PG. Stem cell differentiation as a many-body problem. *Proc Natl Acad Sci*. 2014;**111**(28):10185–190.
24. Hu G. *Stochastic forces and nonlinear systems*. Shanghai: Shanghai Scientific and Technological Education Publishing House; 1994.
25. Van Kampen NG. *Stochastic processes in physics and chemistry*. 3rd *ed*. World Book Inc; New York, 2010.
26. Zhou X, Ren WQ, Weinan E. Adaptive minimum action method for the study of rare events. *J Chem Phys*. 2008;**128**(10):104111.
27. Li C, Balazsi G. A landscape view on the interplay between EMT and cancer metastasis. *NPJ Syst Biol Appl*. 2018;**4**:34.
28. Freidlin M, Weber M. Random perturbations of dynamical systems and diffusion processes with conservation laws. *Probab Theory Relat Fields*. 2004;**128**(3):441–66.
29. Wang J, Zhang K, Wang EK. Kinetic paths, time scale, and underlying landscapes: a path integral framework to study global natures of nonequilibrium systems and networks. *J Chem Phys*. 2010;**133**:125103.
30. Burns FJ, Tannock IF. On the existence of a G0-phase in the cell cycle. *Cell Prolif*. 1970;**3**(4):321–34.
31. Lei J, Mackey MC. Multistability in an age-structured model of hematopoiesis: cyclical neutropenia. *J Theor Biol*. 2011 Feb;**270**(1):143–53.
32. Moustakas A, Pardali K, Gaal A, Heldin CH. Mechanisms of TGF-β signaling in regulation of cell growth and differentiation. *Immunol Lett*. 2002;**82**(12):85–91.
33. Yang L, Pang YL, Moses HL. TGF-β and immune cells: an important regulatory axis in the tumor microenvironment and progression. *Trends Immunol*. 2010;**31**(6):220–27.
34. Samuel Bernard, Jacques Bélair, and Michael C Mackey. Oscillations in cyclical neutropenia: new evidence based on mathematical modeling. *J Theor Biol*. 2003 Aug;**223**(3):283–98.
35. Lei J. A general mathematical framework for understanding the behavior of heterogeneous stem cell regeneration. *J Theor Biol*. 2020 May;**492**:110196.
36. Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell*. 2000 Jan;**100**(1):57–70.
37. Lei J, Levin SA, Nie Q. Mathematical model of adult stem cell regeneration with cross-talk between genetic and epigenetic regulation. *Proc Natl Acad Sci U S A*. 2014 Mar;**111**(10):E880–87. .
38. Probst AV, Dunleavy E, Almouzni G. Epigenetic inheritance during the cell cycle. *Nat Rev Mol Cell Biol*. 2009 Feb;**10**(3):192–206.
39. Schepeler T, Page ME, Jensen KB. Heterogeneity and plasticity of epidermal stem cells. *Development*. 2014 June;**141**(13):2559–67.
40. Serra-Cardona A, Zhang Z. Replication-coupled nucleosome assembly in the passage of epigenetic information and cell identity. *Trends Biochem Sci*. 2018 Feb;**43**(2):136–48.
41. Singer ZS, Yong J, Tischler J, Hackett JA, Altinok A, Surani MA, et al. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol Cell*. 2014 July; **55**(2):319–31.
42. Takaoka K, Hamada H. Origin of cellular asymmetries in the pre-implantation mouse embryo: a hypothesis. *Philos Trans R Soc Lond B Biol Sci*. 2014 Dec;**369**(1657): 20130536.
43. Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell*. 2014 Jan;**156**(1–2):45–68.
44. Lander AD, Gokoffski KK, Wan FYM, Nie Q, Calof AL. Cell lineages and the logic of proliferative control. *PLoS Biol*. Jan;**7**(1):e15.
45. Mangel M, Bonsall MB. Stem cell biology is population biology: differentiation of hematopoietic multipotent progenitors to common lymphoid and myeloid progenitors. *Theor Biol Med Model*. 2012 Dec;**10**(1):5–5.
46. Huang R, Lei J. Dynamics of gene expression with positive feedback to histone modifications at bivalent domains. *Int J Mod Phys B*. 2017 Nov;**4**:1850075.
47. Huang R, Lei J. Cell-type switches induced by stochastic histone modification inheritance. *Discrete Cont Dyn Syst–B*. 2019;**24**:5601–19.
48. Sahoo S, Mishra A, Kaur H, Hari K, Muralidharan S, Mandal S, et al. A mechanistic model captures the emergence and implications of non-genetic heterogeneity and reversible drug resistance in ER+ breast cancer cells. *NAR Cancer*. 2021;**3**(3):zcab027.
49. Lei J. *Systems biology, modeling, analysis, and simulation*. lecture notes on mathematical modelling in the life sciences. Springer; Switzerland, 2021.
50. Cai L, Friedman N, Sunney X. Stochastic protein expression in individual cells at the single molecule level. *Nature*. 2006 March;**440**(7):358–62.
51. Song Y, Yang S, Lei J. ParaCells: a GPU architecture for cell-centered models in computational biology. *IEEE/ACM Trans Comput Biol Bioinform*. 2018 March;**16**(3):994–1006.
52. Brabletz T, Lyden D, Steeg P, Werb Z. Roadblocks to translational advances on metastasis research. *Nat. Med*. 2013;**19**:1104–109.
53. Nieto M. The ins and outs of the epithelial to mesenchymal transition in health and disease. *Annu Rev Cell Dev Biol*. 2011;**27**:347–76.
54. Thiery J, Acloque H, Huang R, Nieto M. Epithelial-mesenchymal transitions in development and disease. *Cell*. 2009;**139**:871–90.
55. Nakaya Y, Sheng G. EMT in developmental morphogenesis. *Cancer Lett*. 2013;**341**:9–15.
56. Jia D, Jolly M, Kulkarni P, Levine H. Phenotypic plasticity and cell fate decisions in cancer: insights from dynamical systems theory. *Cancers*. 2017;**9**:70.
57. Heerboth S. EMT and tumor metastasis. *Clin Transl Med*. 2015;**4**:6.
58. Brabletz T, Kalluri R, Nieto M, Weinberg R. EMT in cancer. *Nat Rev Cancer*. 2018;**18**:128.

59. Lambert A, Pattabiraman D, Weinberg R. Emerging biological principles of metastasis. *Cell*. 2017;**168**:670–91.
60. Gibbons D, Lin W, Creighton C, Rizvi Z, Gregory P, Goodall G, et al. Contextual extracellular cues promote tumor cell EMT and metastasis by regulating mir-200 family expression. *Genes Dev*. 2009;**23**:2140–151.
61. Huang B, Lu M, Jia D, Ben-Jacob E, Levine H, Onuchic JN. Interrogating the topological robustness of gene regulatory circuits by randomization. *PLoS Comput Biol*. 2017;**13**(3):e1005456.
62. Hong T, Watanabe K, Ta CH, Villarreal-Ponce A, Nie Q, Dai X. An ovol2-zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Comput Biol*. 2015 Jan;**11**:e1004569.
63. Li C, Wang J. Quantifying the landscape for development and cancer from a core cancer stem cell circuit. *Cancer Res*. 2015;**75**(13):2067–18.
64. Lang J, Nie Q, Li C. Landscape and kinetic path quantify critical transitions in epithelial-mesenchymal transition. *Biophys J*. 2021;**120**:4484– 500.
65. Kang X, Wang J, Li C. Exposing the underlying relationship of cancer metastasis to metabolism and epithelial-mesenchymal transitions. *iScience*. 2019;**21**:754–72.
66. Wang J, Xu L, Wang E. Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation and coherence of biochemical oscillations. *Proc Natl Acad Sci. U S A*. 2008;**105**:12271–276.
67. Boroughs L, Deberardinis R. Metabolic pathways promoting cancer cell survival and growth. *Nat Cell Biol*. 2015;**17**:351.
68. Yu L, Lu M, Jia D, Ma J, Ben-Jacob E, Levine H, et al. Modeling the genetic regulation of cancer metabolism: interplay between glycolysis and oxidative phosphorylation. *Cancer Res*. 2017;**77**:1564–74.
69. Lee J, Yesilkanal A, Wynne J, Frankenberger C, Liu J, Yan J, et al. Effective breast cancer combination therapy targeting bach1 and mitochondrial metabolism. *Nature*. 2019;**568**:254.
70. Tam WL, Weinberg RA. The epigenetics of epithelial–mesenchymal plasticity in cancer. *Nat Med*. 2013;**19**(11):1438–49.
71. Skrypek N, Goossens S, Smedt ED, Vandamme N, Berx G. Epithelial-to-mesenchymal transition: epigenetic reprogramming driving cellular plasticity. *Trends Genet*. 2017 Sep;**33**(12):943–59.
72. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell*. 2008;**173**:338–54.
73. Liu J, Song Y, Lei J. Single-cell entropy to quantify the cellular order parameter from single-_
74. Zhang Y-D, Lei J. Entropy changes during the proliferation of heterogeneous stem cells. Preprint; 2022.
75. Kim R, Emi M, Tanabe K. Cancer immunoediting from immune surveillance to immune escape. *Immunology*. 2007 May;**121**(1):1–14.
76. Horn LA, Fousek K, Palena C. Tumor Plasticity and Resistance to Immunotherapy. *Trends Cancer*. 2020 May;**6**(5):432–41.
77. Hegde PS, Chen DS. Top 10 challenges in cancer immunotherapy. *Immunity*. 2020 Jan 1;**52**(1):17–35.
78. Zhang C, Shao C, Jiao X, Bai Y, Li M, Shi H, et al. Individual cell-based modeling of tumor cell plasticity-induced immune escape after CART therapy. *Comput Syst Oncol*. 2021;**1**:e21029.

8

The road to cancer and back: A thermodynamic point of view

Arnab Barua and Haralampos Hatzikirou

8.1. Introduction

The process of decision-making involves selecting important choices and responses based on specific criteria [1]. In the realm of cellular biology, cell decision-making refers to the process by which cells choose a new state, such as cell fates or phenotypes, in response to their microenvironmental surroundings [2,3]. This process is crucial in understanding cancer cell differentiation, in which cancer stem cells differentiate into various specialized cell types in an appropriate microenvironment. However, the differentiation process is complex and involves an intricate interplay between various factors, including cell intrinsic dynamics and extrinsic microenvironmental signals. Despite significant progress, there are still limited insights into how progenitor cells encode and process these factors to generate different cell types. Despite significant progress, our understanding of how these factors are encoded and processed by progenitor cells to generate different cell types remains limited.

In 2006, scientists discovered that almost any normal cell can be sent back to a state of pluripotency by expressing appropriate transcription factors [4]. This process of somatic reprogramming using *Yamanaka factors* offers insights into how cancer stem cells may originate. A prime example of this process is neurological cancers, such as primary glioblastomas [5] and retinoblastomas [6], which result from the de-differentiation of glial and photoreceptor cells, respectively. Recently, de-differentiation has been recognized as a hallmark of cancer under the general concept of phenotypic plasticity [7] where experimental and theoretical approaches have been developed to study it [8]. Physiological tissue differentiation is a complex process that involves a delicate balance of intrinsic and extrinsic factors that result in coherent fate decisions. Here, we focus on the question of how probable is the reversal of normal tissues back to pluripotency, which is associated with cancer.

The theory of cell differentiation, formulated by Waddington [9,10,11], provides a framework for modelling single-cell fate decisions within a dynamical systems approach [12,13,14]. In this perspective, cell states are represented as vectors of molecular expressions that can be measured through experimental techniques, such as high-throughput -omic data, flow cytometry data, biopsy markers, etc. [10,15,16]. Normal states correspond to fixed points of microstate attractors, which are associated with a probability distribution peaked around the fixed point [17,18]. On the other hand, cancer and stem-cell-like states can be viewed as limit-cycle attractors, as they do not correspond to stable fixed points. A cancer state can be compatible with various interchangeable molecular microstates [19]. However, the Waddington theory does not consider cell sensing and the corresponding interactions that occur within the tissue. Therefore, further theoretical developments are necessary to address these limitations and answer outstanding questions in this field.

In this chapter, we present the least microenvironmental uncertainty principle (LEUP) [20,21,22] (for details please see **Box 8.2**), a statistical mechanics motivated theory that models cell decision-making in a multicellular environment. The LEUP draws inspiration from various dynamic Bayesian inference theories, including the Bayesian brain hypothesis [23], the free-energy principle [24], and Bialek's work [25]. Specifically, the LEUP proposes that cells acquire knowledge about their microenvironment through various sensing mechanisms, such as receptor–ligand binding [26], pseudopodia extension [27], mechanosensing [28], proton-pump channels [29], gap junctions, among others. We postulate that cell sensing informs cell decisions, and the LEUP is founded on the idea that cells adapt to sensed microenvironmental data to make decisions, resulting in a decrease in the local microenvironmental entropy of the cell decision-maker.

In this study, we aim to develop a thermodynamic-like theory for a generic cell differentiation process using the tools of stochastic thermodynamics to address our research question. Stochastic thermodynamics is a powerful tool for systems where small-scale dynamics are significant [30,31–33,34,35]. Our main biological assumption is that normal cells can revert to an abnormal state through the process of carcinogenesis [36,6,37]. Stochastic thermodynamics enables us to identify the conditions where single-cell de-differentiation is possible, i.e. microscopic reversibility, while the tissue remains thermodynamically robust, i.e. macroscopically

irreversible. To describe the cell-level de-differentiation process, we employ Crook's fluctuation theorem [38,39], a central tenet of stochastic thermodynamics.

The chapter is organized as follows: in Section 8.2, we outline the basic features and concepts of LEUP in connection to cell decision-making. In Section 8.3, we show the main steps in deriving a thermodynamic point of view for tissue cancerization based on cell de-differentiation. In Section 8.4, we discuss the interplay of cancer hallmarks that appear to influence the probability of turning tissue into cancer. We conclude and discuss our results in Section 8.5. Finally, we would like to state that the details for most of our theoretical demonstrations can be found in [40].

8.2. A potential cell decision-making principle

In this section, we focus on the idea of cell decision-making from the point of view of a principle, which is recently proposed as the LEUP as in ref. [22].

8.2.1. Relation to Bayesian learning

Let us define the internal variables of the nth cell as x_n and external variables surrounding the nth cell as y_n which contains the information within a particular interaction radius ℓ around the nth cell (see **Box 8.1** for our definition of microenvironment). Note that cells can reshape the interaction radius while changing internal biophysical mechanisms. Moreover, we assume that cell decisions, interpreted as changes in the cellular internal states x_n within a decision time τ, are the fusion of (i) sensing their microenvironment and (ii) an existing predisposition about their internal state. In a Bayesian language, the former can be interpreted as the empirical likelihood $P(y_n|x_n)$ and the latter as the prior distribution $P(x_n)$. We propose that cells evolve the distribution of their internal states through Bayesian learning. This means that after each decision, the cell updates its internal state distribution from $P_t(x_n)$ to $P_{t+\tau}(x_n)$. The cell is effectively trying to develop more informative priors over time to minimize the energy cost associated with sampling its microenvironment.

According to Bayesian learning, the posterior of the previous time $P_t(x_n|y_n)$ becomes prior to the next time step, i.e. $P_{t+T}(x_n) = P_t(x_n \mid y_n)$.

Therefore, the Bayesian learning dynamics read

$$P_{t+T}(x_n) = \frac{P_t(y_n|x_n)P_t(x_n)}{P_t(y_n)},$$

Box 8.1 What is microenvironment?

A composition of neighbourhood cells, extra-cellular matrix, chemical signals, and ligands inside a finite area around a cell that influences cellular functions and behaviour.

The goal of this principle is to understand cellular decision-making from the perspective of informative decision-makers that increase the microenvironmental information over time (see Box 8.2). LEUP idea was inspired by observing developing tissues or, generally, multicellular systems, where cells and tissues acquire specific differentiated phenotypes and organized/low entropy patterns, respectively. The challenges that LEUP aspires to confront are summarized in Box 8.3. In this section, we discuss the main concepts and some technical aspects of LEUP.

Box 8.2 What is the least microenvironmental uncertainty principle (LEUP)?

During development cells decide over their fate by adapting to their available microenvironmental information. These decisions lead to the higher organization of tissues. LEUP is a mesoscopic theory postulating that during the above process the rate of the microenvironmental entropy (uncertainty) is negative.

$$\Rightarrow \ln\frac{P_{t+T}(x_n)}{P_t(x_n)} = \ln\frac{P_t(y_n \mid x_n)}{P_t(y_n)}. \tag{8.1}$$

At this point, we assume that there exists an attainable steady-state internal state probability density function (pdf) $P(x_n) \geq 0$. Now multiplying and dividing the first part of Equation (8.1) with the steady-state pdf, splitting the logarithm into a sum and averaging both sides equilibrium distribution $P(x_n)$, we obtain

$$D(P\,||\,P_t) - D(P\,||\,P_{t+T}) = \ln\frac{P_t(y_n|x_n)}{P_t(y_n)}, \tag{8.2}$$

where $D(Q\,||\,P) = \int dx Q(x)\ln\frac{Q(x)}{p(x)}$ denotes the Kullback–Leibler divergence of two probability distributions. In turn, we average Equation (8.2) with the joint pdf $P_t(y_n, x_n)$ and we obtain the following:

$$D(P\,||\,P_t) - D(P\,||\,P_{t+T}) = I_t(x_n, y_n)_p = I_t(x_n, y_n), \tag{8.3}$$

where $I(x_n, y_n)$ is the mutual information between internal and external cell states at time t. On the right-hand side, the averaging has no effect on the mutual information since it is a scalar. Approaching the system to the equilibrium pdf and using a well-established result from information theory [41], the left-hand side of the above is always positive. This is a requirement since mutual information should always be a positive quantity.

At this point, it is important to state that Bayesian learning is an oversimplification of reality. Also, it is known that biological processes are typically out of equilibrium. To improve the realism of our model, we introduce an arbitrary parameter $\beta \in R$ that measures the characteristic timescale of the equilibration process. Since we have assumed a mesoscopic point of view for the cell decision-making process, it is reasonable to consider an information source/sink term $\eta(t)$, which is of deterministic or stochastic nature. Then, the Bayesian learning model reads

$$\Delta D = \tilde{\beta} I_T(x_n, y_n) + \eta(t). \tag{8.4}$$

In the following, we assume the term $\eta(t) = 0$. Since $\Delta D \geq 0$, it can approach zero only when its time derivative is negative.

Then, this implies

$$\tilde{\beta}\frac{dI(y_n, x_n)}{dt} \leq 0. \tag{8.5}$$

At this point, can see the critical role of the parameter $\tilde{\beta}$ since when $\tilde{\beta} < 0$ it maximizes and for $\tilde{\beta} \geq 0$ it minimizes in time mutual

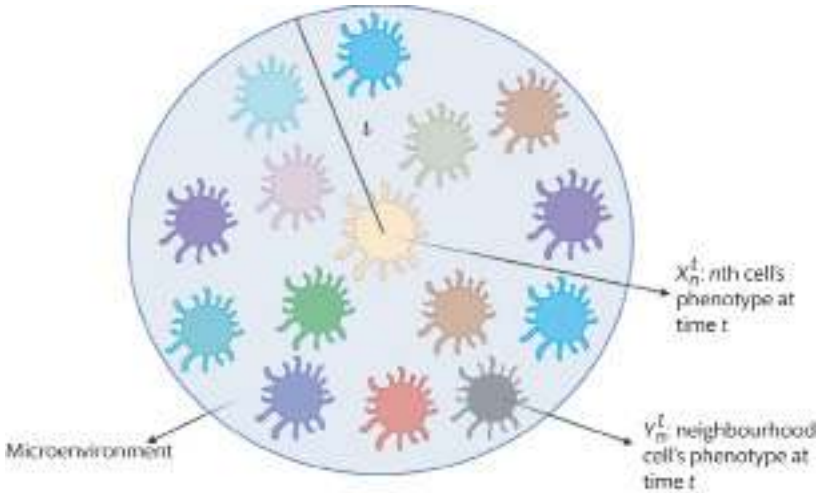


Figure 8.1. The schematic diagram of the microenvironment of a cell decision maker, where ℓ is the maximum microenvironmental sensing radius. The microenvironment is a composition of cells with different phenotypes.

information. When the system maximizes the mutual information between the microenvironment and the cell variables, and assuming that the microenvironmental distribution $P(y_n)$ stays almost invariant, then we can deduce that the microenvironmental entropy decays

$$\tilde{\beta} \le 0 : \tilde{\beta} \frac{dS(y_n|x_n)}{dt} \le 0, \tag{8.6}$$

which is exactly the LEUP premise (for further details about LEUP challenges please see **Box 8.3**). This is in concordance with Bialek's proposition that biological systems follow a maximum information principle [25].

8.2.2. Equilibrium distribution

At this point, we are interested in calculating the steady-state distribution $P(x_n)$. As we have assumed a timescale separation between internal and external variables, we can now formulate the equilibrium distribution calculation as a variational problem regarding the maximization of entropy of cell internal states $S(x_n)$. In turn, we constrain our maximization problem by $\overline{I}((y_n, x_n))$ that is the experimentally observed mutual information of a specific tissue. Considering the internal states as a continuous variable, one can write the variational problem as

$$\frac{\delta}{\delta P(x_n)} \left\{ \begin{matrix} S(x_n) + \beta \left[\int dx P(x) \int dy P(y|x) i(y:x) - \overline{I}((x_n, y_n)) \right] \\ -\lambda \left[\int P(x) dx - 1 \right] \end{matrix} \right\} = 0, \tag{8.7}$$

Box 8.3 Challenges addressed by LEUP theory

- *Lack of mechanistic knowledge*: Quantification of the cell state dynamics without the complete knowledge of the underlying intrinsic mechanisms.
- *Cell–tissue feedback*: Dissecting the relative contribution of intrinsic and extrinsic factors of cell decision-making dynamics to the spatio-temporal order in multicellular systems.
- *Multiscale information processing*: Understand the impact of cell-level information processing in the tissue-level dynamics and organization.

where $\delta / \delta P(x_n)$ is the functional derivative. Two Lagrange multipliers are defined as the sensitivity parameter β, and λ is related to a constraint that ensures the normalization of cellular internal states. Interestingly, one can also use other biological partial knowledge in terms of various constraints subject to Equation (8.7). Now, we can write the solution of Equation (8.7) as

$$P_1(x_n) = \frac{e^{\beta I(y_n, x_n)}}{\int dx_n' e^{\beta I(y_n, x_n')}} = \frac{e^{-\beta S(y_n|x_n)}}{Z_1}, \tag{8.8}$$

where $Z_1(\beta) = \int e^{-\beta S(y_n|x_n')} dx_n'$ is a normalization constant. In the literature, such distributions are called entropic priors [42].

At this point, we want to generalize (8.8) to a biologically relevant scenario of cell differentiation where we include the cell division phenomena. The differentiation process occurs in the case of asymmetric cancer progenitor cells [43]. We assume that asymmetric proliferation follows a Binomial distribution where $N(y_n|x_n)$ in the microenvironment proliferate at a rate $\mu \propto_\tau \mathrm{T}_{\mathrm{div}}^{-1}$. Assuming small proliferation rates, the proliferation distribution converges to a Poisson process with a rate $\mu N(y_n|x_n)$. So, the probability of the internal cellular state of the central nth cell can be expressed as

$$\begin{aligned} P(x_n) &\propto \frac{1}{N(y_n|x_n)} \times N(y_n|x_n) e^{-\mu N(y_n|x_n)} \times e^{-\beta S(y_n|x_n)} \\ &= \frac{e^{-\beta S(y_n|x_n) - \mu N(y_n|x_n)}}{Z(\beta, \mu)}, \end{aligned} \tag{8.9}$$

where the new normalization factor of $P(x_n)$ is defined $Z(\beta,\mu) = \int e^{-\beta S(y_n|x_n) - \mu N(y_n|x_n)} dx_n$. The first product term is the probability of a cell from the microenvironment to proliferate. Now, using Equation (8.9), we can say that the entropy of internal variable is

$$\begin{aligned} S(x_n) &= -\int P(x_n) \ln P(x_n) dx_n \\ &= \beta \langle S(y_n|x_n) \rangle_{x_n} + \mu \langle N(y_n|x_n) \rangle_{x_n} + \ln Z, \end{aligned} \tag{8.10}$$

where $\langle \dots \rangle$ is the average value with respect to the internal cellular states.

8.3. How probable is it for normal tissues to become cancerous?

Let us now focus on the de-differentiation process. A cellular *microstate* corresponds to a cell phenotype that lives in a tissue, which could be gene expression, RNA molecules, receptor distribution, etc. A microstate is typically considered as a stable *attractor* of the internal variables [17]. Here, we denote these steady states of internal variables as x_s and x_d corresponding to differentiated and pluripotent/cancer cells, respectively. In turn, we define a cellular *macrostate* as a statistical observable (e.g. average) related to a cellular ensemble involving different phenotypes. Macrostates contain information about external variables, within a finite neighbourhood around a cell, here labelled as y_s and y_d. The macrostate y_s describes the microenvironment of pluripotent cells characterized by the microstate x_s; the macrostate y_d accordingly denotes the microenvironment of differentiated cells characterized by the microstate x_d. We denote the number of pluripotent/cancer cells neighbouring a cell of microstate x_s as $N(y_s|x_s)$ and the number of differentiated cells neighbouring a cell of microstate x_d as $N(y_d|x_d)$. The total number of pluripotent and specialized cells inside the system is denoted as $N(s)$ and $N(d)$, respectively. Based on (8.9), we can write the probability of the cell to be in the specific microstate x_i, where $i = \{s, d\}$, as

$$P(x_i) = \frac{e^{-\beta_i S(y_i|x_i) - \mu_i N(y_i|x_i)}}{Z_i}. \tag{8.11}$$

8.3.1. Microreversibility and de-differentiation

Now we discuss, using the thermodynamic lens, how probable is that cells de-differentiate. Let the technical assumptions of Crook's theorem being fulfilled [38,39], then we can write

$$\beta' \Delta Q = \ln\left[\frac{w\left[x^{(k)}(x_s \to x_d)\right]}{w\left[x^{(k)}(x_d \to x_s)\right]}\right], \tag{8.12}$$

where $\beta' \equiv \frac{1}{T}$, in which T is the temperature of the heat bath and ΔQ is the total heat released into the bath over the course of a particular $x_s \to x_d$ transition. The corresponding transition probability is $w[x^{(k)}(x_s \to x_d)]$ with k denoting one of the possible paths of this transition. The de-differentiation probability, which is the reversal of x_d along the exact same path k, is denoted as $w[x^{(k)}(x_d \to x_s)$. Equation (8.12) states that when a cell differentiates (forward transition) some of the metabolic energy used is transformed into heat. Therefore, the probability of reversing this transition to the same pluripotent phenotype has a much lower probability to occur. Thus, Equation (8.12) substantiates a relation between metabolic needs and microreversibility at the cellular level. These arguments are in line with the ideas presented in [5], where differentiation is regarded as a series of reversible transitions through many microstates. In particular, stem cell phenotypes exhibit reversible oscillations until an attractor drives them towards a differentiated state. Within this picture, de-differentiation is likely to occur only as a low-probability series of microstate transitions. In our model, such changes in phenotypes are interpreted as Brownian jumps and the associated heat losses are assumed to be due to potential changes in cell metabolism. Finally, a significant source of heat losses comes from cell proliferation, which is required during the differentiation process [44]. However, there are other minor heat dissipation sources that are disregarded since they act on shorter timescales, such as physical friction, changes in the cytoskeleton, etc.

The Crooks' theorem in (8.12) is valid for a single differentiation trajectory that connects the emergence of a single pluripotent state to a particular differentiation transition. However, there are multiple ways that a differentiation and de-differentiation process can take place. As shown in [40], we can derive a general reversibility relationship that reads

$$\frac{w(x_s \to x_d)}{w(x_d \to x_s)} = \left\langle \exp\left[\beta' \Delta Q^T_{x_s \to x_d}\right]\right\rangle x_d \to x_s, \tag{8.13}$$

where $w(x_s \to x_d)$ is the differentiation transition probability and $w(x_d \to x_s)$ is the de-differentiation one (reverse). The above version of Crooks' theorem in (8.13) averages overall paths leading to differentiation and also to de-differentiation.

8.3.2. Macroreversibility and tissue cancerization

Having established the cell-level reversibility relation in Equation (8.13), we can draw our attention to its impact on macroscopic (tissue) transitions. At this point, let us assume that tissue dynamics follow a Markov process. The transition probability from a pluripotent tissue state to a normal tissue reads

$$W(s \to d) = \int_d dx_d P(x_s \,|\, s) w(x_s \to x_d), \tag{8.14}$$

and the transition probability back to pluripotency, associated with tissue cancerization, reads

$$W(d \to s) = \int_d dx_s \int_d dx_d P(x_d \,|\, d) w(x_d \to x_s), \tag{8.15}$$

where $P(x_s \,|\, s)$ is the probability that the system is in the microstate x_s, given that it is observed in the macrostate s, and $w(x_s \to x_d)$ is defined as before. These processes are illustrated in **Figure 8.2**. In turn, one can write the expression of total entropy production of such transitions as

$$\Delta f = \left\langle \beta' \Delta Q^T_{x_s \to x_d} \right\rangle_{s \to d} + \left\langle \Delta S_{\text{LEUP}} \right\rangle_{s \to d} - \ln\left[\frac{Z_s}{Z_d}\right]. \tag{8.16}$$

In other words, this expression of entropy production can be considered as a generalization of the second law of thermodynamics (see also ref. [39]).

The pluripotent tissue differentiation transition $s \to d$ corresponds to the probability $W^+ = W(s \to d)$. On the other hand, the transition $d \to s$ denotes that healthy tissue de-differentiates into a cancerous tissue and it is quantified by the backward transition probability of $W^- = W(d \to s)$. At this point, we can rewrite the fluctuation theorem in terms of W^+ and W^- as

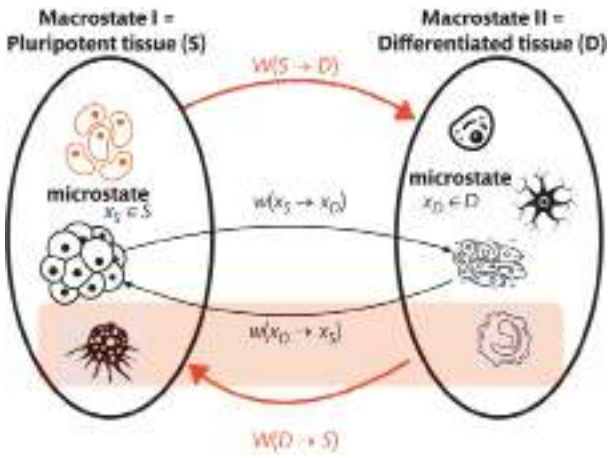


Figure 8.2. Microscopic/macroscopic transitions between two distinct cell/tissue types. Source: Adapted from [3].

$$\frac{W^{+}}{W^{-}} \leq \exp(\Delta f). \tag{8.17}$$

To understand the thermodynamic constraint that ensures the robustness of differentiated tissues, we assume that there exists a *maximum* forward transition probability from s to d in such a way that

$$\frac{W^{+}_{\max}}{W^{-}} = \exp\left[\left\langle \beta' \Delta Q^{\tau}_{x_s \to x_d}\right\rangle_{s\to d} + \left\langle \Delta S_{\mathrm{LEUP}}\right\rangle_{s\to d} + \left\langle \Delta N\right\rangle_{s\to d} - \ln\left(\frac{Z_s}{Z_d}\right)\right] \tag{8.18}$$

Equation (8.18) implies that if the entropy production Δf, i.e. the sum of all contributions from different biological processes, is positive then the tissue remains physiological. If $\Delta f < 0$, then a differentiated tissue can be destabilized and become cancerous.

The last relation is central to our work since it connects the probability of tissue cancerization to four cell processes that have been identified as *hallmarks of cancer* [7,45] (see **Figures 8.3 and 8.4**). We further elaborate on this in the following section.

8.4. The interplay between hallmarks of cancer

In the previous section, we have shown that the tissue reversibility probability depends on changes in the following four cell processes: metabolism, proliferation, epigenetic regulation, and cell decision-making. These correspond to hallmarks of cancer as conceptualized by Hanahan and Weinberg [7,45].

8.4.1. Increased net cell proliferation

Cell proliferation is a fundamental process in the development of organisms, allowing for the creation of new cells that form tissues and organs. This process is typically regulated by nutrient availability in the microenvironment and gene regulatory networks that are influenced by neighbouring cells. While cell proliferation is essential for survival in normal cells, in cancer cells, it becomes uncontrolled and can spread rapidly. This *uncontrolled net proliferation* is the most common hallmark of cancer, leading to the formation of tumours and other abnormal neoplasia. In particular, increased net proliferation is a combination of the following hallmarks: sustained proliferative signalling, evasion of growth suppressors, and resisting cell death. The proliferation of cancer cells is typically much greater than that of normal cells, leading to a *negative* $N_{s\to d} < C$ term in entropy production (Equation (8.16)).

8.4.2. Deregulation of cell metabolism

Metabolism is a critical set of chemical reactions that enable cells to survive in their microenvironment. It encompasses two types of reactions: catabolic reactions, which break down nutrients from the microenvironment, and anabolic reactions, which create essential molecules such as proteins, carbohydrates, and fats. In cancer cells, changes in metabolism are a hallmark that distinguishes them from normal cells. Specifically, the glycolysis pathway, a subset of catabolic reactions, operates differently in cancer cells. While normal cells utilize aerobic glycolysis to break down glucose into carbon dioxide and water, cancer cells rely on anaerobic glycolysis or fermentation to transform glucose into lactate. This difference in metabolic pathways affects the amount of heat dissipation during glycolysis,

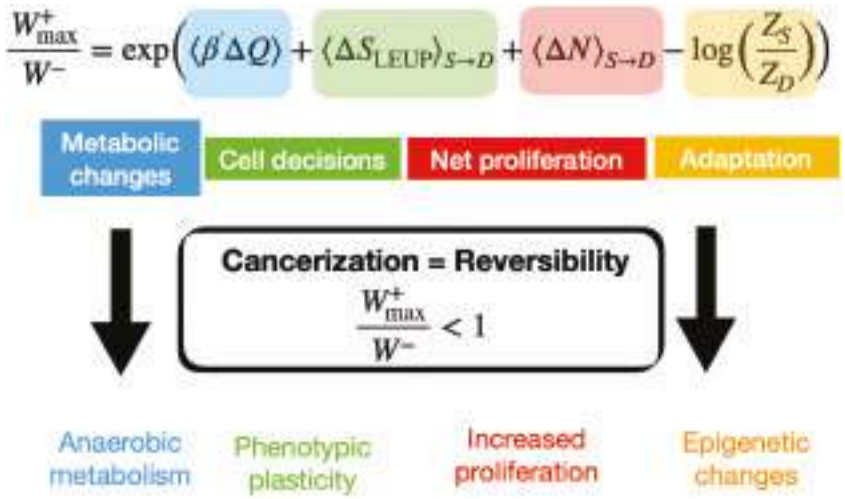


Figure 8.3. The probability of issue cancerization is a function of changes in four different biological processes: metabolism, proliferation, epigenetic adaptations, and phenotypic plasticity. These processes are identified as hallmarks of cancer.

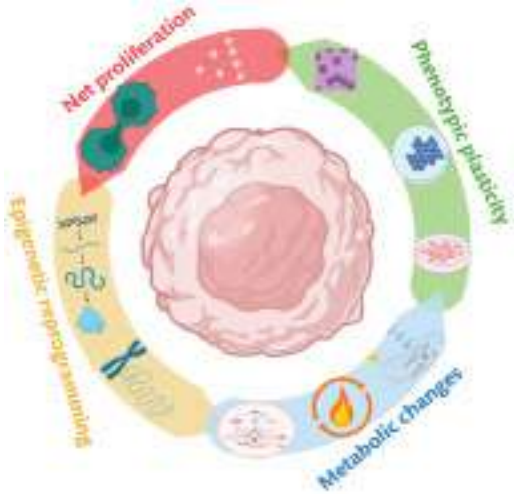


Figure 8.4. A modified representation of the hallmarks of cancer cells (inspired by [22,23]).

with anaerobic glycolysis releasing slightly more heat than aerobic glycolysis. In [46], the authors calculated that the amount of heat released in aerobic glycolysis is around 2,820 kJ/mol, and in anaerobic glycolysis, the number is around 2,929.4 kJ/mol. So, the difference in heat dissipation $\Delta Q^{T}_{x_s \to x_d\, s \to d}$ is 109.4 [40]. In Equation (8.17), the first term in the entropy production equation reflects the contribution of metabolic changes during the transition from pluripotent cells to normal cells, which is *positive* in nature. In the case of transition to cancerous tissue, the sign of the term is expected to be inverted.

8.4.3. Epigenetic reprogramming and genome instability

Epigenetics is the study of how genes interact with the complex world of inter-cellular and microenvironmental factors. The genetic information passed down from one generation to the next is stored in the DNA sequence, acting as a blueprint or memory. However, this process is also influenced by epigenetic changes that affect how genes are turned on or off, such as DNA methylation and histone modifications.

In cancer, *epigenetic reprogramming* is a hallmark of the disease. Usually, this process is mediated by DNA methylation [47], where methyl groups impact to turn ON/OFF the gene or histone modification [48] and the genetic switch is modified by the tightness properties of histone proteins around the DNA.

Epigenetic changes towards pluripotency or differentiation imply weak or strong constraints on the mRNA/protein expression giving rise to associated entropy changes [16]. This implies that the available phase space of admissible phenotypes increases for pluripotent cells and decreases in the case of differentiated ones. Therefore, the origin of the term $\ln\left[\frac{Z_s}{Z_d}\right]$ in the entropy production (Equation (8.16)) shows epigenetically mediated change in the phase space and consequently the sign of the term becomes *positive* since $Z_s > Z_d$.

8.4.4. Abnormal cell decision-making and phenotypic plasticity

Phenotypic plasticity is a hallmark cancer capability that allows cells to adapt to different microenvironments and stimuli, resulting in various disruptions of cellular differentiation. This includes the ability to de-differentiate from a mature state to a progenitor state, block terminal differentiation from progenitor cell states, and transdifferentiate into different cell lineages. As we have argued before, phenotypic plasticity can be regarded as a form of cell decision-making.

A cellular decision-maker needs to sense and gather knowledge about its surroundings, which can then influence the microenvironment's structure and organization over time through a feedback loop. Interestingly, in cancer cells, the microenvironmental architecture has little structure, while in normal cells the microenvironment exhibits specific patterns or expected spatial arrangements (such as the Notch-delta pattern described in [49]). The second term in the entropy production (Equation (8.17)) precisely quantifies the net change in microenvironmental entropy during the transition from a cancer cell to a normal cell, which represents information about the overall architectural change of the microenvironment. The value of the term $\Delta S_{LEUP\, s \to d}$ can be positive or negative depending on the microenvironmental distribution, which is related to the sensing radius [40]. During development, the distribution of a pluripotent is assumed to be a random Poisson distribution, while in differentiated tissues the microenvironmental distribution belongs to a class of hyperuniform structures [40,50]. The latter implies that the term $\langle \Delta S_{\mathrm{LEUP}} \rangle_{s \to d}$ is *positive*.

8.5. Discussion

In this chapter, we investigated the question of how probable is tissue cancerization under the assumption of mature cell de-differentiation. To tackle this problem, we have employed a combination of stochastic thermodynamics and a recently proposed cell decision-making theory, the so-called LEUP. The main result of this theoretical treatment is Equation (8.18) which compares the probabilities of a tissue becoming/staying normal versus returning to a pluripotent/cancerous state. Interestingly, our analysis shows that tumorigenesis can be a result of four processes that have been recognized as cancer hallmarks: metabolism, proliferation, epigenetic regulation, and cell plasticity.

Our proposed theory has important and interesting implications for cancer research and therapy. In clinical practice, cancer therapies have primarily focused on anti-proliferative strategies, such as chemotherapy and radiotherapy. Although metabolic reprogramming has also been considered as a therapeutic target, it has not been mainstream used in clinics [51]. Also, epigenetic therapies hold the promise of being effective, but they are still tested in clinical trials [52]. However, recent research by West et al. [53] has identified the critical role that changes in tissue organization play in tumour evolution. Our theory establishes a mechanistic connection between tissue architecture and cell decision/sensing mechanisms. Notably, the experimental work by Levin's group [54] demonstrates that disrupting ion channel sensing in tissue can induce tumorigenesis. We suggest that investigating changes in cell sensory processes is crucial for effective cancer treatment.

Our theory may contribute to stepping forward in our understanding of cancer development and treatment. It provides a quantitative framework that takes into account the complex interplay of

metabolic, proliferative, epigenetic, and plasticity changes that can lead to cancer or its therapy. With our theory, we can envision quantifying the impact of specific therapeutic modalities on increasing entropy production and assess how many different therapies must be combined to achieve a full reversal of cancerous tissue back to a healthy physiological state. Of course, achieving this goal still requires further research in the context of specific cancers.

REFERENCES

1. Simon HA. The new science of management decision. New York: Harper & Brothers; November 1960. URL: http://content.apa.org/books/13978-000.
2. Alberts B, Alexander J, Lewis J, Martin R, Keith R, Peter W, et al. Molecular biology of the cell. 6th ed. New York City, NY: W. W. Norton & Company; 2014.
3. Slack JMW. Metaplasia and transdifferentiation: from pure biology to the clinic. Nat Rev Mol Cell Biol. 2007;**8**(5):369–78. doi: 10.1038/nrm2146.
4. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;**126**(4):663–76. doi: 10.1016/j.cell.2006.07.024.
5. Prager BC, Bhargava S, Mahadev V, Hubert CG, Rich JN. Glioblastoma stem cells: driving resilience through chaos. Trends Cancer. 2020;**6**(3):223–35. doi: 10.1016/j.trecan.2020.01.009.
6. Kooi IE, Mol BM, Moll AC, van der Valk P, de Jong MC, de Graaf P, et al. Loss of photoreceptorness and gain of genomic alterations in retinoblastoma reveal tumor progression. EBioMed. 2015;**2**(7):660–70. doi: 10.1016/j.ebiom.2015.06.022.
7. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;**12**(1):31–46. ISSN: 21598290. doi: 10.1158/2159-8290.CD-21-1059.
8. Foo J, Basanta D, Rockne RC, Strelez C, Shah C, Ghaffarian K, et al. Roadmap on plasticity and epigenetics in cancer. Phys Biol. 2022;**19**(3): 031501. ISSN: 1478-3967. doi: 10.1088/1478-3975/ac4ee2.
9. Allen M. Compelled by the diagram: thinking through C. H. Waddington's epigenetic landscape. Contemporaneity. 2015;**4**:119–42. doi: 10.5195/contemp. 2015.143.
10. Casey MJ, Stumpf PS, MacArthur BD. Theory of cell fate. Wiley Interdiscip Rev: Syst Biol Med. 2020;**12**(2). doi: 10.1002/wsbm. 1471.
11. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell. 2007;**128**(4):635–38. doi: 10.1016/j.cell.2007.02.006.
12. Ferrell JE. Bistability, bifurcations, and Waddington's epigenetic landscape. Curr. Biol. 2012;**22**(11):R458–66. doi: 10.1016/j.cub.2012.03.045.
13. Paździorek PR. Mathematical model of stem cell differentiation and tissue regeneration with stochastic noise. Bull Math Biol. 2014;**76**(7):1642–69. doi: 10.1007/s11538-014-9971-5.
14. Jilkine A. Mathematical models of stem cell differentiation and dedifferentiation. Curr. Stem Cell Rep. 2019;**5**(2):66–72. doi: 10.1007/s40778-019-00156-z.
15. Garcia-Ojalvo J, Martinez Arias A. Towards a statistical mechanics of cell fate decisions. Curr Opin Genet Dev. 2012;**22**(6):619–26. doi: 10.1016/j.gde.2012.10.004.
16. Macarthur BD, Lemischka IR. Statistical mechanics of pluripotency. Cell. 2013;**154**(3): 484–89. doi: 10.1016/j.cell. 2013.07.024. URL: http://dx.doi.org/10.1016/j.cell.2013.07. 024.
17. Huang S, Eichler G, Bar-Yam Y, Ingber DE. Cell fates as high-dimensional attractor states of a complex gene regulatory network. Phys Rev Lett. 2005;**94**:1–4. doi: 10.1103/PhysRevLett.94.128701.
18. Mojtahedi M, Skupin A, Zhou J, Castaño IG, Leong-Quong RYY, Chang H, et al. Cell fate decision as high-dimensional critical state transition. PLoS Biol. 2016;**14**(12):1–28. doi: 10.1371/journal.pbio.2000640.
19. Furusawa C, Kaneko K. a dynamical-systems view of stem cell biology. Science. 2012;**338**(6104):215–17. doi: 10.1126/science.1224311.
20. Barua A, Nava-Sedeño JM, Meyer-Hermann M, Hatzikirou H. A least microenvironmental uncertainty principle (LEUP) as a generative model of collective cell migration mechanisms. Sci Rep. 2020;**10**:22371. doi: 10.1038/s41598-020-79119-y.
21. Barua A, Syga S, Mascheroni P, Kavallaris N, Meyer-Hermann M, Deutsch A, et al. Entropy-driven cell decision-making predicts 'fluid-to-solid' transition in multicellular systems. New J Phys. 2020;**22**:123034. doi: 10.1088/1367-2630/abcb2e.
22. Hatzikirou H. Statistical mechanics of cell decision-making: the cell migration force distribution. J Mech Behav Mater. 2018 Apr;**27**(1–2):1–7. ISSN: 2191-0243. doi: 10.1515/jmbm-2018-0001. URL: http://www.degruyter.com/view/j/jmbm.2018.27.issue-1-2/jmbm-2018-0001/jmbm-2018-0001.xml.
23. Knill DC, Pouget A. The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci. 2004;**27**(12):712–19. doi: 10.1016/j.tins.2004.10.007.
24. Friston K. Is the free-energy principle neurocentric? Nat Rev Neurosci. 2010;**11**(8):605. doi: 10.1038/nrn2787-c2.
25. Bialek W. Biophysics: Searching for principles. New Jersey: Princeton University Press; 2012. ISBN: 9780691138916.
26. Lauffenburger DA, Linderman JJ. Receptors: models for binding, trafficking, and signaling. New York: Oxford University Press; 1996.
27. King SJ, Asokan SB, Haynes EM, Zimmerman SP, Rotty JD, Alb JG Jr, et al. Lamellipodia are crucial for haptotactic sensing and response. J Cell Sci 2016;**129**(12):2329–42. doi: 10.1242/jcs.184507.
28. Martino F, Perestrelo AR, Vinarský V, Pagliari S, Forte G. Cellular mechanotransduction: from tension to function. Front Physiol. 2018;**9**:1–21. doi: 10.3389/fphys.2018.00824.
29. Li P, Gu M, Xu H. Lysosomal ion channels as decoders of cellular signals. Trends Biochem Sci. 2019;**44**(2):110–24. doi: 10.1016/j.tibs.2018.10.006.
30. Klages R, Just W, Jarzynski C, eds. Nonequilibrium statistical physics of small systems: fluctuation relations and beyond. Reviews of nonlinear dynamics and complexity. Weinheim, Germany: Wiley-VCH; 2013.
31. Seifert U. Stochastic thermodynamics, fluctuation theorems and molecular machines. Rep. Prog. Phys. 2012;**75**(12):126001. doi: 10.1088/0034-4885/75/12/126001.
32. Seifert U. Stochastic thermodynamics: an introduction. In: AIP Conf Proc. 2011;**1332**: 56. doi: 10.1063/1.3569487.
33. Seifert U. Stochastic thermodynamics: from principles to the cost of precision. Physica A. 2018;**504**:176–91. doi: 10.1016/j.physa.2017.10.024.
34. Van den Broeck C. Stochastic thermodynamics: a brief introduction. Proc Int Sch Phys 'Enrico Fermi'. 2013;**184**:155–93. doi: 10.3254/978-1-61499-2783-155.
35. Van den Broeck C, Esposito M. Ensemble and trajectory thermodynamics: a brief introduction. Physica A. 2015;**418**:6–16. doi: 10.1016/j.physa.2014.04.035.

36. Cobrinik D. Retinoblastoma progression. EBioMed. 2015;**2**(7):623–24. doi: 10.1016/j.ebiom.2015.07.023.
37. Xu XL, Singh HP, Wang L, Qi DL, Poulos BK, Abramson DH, et al. Rb suppresses human cone-precursor-derived retinoblastoma tumours. Nature. 2014;**514**(7522):385–88. doi: 10.1038/nature13813.
38. Crooks GE. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys Rev E. 1999;**60**(3):2721–26. doi: 10.1103/PhysRevE.60.2721.
39. England JL. Statistical physics of self-replication. J Chem Phys. 2013;**139**(12):121923. doi: 10.1063/1.4818538.
40. Barua A, Beygi A, Hatzikirou H. Close to optimal cell sensing ensures the robustness of tissue differentiation process: the avian photoreceptor mosaic case. Entropy. 2021. doi: 10.3390/e23070867. URL: https://doi.org/10.3390/e23070867.
41. Cover TM, Thomas JA. Elements of information theory. 2nd ed. (Wiley Series in Telecommunications and Signal Processing). Hoboken, New Jersey: Wiley-Interscience; June 2006. ISBN: 0471241954.
42. Caticha A, Preuss R. Entropic priors. AIP Conference Proc. 2004 Apr;**707**(1):371–80. ISSN: 0094-243X. doi: 10.1063/1.1751380. eprint: https://pubs.aip.org/aip/acp/article-pdf/707/1/371/11557510/371_1_online.pdf. URL: https://doi.org/10.1063/1.1751380.
43. Morrison SJ, Kimble J. Asymmetric and symmetric stem cell divisions in development and cancer. Nature. 2006;**441**:1068–74. doi: 10.1038/nature04956.
44. Rodenfels J, Neugebauer KM, Howard J. Heat oscillations driven by the embryonic cell cycle reveal the energetic costs of signaling. Dev. Cell. 2019;**48**(5):646–58.e6. doi: 10.1016/j.devcel.2018.12.024.
45. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;**100**(1):57–70. ISSN: 0092-8674. doi: 10.1016/S0092-8674(00)81683-9. URL: https://doi.org/10.1016/S0092-8674(00)81683-9.
46. Marıén D, Sabater B. The cancer Warburg effect may be a testable example of the minimum entropy production rate principle. Phys Biol. 2017;**14**(2):024001. doi: 10.1088/1478-3975/aa64a7.
47. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013 Jan;**38**(1):23–38. ISSN: 1740-634X. doi: 10.1038/npp.2012.112. URL: https://doi.org/10.1038/npp.2012.112.
48. Lawrence M, Daujat S, Schneider R. Lateral thinking: how histone modifications regulate gene expression. Trends Genet. 2016 Jan;**32**(1):42–56. ISSN: 0168-9525. URL: https://www.sciencedirect.com/science/article/pii/S0168952515001936.
49. Bocci F, Onuchic JN, Jolly MK. Understanding the principles of pattern formation driven by notch signaling by integrating experiments and theoretical models. Front Physiol. 2020;**11**:929. ISSN: 1664-042X. DOI: 10.3389/fphys.2020.00929. URL: https://www.frontiersin.org/article/10.3389/fphys.2020.00929.
50. Jiao Y, Berman H, Kiehl TR, Torquato S. Spatial organization and correlations of cell nuclei in brain tumors. PLoS ONE. 2011Nov;**6**(11):1–9. doi: 10.1371/journal.pone.0027323. URL: https://doi.org/10.1371/journal.pone.0027323.
51. Faubert B, Solmonson A, DeBerardinis RJ. Metabolic reprogramming and cancer progression. Science. 2020;**368**(6487):eaaw5473. doi: 10.1126/science.aaw5473. eprint: https://www.science.org/doi/pdf/10.1126/science.aaw5473. URL: https://www.science.org/doi/abs/10.1126/science.aaw5473.
52. Cheng Y, He C, Wang M, Ma X, Mo F, Yang S, et al. Targeting epigenetic regulators for cancer therapy: Mechanisms and advances in clinical trials. Signal Transduct Target Ther. 2019;**4**(1). ISSN: 20593635. doi: 10.1038/s41392-019-0095-0. URL: http://dx.doi.org/10.1038/s41392-019-0095-0.
53. West J, Schenck RO, Gatenbee C, Robertson-Tessi M, Anderson ARA. Normal tissue architecture determines the evolutionary course of cancer. Nat Commun. 2021;**12**:1–9. doi: 10.1038/s41467-021-22123-1.
54. Moore D, Walker SI, Levin M. Cancer as a disorder of patterning information: computational and biophysical perspectives on the cancer problem. Converg Sci Phys Oncol. 2017;**3**(4):043001. doi: 10.1088/2057-1739/aa8548.

9

Cellular plasticity as emerging target against dynamic complexity in cancer

Paromita Mitra, Uday Saha, Subhashis Ghosh, and Sandeep Singh

Considering the microevolutionary perspectives, a tumour is composed of heterogeneous cell populations. Tumour cell heterogeneity is being well characterized by recent advancements in single-cell multi-omic techniques. Increasing reports are revealing that heterogeneous tumours show the existence of different subtypes of cancer cell populations at both phenotypic and genotypic levels [1]. Acquisition of certain driver gene mutations by the cell of origin and the influence of tissue microenvironment over these cells are the major events that influence cancer initiation. These factors are vital to generate the initial mixed cell population with varied phenotypic and genotypic properties. Eventually, a group of cancer cells get the advantage of genetic evolution to act as clonal subpopulation. A number of recent studies have unfolded the characteristics of the cancer cells at the levels of genetic and epigenetic alterations [2]. These subpopulations dynamically interact and give rise to functional diversity, such as phenotypic heterogeneity, evasion of the immune response, and therapeutic stress for survival and metastasis [3]. Poorly differentiated high-grade tumours show drastic alterations in tissue histology and architecture, suggesting the existence of multiple cellular states or lineages of cells, whereas low-grade tumours exhibit relatively moderate or well-differentiated histology. Prognostically, the poorly differentiated high-grade tumours show worse outcomes compared to the well-differentiated low-grade tumours. Therefore, addressing various potential questions related to tumour evolution and its functional implications will have huge clinical implications. Approaches are needed to simultaneously measure the cell state and lineage in aggressive cancers. Addressing these aggressive features, the term 'cellular plasticity' has been used to refer to the shifting in cellular identity following a well-regulated process within genetic clones. Cellular plasticity in cancer is often found similar to the process activated during embryonic development to form the whole organism or during wound healing and repair of adult tissues through the dynamic levels of some specific molecular regulators of lineage identity and differentiation states [4]. Moreover, this aspect of cancer biology is majorly ignored due to the assumption that all phenotypic changes must be caused by genetic mutations. However, against this prevailing paradigm where somatic evolution is responsible for cancer progression, the evidence of cellular plasticity and non-genetic heterogeneity establishes the major paradigm shift in cancer biology.

Epithelial-to-mesenchymal transition (EMT) and its inverse process mesenchymal-to-epithelial transition are two reversible and dynamic mechanisms involved in embryonic development of vertebrates. The physiologic significance of EMT has been explored by both *in vitro* and *in vivo* studies. The well-orchestrated/organized pattern of EMT develops the delicate and complex multicellular organs through the cellular terminal differentiation. EMT is also functionally visible at the time of wound healing of adult skin. During the course of wound healing, epithelial keratinocytes from the edge of the wound attain a migratory nature to cover up the wound. This is a prime example of the programmed nature of EMT process. Talking about the pathological significance, EMT program is quite evident in epithelial-originated carcinomas and often be correlated with the histology and tumour grade. The migratory behaviour of cancer cells is explained by reprogrammed EMT. Being a cellular program, EMT is not a restricted phenomenon; rather, the evolving studies are revealing that the cancer cells may exist in a range of degrees of transitions between epithelial and mesenchymal states [5].

Often linked to EMT, stemness or stem-cell-like properties in cancer cells are explained by cancer stem cell model where cells are organized as hierarchical structures in which only stem-like cancer cells (SLCCs) or cancer stem cells (CSCs) are having the ability to initiate cancer and therefore, based on this potential, they occupy the top of the hierarchy. CSCs asymmetrically divide to generate themselves and transient amplifying cells with high proliferative capacity as intermediate populations. The differentiated cells with non-tumorigenic ability are the bulk of the cancer cell population and, therefore, are placed in the bottom position [6]. However, the recent pieces of evidence suggest that stemness in cancer cells can be achieved over time, even by differentiated cells, suggesting cellular plasticity [7]. Hence, the unidirectional rigid hierarchies of SLCC models are widely debated.

Due to their long-term tumorigenic ability, cancer cells with stem-cell-like abilities are reported to be responsible for cancer initiation, progression, metastasis, and resistance to therapy. Thus, the evolving concepts of plasticity of CSCs and interconversion between

CSCs and non-CSCs have further fuelled the complexity of cancer. Mathematical models in plasticity and non-genetic heterogeneity provide theoretical frameworks to assess cellular transition dynamics and predict tumour behaviour and response to therapy [8]. Several models predict the associated regulatory networks in EMT and have shown how the cellular plasticity maintains a stable hybrid type of cell to drive a non-genetic heterogeneity in tumour [9,10]. Bidirectional transition between CSCs and non-CSCs has caught the attention to develop suitable models incorporating the prediction on the degree of plasticity. The nonhierarchical model to quantitatively determine cellular switching that often requires self-renewal of CSCs and transition rates of the non-CSC population are thus more reliable to explain the plasticity in cancer [11]. In a recent report, we have revealed the phenotypic diversity among SLCCs in multiple oral tumour tissues and cell lines. By multiplexing putative cell surface stemness markers, CD24, CD44, and endogenous aldehyde dehydrogenase (ALDH), we have demonstrated the stochastic and cisplatin-induced bidirectional interconversions on the 'ALDH-axis' and unidirectional non-convertible on the 'CD24-axis'. A transition map connecting these states (subpopulations) and transition probabilities was created using a discrete time Markov chain model. These subpopulations showed distinct transcriptional make-up and maintenance of stemness among differentiating populations as hybrid states of SLCCs [12]. The key role of SOX2 as a transcription factor in regulating stemness has been well demonstrated in multiple cancer models [13]. SOX2 and SOX9 axes have been reported to be responsible for maintaining progenitor-like states of alveolar cells in lung adenocarcinoma [14]. Finding such molecular switches responsible for the dynamic changes between tumorigenic and non-tumorigenic cancer cells will provide rationale for adoptive therapy.

Resistance towards the therapy is the major cause of relapse, recurrence, and cancer-related mortality. Upon treatment with chemotherapy, radiotherapy, and targeted therapies, such as EGFR tyrosine kinase inhibitors in lung cancer, anti-HER2 therapies in breast cancer, BRAF inhibitors in melanoma, and even due to the immunotherapy, there is the emergence of the resistant cells [15]. It was initially proposed to be similar to the Darwinian selection of the cell population with the advantage of the genetic alterations that edge them to therapy. However, landmark studies have demonstrated the emergence of therapy resistance among genetically identical clones [16]. Therefore, apart from the genetic factors, non-genetic events, such as chromatin remodelling, tumour microenvironmental cues, and activation of the signalling pathways, also play pivotal roles in drug tolerance state (**Figure 9.1**). Recent studies are revealing the evidence of both intrinsic and adaptive or acquired mechanisms of resistance [17]. Most of the adaptive resistances are governed by non-genetic chromosomal and epigenetic changes (**Figure 9.1**).

The evolving studies are strongly correlating cancer cell plasticity with therapy resistance [7]. The cancer cells escape the conventional chemotherapy and targeted therapies by adapting EMT, dedifferentiation, or signal specific modulations [18]. Cancer cells exhibiting CSCs or mesenchymal-like properties show enhanced resistance to conventional chemotherapeutic agents compared to more differentiated or epithelial-like cancer cells [19]. Furthermore, increased stemness in cancer cells, post-treatment, could be due to adaptation of non-CSCs to the CSC state in response to the therapeutic treatment stress. For example, temozolomide (TZM) treatment in glioblastoma resulted in expression of stemness markers, SOX2, OCT4, and Nestin [20]. Similarly, radiation therapy has also shown to increase stemness. Fractionized ionizing radiation induced loss of adhesion in fractions of human metastatic pancreatic cancer cells. The non-adherent cells exhibited the properties of SLCCs with active NOTCH signalling and stem cell markers, such as CD133, Oct-4, Sox2, and Nanog expression [21]. Most often the therapy-tolerant states are characterized by slow proliferation, endowed with the property of quiescence, and are able to maintain the viability in therapeutic stress, where other cancer cells are killed [22]. The therapy-tolerant state may be a transient state that can be reverted after removal of the therapeutic pressure, but if the therapeutic stress continues, it will generate a more stable resistant state [23]. These observations are suggesting that drug resistance is a complex state of a heterogeneous cancer cell population. A number of studies have demonstrated that state transition is a random process, generating a high level of heterogeneity that is responsible for the therapy resistance [24].

With the change in environment, cancer cells constantly adapt to the environmental hypoxic and low nutrient conditions by reprogramming their metabolism to support their rapid growth and survival. Thus, the phenomenon of 'metabolic reprogramming' can be one of the crucial mechanisms of cellular plasticity and adaptation [25]. Metabolic plasticity refers to the dynamic transition of the metabolism of cancer cells in order to maintain its fitness to a hostile environment, more specifically to shift between distinct metabolic states [26]. From the very beginning, the metabolic reprogramming in cancer cells was mainly described as the 'Warburg effect', i.e. the use of glycolysis to produce lactate under a normoxic condition rather than going into the TCA cycle or oxidative phosphorylation (OXPHOS) [27]. Classical oncogenic drivers, such as KRAS and c-MYC, induce expressions of the glycolytic enzymes and also uptake of glucose-by-glucose transporter up-regulation from the external matrix. KRAS can activate their downstream activator AKT/mTOR axis, and mTOR alone can stimulate anabolic processes, and also under normoxic conditions, it can stimulate c-MYC and hypoxia-inducible factor 1α (HIF-1α) expression [28]. Under hypoxic condition, HIF-1α is known to induce the glycolysis pathway by up-regulating different glucose transporters and also the necessary enzymes for the glycolysis. A recent study has found that the cancer cells modulate the tumour stroma, specifically cancer-associated fibroblasts (CAFs) to adopt the aerobic glycolysis, that can favour the cancer cells to take up secreted by-products such as lactate and pyruvate for energy production directly via mitochondrial OXPHOS, a phenomenon named 'reverse Warburg effect' [29]. Thus, cancer cells efficiently tune their energy production by differential uptake of the metabolic substrates as well as the uptake of enzymes related to metabolic activity. It has also been observed that the tumour cells can exist in an intermediated or hybrid type of metabolic phenotype [30]. In 2019, Jia et al. constructed a mathematical model for analysing the temporal dynamics of the production and degradation of different genes as a measure of plasticity regulation and comparing them with metabolic pathway activity [31]. *In vitro* study with TNBC cell lines confirmed the presence of a hybrid metabolic state with the metabolic pathway activities. Disrupting either pathway was not sufficient to eliminate those hybrid breast cancer cells [31]

Extensive metabolic plasticity is displayed by the SLCCs to maintain the self-renewal and chemotherapy resistance [32]. It has been

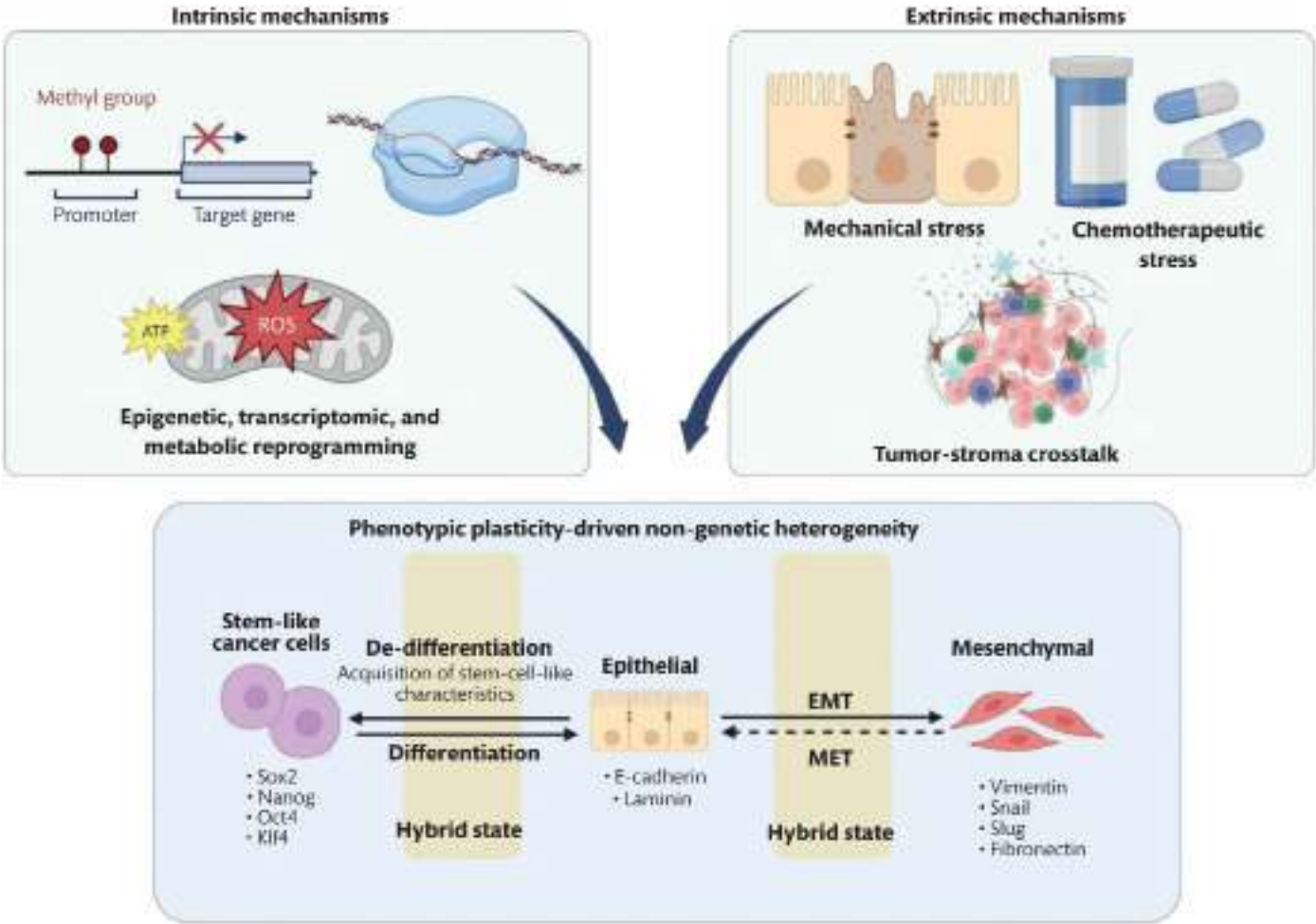


Figure 9.1. Phenotypic plasticity driven non-genetic heterogeneity in cancer cells. Intrinsic factors, such as metabolic, epigenetic, and transcriptional regulators, induce phenotypic plasticity of the cancer cells. Similarly, extrinsic regulators, such as mechanical stress, chemotherapeutic stress, and stromal factors, contribute to the emergence of diverse molecular states of cells via EMT, MET, differentiation, and de-differentiation in cancer cells.

evident that the metabolic plasticity can provide insights into the CSC plasticity as the non-CSCs can gain stem-cell-like characteristics by altering their metabolic state. For instance, WNT/FGF3 signalling in MCF7, breast cancer cell line, increases the mitochondrial biogenesis that has been associated with an increase in stem-cell-like property in differentiated populations of cells [33]. Similarly, increased expressions of Hexokinase 1, Hexokinase 2, and PDK1 have been shown to increase the SLCC population in lung and colon cancer [34]. Pancreatic CSCs maintain a distinct OXPHOS-dependent state, whereas the differentiated progenies are more glycolytic [35]. However, metformin-targeted inhibition of the mitochondrial mass by altering MYC/PGC 1-alpha balance has shown depletion in the pancreatic-CSC progeny and eventually changed into an intermediate hybrid glycolysis/OXPHOS state resisting the metformin effects [36]. Gemcitabine-resistant pancreatic ductal adenocarcinoma (PDAC) CSCs largely rely on high glucose turnover; therefore, the use of 3-bromopyruvate as a glycolysis inhibitor enhances the proliferation of the cells by blocking stem cell features and self-renewal ability conferring sensitivity towards gemcitabine treatment [37]. Non-CSCs in basal-like breast cancer can reprogram their metabolic phenotype by promoter suppression of fructose bisphosphate-1 by Snail-G9a-DNMT complex to acquire a less oxygen-dependent glycolytic state and a low-ROS producing state. This epigenetic reprogramming resulted in a transformation of luminal cells to a basal-cell-like phenotype (CD44 high/CD24low/EpCAM+) and induced EMT by E-cadherin silencing [38].

Based on the location of the cancer cells in the tumour, hypoxia or pH, the cells may exert different metabolic states. Oxygen gradient plays a vital role in metabolic heterogeneity in the cancer cells close to the blood vessel and the cancer cells remaining in a hypoxic condition. In non-small-cell lung cancers (NSCLCs), the distant site makes the energy from glucose catabolism, but the blood vessel proximal areas utilize fatty acids, ketones, and lactate to produce energy [39]. A symbiotic relation is found between these two types of cells in a tumour. Lactate, made from the glycolytic cells under hypoxic conditions, fuels ATP production for the cells in OXPHOS metabolic state under normoxic condition [40]. Overall, metabolic plasticity helps the cancer cells to gain adaptive advantages, state transition, and acquisition of drug resistance. Thus, in different types of stresses in the microenvironment, cancer cells can survive by selecting a broad spectrum of metabolic flexibility due to high degree of intrinsic metabolic

plasticity. Accumulated evidences suggest that metabolic flexibility or the metabolic plasticity has been tightly regulated by multiple factors, but the underlying mechanism is insufficiently understood. This sector in cancer biology has gained more attention from the past decade, and efforts are underway to elucidate the least metabolically flexible stage of the cancer cells for better therapeutic opportunities.

Collectively, the pieces of evidence discussed above have strongly suggested that stemness in cancer cells is a transient state driven by the plasticity in cancerous cells. Adding to the complexity, next we discuss the evidence of cellular plasticity in non-cancerous cells in the tumour microenvironment and its role in driving overall tumour heterogeneity. CAFs are the most abundant stromal cells in the tumour microenvironment. During the early stage of cancer progression, fibroblasts differentiate into more active myofibroblastic states [41]. Myofibroblasts are spindle-shaped cells characterized by stress fibre of alpha smooth muscle actin (αSMA) with contractile nature. CAFs are reported to be heterogeneous, suggesting inter-CAF heterogeneity among different patients. In pancreatic cancer, CAFs that reside in proximity of cancer cells, exhibiting higher expression of αSMA driven by TGFβ, are known as myofibroblastic CAFs (myCAFs). The CAFs reside distal to the cancer core, exhibit interleukin-6 (IL-6) expression, and are known as inflammatory CAFs (iCAF). myCAFs are characterized by their contractile property and matrix remodelling ability, whereas iCAFs showed an immunomodulatory phenotype. myCAFs possess higher expression of fibroblast-activated protein (FAP). In breast cancer, FAP+ CAFs are responsible for T-cell-mediated immunosuppression and poor prognosis [42]. Also, a subset of FAP positive CAFs, FAP+ve/PDPN+ve CAFs, suppress effector T- lymphocyte proliferation and induce immunosuppression, whereas FAP+ve/PDPN–ve CAFs fail to do so in breast cancer [43]. FAP+ CAFs play a crucial role in recruiting T lymphocytes and its subsequent differentiation into immunosuppressive CD4+/CD25+/FOXP3+ regulatory T cells. FAP+ myCAFs have higher TGFβ expression that is correlated with the infiltration of CTLA4+ regulatory T cells in breast cancer and responsible for immunotherapy failure in melanoma and NSCLC [44,45]. Costa et al. have identified four subpopulations (CAF S1–S4) of cancer-associated fibroblasts with differential tumour-promoting ability in breast cancer. CAF-S1 promotes an immunosuppressive microenvironment, whereas CAF-S4 devoid of this function indicates functional intra-CAF heterogeneity [46]. Similarly, two different CAF subpopulations are reported to be present in oral cancer, named CAF-N and CAF-D. CAF-N and CAF-D have differential tumour-promoting and invasive properties [47]. We have reported the heterogeneity of oral tumour fibroblasts and that C2-type CAFs with high αSMA expression support the enrichment of oral SLCCs as compared to the C1-type CAFs with lower αSMA expression [48]. However, in colorectal cancer, the depletion of myCAFs lead to increased invasiveness, lymph node metastasis, and poor prognosis in mouse models. Mechanistically, the depletion of αSMA positive CAFs was found to be responsible for reduced BMP4 secretion and increased abundance of Lgr5+ CSCs [49]. The role of BMP4 in restraining stemness is also found to be associated with C1-type CAFs in oral cancer though these cells express lower levels of αSMA [48]. These findings are suggesting that the heterogeneity and plasticity of the CAFs might also be organ specific. CAF heterogeneity is also induced by matrix stiffness. Increased matrix stiffening promotes nuclear localization of YAP and its co-activator TAZ. YAP and TAZ, the effectors of Hippo pathway, act as a mechano-transducers that sense and convey mechanical stimuli to cell's intrinsic transcriptional program for increased myofibroblastic differentiation [50].

Similar to the CAFs, macrophages also possess highly diverse molecular and phenotypic states in both the primary and metastatic tumours. Zhu et al. identified that in PDAC, the bone-marrow-derived monocytes and yolk-sac-derived macrophages coexist and exert different functions. Bone-marrow-derived monocytes enhance antigen presentation while the macrophages had profibrotic properties that drive tumour progression. This indicates macrophage plasticity in the tumour microenvironment. Macrophages can be polarized to either M1- or M2-type depending on the cues received. M1 macrophages release pro-inflammatory cytokines, such as interleukins IL-6, IL-12, IL-23, and TNF-α, and exert anti-tumorigenic activity by activating type-I T-cell responses [51]. On the contrary, M2 macrophages, also known as tumour-associated macrophage (TAM), activate type-II T-cell responses that have pro-tumorigenic activity. M2 macrophages have high IL-10, IL-4, and IL-13 and low levels of IL-23 [52]. But these classical markers are often expressed by a variety of macrophage subset such as M2 macrophage express M1 markers at a low level and vice versa [53]. Nitric oxide (NO) is a critical regulator of macrophage polarization. Depletion of IL-4 and IL-13 leads to inhibition of NO and polarizes macrophage to M2 phenotype, whereas lipopolysaccharide and IL-4 induce macrophage polarization into M1 phenotype by targeting Akt and mTOR pathways [54]. M2-type macrophages promote angiogenesis, and its involvement is further categorized into M2a and M2c phenotypes. Jetten et al. demonstrated that IL-4 induces M2a in a fibroblast growth factor (FGF) and IL-10 induces M2c phenotype placental growth factor in a signalling-dependent manner [55]. M2 macrophages thereby play a key role in neo-angiogenesis, metastasis, and are responsible for poor disease-specific survival, whereas high M1/M2 ratio favoured disease outcome.

Neutrophils are also known for their involvement in pre-metastatic niche formation. Tumour-associated neutrophils (TANs) with receptor CXCR1 and CXCR2 expressions are attracted to tumour microenvironment where cancer cells as well as stromal cells, such as infiltrating immune cells, fibroblasts, and endothelial cells, produce their ligand CXCL1, CXCL2, CXCL5, CXCL6, and CXCL8 [56]. Like macrophages, the nomenclature N1 and N2 was given to neutrophils having two distinct phenotypic states with anti-tumorigenic and pro-tumorigenic effects. TGFβ induce polarization of NK as evidenced by higher expression of ARG1, CCL17, and CXCL14 and low expression of CXCL10, CXCL13, and CCL6 [57], whereas interferon-β (IFNβ) alone or a combination treatment of IFNγ and GM-CSF could drive neutrophil polarization towards an antitumour state [58]. N2 TANs directly or indirectly induce tumour growth, dissemination of cancer cells by secreting ECM remodelling enzymes and other pro-angiogenic factors that promote tumour metastasis and angiogenesis at a distant site. In an induced sarcoma model by 3-MCA, TANs act as a hybrid cell type between N1 and N2 neutrophils [59].

As mentioned earlier, a number of recent studies revealed that tumor microenvironment (TME) is the crucial factor in induction or maintenance of various cellular states in cancerous cells. The

TME-released soluble factors have capability to initiate the CSC-like features, was first reported in brain tumours, where interaction with endothelial cells is pivotal in self-renewal and proliferation of cancer cells [60]. Nitric oxide production from the tumour endothelium has shown to activate NOTCH signalling to promote stemness of glioma cells [61]. CAFs secrete growth factors that induce WNT signalling to promote cancer cell stemness [62]. Tumour-associated macrophages also play a stemness-inducing role in breast and brain CSCs and drug resistance [63,64]. These reports imply the importance of TME in cancer cell plasticity and acquiring stem-cell-like properties and therapy resistance. Thus, targeting the plastic landscape of the tumour ecosystem has emerged as a key paradigm in patient survival. Directly inhibiting the plasticity by blocking the state switching and by reversal of the states in drug-induced condition can be the possible therapeutic measures.

Understanding of the molecular mechanisms underlying the cellular and molecular plasticity in cancer cells is being expanded. Therefore, the new therapeutic strategies along with pre-existing anticancer therapies may provide better clinical outcomes. EMT, being a crucial representative of cellular plasticity, is of a great therapeutic interest to regress the metastatic cancer in highly aggressive cases or to prevent the dissemination of the tumour. Reverting the mesenchymal state to adenylate cyclase/PKA-axis-dependent epithelial state by cholera toxin and forskolin has shown improvements in the chemotherapy treatment [65]. Similarly, cell intrinsic mechanisms of plasticity-driven enrichment of CSC from non-CSCs are influenced by multiple transcription factors, such as the SOX gene family, and signalling pathways, such as NOTCH, WNT, or BMP signalling. Therefore, inhibiting these dominant drivers of stemness may provide potential therapeutic benefits. For instance, ionizing radiation induced a molecular switch in the breast cancer non-stem cells and generated NOTCH-dependent breast CSCs. The phenomenon was partially prevented by NOTCH inhibition [66]. Gamma-secretase inhibitors and the antibodies against the NOTCH receptor or ligand are clinically used as a potential therapeutic strategy against NOTCH up-regulated solid cancers (ClinicalTrials.gov Identifier: NCT01096355). Some epigenetic regulators such as EZH2 and REST are involved in transdifferentiation and show resistance to routine chemotherapy in lung and prostate cancer [67]. Androgen deprivation therapy has been successful in experimental models of prostate cancer by targeting the lineage switch of luminal cells by inhibiting EZH2 using different EZH2 inhibitors [68]. Metabolic reprogramming of the breast cancer cells in response to taxens conferred tolerance against anthracycline, and thus induced a $CD44^{high}$ phenotype. Inhibitor of glucose 6 phosphate dehydrogenase along with the aforementioned drugs improved the survival in mouse models and reduced the viability of patient-derived tissue explants [69]. Arsenic trioxide in combination with all trans-retinoic acid (ATRA) was found to inhibit the isomerase Pin1 [70], as well as ATRA in combination with gamma-secretase inhibitors was found to block NOTCH pathway [71], both resulted in suppression stemness and inhibit breast cancer growth. Targeting SLCCs with the FDA-approved antibiotic Salinomycin has provided promising evidence of controlling cellular plasticity and better efficacy of chemotherapeutic agents [72].

Targeting TME may also be adapted as an alternative strategy to combat cellular plasticity in TME and its associated effect on cancer cells. Being the most abundant cell type in TME, CAFs are the most promising therapeutic target in different cancer types. AMD3100 targeting CXCL12-CXCR4 interaction (preclinical trial), NIS793 and ABBV151 (under phase I clinical trial) targeting pan TGFβ and GARP in combination with anti PD-L1 therapy, reversed immunosuppression in breast, lung, colorectal, and pancreatic cancer. Ruxolitinib, JAK-STAT inhibitor (under phase II clinical trial), in combination with Capecitabine also showed anti-tumorigenic effect in metastatic pancreatic cancer [73]. Although therapeutic approaches targeting CCL2, CSF-1 in TAM showed anti-tumour effects, TAM targeting leads to therapeutic failure and recurrence. Recent advances in cancer therapy suggested that the depletion of heterogeneous macrophage population may not be beneficial; rather reprogramming of TAM from pro-tumorigenic to anti-tumorigenic M1 macrophages is a promising targeting strategy in cancer therapy. One study has reported that the structural confirmation of the mannose receptor (CD-206) expressed on M2 macrophage can be remodelled by RP-182 that switches the macrophage phenotype into M1 types. RIP1 (receptor-interacting serine/threonine protein kinase 1) is a checkpoint kinase particularly up-regulated in TAMs in PDAC. Inhibition of RIP1 by small molecule inhibitor can reprogram M2-type macrophage into an $MHCII^{hi}TNF\alpha^{+}IFN\gamma^{+}$ phenotype in a STAT1-dependent manner. RIP1 inhibition in M2 macrophage resulted in T helper cell differentiation and cytotoxic T-cell activation leading to tumour immunity in mice and organotypic models of human PDAC [74]. Gamma isoform of phosphoinositide 3-kinase (PI3Kγ) is a highly expressed molecule in TAMs. NCT02637531 (phase I clinical trial), a small molecule inhibitor targeting PI3Kγ, induces the switch of M2 macrophage phenotype into M1 phenotype [75] (**Figure 9.2**). These results have provided promising evidence of harnessing plasticity between dynamic cellular states to revert the phenotypes to a more stable and less aggressive state both in cancer cells and tumour stroma.

Cellular plasticity has emerged as one of the key hallmarks of cancer responsible for driving non-genetic heterogeneity and aggressive features of cancer. Although the mechanisms governing cellular plasticity are still not fully understood, the molecular pathways related to development and tissue regeneration are mainly linked with this property. Evidence is accumulating where plasticity in cancer cells results in hybrid states of cells with more deleterious properties, such as stemness, metastasis, and drug tolerance, and therefore needs immediate attention for interventions. The intrinsic mechanisms related to the metabolic, epigenetic, or transcriptional states and the extrinsic factors contributed by TME both may also serve as potential targets. More importantly, plasticity is not only exhibited by neoplastic cells but also by the tumour-associated cells in TME, thus adding to the overall complexity. The interactions between diverse cells of TME and cancer cells can be highly dynamic and provide stochastic outcomes. Mathematical modelling and its validation with appropriate cellular models are needed to be developed as experimental tools to demonstrate the complex and ever-evolving plastic ecosystem in cancer. This will facilitate discoveries to translate as therapeutic possibilities against cellular plasticity in aggressive cancer.

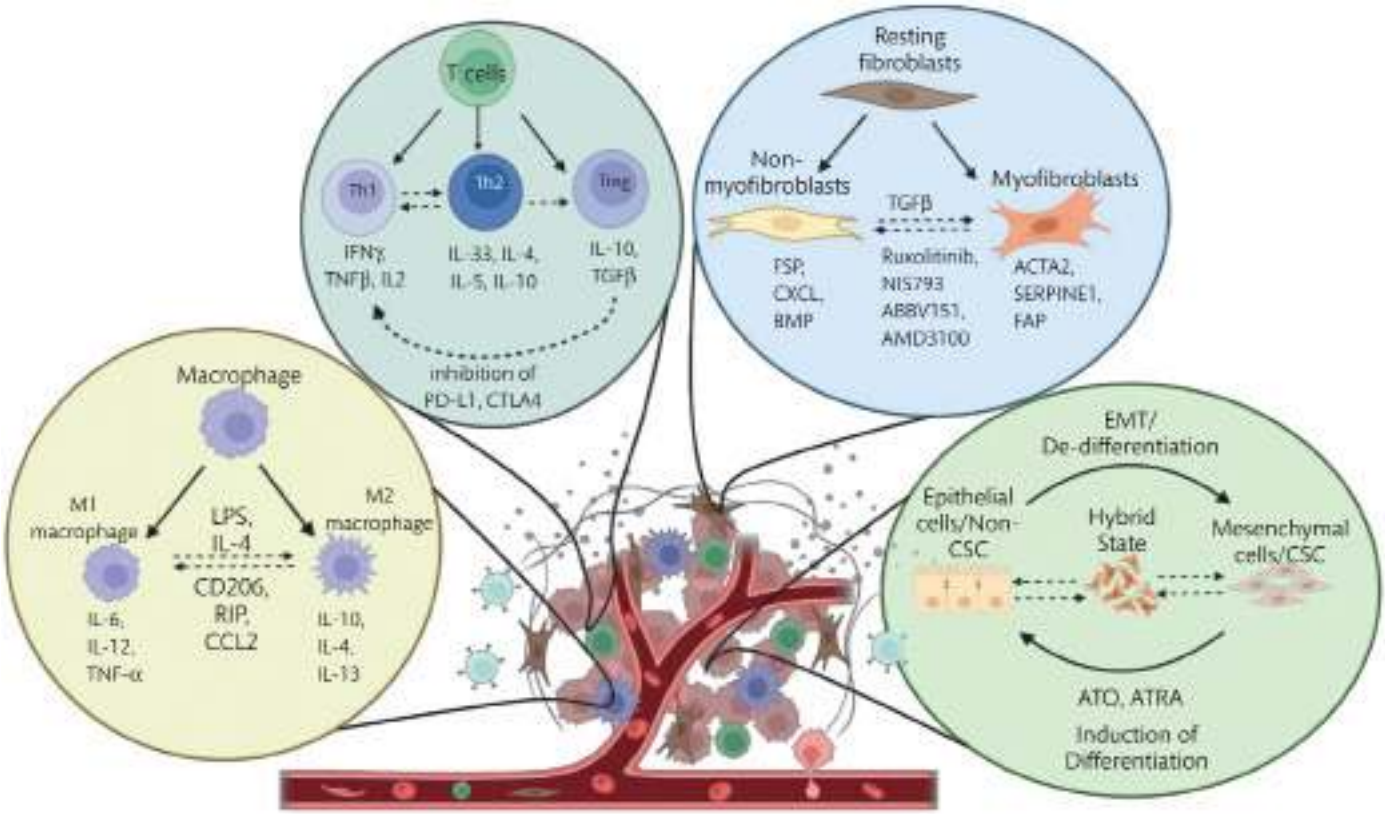


Figure 9.2. A schematic diagram of cellular plasticity in tumour microenvironment. In tumours, neoplastic cells coexist with stromal cells and immune cells. T cells, macrophages, and the resident fibroblasts are highly heterogeneous, and their plasticity is responsible for forming an immunosuppressive niche. The intrinsic and extrinsic cues drive cellular plasticity in TME resulting in dynamic complex interactions. These mechanisms can be explored for possible molecular targets against plasticity-driven cellular reprogramming.

Acknowledgements

This work was supported by the grant received from DST-SERB, India (CRG/2021/005556). P.M. acknowledges CSIR, India, U.S. acknowledges BRIC-NIBMG and UGC, India, and S.G. acknowledges DBT, India, for fellowship support.

REFERENCES

1. Contreras-Trujillo H, Eerdeng J, Akre S, Jiang D, Contreras J, Gala B, et al. Deciphering intratumoral heterogeneity using integrated clonal tracking and single-cell transcriptome analyses. Nat Commun. 2021;**12**:1 [Internet]. 2021 Nov 11 [cited 2022 Nov 30];12(1):1–14. Available from: https://www.nature.com/articles/s41467-021-26771-1
2. Salehi S, Kabeer F, Ceglia N, Andronescu M, Williams MJ, Campbell KR, et al. Clonal fitness inferred from time-series modelling of single-cell cancer genomes. Nature. 2021;**595**:7868 [Internet]. 2021 Jun 23 [cited 2022 Nov 30];595(7868):585–90. Available from: https://www.nature.com/articles/s41586-021-03648-3
3. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. Science. 2013 Mar 29;**340**(6127):1546–58.
4. Torborg SR, Li Z, Chan JE, Tammela T. Cellular and molecular mechanisms of plasticity in cancer. Trends Cancer. 2022 Sep 1;**8**(9):735–46.
5. Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M, et al. Guidelines and definitions for research on epithelial–mesenchymal transition. Nat Rev Mol Cell Biol. 2020;**21**:6 [Internet]. 2020 Apr 16 [cited 2022 Nov 30];21(6):341–52. Available from: https://www.nature.com/articles/s41580-020-0237-9
6. Prasetyanti PR, Medema JP. Intra-tumor heterogeneity from a cancer stem cell perspective. Mol Cancer. [Internet]. 2017 Feb 16 [cited 2022 Nov 30];**16**(1):1–9. Available from: https://molecular-cancer.biomedcentral.com/articles/10.1186/s12943-017-0600-4
7. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic plasticity: driver of cancer initiation, progression, and therapy resistance. Cell Stem Cell. 2019 Jan 3;**24**(1):65–78.
8. Weera singhe HN, Burrage PM, Burrage K, Nicolau DV. Mathematical models of cancer cell plasticity. J Oncol. 2019;**2019**(1):2403483.
9. Font-Clos F, Zapperi S, la Porta CAM. Topography of epithelial–mesenchymal plasticity. Proc Natl Acad Sci U S A [Internet]. 2018 Jun 5 [cited 2022 Nov 30];**115**(23):5902–7. Available from: https://pubmed.ncbi.nlm.nih.gov/29784817/
10. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. Nature. 2018; **556**:7702 [Internet]. 2018 Apr 18 [cited 2022 Nov 30];556(7702):463–8. Available from: https://www.nature.com/articles/s41586-018-0040-3
11. Weekes SL, Barker B, Bober S, Cisneros K, Cline J, Thompson A, et al. A Multicompartment mathematical model of cancer stem cell-driven tumor growth dynamics. Bull Math Biol. [Internet]. 2014 May 20 [cited 2022 Nov 30];**76**(7):1762–82.

Available from: https://link.springer.com/article/10.1007/s11538-014-9976-0
12. Vipparthi K, Hari K, Chakraborty P, Ghosh S, Patel AK, Ghosh A, et al. Emergence of hybrid states of stem-like cancer cells correlates with poor prognosis in oral cancer. iScience. 2022 May 20;**25**(5):104317.
13. Boumahdi S, Driessens G, Lapouge G, Rorive S, Nassar D, le Mercier M, et al. SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. Nature. 2014;**511**:7508 [Internet]. 2014 Jun 8 [cited 2022 Nov 30];511(7508):246–50. Available from: https://www.nature.com/articles/nature13305
14. Laughney AM, Hu J, Campbell NR, Bakhoum SF, Setty M, Lavallée VP, et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. Nat Med. 2020;**26**:2 [Internet]. 2020 Feb 10 [cited 2022 Nov 30];26(2):259–69. Available from: https://www.nature.com/articles/s41591-019-0750-6
15. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A. Primary, adaptive, and acquired resistance to cancer immunotherapy. Cell. 2017 Feb 9;**168**(4):707–23.
16. Kreso A, O'Brien CA, van Galen P, Gan OI, Notta F, Brown AMK, et al. Variable clonal repopulation dynamics influence chemotherapy response in colorectal cancer. Science (1979) [Internet]. 2013 Feb 1 [cited 2022 Nov 30];339(6119):543–8. Available from: https://www.science.org/doi/10.1126/science.1227670
17. Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell. 2010 Apr 2;**141**(1):69–80.
18. Labelle M, Begum S, Hynes RO. Direct signaling between platelets and cancer cells induces an epithelial–mesenchymal-like transition and promotes metastasis. Cancer Cell. 2011 Nov 15;**20**(5):576–90.
19. Gupta PB, Onder TT, Jiang G, Tao K, Kuperwasser C, Weinberg RA, et al. Identification of selective inhibitors of cancer stem cells by high-throughput screening. Cell. 2009 Aug 21;**138**(4):645–59.
20. Auffinger B, Tobias AL, Han Y, Lee G, Guo D, Dey M, et al. Conversion of differentiated cancer cells into cancer stem-like cells in a glioblastoma model after primary chemotherapy. Cell Death Differ. 2014;**21**:7 [Internet]. 2014 Mar 7 [cited 2022 Nov 30];21(7):1119–31. Available from: https://www.nature.com/articles/cdd201431
21. Kyjacova L, Hubackova S, Krejcikova K, Strauss R, Hanzlikova H, Dzijak R, et al. Radiotherapy-induced plasticity of prostate cancer mobilizes stem-like non-adherent, Erk signaling-dependent cells. Cell Death Differ. 2015;**22**:6 [Internet]. 2014 Jul 11 [cited 2022 Nov 30];22(6):898–911. Available from: https://www.nature.com/articles/cdd201497
22. Sharma SV., Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell [Internet]. 2010 [cited 2022 Nov 30];**141**(1):69–80. Available from: https://pubmed.ncbi.nlm.nih.gov/20371346/
23. Pisco AO, Huang S. Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse: 'What does not kill me strengthens me.' Br J Cancer. 2015 **112**:11 [Internet]. 2015 May 12 [cited 2022 Nov 30];112(11):1725–32. Available from: https://www.nature.com/articles/bjc2015146
24. Holohan C, van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. Nat Rev Cancer. 2013;**13**:10 [Internet]. 2013 Sep 24 [cited 2022 Nov 30];13(10):714–26. Available from: https://www.nature.com/articles/nrc3599
25. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell [Internet]. 2011 Mar 4 [cited 2022 Nov 30];**144**(5):646–74. Available from: https://pubmed.ncbi.nlm.nih.gov/21376230/
26. Al-Masri M, Paliotti K, Tran R, Halaoui R, Lelarge V, Chatterjee S, et al. Architectural control of metabolic plasticity in epithelial cancer cells. Commun Biol. 2021;**4**:1 [Internet]. 2021 Mar 19 [cited 2022 Nov 30];4(1):1–16. Available from: https://www.nature.com/articles/s42003-021-01899-4
27. DeBerardinis RJ, Chandel NS. We need to talk about the Warburg effect. Nat Metab [Internet]. 2020 Feb 1 [cited 2022 Nov 30];**2**(2):127–9. Available from: https://pubmed.ncbi.nlm.nih.gov/32694689/
28. Cantor JR, Sabatini DM. Cancer cell metabolism: One hallmark, many faces. Cancer Discov [Internet]. 2012 Oct 1 [cited 2022 Nov 30];**2**(10):881–98. Available from: https://aacrjournals.org/cancerdiscovery/article/2/10/881/2933/Cancer-Cell-Metabolism-One-Hallmark-Many
29. Pavlides S, Whitaker-Menezes D, Castello-Cros R, Flomenberg N, Witkiewicz AK, Frank PG, et al. The reverse Warburg effect: Aerobic glycolysis in cancer associated fibroblasts and the tumor stroma. http://dx.doi.org/104161/cc82310238 [Internet]. 2009 Dec 1 [cited 2022 Nov 30];8(23):3984–4001. Available from: https://www.tandfonline.com/doi/abs/10.4161/cc.8.23.10238
30. Bishal Paudel B, Quaranta V. Metabolic plasticity meets gene regulation. Proc Natl Acad Sci U S A [Internet]. 2019 Feb 26 [cited 2022 Nov 30];**116**(9):3370–2. Available from: https://pubmed.ncbi.nlm.nih.gov/30737291/
31. Jia D, Lu M, Jung KH, Park JH, Yu L, Onuchic JN, et al. Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways. Proc Natl Acad Sci U S A [Internet]. 2019 Feb 26 [cited 2022 Nov 30];**116**(9):3909–18. Available from: https://pubmed.ncbi.nlm.nih.gov/30733294/
32. Sancho P, Barneda D, Heeschen C. Hallmarks of cancer stem cell metabolism. Br J Cancer [Internet]. 2016 Jun 14 [cited 2022 Nov 30];**114**(12):1305–12. Available from: https://pubmed.ncbi.nlm.nih.gov/27219018/
33. Lamb R, Bonuccelli G, Ozsvári B, Peiris-Pagès M, Fiorillo M, Smith DL, et al. Mitochondrial mass, a new metabolic biomarker for stem-like cancer cells: Understanding WNT/FGF-driven anabolic signaling. Oncotarget. [Internet]. 2015 Sep 28 [cited 2022 Nov 30];**6**(31):30453–71. Available from: https://www.oncotarget.com/article/5852/text/
34. Liu PP, Liao J, Tang ZJ, Wu WJ, Yang J, Zeng ZL, et al. Metabolic regulation of cancer cell side population by glucose through activation of the Akt pathway. Cell Death Differ. [Internet]. 2014 Jan [cited 2022 Nov 30];**21**(1):124–35. Available from:https://pubmed.ncbi.nlm.nih.gov/24096870/
35. Sancho P, Burgos-Ramos E, Tavera A, Bou Kheir T, Jagust P, Schoenhals M, et al. MYC/PGC-1α balance determines the metabolic phenotype and plasticity of pancreatic cancer stem cells. Cell Metab. [Internet]. 2015 Oct 6 [cited 2022 Nov 30];**22**(4):590–605. Available from: https://pubmed.ncbi.nlm.nih.gov/26365176/
36. Sancho P, Burgos-Ramos E, Tavera A, Bou Kheir T, Jagust P, Schoenhals M, et al. MYC/PGC-1α balance determines the metabolic phenotype and plasticity of pancreatic cancer stem cells. Cell Metab. [Internet]. 2015 Oct 6 [cited 2022 Nov 30];**22**(4):590–605. Available from: https://pubmed.ncbi.nlm.nih.gov/26365176/

37. Isayev O, Rausch V, Bauer N, Liu L, Fan P, Zhang Y, et al. Inhibition of glucose turnover by 3-bromopyruvate counteracts pancreatic cancer stem cell features and sensitizes cells to gemcitabine. Oncotarget. 2014;**5**(13):5177–89.
38. Dong C, Yuan T, Wu Y, Wang Y, Fan TWM, Miriyala S, et al. Loss of FBP1 by Snail-mediated repression provides metabolic advantages in basal-like breast cancer. Cancer Cell. [Internet]. 2013 Mar 18 [cited 2022 Nov 30];**23**(3):316–31. Available from: https://pubmed.ncbi.nlm.nih.gov/23453623/
39. Hensley CT, Faubert B, Yuan Q, Lev-Cohain N, Jin E, Kim J, et al. Metabolic heterogeneity in human lung tumors. Cell. [Internet]. 2016 Feb 11 [cited 2022 Nov 30];**164**(4):681–94. Available from: https://pubmed.ncbi.nlm.nih.gov/26853473/
40. Nakajima EC, van Houten B. Metabolic symbiosis in cancer: Refocusing the Warburg lens. Mol Carcinog. [Internet]. 2013 May 1 [cited 2022 Nov 30];**52**(5):329–37. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/mc.21863
41. Sahai E, Astsaturov I, Cukierman E, DeNardo DG, Egeblad M, Evans RM, et al. A framework for advancing our understanding of cancer-associated fibroblasts. Nat Rev Cancer. 2020;**20**:3 [Internet]. 2020 Jan 24 [cited 2022 Nov 30];20(3):174–86. Available from: https://www.nature.com/articles/s41568-019-0238-1
42. Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, et al. Fibroblast heterogeneity and immunosuppressive environment in human breast cancer. Cancer Cell. 2018 Mar 12;**33**(3):463–79.e10.
43. Cremasco V, Astarita JL, Grauel AL, Keerthivasan S, MacIsaac K, Woodruff MC, et al. FAP delineates heterogeneous and functionally divergent stromal cells in immune-excluded breast tumors. Cancer Immunol Res. [Internet]. 2018 Dec 1 [cited 2022 Nov 30];**6**(12):1472–85. Available from: https://aacrjournals.org/cancerimmunolres/article/6/12/1472/468926/FAP-Delineates-Heterogeneous-and-Functionally
44. Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, et al. Fibroblast heterogeneity and immunosuppressive environment in human breast cancer. Cancer Cell. 2018 Mar 12;**33**(3):463–79.e10.
45. Kieffer Y, Hocine HR, Gentric G, Pelon F, Bernard C, Bourachot B, et al. Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. Cancer Discov. [Internet]. 2020 Sep 1 [cited 2022 Nov 30];**10**(9):1330–51. Available from: https://aacrjournals.org/cancerdiscovery/article/10/9/1330/2752/Single-Cell-Analysis-Reveals-Fibroblast-Clusters
46. Costa A, Kieffer Y, Scholer-Dahirel A, Pelon F, Bourachot B, Cardon M, et al. Fibroblast heterogeneity and immunosuppressive environment in human breast cancer. Cancer Cell. 2018 Mar 12;**33**(3):463–79.e10.
47. Costea DE, Hills A, Osman AH, Thurlow J, Kalna G, Huang X, et al. Identification of two distinct carcinoma-associated fibroblast subtypes with differential tumor-promoting abilities in oral squamous cell carcinoma. Cancer Res [Internet]. 2013 Jul 1 [cited 2022 Nov 30];**73**(13):3888–901. Available from: https://aacrjournals.org/cancerres/article/73/13/3888/584707/Identification-of-Two-Distinct-Carcinoma
48. Patel AK, Vipparthi K, Thatikonda V, Arun I, Bhattacharjee S, Sharan R, et al. A subtype of cancer-associated fibroblasts with lower expression of alpha-smooth muscle actin suppresses stemness through BMP4 in oral carcinoma. Oncogenesis. 2018;**7**:10 [Internet]. 2018 Oct 5 [cited 2022 Nov 30];7(10):1–15. Available from: https://www.nature.com/articles/s41389-018-0087-x
49. McAndrews KM, Vázquez-Arreguín K, Kwak C, Sugimoto H, Zheng X, Li B, et al. αSMA+ fibroblasts suppress Lgr5+ cancer stem cells and restrain colorectal cancer progression. Oncogene. 2021;**40**:26 [Internet]. 2021 Jun 9 [cited 2022 Nov 30];40(26):4440–52. Available from: https://www.nature.com/articles/s41388-021-01866-7
50. Calvo F, Ege N, Grande-Garcia A, Hooper S, Jenkins RP, Chaudhry SI, et al. Mechanotransduction and YAP-dependent matrix remodelling is required for the generation and maintenance of cancer-associated fibroblasts. 2013;
51. Verreck FAW, de Boer T, Langenberg DML, Hoeve MA, Kramer M, Vaisberg E, et al. Human IL-23-producing type 1 macrophages promote but IL-10-producing type 2 macrophages subvert immunity to (myco)bacteria. Proc Natl Acad Sci U S A [Internet]. 2004 Mar 30 [cited 2022 Nov 30];**101**(13):4560–5. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.0400983101
52. Allavena P, Sica A, Solinas G, Porta C, Mantovani A. The inflammatory micro-environment in tumor progression: The role of tumor-associated macrophages. Crit Rev Oncol Hematol. 2008 Apr 1;**66**(1):1–9.
53. Orecchioni M, Ghosheh Y, Pramod AB, Ley K. Macrophage polarization: different gene signatures in M1(Lps+) vs. classically and M2(LPS-) vs. Alternatively activated macrophages. Front Immunol. 2019 May;**10**:1084.
54. Covarrubias AJ, Aksoylar HI, Horng T. Control of macrophage metabolism and activation by mTOR and Akt signaling. Semin Immunol. 2015 Aug 1;**27**(4):286–96.
55. Jetten N, Verbruggen S, Gijbels MJ, Post MJ, de Winther MPJ, Donners MMPC. Anti-inflammatory M2, but not pro-inflammatory M1 macrophages promote angiogenesis in vivo. Angiogenesis 2013;**17**:1 [Internet]. 2013 Sep 8 [cited 2022 Nov 30];17(1):109–18. Available from: https://link.springer.com/article/10.1007/s10456-013-9381-6
56. Capucetti A, Albano F, Bonecchi R. Multiple roles for chemokines in neutrophil biology. Front Immunol. 2020 Jul 9;**11**:1259.
57. Masucci MT, Minopoli M, Carriero MV. Tumor associated neutrophils. Their role in tumorigenesis, metastasis, prognosis and therapy. Front Oncol. 2019 Nov 15;**9**:1146.
58. Singhal S, Bhojnagarwala PS, O'Brien S, Moon EK, Garfall AL, Rao AS, et al. Origin and role of a subset of tumor-associated neutrophils with antigen-presenting cell features in early-stage human lung cancer. Cancer Cell. 2016 Jul 11;**30**(1):120–35.
59. Ponzetta A, Carriero R, Carnevale S, Barbagallo M, Molgora M, Perucchini C, et al. Neutrophils driving unconventional t cells mediate resistance against murine sarcomas and selected human tumors. Cell. 2019 Jul 11;**178**(2):346–60.e24.
60. Calabrese C, Poppleton H, Kocak M, Hogg TL, Fuller C, Hamner B, et al. A perivascular niche for brain tumor stem cells. Cancer Cell. 2007 Jan 1;**11**(1):69–82.
61. Charles N, Ozawa T, Squatrito M, Bleau AM, Brennan CW, Hambardzumyan D, et al. Perivascular nitric oxide activates notch signaling and promotes stem-like character in PDGF-induced glioma cells. Cell Stem Cell. 2010 Feb 5;**6**(2):141–52.
62. Vermeulen L, de Sousa E Melo F, van der Heijden M, Cameron K, de Jong JH, Borovski T, et al. Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. Nat Cell Biol. 2010;**12**:5 [Internet]. 2010 Apr 25 [cited 2022 Nov 30];12(5):468–76. Available from: https://www.nature.com/articles/ncb2048
63. Lu H, Clauser KR, Tam WL, Fröse J, Ye X, Eaton EN, et al. A breast cancer stem cell niche supported by juxtacrine signalling

from monocytes and macrophages. Nat Cell Biol. 2014;**16**:11 [Internet]. 2014 Sep 28 [cited 2022 Nov 30];16(11):1105–17. Available from: https://www.nature.com/articles/ncb3041
64. Zhou W, Ke SQ, Huang Z, Flavahan W, Fang X, Paul J, et al. Periostin secreted by glioblastoma stem cells recruits M2 tumor-associated macrophages and promotes malignant growth. Nat Cell Biol [Internet]. 2015 Feb 17 [cited 2022 Nov 30];**17**(2):170. Available from: /pmc/articles/PMC4312504/
65. Pattabiraman DR, Bierie B, Kober KI, Thiru P, Krall JA, Zill C, et al. Activation of PKA leads to mesenchymal-to-epithelial transition and loss of tumor-initiating ability. Science. [Internet]. 2016 Mar 4 [cited 2022 Nov 30];**351**(6277). Available from: https://www.science.org/doi/10.1126/science.aad3680
66. Lagadec C, Vlashi E, della Donna L, Dekmezian C, Pajonk F. Radiation-induced reprogramming of breast cancer cells. Stem Cells. [Internet]. 2012 May [cited 2022 Nov 30];**30**(5):833–44. Available from: https://pubmed.ncbi.nlm.nih.gov/22489015/
67. Boumahdi S, de Sauvage FJ. The great escape: tumour cell plasticity in resistance to targeted therapy. Nat Rev Drug Discov. 2019;**19**:1 [Internet]. 2019 Oct 10 [cited 2022 Nov 30];19(1):39–56. Available from: https://www.nature.com/articles/s41573-019-0044-1
68. Ku SY, Rosario S, Wang Y, Mu P, Seshadri M, Goodrich ZW, et al. Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. Science. [Internet]. 2017 Jan 6 [cited 2022 Nov 30];**355**(6320). Available from: https://pubmed.ncbi.nlm.nih.gov/28059767/
69. Goldman A, Khiste S, Freinkman E, Dhawan A, Majumder B, Mondal J, et al. Targeting tumor phenotypic plasticity and metabolic remodeling in adaptive cross-drug tolerance. Sci Signal. [Internet]. 2019 Aug 20 [cited 2022 Nov 30];**12**(595). Available from: https://www.science.org/doi/10.1126/scisignal.aas8779
70. Kozono S, Lin YM, Seo HS, Pinch B, Lian X, Qiu C, et al. Arsenic targets Pin1 and cooperates with retinoic acid to inhibit cancer-driving pathways and tumor-initiating cells. Nat Commun. [Internet]. 2018 Aug 9 [cited 2022 Nov 30];**9**(1):1–17. Available from: https://www.nature.com/articles/s41467-018-05402-2
71. Zanetti A, Affatato R, Centritto F, Fratelli M, Kurosaki M, Barzago MM, et al. All-trans-retinoic acid modulates the plasticity and inhibits the motility of breast cancer cells: role of notch1 and transforming growth factor (TGFβ). J Biol Chem. 2015 Jul 17;**290**(29):17690–709.
72. Cui Y, Zhao M, Yang Y, Xu R, Tong L, Liang J, et al. Reversal of epithelial–mesenchymal transition and inhibition of tumor stemness of breast cancer cells through advanced combined chemotherapy. Acta Biomater. 2022 Oct 15;**152**:380–92.
73. Liu T, Han C, Wang S, Fang P, Ma Z, Xu L, et al. Cancer-associated fibroblasts: an emerging target of anti-cancer immunotherapy. J Hematol Oncol. [Internet]. 2019 Aug 28 [cited 2022 Nov 30];**12**(1):1–15. Available from: https://jhoonline.biomedcentral.com/articles/10.1186/s13045-019-0770-1
74. Wang W, Marinis JM, Beal AM, Savadkar S, Wu Y, Khan M, et al. RIP1 kinase drives macrophage-mediated adaptive immune tolerance in pancreatic cancer. Cancer Cell. 2018 Nov 12;**34**(5):757–74.e7.
75. de Henau O, Rausch M, Winkler D, Campesato LF, Liu C, Cymerman DH, et al. Overcoming resistance to checkpoint blockade therapy by targeting PI3Kγ in myeloid cells. Nature. 2016 Nov 9;**539**:7629 [Internet]. [cited 2022 Nov 30];539(7629):443–7. Available from: https://www.nature.com/articles/nature20554

10

Modelling phenotypic heterogeneity and cell-state transitions during cancer progression

Vishaka Gopalan, Sidhartha Goyal, and Sumaiyah Rehman

10.1. Introduction

It is now nearly 15 years since the first publication from The Cancer Genome Atlas consortium that reported a detailed molecular genetic characterization of hundreds of primary glioblastoma tumour samples before going on to characterize transcriptomes and mutational profiles from a wide variety of cancer types [1]. This effort represents the pinnacle of the molecular view of cancer, which is an attempt at providing a unified view of cancer progression, independent of cancer type, through the lens of natural selection based on the accumulation of driver, passenger, and tumour suppressor mutations that impinge on key pathways in cancer cells.

However, some key observations of tumour biology, and in particular responses (or lack thereof) of tumours to therapies in the clinic, can be understood by studying the evolution of cancer cell phenotypes via changes in cellular phenotypes without acquiring mutations. The key features that differentiate evolution of cell states, or cell-state switching, from mutational accumulation are reversibility among the states and the timescales over which these changes manifest. Once a cell accumulates a mutation at a genomic site, the probability of the mutation at that site reverting to the ancestral state of the parental cell is negligible. On the other hand, the phenotype of descendant cells is plastic and can switch back and forth. Phenotypic transitions can also be relatively rapid and can happen asynchronously across a population in contrast to a clonal selection sweep that is initiated by individual cells. To build models of phenotypic state evolution that account for these possibilities, we draw upon a rich tradition and body of theoretical and experimental knowledge in embryonic development, bacterial physiology, and homeostasis in organ systems such as the haematopoietic system.

Cell-state transitions were initially studied in the context of developing systems. The idea of cells rolling down an epigenetic landscape in Conrad Waddington's work [2] is a dominant metaphor through which we understand how a single cell and a single genome can, through multiple rounds of proliferation, give rise to the cell types that constitute a multicellular adult with remarkable precision and tolerance to environmental fluctuations. Early theories of cell-fate specification understood cell phenotype and state evolution as a result of bifurcations of equilibrium states of gene regulatory networks. More recently, a deeper understanding of gene expression noise has led to more refined notions of gene expression macro- and micro-heterogeneity. Here, a cell in a tissue exists in one of many possible distinct phenotypic states, but where noise may permit a cell to switch from one state to another. While such noisy transitions may be undesirable in many developmental contexts as they risk leading to body plan misspecification, they may promote survival in the face of fluctuating environments. In the context of cancer, noisy transitions can promote metastasis [3] and drug resistance [4].

The link between development and cancer has a storied history going back over 150 years (reviewed in [5]), with the recent advent of single-cell RNA-sequencing (scRNA-seq) technologies showing that the multiple cell states present in a tumour bear a transcriptional resemblance to both epithelial cell types in the tissue of origin of the tumour and the embryonic precursors of the tissue itself [6]. This idea lends itself to the metaphor of a tumour as a corruption of the Waddington landscape that represents its tissue of origin. Nonetheless, there are key differences between developing systems and tumours. There are asymmetries in cell-state switches during development which give rise to the notion of a cell-state hierarchy where state A that can give rise to state B, but not vice versa, is referred to as a stem-like state. While tumours are thought to possess similar hierarchies [7,8], much evidence suggests that such hierarchies are likely not as rigidly maintained in tumours. Outside of development and cancer, cell-state switches are a crucial component of homeostasis and recovery from injury in organs such as the skin, liver, and pancreas [9–11]. However, a key implication of links between development and cancer is that studies of cell-state switching during development hold great clinical relevance, especially when attempting to understand the emergence of tumour relapse and eventual drug resistance through non-genetic means.

This chapter is organized into two parts. The first deals with mathematical formulations that have modelled these phenomena across diverse systems. The second deals with experimental studies of tumour initiation, progression, and drug resistance that can be understood through the lens of phenotypic evolution and cell-state switches. We then summarize the state of both our theoretical and experimental understanding of cell-state switching in cancer and speculate on future directions of research.

10.2. Cell-state transitions

Developing embryos manifestly require new cell types to emerge from the old ones. Waddington laid out a framework where a myriad of underlying molecules conspire together to accomplish these cell-state transitions [2,12]. This led to some early work on applying catastrophe theory as a way to conceptually understand cell-fate changes and morphogenesis [13]. Recent work shows that quantitative descriptions of data is now possible using such frameworks [14,15]. These papers represent a phenomenological approach that uses landscape models to explain cell fates and their transitions in response to external cues but does not specify specific genes nor does it rely on them. While it is liberating to not consider the thousands of underlying genes, it also leaves the molecular description that may allow us to engineer and alter cell fates for therapy.

ScRNA-seq of developing embryos is now making it possible to unpack these cell-fate transitions at the genetic level. Since the current techniques of sequencing are destructive, trajectories of individual cells cannot be tracked and thus necessitate new computational techniques and inference methods to map out trajectories from snapshot data [16,17]. These single-cell data techniques have been recently combined with lineage tracing and clonal analysis that has been another lens that has provided important insights into cell-fate dynamics through quantitative modelling [18]. The central idea is to define a 'distance' between cells based on their gene expression states and build a map of cellular trajectories as they switch fates.

The strength of landscape models lies in their low dimensionality and their ability to predict response of cell-fate changes under changing signalling cues that are often much fewer than the many genes that vary during a cell-fate change. Their predictive power comes from the nature of the data that is used to train them, which so far has been limited to *in vitro* systems. In contrast, single-cell approaches have opened up enquiry into the *in vivo* setting and provide a detailed view of the many hundreds of genes as they change with cell fates. The manifestly high dimensionality of the single-cell data makes predictions challenging and rather provides a description of cell-fate changes. The frontier lies in bridging this gap between the two approaches by building landscape-like predictive models from single-cell data with an eye towards molecular players that control fate changes.

10.2.1. Landscape models

Bifurcation or related catastrophe theory of dynamical systems has offered a natural framework to understand morphogenesis and cell-fate transitions. The central connection is the non-linear change in the state of the system with a small change in some parameter that biological phenomena such as cell-fate transitions exhibit. Other features, such as irreversibility and multi-stability, are also natural to dynamical systems near bifurcations.

A classic example from microbial physiology is that of nutrient choice. The expression of lac operon proteins in bacteria undergoes a saddle node bifurcation when cells shift from consuming glucose to lactose [19], leading to a mixed population of lactose consuming and non-consuming bacteria. Another contrasting example from microbes is that of competence, where cells transiently switch to a state that allows cells to take up exogenous DNA. Here, the dynamics is driven by an excitable dynamical system [20]. Now, similar dynamical systems thinking is being applied to microbial communities [21–23], where non-linear interactions between different microbial species and strains can result in distinct ecological states, represented by abundance differences, under the same environmental conditions.

This work on microbes was driven by decades of work where the identities of the genes and their interactions were mapped, which allowed mechanistic dynamical models to be built and tested. Some of this has been replicated in the mammalian context [24,25], but a more geometric approach has recently been used. The geometry here refers to the constraints that low-dimensional models have to satisfy to embed multiple stable and unstable states that capture the different cellular fates seen in an experiment. The variables are abstract and not related to individual genes or molecules. The models, by construction, are lower dimensional given the low dimensionality of observations, i.e. the number of cell types and their transitions. For example, Sáez et al. [15] developed a model for cell-fate dynamics during gastrulation in mouse embryos using an *in vitro* system, where they can predict cell-fate changes in response to temporally changing input signalling cues. They find that a two-dimensional model can be inferred that can predict early cell-fate changes among the five cell types that they could capture in their *in vitro* experiments. Notably, the two dimensions of the model have no connection to the underlying genes that change during these fate changes.

10.2.2. Trajectories in genetic space

Can such a geometric approach be extended to the genetic space? ScRNA-seq technology provides a snapshot RNA expression from an ensemble of cells. Since individual cells cannot be tracked as new cell types emerge, building cell-fate trajectories from ensemble data provides a way to discover the genetic underpinnings of cell-fate transitions.

The availability of single-cell data has led to a flood of computational methods to analyse these large data sets. While the early methods focused on finding new cell types [26], the current focus has shifted to the dynamics of cellular fates as they transition from one cell type to another. The mechanistic goal of these methods has been to find genes (or pathways) that drive cell-fate changes [27,28] but has also initiated deeper questions about underlying dynamics of cell-fate transitions. For example, is there a bias or memory in fate dynamics [29]? What is the nature of the landscape on which the cell-fate dynamics unfolds [17]?

Beyond global trajectories of cell fates, with the goal of building a more fine-grained understanding of cell-fate dynamics, there has been work to leverage local information that the single-cell data provides on local 'velocity' of individual cells. This has proven to be challenging [30,31], and attempts are now being made to build vector maps that may predict the fate choice of a cell given its transcriptome

[17,32]. How these vector maps connect to low-dimensional landscape models remains to be seen.

10.2.3. Lineage tracing and clonal dynamics

Another powerful tool used to understand cell-fate dynamics has been lineage tracing and clonal dynamics. This idea is old and can be traced back to early blood transplant studies and to the discovery of stem cells [33,34]. Over the past two decades, there has been a flurry of work on stem cell fates [35] and beyond. The central idea is to track the lineage of individual cells or the number of cells belonging to a given lineage, called the clone size, and the statistics that these clone sizes exhibit across many lineages then provide important clues about the underlying cell-fate dynamics.

An interesting example of such a use of clonal analysis is work by Shakiba et al. on challenging the central hypothesis of equipotency in the ability of cells to reprogram [36]. Cellular reprogramming was discovered by the seminal work of Takahashi and Yamanaka that showed that differentiated mouse and human fibroblasts could be reprogrammed into induced pluripotent stem cells, holding significant promise for regenerative medicine [37,38]. However, understanding the factors and steps that underlie reprogramming, and being able to reliably model these outcomes, have broad implications in regenerative medicine.

Analyses of single cells undergoing reprogramming had suggested that every somatic cell has the ability to reprogram and also provided some evidence for equipotency [39]. It had been further hypothesized that genetically identical cells may be equipotent in their ability to reprogram [40]. However, while population analyses provide the perception of a synchronous reprogramming schedule of a cell population, single-cell analyses provide insight into the stepwise and asynchronous nature of clonal reprogramming. Shakiba et al. compared cell population outcomes and single-cell reprogramming events and reported the existence of a subpopulation of 'elite' clones within a population of somatic cells [36].

A population of mouse embryonic fibroblasts were tagged with unique genetic barcodes to track their lineage after induction of reprogramming. Among the 10% of surviving clones after 30 days of reprogramming initiation, there was large variability across the clone sizes. Variability across clones is expected due to inherent stochasticity of birth and death. However, the observed clone size distribution was bimodal and allowed the detection of dominant clones that challenged the hypothesis of clonal equipotency, which would have yielded a unimodal clone size distribution if it were true.

The nature of the competition between clones—indirect (limited nutrition or space) or direct (between reprogramming clones) —remains to be fully tested. The varying reprogramming potential of cells in the population versus single cells suggested that competitive interactions between the clones play a significant role in defining population dynamics. A more quantitative analysis of clone size distribution provided a more concrete quantification of the clonal selection. To further determine when in time this selection operates, temporal analysis was performed of clone size dynamics. Although there were some prior differences in reprogramming potential, the elite clones largely emerged after the reprogramming protocol was initiated, i.e. the clones that were large by days 8–14 demonstrated dominance at day 30. In contrast, clones that were large at the start of reprogramming (due to differences in barcoding) did not show any bias. This suggests an early stochastic transition, likely due to broader epigenetic heterogeneity among cells, that governs the reprogramming fate of individual, or a group of interacting, fibroblasts.

Understanding the competitive ability of cell undergoing reprogramming will help (1) understand how these cells, when injected, may interact with endogenous cells in a patient, (2) determine the factors that would facilitate their survival to (3) devise methods by which to control cellular competition and (4) enable prediction of cell reprogramming outcomes in patients.

10.3. Reconciling cell states with the somatic mutation theory of cancer

The somatic mutation theory (SMT) view of cancer is typically the first paradigm that one encounters in a study of cancer biology. SMT posits that cancer arises from the presence of mutations in the epithelial cells (or, more generally, a cell of origin) of a tissue. The accumulation of these mutations is accompanied by the transformation of the tissue into a tumour via intermediate stages of pre-cancerous transformation. Each step of evolution involves Darwinian selection of cells bearing mutations best adapted to their environment [41]. The theory explains cancer epidemiology data such as the presence of germline mutations that increase cancer risk, the repeated occurrence of the same mutation across patient tumours of a given cancer type, and the increase in cancer risk with age [42,43].

However, there are key problems with the SMT paradigm [44,45] when studying both tumour initiation and tumour progression. Some of these issues largely arise from a malignant cell-intrinsic view of cancer where the microenvironment is treated as a passive bystander rather than an active participant in tumour progression and control. But of more relevance to this chapter is that the SMT presents a simplified one-to-one map between genotype and phenotype, where the presence of mutations implies that a cell must be malignant and the absence of mutations implies a normal cell. This is vastly at odds with the increasingly accepted view of the one-to-many nature of the genotype–phenotype map discussed above in the context of development and as evidenced by the inclusion of plasticity in the recent update to Hanahan and Weinberg's 'Hallmarks of Cancer' series [46].

As a result, the SMT view alone fails to explain several key observations such as the fact that mutations can persist in histologically normal tissues without any evidence of cancer [47], [48], premalignant lesions can regress towards a low-grade lesion [49], there can be reversion of established tumours altogether despite the continued presence of oncogenic mutations [50], tumours with fewer driver mutations seemingly occur later in life than those that require more driver mutations [51], and the lack of metastatic driver mutations despite extensive searches [52–54]. Going beyond SMT, a cell-state-centric view of cancer not only helps understand these observations but points to future directions of mathematical modelling and cancer research to incorporate co-evolution between microenvironmental and epithelial cells that are key to tumour initiation and progression.

The pervasiveness of non-genetic transcriptional variation in healthy tissues provides a starting point to explain the above observations. Transitions between cell states are frequent during development and during tissue homeostasis in adult humans, with differentiation and de-differentiation being common across various

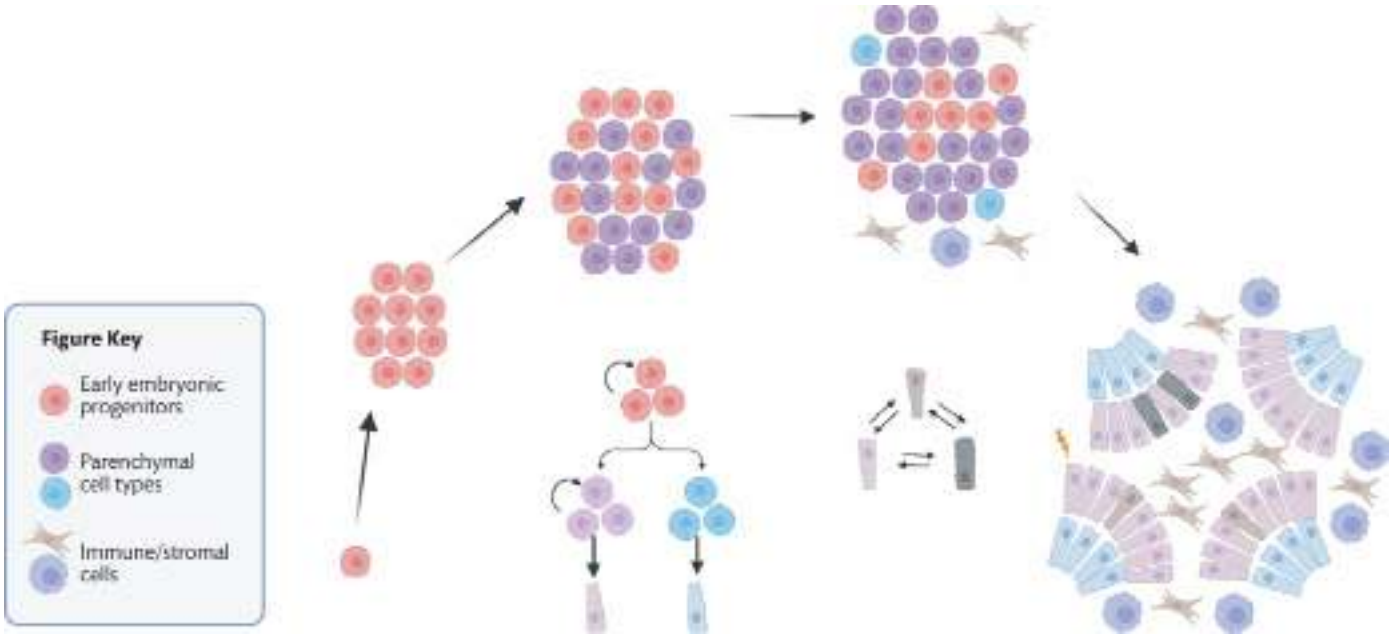


Figure 10.1. Schematic of cell state transitions in organ development and homeostasis. Organ development proceeds through the proliferation of early embryonic progenitors (red) and differentiation into the parenchymal cell types (purple and blue) that will eventually make up the adult organ. The function of these parenchymal cells is supported by other stromal and immune populations. At any point during development and organ homeostasis, every cell of a given type can exist in one of many states. Interconversion between these states may occur both due to microenvironmental changes and noise-induced transitions. The hierarchy of cell types shown represents a sense of terminal differentiation during organ homeostasis. During wound repair and carcinogenesis, however, this hierarchy becomes less rigid and more plastic. This is supported by observations that cell states across many solid tumours resemble the embryonic counterparts of the tissue of origin of the tumour. *Source:* Figure created with BioRender.com

organ tissues (Figure 10.1). Cell types can be considered as a collection of *attractor* cell states [6], where a cell transitions between states in response to gene expression and/or microenvironmental changes [55]. Thus, mutations could either influence the rates of transitions between states or alter the stability of an attractor state [56,57]. In this schema, malignancy represents a distinct set of attractor states where malignant transformation involves (a) a transition from normal to malignant attractor state via pre-malignant (benign) attractor states and (b) the maintenance of cells in a malignant attractor state, likely because its stability is increased by the accumulation of oncogenic mutations.

A key question about tumour initiation that arises in this framework is whether certain attractor states representing a normal cell type have a higher probability of transitioning towards a (pre-) malignant cell state than others. Experimentally, this is akin to asking, whether there exists a subpopulation of cells that is easier to transform into a malignancy than other cells for any given tissue. Normal human mammary epithelial cells that express the CD44 protein are better able to initiate tumours than non-CD44 expressing cells, with cells switching between CD44 expressing and non-expressing states [56]. Differentiated melanocytes give rise to melanomas more easily than melanocytic stem cells [58], while multiple cell types in the brain, breast, and pancreas can give rise to tumours with clinical characteristics varying with the cell of origin [59–61]. Carcinogens can induce cancer without introducing mutations [62,63], suggesting that mutations may act on pre-existing phenotypic variation among epithelial cells. In a remarkable study, lung epithelial cells exposed to cigarette smoke were epigenetically primed for tumour initiation before they accumulated mutations [64]. ScRNA-seq studies allow a deeper analysis of cell-state variation present prior to malignancy. A re-analysis of scRNA-seq datasets from the pancreas found an epithelial subpopulation that was not clonally derived nor contained any oncogenic mutations but expressed several gene characteristics of pancreatic adenocarcinoma (PDAC) [65]. PDAC is driven by a G12D mutation in the KRAS protein, and $Kras^{G12D}$-bearing mice contained a higher fraction of acinar 'edge' cells than $Kras^{wt}$ mice [65]. Thus, $Kras^{G12D}$ induction promotes transitions towards the edge state possibly via both cell-autonomous and non-cell-autonomous factors. A time-course scRNA-seq study of lung cancer initiation through $Kras^{G12D}$ induction (and *Trp53* deletion) in lung alveolar cells revealed the presence of a cell state termed the high plasticity cell state that expressed genes of multiple epithelial lineages and appeared to be a precursor to malignant transformation [66]. These studies suggest not only that some cell states are potentially easier to undergo malignant transformation than others, but also that the pre-malignant cell state may be reached from multiple initial epithelial cell states. A major component missing in these studies, however, is the spatial component to variation in gene expression. For instance, zonation patterns influence epithelial gene expression in the liver [67], pancreatic cancers are more likely in the head of the pancreas than the tail [68], and it is unclear how much of the cell-state variation relevant to tumour initiation is driven by spatial factors across tissues.

10.3.1. The pre-malignant cell state

The transformation from a normal to a malignant cell goes through multiple stages of pre-malignancy [69] that can take years or even decades. Patients with pre-malignant lesions progress at widely

different rates to malignancy, suggesting that microenvironmental factors play a key role in restraining pre-malignant lesions. For instance, nearly a third of women in the ages of 40–50 years had pre-malignant lesions in their breast [70]. In another study, the fraction of tissue donors with pre-malignant lesions in their thyroid was so large that lesions are now considered a normal feature of the ageing thyroid gland [71]. Pre-malignant lesions progress to malignancy at widely different rates and can often regress. Close to a third of pre-malignant lesions in the lung spontaneously regress to a lower grade [72], with regressing lesions tending to have a higher cytotoxic T-cell infiltration [73] and a higher mutation burden than progressing lesions. Regression of pre-malignancy need not require the immune system, however, as pre-malignant cells in the oesophagus and skin are outcompeted by wild-type epithelium cells without any involvement from the adaptive immune system [74,75]. Nonetheless, excluding the adaptive immune system from the microenvironment seems an early step in tumour initiation in both lung [76] and colorectal cancer [71], potentially allowing cells to sample many more malignant cell states without any selective pressure from the immune system. Viewing pre-malignant cells as a distinct cell state allows for the possibility of reversion towards a normal cell state (or a less aggressive pre-malignant cell state) under the influence of the microenvironment [77].

The extent to which pre-malignant cell states that are successful in achieving a transition to malignancy are retained in a progressing tumour is unclear [78]. There is considerable phenotypic heterogeneity in a primary or metastatic tumour that can be attributed to non-genetic/genetic variation and cell-autonomous/non-cell-autonomous factors. ScRNA-seq studies have shown that, across tumour types, malignant cell transcriptional states resemble transcriptional states seen in embryonic versions of the host tissue [79,80]. However, basic scRNA-seq protocols do not allow cell lineages to be tracked without additional barcoding of tumour cells which makes it impossible to determine whether malignant cells in one state can switch to another one. Lineage-barcoding approaches have established such lineage switches to be the case in glioblastoma, pancreatic adenocarcinoma, and breast cancer mouse models [81]. Interconnection between the views of the tumour as a collection of cell states versus a collection of genetic sub-clones may be resolved by considering that mutations may alter the stability of existing cell states or alter the rates of transitions between them [82].

10.3.2. The metastatic state

Metastasis accounts for most cancer deaths and is a common occurrence across many cancer types [83]. There are two dominant paradigms for metastatic progression in cancer—serial progression and parallel progression [84]. The serial progression paradigm posits that metastasis is an event that occurs late during primary tumour growth, while in parallel progression metastatic seeding occurs early in tumour evolution. Parallel progression posits that metastasis occurs early, and the metastatic tumour progresses in parallel to the primary tumour. In terms of cell states, the differences between paradigms seemingly occur from the question of (1) whether tumour cell states required for survival in metastatic environments are achieved early or late in tumour progression and (2) the extent to which further mutations drive the acquisition of metastatic competent cell states within the primary tumour. The search for mutations that drive a metastatic phenotype, i.e. mutations that are enriched in metastatic tumours but not primary tumours, has not been successful [85,86] although chromosomal-scale losses and gains have been noted between primary and metastatic tumours [87]. This likely reflects the differences between microenvironments across organ sites, where different transcriptomic states and adaptations are required by metastatic cells to survive in different organs. Through whole-exome sequencing of paired primary and metastatic tumours from the same patient, studies find support for both parallel and serial progression models [85], with germline variation potentially influencing the rate of metastatic spread. A similar study carried out on a larger cohort of lung, breast, and colorectal cancers [88] suggested that parallel progression was far more common in lung and colorectal cancer samples than in breast.

These observations suggest that transcriptionally, metastatic competence may be generated relatively earlier during primary tumour evolution than has been traditionally understood. In a mouse tail-vein injection experiment, cell lines with high transcriptional variation generated metastases far more efficiently than those with low variation [83]. While this differs from natural patient courses of metastasis, a study of human pancreatic cancer patients and mouse models [89,90] found circulating pancreatic cells even before a primary tumour could be clinically detected. It is unknown as to how efficient these cells are in actually generating a metastatic tumour, but their early appearance suggests that a metastasis-competent cell state may be acquired far earlier than once assumed.

A few studies currently address similarities of cell states found in paired primary and metastatic tumours, and whether they can be mapped to different embryonic cell states. A scRNA-seq study of lung adenocarcinoma metastases [91] reveals markers of embryonic lung developmental states in metastatic biopsies though this study does not comment on whether such states were already present in the primary tumour or represent an adaptation to its metastatic microenvironment. A study where pancreatic cancer cell lineages could be traced *in vivo* in mice during metastatic colonization from xenografts [92] maps epithelial-to-mesenchymal transition (EMT) signatures onto metastatic cells but does not compare metastatic cells to developmental pancreatic cell states. A similar lineage-tracing study of a lung adenocarcinoma xenograft [93] generated multiple metastases but the metastatic cell transcriptomes were not mapped onto lung development.

The intriguing observation that normal epithelial cells that lack mutations grow into organoids when implanted in lymph nodes [94] suggests that certain combinations of organ sites and epithelial cell states allow for metastatic spread more easily than others. Tumour cells in the pre-malignant state can last years, and the accompanying transcriptional and/or microenvironmental changes may promote the early acquisition of a heritable metastatic cell state early during primary tumour progression in some patients. Thus, a cell-state view of cancer progression that considers transcriptional variation across cells would suggest that parallel versus serial progression paradigm could be a false dichotomy. Fitting a cell-state model of tumour progression to clinical data calls for a deep collaboration between experimenters, data scientists, and systems biologists. Such a model-fitting exercise would potentially require a set of markers that can reliably classify cell states in scRNA-seq data, mathematical models that can compute rates of transitions between cell states based on the mutational composition of a cell, and clinical follow-ups of patients to try and estimate rates of metastatic spread and disease burden growth.

10.4. Tumour heterogeneity and identifying the tumour persister state

Various mechanisms of therapeutic resistance and/or relapse have been identified in a range of cancers that were found to be driven by a subpopulation of cells within the tumour. Extensive studies revealed genetic mechanisms as the driving force of therapeutic resistance in some tumours [95]. Alternatively, tumour cells can stochastically acquire therapeutic resistance through genetic mutations in response to treatment, determined by the Darwinian clonal evolution of tumour cells (**Figure 10.2**). The subsequent development of targeted therapies resulted in these once untreatable cancers to be treatable. A seminal example of this is the treatment of Her2 overexpressing breast cancer patients with Trastuzumab, a monoclonal antibody targeting Her2 [96]. However, not all therapeutic resistance can be solely explained by genetic mutations, highlighting how non-genetic mechanisms of drug resistance present a key obstacle for successful therapeutic outcomes [97].

The accumulating evidence of cancer cell heterogeneity within a tumour mass has been strongly tied to non-genetic mechanisms of drug resistance [98,99]. With the explosion of single-cell sequencing and the identification of extensive intratumoral heterogeneity (ITH), phenotypic and genetic ITH is being increasingly appreciated as a determining factor of therapeutic failure and tumour relapse [100]. Many studies have identified a subpopulation of cancer stem cells as contributing to ITH and being responsible for therapeutic evasion and tumour progression [101]. Patient-derived tumour cells were barcoded, injected into mice to generate xenografts, and subjected to standard-of-care therapeutic response studies *in vivo* [102–104]. However, tumours that regrow upon treatment cessation exhibited no change in barcode complexity indicating no genetic subpopulation enrichment that facilitated the tumour resistance [102–104]. Others have uncovered additional reversible non-genetic mechanisms of therapeutic response (such as epigenetic) as the driving force of relapse in heterogeneous tumours [99,101].

Seminal studies by Sharma et al. identified a subpopulation of cells in cancer cell lines that transiently acquired and relinquished the drug-resistant phenotype by individual cells within the cancer cell population in response to targeted therapy [98]. These findings exemplify the dynamic regulation of phenotypic heterogeneity and cell states, and that cancer cell populations can exhibit reversible tolerance to drugs to persist in stressful environments. This subpopulation was termed as drug-tolerant persisters (DTPs) and launched the field of tumour cells in the persister state. DTPs have since been recognized as significant players in the field of non-genetic tumour heterogeneity and have been identified across a wide range of tumours in response to therapy [98,99,101,103,105,106]. In cancer, DTPs represent a quiescent or slow-cycling reversible state where the tumour initially responds to treatment but then regrows upon treatment cessation, often referred to as drug holiday, while

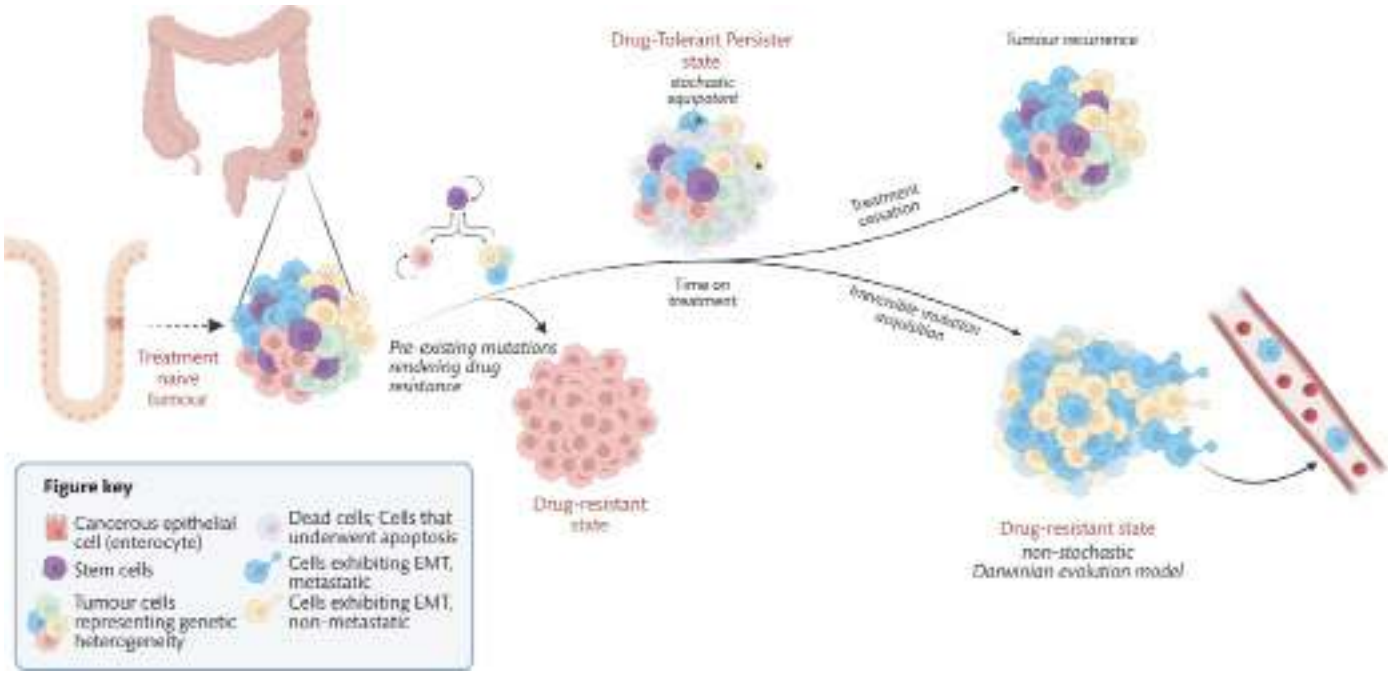


Figure 10.2. Schematic representation of tumour cell population dynamics in response to treatment. Tumours are made up of a heterogeneous population of cells, including cancer stem cells (purple) and their progeny (stem cells—pink; differentiated cells—yellow, green, and blue). Transitions between these cell states often resemble state transitions observed during embryonic development of the tissue of origin of the tumour. When cancer cells are exposed to therapy, some cells in the tumour may have pre-existing mutations that render drug resistance and tumour growth while on treatment. Alternatively, the number of viable cells may start to decline as the population of sensitive cells (grey) undergo apoptosis (cell death), and some cells survive drug treatment due to the transition to a drug-tolerant persister state (stochastic equipotency). Upon stopping treatment, these drug-tolerant persisters drive tumour regrowth and tumour relapse. While in the persister state, some cells may acquire irreversible genetic mutations (blue and yellow cells with *) rendering these cells resistant to therapy. These cells can drive drug resistance state and tumour growth on treatment. Cells may additionally acquire the epithelial-to-mesenchymal (EMT) state, facilitating metastasis (blue cells). *Source:* Figure created with BioRender.com

maintaining sensitivity to retreatment, indicative of a reversible 'drug-tolerant' cancer cell state [107]. DTPs represent a novel therapeutic vulnerability for strategic intervention prior to the development of irreversible genetic mutation-driven drug resistance.

The 'persister' concept originates from the field of microbiology. While the treatment of bacterial infections with antibiotics is known to reduce the bacterial burden, it is well established that in some instances the treatment fails to eliminate refractory bacteria [108,109]. In response to therapeutic stress, bacterial persisters are described as phenotypic variants that transiently tolerate tremendous amounts of antibiotics yet remain genetically sensitive to the treatment [108,110–112]. Studying and understanding bacterial biology and pathophysiology has provided an unparalleled insight into the workings of more complex systems such as mammalian cells. Akin to bacterial persisters, the ability of DTPs to tolerate chemotherapy cannot be explained by genetic mutations [107].

10.4.1. DTPs recapitulate evolutionary conserved embryonic survival strategy of diapause

Having gained significant traction since their first identification by Sharma et al., persisters have been extensively modelled in response to various targeted therapies and chemotherapies to determine their drivers and characteristics.

Since DTPs represent a non-genetic reversible state that cancer cells enter to survive the harsh environment created by chemotherapy, like bacterial species surviving antibiotic treatment, many groups have explored the hypothesis of whether tumours are hijacking an evolutionarily conserved survival strategy. Torpor, estivation, and diapause are among the few/handful conserved survival strategies that organisms employ to survive hostile environments that are conserved across the animal kingdom. Torpor and estivation enable animals to survive environments of extreme cold and heat, respectively. Diapause is a physiological reproductive survival strategy utilized across the animal kingdom, including some mammals, to survive stressful environments. Diapause is defined as a reversible state of suspended embryonic development triggered by unfavourable environmental conditions including nutrient deprivation [113,114].

Interestingly, a few groups have uncovered that the DTPs in various tumour models recapitulate the evolutionary conserved mechanism of diapause to survive the environmental stress associate with chemotherapy exposure [102,103,105,115,116]. Comparative transcriptomic analysis of vehicle-treated tumours, tumours in the DTP state, and chemotherapy-treated regrowth tumours to published gene expression data of *in vivo* diapause embryos found that DTPs shared significant similarities with this model [103,105,117–119]. These similarities were further corroborated by analyses of expression signatures of specific signalling pathways associated with embryonic diapause, such as down-regulation of Myc and mTORc1, and crucial cellular growth processes (cell cycle, translation, and RNA processing) [103,105,115].

10.4.2. Modelling the persister state

Many studies have employed the use of barcoding technology and single-cell sequencing to assess tumour population dynamics to illustrate how tumours respond to therapeutic insult. To probe the clonal dynamics of the tumour DTP state *in vivo* and/or *in vitro*, a barcoding strategy was adapted to track individual cells within tumours [102,105,120]. Comparative analysis of barcode abundance at various stages of the tumour's therapeutic response identified no significant alteration in barcode complexity [102–105], indicating there was no selection of a pre-existing cell subpopulation that gave rise to DTPs. *Mathematical modelling* of the distribution of clone sizes (measured using barcode abundance) further identified that tumour cells possess an equipotent capacity to enter the persister state [103].

Interestingly, neither bacterial persisters, in response to antibiotic treatment, nor embryonic cells, entering diapause in response to nutrient deprivation, exhibit enrichment of a cellular subpopulation in response to stress [119]. Rather, every cell reversibly enters the state, and once the environmental insult resolves, the bacterial persisters or embryonic cells resume normal development. These results further strengthened the argument that DTPs were employing an evolutionary conserved survival cell state to survive chemotherapy.

10.4.3. Modelling mutation rates in persisters

The ability of cancer cells to survive treatment while remaining sensitive to drug may potentially enable DTP cells to evolve under selective therapeutic pressure to acquire resistance-conferring genetic mutations [121]. In response to therapeutic exposure, cells undergo DNA damage and impairment of DNA repair pathways [122–124], leading to error-prone DNA replication. Given the difficulty in lineage tracking of heterogenous tumour cells while measuring mutational processes by DNA sequencing [123], a complementary strategy, the 'fluctuation test', was developed by Luria and Delbrück to characterize the onset of resistance in bacterial populations [125]. This assay exploited multiple replicates of clonal populations to bypass lineage-tracking issues and provided a strategy to estimate mutation rates.

The fluctuation test was modified to deduce the acquisition of drug resistance in tumours [126–129], to estimate the mutation rate of cancer cells in basal conditions, and to understand the evolution of pre-existing resistant cells before treatment [130]. However, it was not designed to quantify mutation rates in DTPs during treatment. Russo et al. developed a novel approach to measure the mutation rate in the presence of anti-cancer drugs, known as the mammalian-cell-Lauria–Delbrück (MC-LD) model, referring to a stochastic birth–death branching process describing the growth of resistant cells before and during drug treatment [116].

Russo et al. present a general quantitative method to characterize the transition of cancer cells to the persister state and to measure their population dynamics during the course of therapeutic treatment. Using a two-step fluctuation test, they quantify phenotypic mutation rates of colorectal cancer cells with the ability to distinguish between pre-existing resistant clones and persister-derived resistant clones, facilitating the quantification of spontaneous and drug-induced mutation rates.

Analyses from the MC-LD model identified that (1) resistant clones that become visible within the first 4 weeks of treatment represent pre-existing resistant cells consisting of cells with spontaneous mutations that emerged during the initial tumour growth phase, (2) resistant clones that emerge more than 4 weeks of treatment are not likely to originate from pre-existing resistant cells, and (3) resistant colonies that emerged more than 10 weeks after drug treatment in persister-containing wells may have developed mutations

that confer therapeutic resistance [116]. Interestingly, cancer cells in the persister state exhibited an increased mutation rate (7–50 fold) in the presence of therapies that were lethal for most of the parental population without any effect of clonal bias. In the case of the CRC tumours assessed, molecular profiling of persister-derived resistant clones revealed acquisition of single nucleotide variants (SNVs) or copy number aberrations (CNAs) in the RAS-MEK pathway genes that are known drivers of resistance to anti-EGFR/anti-BRAF therapies [116].

10.5. Conclusions and future directions

In this chapter, we have explored the use of mathematical models to investigate the heterogeneity of cell states during cancer progression. While the genetic view of cancer has been instrumental in understanding tumour initiation and the development of targeted therapies, models of gene regulation and embryonic development have provided key insights into the ability of a given genotype to yield diverse cellular phenotypes. Our review of the landscape of mathematical modelling approaches used to study cell states in developing systems has revealed that many of these approaches have also been successfully applied to the study of cancer and has yielded many breakthroughs.

The synthesis of computational and experimental work has produced many fruitful directions for future research. One important area of focus should be the development of new mathematical models that can capture the complex interplay between genetic, non-genetic, and environmental factors that contribute to the heterogeneity of cancer cell states. Furthermore, as our understanding of the complexity of cancer increases, models that can capture the multi-scale nature of cancer will need to be developed, from the molecular level up to the tissue and organ levels. For one, the role of environmental factors in cancer progression and heterogeneity is being increasingly recognized, and future models would need to account for the effects of the tumour microenvironment and other external factors.

Another promising area of research should be the application of single-cell sequencing technologies and other high-throughput experimental techniques to further investigate the molecular basis of cell-state transitions during cancer progression. With the increasing availability of high-throughput multi-omic data, we may see more efforts to integrate these data into mathematical models to gain a more comprehensive understanding of the molecular mechanisms underlying cancer cell-state transitions.

Ultimately, the goal in this field is to develop new therapeutic strategies that can target the heterogeneous cell states that contribute to cancer progression, therapeutic response, cancer relapse, and resistance to therapy. Continued collaboration between computational and experimental researchers will be key to making progress in our understanding of the heterogeneity of cancer cell states and the development of new therapeutic strategies that can target these states more effectively. With the increasing availability of patient-specific data, more efforts should be made to develop personalized models of cancer that can predict disease progression and response to therapy for individual patients. This is of particular relevance as the FDA will no longer require animal tests of drug candidates prior to clinical trials, thus driving the need for better computational and *in vitro* models that better mimic clinical cancer phenotypes. The field of mathematical modelling of phenotypic heterogeneity and cell-state transitions in cancer is poised for exponential growth and innovation, with many exciting developments in the years to come.

REFERENCES

1. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008 Oct;**455**(7216):1061–68. doi: 10.1038/nature07385.
2. Waddington CH. Canalization of development and the inheritance of acquired characters. Nature. **150**(3811):563–565, Nov. 1942, doi: 10.1038/150563a0.
3. Nguyen A, Yoshida M, Goodarzi H, Tavazoie SF. Highly variable cancer subpopulations that exhibit enhanced transcriptome variability and metastatic fitness. Nat. Commun. 2016 May;**7**:11246. doi: 10.1038/ncomms11246.
4. Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature. 2017 Jun;**546**(7658):431–35. doi: 10.1038/nature22794.
5. Krebs ET. Cancer and the embryonal hypothesis. Calif Med. 1947 Apr;**66**(4):270–71.
6. Huang S. On the intrinsic inevitability of cancer: from foetal to fatal attraction. Semin. Cancer Biol. 2011 Jun;**21**(3):183–99. doi: 10.1016/j.semcancer.2011.05.003.
7. Lan X, Jörg DJ, Cavalli FMG, Richards LM, Nguyen LV, Vanner RJ, et al. Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. Nature. 2017 Sep;**549**(7671):227–32. doi: 10.1038/nature23666.
8. Batlle E, Clevers H. Cancer stem cells revisited. Nat. Med. 2017 Oct;**23**(10):1124–34. doi: 10.1038/nm.4409.
9. Rognoni E, Watt FM. Skin cell heterogeneity in development, wound healing, and cancer. Trends Cell Biol. 2018 Sep;**28**(9):709–22. doi: 10.1016/j.tcb.2018.05.002.
10. Michalopoulos GK, Bhushan B. Liver regeneration: biological and pathological mechanisms and implications. Nat. Rev. Gastroenterol. Hepatol. 2021 Jan;**18**(1):40–55. doi: 10.1038/s41575-020-0342-4.
11. Murtaugh LC, Keefe MD. Regeneration and repair of the exocrine pancreas. Annu. Rev. Physiol. 2015;**77**:229–49. doi: 10.1146/annurev-physiol-021014-071727.
12. Waddington CH. Organisers and Genes by C. H. Waddington. The University Press, 1940.
13. Thom R. Topological models in biology. Topology. 1969 Jul;**8**(3):313–35. doi: 10.1016/0040-9383(69)90018-4.
14. Rand DA, Raju A, Sáez M, Corson F, Siggia ED. Geometry of gene regulatory dynamics. Proc Natl Acad Sci U S A. 2021 Sep;**118**(38):e2109729118. doi: 10.1073/pnas.2109729118.
15. Sáez M, Blassberg R, Camacho-Aguilar E, Siggia ED, Rand DA, Briscoe J. Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions. Cell Syst. 2022 Jan;**13**(1):12–28.e3. doi: 10.1016/j.cels.2021.08.013.
16. Lange M, Bergen V, Klein M, Setty M, Reuter R, Bakhti M, et al. CellRank for directed single-cell fate mapping. Nat Methods. 2022 Feb; **19**(2):159–70. doi: 10.1038/s41592-021-01346-6.
17. Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, et al. Mapping transcriptomic vector fields of single cells. Cell. 2022 Feb; **185**(4):690–711.e45. doi: 10.1016/j.cell.2021.12.045.

18. Wagner DE, Klein AM. Lineage tracing meets single-cell omics: opportunities and challenges. Nat Rev Genet. 2020 Jul;**21**(7):410–27. doi: 10.1038/s41576-020-0223-2.
19. Novick A, Weiner M. Enzyme induction as an all-or-none phenomenon. Proc Natl Acad Sci U S A. 1957 Jul;**43**(7):553–66. doi: 10.1073/pnas.43.7.553.
20. Süel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB. An excitable gene regulatory circuit induces transient cellular differentiation. Nature. 2006 Mar;**440**(7083):545–50. doi: 10.1038/nature04588.
21. Sanchez A, Bajic D, Diaz-Colunga J, Skwara A, Vila JCC, Kuehn S. The community-function landscape of microbial consortia. Cell Syst. 2023 Feb;**14**(2):122–34. doi: 10.1016/j.cels.2022.12.011.
22. Hu J, Amor DR, Barbier M, Bunin G, Gore J. Emergent phases of ecological diversity and dynamics mapped in microcosms. Science. 2022 Oct;**378**(6615):85–9. doi: 10.1126/science.abm7841.
23. Leibovich N, Rothschild J, Goyal S, Zilman A. Phenomenology and dynamics of competitive ecosystems beyond the niche-neutral regimes. Proc Natl Acad Sci U S A. 2022 Oct;**119**(43):e2204394119. doi: 10.1073/pnas.2204394119.
24. Dullweber T, Erzberger A. Mechanochemical feedback loops in contact-dependent fate patterning. Curr Opin Syst Biol. 2023 Mar;**32–33**:100445. doi: 10.1016/j.coisb.2023.100445.
25. Cohen R, Taiber S, Loza O, Kasirer S, Woland S, Sprinzak D. Precise alternating cellular pattern in the inner ear by coordinated hopping intercalations and delaminations. Sci. Adv. 2023 Feb;**9**(8):eadd2157. doi: 10.1126/sciadv.add2157.
26. Dasgupta S, Bader GD, Goyal S. Single-cell RNA sequencing: a new window into cell scale dynamics. Biophys. J. 2018 Aug;**115**(3):429–35. doi: 10.1016/j.bpj.2018.07.003.
27. Furchtgott LA, Melton S, Menon V, Ramanathan S. Discovering sparse transcription factor codes for cell states and state transitions during development. eLife. 2017 Mar; **6**:e20488. doi: 10.7554/eLife.20488.
28. Lang AH, Li H, Collins JJ, Mehta P. Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes. PLoS Comput. Biol. 2014 Aug;**10**(8):e1003734. doi: 10.1371/journal.pcbi.1003734.
29. Wang SW, Herriges MJ, Hurley K, Kotton DN, and Klein AM. CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information. Nat Biotechnol. 2022 Jul;**40**(7):1066–74. doi: 10.1038/s41587-022-01209-1.
30. Gorin G, Fang M, Chari T, Pachter L. RNA velocity unraveled. PLoS Comput Biol. 2022 Sep;**18**(9):e1010492. doi: 10.1371/journal.pcbi.1010492.
31. Soneson C, Srivastava A, Patro R, Stadler MB. Preprocessing choices affect RNA velocity results for droplet scRNA-seq data. PLoS Comput. Biol. 2021 Jan; **17**(1):e1008585. doi: 10.1371/journal.pcbi.1008585.
32. Farrell S, Mani M, Goyal S. Inferring single-cell transcriptomic dynamics with structured latent gene expression dynamics. bioRxiv. 2022.08.22.504858, Dec. 01, 2022. doi: 10.1101/2022.08.22.504858.
33. Becker AJ, McCulloch EA, Till JE. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. Nature. 1963 Feb;**197**:452–54. doi: 10.1038/197452a0.
34. Till JE, McCulloch EA. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. Radiat. Res. 1961 Feb;**14**:213–22.
35. Chatzeli L, Simons BD. Tracing the dynamics of stem cell fate. Cold Spring Harb. Perspect. Biol. 2020 Jun;**12**(6):a036202. doi: 10.1101/cshperspect.a036202.
36. Shakiba N, Fahmy A, Jayakumaran G, McGibbon S, David L, Trcka D, et al. Cell competition during reprogramming gives rise to dominant clones. Science. 2019 Apr;**364**(6438):eaan0925. doi: 10.1126/science.aan0925.
37. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006 Aug;**126**(4):663–76. doi: 10.1016/j.cell.2006.07.024.
38. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007 Nov;**131**(5):861–72. doi: 10.1016/j.cell.2007.11.019.
39. Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, et al. Direct cell reprogramming is a stochastic process amenable to acceleration. Nature. 2009 Dec;**462**(7273):595–601. doi: 10.1038/nature08592.
40. Yamanaka S. Elite and stochastic models for induced pluripotent stem cell generation. Nature. 2009 Jul;**460**(7251):49–52. doi: 10.1038/nature08180.
41. Nowell PC. Tumor progression: a brief historical perspective. Semin Cancer Biol. 2002 Aug;**12**(4):261–66. doi: 10.1016/s1044-579x(02)00012-3.
42. Martínez-Jiménez F, Muiños F, Sentís I, Deu-Pons J, Reyes-Salazar I, Arnedo-Pac C, et al. A compendium of mutational cancer driver genes. Nat Rev Cancer. 2020 Oct;**20**(10):555–72. doi: 10.1038/s41568-020-0290-x.
43. Huang K-L, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, et al. Pathogenic germline variants in 10,389 adult cancers. Cell. 2018 Apr; **173**(2):355–70.e14. doi: 10.1016/j.cell.2018.03.039.
44. Soto AM, Sonnenschein C. The tissue organization field theory of cancer: a testable replacement for the somatic mutation theory. BioEssays News Rev Mol Cell Dev Biol. 2011 May;**33**(5):332–40. doi: 10.1002/bies.201100025.
45. Rozhok AI, DeGregori J. Toward an evolutionary model of cancer: Considering the mechanisms that govern the fate of somatic mutations. Proc Natl Acad Sci U S A. 2015 Jul;**112**(29):8914–21. doi: 10.1073/pnas.1501713112.
46. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022 Jan;**12**(1):31–46. doi: 10.1158/2159-8290.CD-21-1059.
47. Martincorena I, Roshan A, Gerstung M, Ellis P, Loo PV, McLaren S, et al. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. Science. 2015 May;**348**(6237):880–86. doi: 10.1126/science.aaa6806.
48. Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall MWJ, et al. Somatic mutant clones colonize the human esophagus with age. Science. 2018 Nov;**362**(6417):911–17. doi: 10.1126/science.aau3879.
49. Krysan K, Tran LM, Dubinett SM. Immunosurveillance and regression in the context of squamous pulmonary premalignancy. Cancer Discov. 2020 Oct;**10**(10):1442–44. doi: 10.1158/2159-8290.CD-20-1087.
50. Brodeur GM, Bagatell R. Mechanisms of neuroblastoma regression. Nat Rev Clin Oncol. 2014 Dec;**11**(12):704–13. doi: 10.1038/nrclinonc.2014.168.
51. Rozhok A, DeGregori J. A generalized theory of age-dependent carcinogenesis. eLife. 2019 Apr;**8**:e39950. doi: 10.7554/eLife.39950.

52. Hunter KW, Amin R, Deasy S, Ha N-H, Wakefield L. Genetic insights into the morass of metastatic heterogeneity. Nat Rev Cancer. 2018 Apr;**18**(4):211–23. doi: 10.1038/nrc.2017.126.
53. Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, Kohutek ZA, et al. Minimal functional driver gene heterogeneity among untreated metastases. Science. 2018 Sep;**361**(6406):1033–37. doi: 10.1126/science.aat7171.
54. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LAJ, Kinzler KW. Cancer genome landscapes. Science. 2013 Mar;**339**(6127):1546–58. doi: 10.1126/science.1235122.
55. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. Dev Camb Engl. 2009 Dec;**136**(23):3853–62. doi: 10.1242/dev.035139.
56. Chaffer CL, Brueckmann I, Scheel C, Kaestli AJ, Wiggins PA, Rodrigues LO, et al. Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proc Natl Acad Sci U S A. 2011 May;**108**(19):7950–55. doi: 10.1073/pnas.1102454108.
57. Nam AS, Chaligne, Landau DA. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. Nat Rev Genet. 2021 Jan;**22**(1):3–18. doi: 10.1038/s41576-020-0265-5.
58. Köhler C, Nittner D, Rambow F, Radaelli E, Stanchi F, Vandamme N, et al. Mouse cutaneous melanoma induced by mutant BRaf arises from expansion and dedifferentiation of mature pigmented melanocytes. Cell Stem Cell. 2017 Nov;**21**(5):679–93.e6. doi: 10.1016/j.stem.2017.08.003.
59. Alcantara Llaguno SR, Wang Z, Sun D, Chen J, Xu J, Kim E, et al. Adult lineage-restricted CNS progenitors specify distinct glioblastoma subtypes. Cancer Cell. 2015 Oct;**28**(4):429–40. doi: 10.1016/j.ccell.2015.09.007.
60. Koren S, Reavie L, Couto JP, De Silva D, Stadler MB, Roloff T, et al. PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. Nature. 2015 Sep;**525**(7567):114–18. doi: 10.1038/nature14669.
61. Flowers BM, Xu H, Mulligan AS, Hanson KJ, Seoane JA, Vogel H, et al. Cell of origin influences pancreatic cancer subtype. Cancer Discov. 2021 Mar;**11**(3):660–77. doi: 10.1158/2159-8290.CD-20-0633.
62. Chen QY, DesMarais T, Costa M. Metals and mechanisms of carcinogenesis. Annu Rev Pharmacol Toxicol. 2019 Jan;**59**:537–54. doi: 10.1146/annurev-pharmtox-010818-021031.
63. Riva L, Pandiri AR, Li YR, Droop A, Hewinson J, Quail MA, et al. The mutational signature profile of known and suspected human carcinogens in mice. Nat Genet. 2020 Nov;**52**(11):1189–97. doi: 10.1038/s41588-020-0692-4.
64. Vaz M, Hwang SY, Kagiampakis I, Phallen J, Patil A, O'Hagan HM, et al. Chronic cigarette smoke-induced epigenomic changes precede sensitization of bronchial epithelial cells to single-step transformation by KRAS mutations. Cancer Cell. 2017 Sep;**32**(3):360–76.e6. doi: 10.1016/j.ccell.2017.08.006.
65. Gopalan V, Singh A, Mehrabadi FR, Wang L, Ruppin E, Arda HE, et al. A Transcriptionally distinct subpopulation of healthy acinar cells exhibit features of pancreatic progenitors and PDAC. Cancer Res. 2021 Aug;**81**(15):3958–70. doi: 10.1158/0008-5472.CAN-21-0427.
66. Marjanovic ND, Hofree M, Chan JE, Canner D, Wu K, Trakala M, et al. Emergence of a high-plasticity cell state during lung cancer evolution. Cancer Cell. 2020 Aug;**38**(2):229–46.e13. doi: 10.1016/j.ccell.2020.06.012.
67. Halpern KB, Shenhav R, Matcovitch-Natan O, Tóth B, Lemze D, Golan M, et al. Single-cell spatial reconstruction reveals global division of labour in the mammalian liver. Nature. 2017 Feb;**542**(7641):352–56. doi: 10.1038/nature21065.
68. Jiang Z, White RA, Wang TC. Adult pancreatic acinar progenitor-like populations in regeneration and cancer. Trends Mol Med. 2020 Aug;**26**(8):758–67. doi: 10.1016/j.molmed.2020.04.003.
69. Clark WH. Tumour progression and the nature of cancer. Br J Cancer. 1991 Oct; **64**(4):631–44. doi: 10.1038/bjc.1991.375.
70. Gonzalez F, Sanchez-Salorio M, Pacheco P. Simultaneous bilateral 'malignant glaucoma' attack in a patient with no antecedent eye surgery or miotics. Eur J Ophthalmol. 1992;**2**(2):91–3. doi: 10.1177/112067219200200208.
71. Bissell MJ, Hines WC. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. Nat Med. 2011 Mar;**17**(3):320–29. doi: 10.1038/nm.2328.
72. Teixeira VH, Pipinikas CP, Pennycuick A, Lee-Six H, Chandrasekharan D, Beane J, et al. Deciphering the genomic, epigenomic, and transcriptomic landscapes of pre-invasive lung cancer lesions. Nat Med. 2019 Mar;**25**(3):517–25. doi: 10.1038/s41591-018-0323-0.
73. Immune surveillance in clinical regression of preinvasive squamous cell lung cancer—PubMed. https://pubmed.ncbi.nlm.nih.gov/32690541/ (accessed Mar. 30, 2023).
74. Colom B, Herms A, Hall MWJ, Dentro SC, King C, Sood RK, et al. Mutant clones in normal epithelium outcompete and eliminate emerging tumours. Nature. 2021 Oct;**598**(7881):510–14. doi: 10.1038/s41586-021-03965-7.
75. Brown S, Pineda CM, Xin T, Boucher J, Suozzi KC, Park S, et al. Correction of aberrant growth preserves tissue homeostasis. Nature. 2017 Aug;**548**(7667):334–37. doi: 10.1038/nature23304.
76. Mascaux C, Angelova M, Vasaturo A, Beane J, Hijazi K, Anthoine G, et al. Immune evasion before tumour invasion in early lung squamous carcinogenesis. Nature. 2019 Jul;**571**(7766):570–75. doi: 10.1038/s41586-019-1330-0.
77. Clark WH. Tumour progression and the nature of cancer. Br J Cancer. 1991 Oct; **64**(4):631–44. doi: 10.1038/bjc.1991.375.
78. Capp J-P, DeGregori J, Nedelcu AM, Dujon AM, Boutry J, Pujol P, et al. Group phenotypic composition in cancer. eLife. Mar. 2021;**10**:e63518. doi: 10.7554/eLife.63518.
79. Baron M, Tagore M, Hunter MV, Kim IS, Moncada R, Yan Y, et al. The stress-like cancer cell state is a consistent component of tumorigenesis. Cell Syst. 2020 Nov;**11**(5):536–46.e7. doi: 10.1016/j.cels.2020.08.018.
80. Rambow F, Marine J-C, Goding CR. Melanoma plasticity and phenotypic diversity: therapeutic barriers and opportunities. Genes Dev. 2019 Oct;**33**(19–20):1295–1318. doi: 10.1101/gad.329771.119.
81. Thankamony AP, Subbalakshmi AR, Jolly MK, Nair R. Lineage plasticity in cancer: the tale of a skin-walker. Cancers. 2021 Jul;**13**:14. doi: 10.3390/cancers13143602.
82. Nam AS, Chaligne R, Landau DA. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. Nat Rev Genet. 2021 Jan;**22**(1):3–18. doi: 10.1038/s41576-020-0265-5.
83. Dillekås H, Rogers MS, Straume O. Are 90% of deaths from cancer caused by metastases?. Cancer Med. 2019 Sep;**8**(12):5574–76. doi: 10.1002/cam4.2474.
84. Klein CA. Parallel progression of primary tumours and metastases. Nat Rev Cancer. 2009 Apr;**9**(4):302–12. doi: 10.1038/nrc2627.

85. Hunter KW, Amin R, Deasy S, Ha N-H, Wakefield L. Genetic insights into the morass of metastatic heterogeneity. Nat Rev Cancer. 2018 Apr;**18**(4):211–23. doi: 10.1038/nrc.2017.126.
86. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LAJ, Kinzler KW. Cancer genome landscapes. Science. 2013 March;**339**(6127):1546–58. doi: 10.1126/science.1235122.
87. Turajlic S, Xu H, Litchfield K, Rowan A, Chambers T, Lopez JI, et al. Tracking cancer evolution reveals constrained routes to metastases: TRACERx renal. Cell. 2018 Apr;**173**(3):581–94.e12. doi: 10.1016/j.cell.2018.03.057.
88. Hu Z, Li Z, Ma Z, Curtis C. Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. Nat. Genet. 2020 Jul;**52**(7):701–8. doi: 10.1038/s41588-020-0628-z.
89. Rhim AD, Thege FI, Santana SM, Lannin TB, Saha TN, Tsai S, et al. Detection of circulating pancreas epithelial cells in patients with pancreatic cystic lesions. Gastroenterology. 2014 March;**146**(3):647–51. doi: 10.1053/j.gastro.2013.12.007.
90. Rhim AD, Mirek ET, Aiello NM, Maitra A, Bailey JM, McAllister F, et al. EMT and dissemination precede pancreatic tumor formation. Cell. 2012 Jan;**148**(1–2):349–61. doi: 10.1016/j.cell.2011.11.025.
91. Laughney AM, Hu J, Campbell NR, Bakhoum SF, Setty M, Lavallée V-P, et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. Nat Med. 2020 Feb;**26**(2):259–69. doi: 10.1038/s41591-019-0750-6.
92. Simeonov KP, Byrns CN, Clark ML, Norgard RJ, Martin B, Stanger BZ, et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. Cancer Cell. 2021 Aug;**39**(8):1150–62.e9. doi: 10.1016/j.ccell.2021.05.005.
93. Quinn JJ, Jones MG, Okimoto RA, Nanjo S, Chan MM, Yosef N, et al. Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. Science. 2021 Feb;**371**:6532. doi: 10.1126/science.abc1944.
94. Komori J, Boone L, DeWard A, Hoppo T, Lagasse E. The mouse lymph node as an ectopic transplantation site for multiple tissues. Nat. Biotechnol. 2012 Oct;**30**(10):976–83. doi: 10.1038/nbt.2379.
95. Redmond KM, Wilson TR, Johnston PG, Longley DB. Resistance mechanisms to cancer chemotherapy. Front Biosci J Virtual Libr. 2008 May;**13**:5138–54. doi: 10.2741/3070.
96. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. New Engl J Med. 2001 Mar; **344**(11):783–92. doi: 10.1056/NEJM200103153441101.
97. Recasens A, Munoz L. Targeting cancer cell dormancy. Trends Pharmacol Sci. 2019 Feb; **40**(2):128–41. doi: 10.1016/j.tips.2018.12.004.
98. Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell. 2010 Apr;**141**(1):69–80. doi: 10.1016/j.cell.2010.02.027.
99. Guler GD, Tindell CA, Pitti R, Wilson C, Nichols K, Cheung TK, et al. Repression of stress-induced LINE-1 expression protects cancer cell subpopulations from lethal drug exposure. Cancer Cell. 2017 Aug;**32**(2):221–237.e13. doi: 10.1016/j.ccell.2017.07.002.
100. Bedard PL, Hansen AR, Ratain MJ, Siu LL. Tumour heterogeneity in the clinic. Nature. 2013 Sep;**501**(7467):355–64. doi: 10.1038/nature12627.
101. Liau BB, Sievers C, Donohue LK, Gillespie SM, Flavahan WA, Miller TE, et al. Adaptive chromatin remodeling drives glioblastoma stem cell plasticity and drug tolerance. Cell Stem Cell. 2017 Feb;**20**(2):233-46.e7. doi: 10.1016/j.stem.2016.11.003.
102. Merino D, Weber TS, Serrano A, Vaillant F, Liu K, Pal B, et al. Barcoding reveals complex clonal behavior in patient-derived xenografts of metastatic triple negative breast cancer. Nat Commun. 2019 Feb: **10**(1):766. doi: 10.1038/s41467-019-08595-2.
103. Rehman SK, Haynes J, Collignon E, Brown KR, Wang Y, Nixon AML, et al. Colorectal cancer cells enter a diapause-like DTP state to survive chemotherapy. Cell. 2021 Jan;**184**(1):226–42.e21. doi: 10.1016/j.cell.2020.11.018.
104. Echeverria GV, Ge Z, Seth S, Zhang X, Jeter-Jones S, Zhou X, et al. Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. Sci Transl Med. 2019 Apr;**11**:488. doi: 10.1126/scitranslmed.aav0936.
105. Dhimolea E, Simoes RM, Kansara D, Al'Khafaji A, Bouyssou J, Weng X, et al. An embryonic diapause-like adaptation with suppressed myc activity enables tumor treatment persistence. Cancer Cell. 2021 Feb;**39**(2):240–56.e11. doi: 10.1016/j.ccell.2020.12.002.
106. Hangauer MJ, Viswanathan VS, Ryan MJ, Bole D, Eaton JK, Matov A, et al. Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition. Nature. 2017 Nov;**551**(7679):247–50. doi: 10.1038/nature24297.
107. Recasens A, Munoz L. Targeting cancer cell dormancy. Trends Pharmacol Sci. 2019 Feb;**40**(2):128–41. doi: 10.1016/j.tips.2018.12.004.
108. Kint CI, Verstraeten N, Fauvart M, Michiels J. New-found fundamentals of bacterial persistence. Trends Microbiol. 2012 Dec;**20**(12):577–85. doi: 10.1016/j.tim.2012.08.009.
109. Van den Bergh B, Fauvart M, Michiels J. Formation, physiology, ecology, evolution and clinical importance of bacterial persisters. FEMS Microbiol Rev. 2017 May;**41**(3):219–51. doi: 10.1093/femsre/fux001.
110. Brauner A, Fridman O, Gefen O, Balaban NQ. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. Nat Rev Microbiol. 2016 Apr;**14**(5):320–30. doi: 10.1038/nrmicro.2016.34.
111. Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S. Bacterial persistence as a phenotypic switch. Science. 2004 Sep;**305**(5690):1622–25. doi: 10.1126/science.1099390.
112. Balaban NQ, Gerdes K, Lewis K, McKinney JD. A problem of persistence: still more questions than answers?. Nat Rev Microbiol. 2013 Aug;**11**(8):587–91. doi: 10.1038/nrmicro3076.
113. Deng L, Li C, Chen L., Y. Liu, R. Hou, and X. Zhou. Research advances on embryonic diapause in mammals. Anim. Reprod. Sci. 2018 Nov;**198**:1–10. doi: 10.1016/j.anireprosci.2018.09.009.
114. Fenelon JC, Renfree MB. The history of the discovery of embryonic diapause in mammals. Biol. Reprod. 2018 Jul;**99**(1):242–51. doi: 10.1093/biolre/ioy112.
115. Scognamiglio R, Cabezas-Wallscheid N, Thier MC, Altamura S, Reyes A, Prendergast AM, et al. Myc depletion induces a pluripotent dormant state mimicking diapause. Cell. 2016 Feb;**164**(4):668–80. doi: 10.1016/j.cell.2015.12.033.
116. Russo M, Pompei S, Sogari A, Corigliano M, Crisafulli G, Puliafito A, et al. A modified fluctuation-test framework characterizes the population dynamics and mutation rate of

colorectal cancer persister cells. Nat Genet. 2022 Jul;**54**(7):976–84. doi: 10.1038/s41588-022-01105-z.
117. Duy C, Li M, Teater M, Meydan C, Garrett-Bakelman FE, Lee TC, et al. Chemotherapy induces senescence-like resilient cells capable of initiating AML recurrence. Cancer Discov. 2021 Jun;**11**(6):1542–1561. doi: 10.1158/2159-8290.CD-20-1375.
118. Boroviak T, Loos R, Lombard P, Okahara J, Behr R, Sasaki E, et al. Lineage-specific profiling delineates the emergence and progression of naive pluripotency in mammalian embryogenesis. Dev. Cell. 2015 Nov;**35**(3):366–82. , doi: 10.1016/j.devcel.2015.10.011.
119. Bulut-Karslioglu A, Biechele S, Jin H, Macrae TA, Hejna M, Gertsenstein M, et al. Inhibition of mTOR induces a paused pluripotent state. Nature. 2016 Dec;**540**(7631):119–23. doi: 10.1038/nature20578.
120. Bhang HC, Ruddy DA, Radhakrishna VK, Caushi JX, Zhao R, Hims MM, et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. Nat. Med. 2015 May;**21**(5):440–48. doi: 10.1038/nm.3841.
121. Hammerlindl H, Schaider H. Tumor cell-intrinsic phenotypic plasticity facilitates adaptive cellular reprogramming driving acquired drug resistance. J Cell Commun Signal. 2018 Mar;**12**(1):133–141, doi: 10.1007/s12079-017-0435-1.
122. Russo M, Crisafulli G, Sogari A, Reilly NM, Arena S, Lamba S, et al. Adaptive mutability of colorectal cancers in response to targeted therapies. Science. 2019 Dec;**366**(6472):1473–80. doi: 10.1126/science.aav4474.
123. Cipponi A, Goode DL, Bedo J, McCabe MJ, Pajic M, Croucher DR, et al. MTOR signaling orchestrates stress-induced mutagenesis, facilitating adaptive evolution in cancer. Science. 2020 Jun;**368**(6495):1127–31. doi: 10.1126/science.aau8768.
124. Stress-induced mutation via DNA breaks in Escherichia coli: a molecular mechanism with implications for evolution and medicine—PubMed. https://pubmed.ncbi.nlm.nih.gov/22911060/ (accessed Mar. 31, 2023).
125. Luria SE, Delbrück M. Mutations of bacteria from virus sensitivity to virus resistance. Genetics. 1943 Nov;**28**(6):491–511. doi: 10.1093/genetics/28.6.491.
126. Iwasa Y, Nowak MA, Michor F. Evolution of resistance during clonal expansion. Genetics. 2006 Apr;**172**(4):2557–66., doi: 10.1534/genetics.105.049791.
127. Komarova NL, Wodarz D. Drug resistance in cancer: principles of emergence and prevention. Proc Natl Acad Sci U S A. 2005 Jul;**102**(27):9714–19. doi: 10.1073/pnas.0501870102.
128. Attolini CS-O, Michor F. Evolutionary theory of cancer. Ann N Y Acad Sci. 2009 Jun; **1168**:23–51. doi: 10.1111/j.1749-6632.2009.04880.x.
129. Kendal WS, Frost P. Pitfalls and practice of Luria-Delbrück fluctuation analysis: a review. Cancer Res. 1988 Mar;**48**(5):1060–65.
130. Bozic I, Reiter JG, Allen B, Antal T, Chatterjee K, Shah P, et al. Evolutionary dynamics of cancer in response to targeted combination therapy. eLife. 2013 Jun;**2**:e00747. doi: 10.7554/eLife.00747.

SECTION 3

Single cell 'Omics' analysis

11

Decoding drug resistance at a single-cell level using systems-level approaches

Benedict Anchang and Loukia G. Karacosta

11.1. Introduction

Despite many advances in cancer research and drug development, the fact remains that across almost all types of cancers, drug resistance inevitably develops leading to cancer progression, metastasis, and ultimately death [1]. A plethora of drug resistance mechanisms have been reported in chemotherapy, targeted therapy, and immunotherapy settings, and although we have deep molecular understanding on the development of drug resistance, we have yet to find ways of predicting and/or counteracting it in clinically impactful ways. It seems that no matter what new avenues of therapy or targeted markers researchers bring to the clinic, cancer cells find ways to outsmart and evade our efforts.

To complicate matters, within any given heterogeneous tumour, there can exist multiple ways of resisting therapy. For years, drug resistance was thought to be driven by phenomena well described in the microbiology fields, whereby pre-existing resistant clones (usually driven by a genetic mutation) are 'selected' via therapy, giving them the opportunity to grow and become the new dominant clone of resistant cells [2]. However, studies have shown that in many cases cells can exhibit non-genetic plasticity, often driven by epigenetic mechanisms through which they can transiently resist to any given drug and survive [2–4]. These two types of resistance (and their subcategories) are not mutually exclusive and can be employed at the same time by different cancer cells of a tumour depending on their surrounding microenvironment, drug treatment, or may take place at different time points during therapy.

Tumour heterogeneity, rarity of resistant cell populations, and the dynamics of emergent cell behaviours during therapy bring forth the importance of using single-cell approaches to help decode drug resistance. Single-cell -omics (which include the study of genomics, transcriptomics, proteomics, metabolomics, or a combination of these) have revolutionized the way we study and understand drug resistance. Specifically, we now have the tools and computational power to perform a variety of tasks such as (1) the in-depth characterization of heterogeneous drug-resistant populations with clustering algorithms [5], (2) the dynamic interrogation of phenotypic state transitions [6], (3) the prediction of resistance and/or promising drug combinations with deep learning [7], and (4) the optimization of drug combinations and therapy [8,9]. However, it is important to note that while single-cell analysis helps us simplify and delineate the complexity of drug resistance down to individual cells, cancer systems approaches help us put the informational pieces back together, towards understanding drug resistance at a systems level, within the ecosystem of a tumour and by extension of the patient [10].

In this chapter, we briefly discuss various mechanisms of drug resistance identified in cancer biology, and how single-cell analysis and computational approaches have helped gain a better understanding of them. Furthermore, we highlight how systems biology approaches can be utilized not only for delineating drug resistance, but, most importantly, how we can leverage them to translate scientific findings to the clinic and get closer to implementing precision oncology for cancer patients.

11.2. Mechanisms of drug resistance identified in cancer biology

11.2.1. Genetic drug resistance

For many years, the prevailing narrative has been that drug resistance in cancer is driven by mutations through which cells genetically inherit the ability to grow in the presence of high drug concentrations. Consequently, these pre-existing mutated cells become the main drivers of tumour progression by virtue of being selected (and expanded) during therapy [2,11] (**Figure 11.1**A). This Darwinian mode of drug resistance is well represented in instances where a point mutation alters a specific drug target, by affecting, for instance, the drug-specific binding site. One such example is the multitude of mutations that have been identified in tyrosine receptor kinases like the epidermal growth factor receptor (EGFR) [12]. However, the observation of extremely diverse cellular phenotypes alongside their frequency and their relatively fast appearance following therapy cannot be explained by mutational/genetic underpinnings alone [13] as some genetic studies have suggested [14]. The fact that drug therapy itself induces therapy resistance and that resistance can be

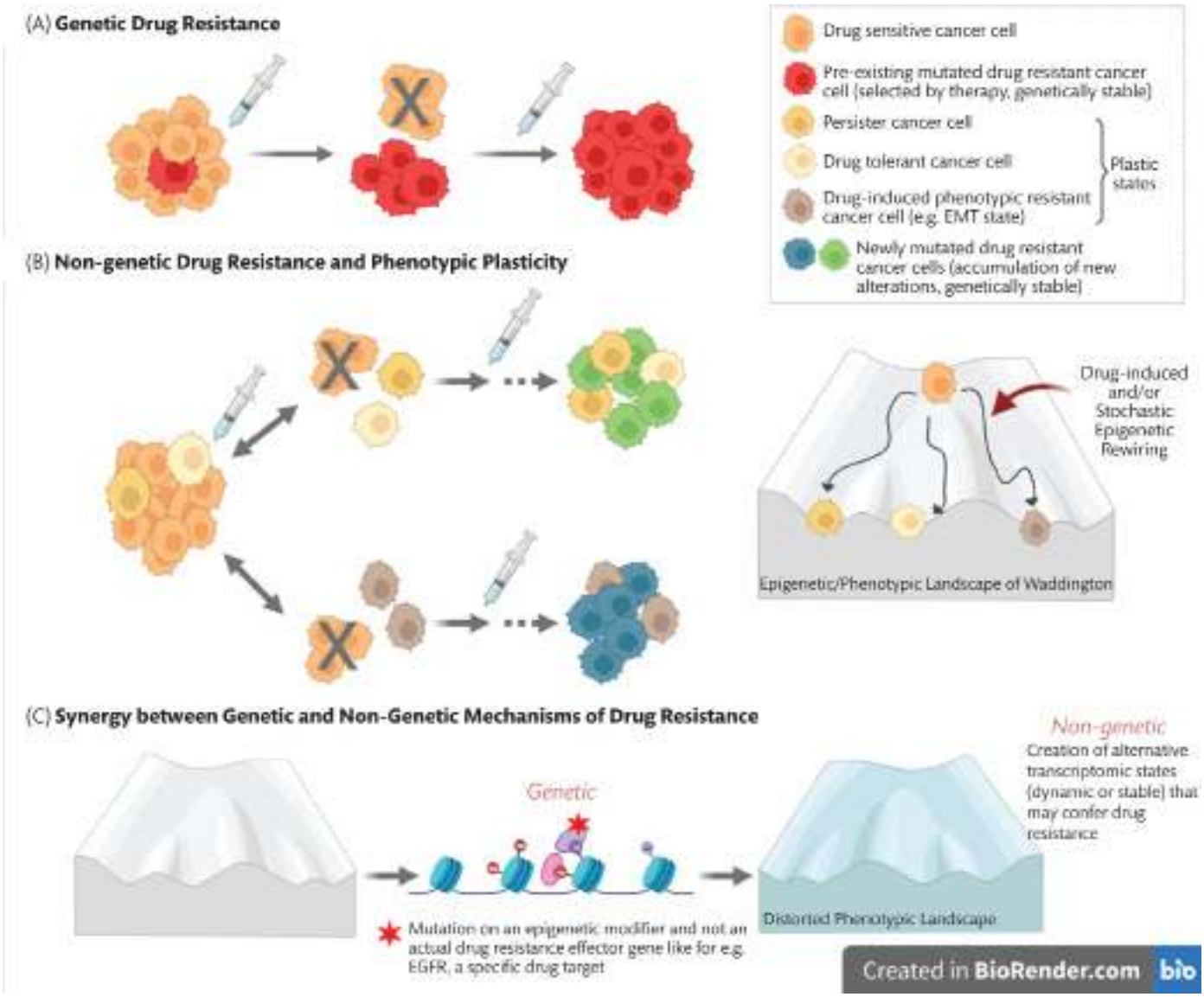


Figure 11.1. Mechanisms of drug resistance identified in cancer biology. (A) Genetic drug resistance schematic. Resistance emerges because of a pre-existing, rare cancer cell of a tumour with a specific drug-resistance-conferring mutation. This mutated cancer cell is 'selected' by drug therapy and subsequently expands to a new tumour mass via the classical Darwinian scheme of selection. (B) Non-genetic drug resistance schematic. Resistance develops due to the existence or induction of cellular phenotypes and/or states (e.g. persister, tolerant, and EMT cancer cells) within the same clonal population of cells in the absence of a resistance conferring mutation on a drug resistance effector gene (e.g., EGFR that is a drug target for many cancer types). Phenotypic states can be plastic in nature or further selected for (Lamarckian scheme) depending on how transcriptionally stable they are. In certain cases, depending on how stable a new state is, with time, new genetic mutations may appear, giving rise to distinct genetically stable cancer cell clones. Cell phenotypes can be the direct result of drug treatment (drug-induced) or stochastic epigenetic rewiring, which enables cancer cells to access heterogeneous, alternative transcriptional/epigenetic states (Waddington landscape). These can be dynamic in nature and confer new drug resistance traits to cancer cells. (C) An example of genetic and non-genetic mechanisms working in synergy and resulting in drug resistance development. A mutation on an epigenetic modifier and not specifically on a drug resistance effector gene may induce epigenetic changes and chromatin modifications that subsequently alter the landscape of phenotypic states available to cancer cells to leverage and resist to therapy. See more details in the main text.

reversible implies that non-genetic mechanisms of phenotypic plasticity also underlie the emergence of drug resistance.

11.2.2. Non-genetic drug resistance and phenotypic plasticity: drug-tolerant and persister cancer cells

Studies have shown that drug treatments can not only 'select' pre-existing resistant clones but also actively induce epigenetic reprogramming and phenotypic plasticity in cancer cells [4], which also explains the presence of reversible drug-*tolerant* cancer cells [3]. *Tolerance* is described as the ability of cells to survive transient exposures to high drug concentrations, and as opposed to *resistance*, it is not thought to be always inherited [15]. This brings forth the notion of a Lamarckian scheme of drug resistance (which is not necessarily mutually exclusive with the Darwinian scheme),

where acquired traits can be inherited and further selected for [13,16,17]. The Lamarckian scheme is feasible due to the existence of alternative, heterogeneous, stable transcriptional states that can produce aggressive phenotypes that can be passed on to multiple cancer cell generations. This has been demonstrated by a number of transcriptomic single-cell studies in recent years, highlighting the importance of leveraging single-cell systems biology approaches for studying drug resistance [6,18,19]. Furthermore, stochasticity also plays an important role in the development of drug resistance. Stochastic rewiring of various transcriptional regulatory networks can consequently activate alternate/latent pathways in response to drug perturbations, and this involves dynamic, epigenetic chromatin modifications [2]. Similar mechanisms take place during developmental processes in mammalian cells, further substantiating the existence of a repertoire of non-genetic 'choices' cancer cells have at their disposal to best survive drug-induced stress signals [20]. Stochastic epigenetic rewiring is also thought to be the mechanism that gives rise to *persister cells*. *Persistence* is defined as the ability of a subpopulation of cells within a clonal population to survive exposure to high drug concentrations; in this case, the majority of the population is rapidly eradicated, yet the rare, persister cells persist for long periods of time, offering a tumour one more trick under its sleeve to survive drug therapy and relapse at a later time [15]. Drug-tolerant and persister cells are in general plastic in nature and depending on the conditions may reverse (e.g. during drug withdrawal) or be further selected for (Lamarckian scheme). Finally, given that drug persister cells can be viable for extended periods of time, they may in turn acquire new mutations and give rise to genetically, stable resistant cancer cells. These cells can be further selected for with the traditional Darwinian scheme, highlighting the complexity of drug resistance in terms of the temporal dynamics and heterogeneity of the tumour system. (**Figure 11.1**B).

11.2.3. Synergy between genetic and non-genetic mechanisms of drug resistance

Although there have been numerous studies that have shown that genetic and non-genetic mechanisms are independently important for the development of drug resistance in cancer, one could envisage a scenario where both coexist in a synergistic manner. It is now easy to visualize how one genotype can map into an entire landscape of phenotypic states, some more stable than others (epigenetic landscape of Waddington [21]). Even though being the result of a single genome, the epigenetic landscape offers a plethora of decisions for a cancer cell to make, whether this is due to stochastic events or environmental cues (like exposure to drugs) [13,16,21], effectively explaining the phenomena such as plasticity and reversibility of phenotypic states (**Figure 11.1**B). Consequently, a resistance-conferring mutation may not be related to an actual effector gene/target (like EGFR), but can affect cell state transition dynamics instead [16]. Such a scenario is realized through many examples, one of them being the association of resistance to targeted therapies against EGFR with epithelial–mesenchymal transition (EMT) phenomena [22], and the other a commonly mutated epigenetic modifiers that can further distort the phenotypic landscape of cancer cells [23] (**Figure 11.1**C). There are many ways in which a specific phenotypic state can confer resistance. It is possible that a state may be defined by altered protein expression of a drug target, a drug efflux pump, or stemness. For example, a specific state of the EMT spectrum that was defined with single-cell mass cytometry (i.e. CyTOF) in lung cancer cells was characterized by depleted levels of pEGFR which would in theory confer resistance to tyrosine kinase inhibitor therapy (in the absence of a EGFR-conferring mutation) [24]. All other dynamic EMT states that were identified had varying levels of pEGFR, highlighting the importance of also understanding the dynamics of these processes.

11.3. How the advances in single-cell technologies can help us further decode cancer resistance

Drug resistance whether it is driven via genetic or non-genetic forces can be manifested in a variety of phenomena described in tumour biology: reversible therapy-induced senescence [25], quiescence [26], dormancy [27], cancer stemness, and EMT [28] to name a few. These phenomena share some characteristics: (1) they can all be induced by therapy and by the surrounding tumour microenvironment [29], (2) they are usually rare in occurrence, (3) they are plastic in nature, and (4) they are in general poorly described at the single-cell level. These common features necessitate the use of single-cell technologies and computational approaches for studying drug resistance. Although single-cell studies have been successful in identifying types of resistance, tackling tumour heterogeneity, and identifying rare drug-resistant populations [4,6,30] single-cell approaches (which are admittedly rapidly evolving) have yet to achieve important strides in predicting and counteracting drug resistance especially in a clinically impactful manner. Next, we summarize an array of single-cell technologies that are available to cancer researchers and how to use computational approaches to best realize what these types of analyses can offer towards decoding drug resistance.

11.4. The era of high-throughput single-cell technologies

A very broad way of describing the major classes of high-throughput single-cell technologies available to cancer biologists is breaking them down to (1) imaging-based techniques (the 'original' way of performing single-cell analysis) with the benefit of retaining spatial information (e.g. merFISH for spatial transcriptomics and imaging mass cytometry (IMC) for spatial proteomics [31]); (2) techniques that are based on flow and/or mass cytometry techniques where one can profile and analyse single cells towards their protein profiles by utilizing target-specific antibodies [32]; and (3) sequencing techniques that allow the quantification of genomes, epigenomes, and transcriptomes at the single-cell level [33]. These technologies have seen exponential growth and marked advancements the last decade alone, towards their multiparametric natures and high-dimensionality data output.

Another way of classifying single-cell technologies is by using the biological output (see review by Chen and Teichmann for more details [34]): single-cell (sc)-genomics (e.g. scDNA-seq), sc-transcriptomics (e.g. scRNA-seq and Split-seq), sc-epigenomics (e.g. scATAC-seq and scChIP-seq through which one can study chromatin modifications and these include sc-methylomic technologies), sc-proteomics

(e.g. flow cytometry and mass cytometry), and single-cell metabolomics via mass spectrometry approaches. In addition, there are now technological platforms that provide parallel quantification of more than one of these outputs: genomics/transcriptomics (G&T-seq [35]), transcriptomics/epigenomics (SNARE-seq and SHARE-seq [36,37]), and transcriptomics/subset of protein epitopes (CITE-seq and REAP-seq [38]).

Some of these technologies have been successfully used in drug resistance studies. For example, scRNA-seq and scATAC-seq have been implemented for studying retinoic acid resistance in leukaemia [39]. scChIP-seq was utilized for delineating chromatin states and drug resistance in breast cancer [40]. Single-cell multiplexed imaging approaches have only recently begun to be utilized for studying drug resistance at a spatial level. CITE-seq was used to longitudinally track leukaemic and immune cells populations during ibrutinib treatment and at relapse [41]. For instance, Bouzekri et al. used IMC to perform spatial proteomic profiling on drug-treated cells [42]. Spatial proteomics are increasingly becoming a point of focus, given that the response of cancer cells to treatment can also be affected by the surrounding microenvironment and spatial arrangements among various cell types. Finally, lineage tracing methodologies [43] can now be effortlessly paired with single-cell technologies, and this provides the opportunity to study the dynamics of emergent drug-resistant phenotypes.

Additional examples of these technologies and the downstream computational analysis that have been implemented for drug resistance studies are described in the following sections and in Table 11.1. It is important to note that we have not discussed in detail the standard computational workflow that typically follows single-cell experiments, which includes dimensionality reduction and visualization tools. For a more detailed description of these, see review by Todorov and Sayes [33].

11.5. Computational methods to study and overcome drug resistance at the single-cell level

Single-cell studies that are utilized to model or overcome drug resistance make use of three major computational types of analyses: clustering, network-structure learning (including trajectory modelling), and predictive modelling that typically involves machine learning approaches. In the following sections, we present a brief description of clustering, trajectory, and predictive modelling approaches with associated computational algorithms that have been/can be applied to high-throughput single-cell technologies to (1) characterize heterogeneous drug-resistant populations, (2) study dynamic phenotypic state transitions and trajectories, (3) predict resistance and promising drug combinations, and (4) optimize drug combinations that selectively kill cancer cells while simultaneously limiting toxic effects (by sparing normal, non-malignant cells).

11.5.1. Clustering single-cell data for identifying therapy-resistant cell populations and states

Computational methods based on clustering single-cell -omic data, such as scRNA-seq and CyTOF data, have been used to distinguish drug-tolerant states and select effective drugs and drug combinations to target resistant cell subpopulations [44]. The goal of clustering is to divide data points into a finite number of groups such that data points (usually high dimensional) in the same group are more similar to those in other groups (**Figure 11.2**). The result of a clustering algorithm can be unique for each data point (hard clustering) or probabilistic (soft clustering). Clustering algorithms make use of 'similarity' metric to measure how close points are to each other in space. This metric can be determined using distances (*k*-means clustering [45] and hierarchical clustering [46]), distributions (mixtures [47]), or network connectivity (graphical clustering [48]). For most of these methods, the number of clusters need to be predefined. For example, the *k*-means algorithm (**Figure 11.2**A) uses the distance between each cell and a centroid to assign $k = 3$ clusters. The agglomerative hierarchical clustering (**Figure 11.2**B) produces a five-level hierarchical tree called a dendrogram, which can be cut to generate four distinct clusters. Although several clustering methods have been developed for single-cell analysis, determining the unknown number of clusters remain a major challenge [49]. Network models try to overcome this problem by modelling data points as a continuous relationship with pairwise nodes representing interactions or similarities between biological entities. These approaches aim to preserve the cluster relatedness of the data. Clustering of these network models is derived manually (SPADE [46]) using an objective function (PAGA [48]) or a model-based strategy (CCAST [50]). At the same time, it is important to remember that understanding the single-cell data from a biological perspective can also help decide the optimal number of clusters; combining biology knowledge with well-informed computational tools promises to produce results that better represent what happens in the tumour microenvironment during treatment.

11.5.2. Network-structure learning and trajectory analysis of therapy-resistant states

Given that drug resistance is often confounded by time, temporal single-cell -omic analyses have been applied to study resistant cell state dynamics during tumour progression and therapy. Interrogating state transitions between therapy-resistant cellular states can be critical towards predicting what new resistant and potentially targetable state may appear upon treatment with any given drug. Learning an unknown cellular network structure over time can be challenging for single-cell analysis. For example, cellular differentiation is typically modelled using pseudotime trajectory, in which cells are ordered by progression instead of time. These models combine data reduction with clustering, and the trajectory is usually inferred on the reduced space (**Figure 11.3**). For example, a pseudotime trajectory model, like monocle for scRNA-seq analysis, first reduces the number of genes using independent component analysis [51]. It next connects all cells using a minimum spanning tree and then uses the longest connected path in that tree to form a pseudotime trajectory. Pseudotime models make use of prior knowledge to drive the developmental trajectory that can be biased. Temporary trajectory models, such as DSFMix [52], CStreet [53], and Tempora [54], order cellular progression deterministically by observed discrete time points. However, these models are not optimal for predictive modelling of new data points. Markov-based methods, such as Waddington-OT [55] and TRACER [24], can model cell growth and state transition rates and statistically compare time-dependent trajectories, respectively. Quantifying

Table 11.1. Examples of drug resistance studies utilizing single-cell technologies and computational approaches for identifying drug resistant states, interrogating trajectories, predicting therapy resistance and optimizing therapy. Respective reference numbers are indicated in parentheses.

Single-cell Platform	Omic Output	Cancer Type	Drug(s)	Objective(s)	Computational Approach(es)	References
scChIP-seq & scRNA-seq	Epigenomics & Transcriptomics	Breast Cancer	Chemotherapy (Cabecitabine, Tamoxifen)	Identification of heterogeneous chromatin states related to chemotherapy resistance	Hierarchical and consensus clustering	Grosselin et al., 2019 [40]
SCATAC-seq & scRNA-seq	Epigenomics & Transcriptomics	Leukemia	Retinoic Acid (RA)	Identification/charaterization of RA resistant cells, Trajectory analysis and scRNA-seq/ scATACseq Integration	Seurat clustering, based on PCA, UMAP. Pseudotime analysis with Single-cell Trajectories Reconstruction, Exploration And Mapping (STREAM algorithm)	Poplineau et al., 2020 [39]
scRNA-seq	Transcriptomics	Non-small Cell Lung Cancer (NSCLC)	Erlotinib	Identification of drug tolerant states and cell subpopulations	Seurat clustering based on t-SNE, UMAP, RNA velocity and GSEA method	Aissa et al., 2021 [44]
scRNA-seq	Transcriptomics	Acute Myeloid Leukemia (AML)	Multiple (approved and investigational)	Prediction of optimal drug combinations towards synergy, efficacy, and toxicity	Pathway analysis, XGBoost and Gene set variation analysis (GSVA)	Ianevski et al., 2021 [67]
scRNA-seq	Transcriptomics	Breast Cancer	Dexamethasone	Identification of heterogeneous populations of cells towards response to glucorticoids	Seurat clustering based on t-SNE and UMAP	Hoffman et al., 2020 [65]
ScRNA-seq & single cell imaging cytometry (for viability readout)	Transcriptomics	Ovarian Cancer	Multiple	Prediction of drug response	Patient specific drug-target interaction networks	He et al., 2021 [66]
ScRNA-seq & live imaging	Transcriptomics	Breast Cancer	Endocrine therapy	Identification of pre-adapted therapy resistant cells and multistep phenotypic adaptations to therapy	Hierarchical clustering, random forests, copy number variation analysis, Seurat clustering, Single-cell gene regulatory network inference	Hong et al., 2019 [64]
SCRNA-seq & lineage tracing	Transcriptomics & genomics	Glioblastoma	Receptor Tyrosine Kinase (RTK) inhibitors, (Dasatinib)	Identification of genetic and epigenetic mechanisms of resistance and persister phenotypes	t-SNE visualization, GSEA method	Eyler et al., 2020 [30]
CYTOF (single-cell mass cytometry)	Proteomics	Acute Lymphoblastic Leukemia (ALL)	Dasatinib, BEZ235, Tofacitinib	Optimization of patient specific drug combinations	Clustering, Bayesian network, Drug-target network with DRUGNEM algorithm	Anchang et al., 2018 [8]
CYTOF	Proteomics	Various cancer cell lines	TNFa-related apoptosis-Inducing ligand (TRAIL)	Identifying TRAIL resistance signaling states and interrogating trajectory of acquired resistance to TRAIL	VISNE and pseudatime analysis with Wanderlust algorithm	Baskar et al., 2019 [63]
CYTOF	Proteomics	NSCLC	TGFβ*	Identifying spectrum of EMT states and interrogating EMT state transition states reflective of tumor progression and therapy resistant features	Clustering with CCAST algorithm, mapping EMT states with PHENOSTAMP (Neural network) at a personalized level, trajectory analysis with TRACER algorithm	Karacosta et al., 2019 [24, 74]
CITE-seq	Transcriptomics & Proteomics	Chronic Lymphocytic Leukemia (CLL)	Ibrutinib	Identifying heterogeneous leukemic B cell populations and other immune populations during treatment	UMAP, Geneset enrichment with Single-cell signature explorer	Cadot et al., 2020 [41]
Imaging Mass Cytometry (IMC)	Single-cell spatial proteomics	Breast Cancer	Multiple	Profiling responses in drug treated cells	Spatial distribution mapping, dimensionality reduction, unsupervised and hierarchical clustering	Bouzekri et al., 2019 [42]
Single-cell Metabolomics	Metabolomics	Chronic Myeloid Leukemia (CML)	Chemotherapy	Prediction of drug resistant phenotypes	Machine learning: Random forest, logistic regression and neural networks	Liu et al., 2019 [58]

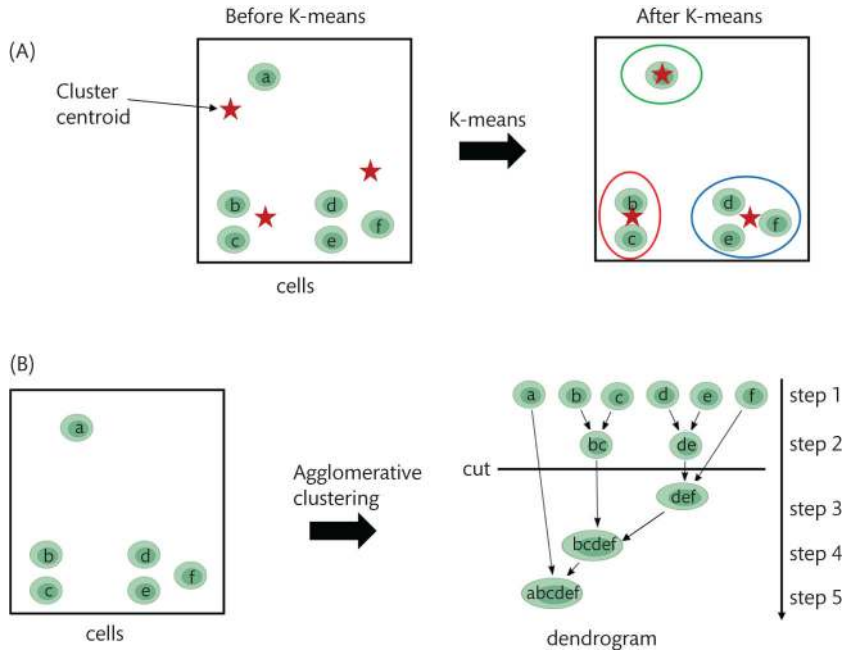


Figure 11.2. A brief overview of the (A) *k*-means and (B) hierarchical clustering respectively of six cells distributed in 2D space.

state transition rates offers the ability to interrogate reversibility among states, whereas pseudotime models provide a predetermined single trajectory of sequential states with no real input on the plasticity that often drives drug resistance phenomena. Finally, mechanistically driven ordinary-differentiation-equation-based models like RNA velocity can be used to predict the future state of individual cells on a timescale of hours driven by transcriptional kinetics [56].

11.5.3. Deep learning/neural networks for predictive modelling

With increase in computational power, deep learning or generative models based on artificial neural networks (ANNs) have also been used in single-cell analysis for predictive modelling. These network models are characterized by input, hidden, and output layers of nodes or neurons [57]. Figure 11.4A represents a simple feedforward neural network with a single hidden layer. Feedforward neural-network-driven approaches like PHENOtypic STAte MaP (PHENOSTAMP) [24] have been used to predict dynamic phenotypic (EMT) states of lung cancer cells using CyTOF training data or to predict drug resistance using mass spectrometry single-cell data [58]. Most popular ANNs for single-cell analysis are autoencoders [59], where the input and output layers are of the same dimension (Figure 11.4B). This is made possible with the help of encoder and decoder functions connected to the hidden layer. They have been used as a deep learning tool for low-dimensional structures for synthetic CyTOF data [59] and scRNA-seq data [60]. The clustering and predictive models highlighted above can be tailored towards identifying biomarkers, characterizing drug resistance states but are not informative on how to control or regulate them.

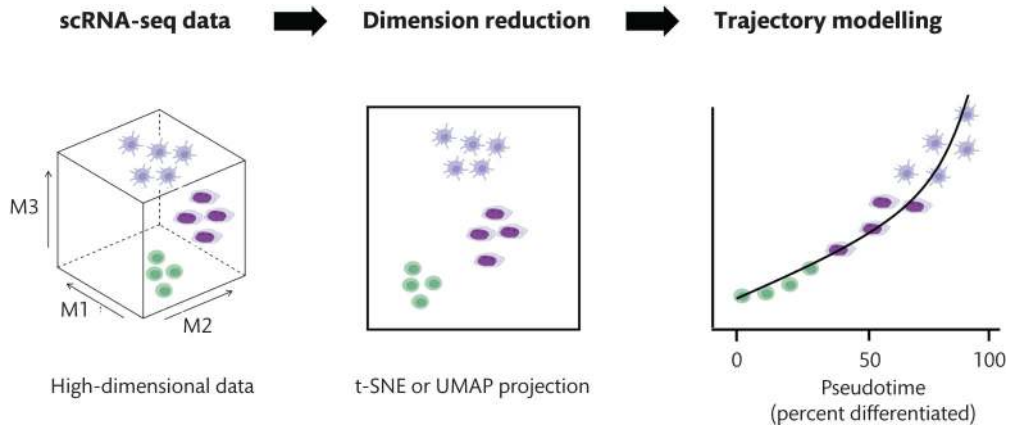


Figure 11.3. A schematic overview of trajectory modelling. Given a single-cell expression data, these models typically combine data reduction with clustering and the trajectory model is applied on the reduced data space.

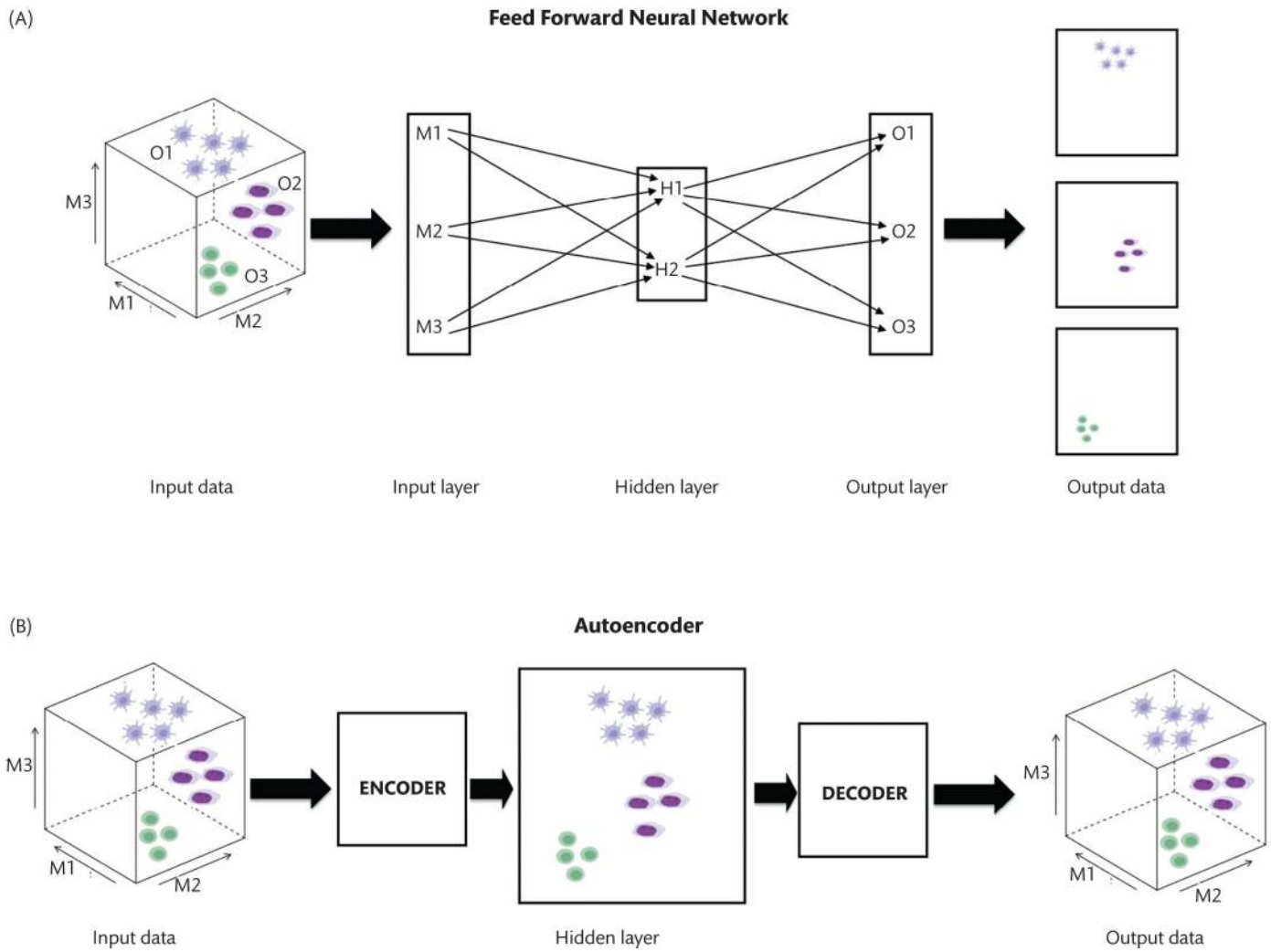


Figure 11.4. Examples of neural networks. (A) Feedforward neural network with a single hidden layer. (B) Autoencoder showing the same data input and output dimensions.

11.5.4. Integrating network analysis and machine learning approaches for predicting and optimizing drug combinations

We now focus on recent computational methods that have the potential to select effective drugs to target both sensitive and resistant cell subpopulations for a given patient, accounting for both the heterogeneity of malignant cells and the possible toxic effects on non-malignant cells, thereby offering a unique opportunity to optimize therapy towards both efficacy and toxicity.

Several tools based on high-throughput genomic and proteomic screens derived from molecular bulk assays have been proposed for combination therapy that leverages the idea that an individual tumour's biology is dominated by a single regulatory network under the control of a limited number of master regulators [61,62]. They make use of network analysis encoding protein–protein interactions, dynamic pathway simulations, and reduced network motifs to identify putative mechanisms of synergistic and antagonistic drug–drug and gene–drug interactions to overcome resistance. These approaches do not account for tumour heterogeneity as highlighted by numerous CyTOF and scRNA-seq *in vitro* studies that have revealed striking cell-to-cell variability to single and/or combinations of drugs [63–65]. This complexity challenges the current field of combination therapy to target cell interactions within the tumour microenvironment. To address this challenge, network-guided identification of selective combinatorial therapies in cancer has been developed to tailor personalized therapeutic regimens that consider both the molecular heterogeneity of cancer cells and the possible nonselective toxic effects of the drug combinations. For example, He et al. implemented a two-phase machine-learning-based platform to identify safe and effective combinatorial treatments that selectively inhibit cancer-related dysfunctions or resistance mechanisms in individual patients [66]. They used a comprehensive drug–target interaction network to improve both combination efficacy and tolerability together with genomic and molecular aberrations to construct patient-specific co-vulnerability networks. Another Bayesian-model-based algorithm like DRUGNEM [8] optimizes combination therapy for an individual patient by (1) identifying the subpopulations that make up the tumour and may respond differently to treatment, (2) reconstructing a drug-nested-effects network model that integrates the drug effects across all subpopulations, and (3) systematically scoring and ranking potential drug combinations that maximize

the desired effects in both sensitive and non-sensitive populations of cells with the minimum number of drugs. DRUGNEM was applied using CyTOF data to individualize therapy for a cohort of 30 paediatric acute lymphoblastic leukaemia (ALL) patients. Ianevski et al. used gradient boosting machine learning approach (XGBoost) and gene set variation analysis to prioritize patient-specific drug combinations by accounting for synergy, efficacy, and toxicity by combining scRNA-seq with *ex vivo* single-agent viability assays in patient-derived primary cells [67]. Specifically, they predicted which drug pair will produce the highest cell killing effect and then selected among the top predictions those with a lower likelihood of toxic effects in four leukaemia patients.

The above studies demonstrate the complexity involved in overcoming drug resistance networks within the tumour microenvironment as we increase the number of drugs. Computational tools focused on optimizing drug combinations while accounting for drug resistance heterogeneity are inherently complex. They combine machine learning, network modelling, and optimization of several objective functions. To make things even more complicated, there is a need to expand these computational frameworks to account for multi-omic data integration, higher-order synergistic interactions beyond pairwise interactions as well as temporally staggered treatments, which are known to affect *in vivo* drug efficacy.

11.6. Systems biology approaches and data integration for decoding drug resistance

There is no doubt that we are now experiencing the golden era of single-cell multi-omic technologies and advanced computational tools in cancer research. Researchers today have an abundance of state-of-the-art technologies at their disposal for decoding drug resistance mechanisms. The challenge however lies in the task of putting the 'single-cell' pieces back together towards achieving a holistic view of drug resistance in the tumour and by extension, the patient *system*. This is where systems biology approaches are critical. Systems biology is a multidisciplinary field that is uniquely suited to study complex-systems-wide behaviours (like drug resistance), by applying computational and mathematical tools specifically developed for analysing high-dimensional data generated by well-designed experiments and publicly available datasets [68]. To best delineate and model drug resistance as a behaviour of the *system*, one key step is to integrate the various -omic molecular layers and scales (e.g. transcriptomic, epigenomic, proteomic, and spatial). This has become increasingly important as more and more technologies offer the capability of quantifying more than one output. Achieving integration however can be quite challenging as different layers of information often require vastly different analytical approaches [53]. Nevertheless, computational frameworks have been developed to harmonize multimodal -omic data. Methods, such as **L**inked **I**nference of **G**enomic **E**xperimental **R**elationships (LIGER) [69] and **J**ointly **S**emi-orthogonal **N**onnegative **M**atrix **F**actorization (JSNMF) [70], have been used to integrate transcriptomic and epigenomic data profiled from the same cell. Bi-CCA allows for combinations of any two single-cell modalities [71]. Finally, there is an increased interest in integrating multimodal -omic data with single-cell multiplexed imaging platforms, and this offers the capability of identifying cell types, their architectural patterns in tissue, and at the same time interrogating their functional traits in the context of their -omic expression profiles [72]. For example, DBiT-seq offers simultaneous interrogation of transcriptomic, proteomic, and spatial data [73].

11.7. Concluding remarks: decoding drug resistance at a personalized level

For summary, we provide Table 11.1 with examples of studies in which single-cell technologies and computational approaches were developed to better understand and treat drug resistance across various cancer models. We highlight the specific technologies and analytical tools utilized and whether the study was focused on identifying resistant populations/states with clustering, interrogating trajectories, predicting therapy response/optimizing therapy, or a combination of these.

The plethora of single-cell technologies and the computational power available to researchers today have greatly advanced our understanding on drug resistance in cancer. What is clearly missing however from the field is a more focused effort towards decoding drug resistance at a personalized level. In this chapter, we primarily discussed approaches to interrogate drug resistance within a heterogeneous tumour. Yet, it is well known that cancers are vastly different from patient to patient. One way to approach this important issue is by utilizing high-dimensional single-cell -omic data and machine learning algorithms to construct reference maps of therapy resistance states observed *in vitro* and in clinical specimens per cancer type. These reference maps can then be used to assess therapy response and drug resistance in longitudinal clinical specimens and tailor treatment at a personalized level [24,74]. Constructing well-informed and clinically applicable reference maps requires performing translational studies that combine and integrate data from *top-down* (e.g. analysis of patient intact tissue pre- and post-therapy with spatial multiplexed imaging) and *bottom-up* approaches (functional, *in vitro* drug treatments and single-cell analysis on patient-derived cells and/or organoids) [75]. This once again highlights the notion of studying drug resistance at multiple levels of the *system* to better understand its development and come up with ways to counteract it.

With all the advances in multi-omics, artificial intelligence, and systems biology approaches we are experiencing today, we are uniquely positioned to begin implementing precision oncology when it comes to predicting and preventing drug resistance. As single-cell technologies and complex computational methods continue to rapidly advance, it is up to cancer researchers to harness the multidisciplinary nature of systems biology to decode drug resistance identified in one cell and patient at a time.

REFERENCES

1. Vasan N, Baselga J, Hyman DM. A view on drug resistance in cancer. Nature. 2019;**575**:299–309.
2. Salgia R, Kulkarni P. The genetic/non-genetic duality of drug 'resistance' in cancer. Trends Cancer. 2018;**4**:110–18.
3. Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell. 2010;**141**:69–80.

4. Shaffer SM, Dunagin M, Torborg S, Torre EA, Emert B, Krepler C, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature. 2017;**546**:431–5.
5. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;**15**:81–94.
6. Su Y. et al. Single-cell analysis resolves the cell state transition and signaling dynamics associated with melanoma drug-induced resistance. Proc Natl Acad Sci U S A. 2017;**114**: 13679–84.
7. Wu Z, Lawrence PJ, Ma A, Zhu J, Xu D, Ma Q. Single-cell techniques and deep learning in predicting drug response. Trends Pharmacol Sci. 2020;**41**:1050–65.
8. Anchang B, Davis KL, Fienberg HG, Williamson BD, Bendall SC, Karacosta LG, et al. DRUG-NEM: optimizing drug combinations using single-cell perturbation response to account for intratumoral heterogeneity. Proc Natl Acad Sci U S A. 2018;**115**:E4294–4303.
9. Shalek AK, Benson M. Single-cell analyses to tailor treatments. Sci Transl Med. 2017;**9**: eaan4730.
10. Cheng H, Fan R, Wei W. Cancer systems biology in the era of single-cell multi-omics. Proteomics. 2020;**20**:1900106.
11. Xue Y, Martelotto L, Baslan T, Baslan T, Vides A, Solomon M, Mai TT, Chaudhary N, et al. An approach to suppress the evolution of resistance in BRAFV600E-mutant cancer. Nat Med. 2017;**23**:929–37.
12. Rosenzweig SA. Acquired resistance to drugs targeting receptor tyrosine kinases. Biochem Pharmacol. 2012;**83**:1041–48.
13. Huang S. Reconciling non-genetic plasticity with somatic evolution in cancer. Trends Cancer. 2021;**7**:309–22.
14. Satgé D. Analysis of somatic mutations in cancer tissues challenges the somatic mutation theory of cancer. eLS (John Wiley & Sons, Ltd, 2013). doi:10.1002/9780470015902.a0024465.
15. Brauner A, Fridman O, Gefen O, Balaban NQ. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. Nat Rev Microbiol. 2016;**14**:320–30.
16. Huang S. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. Cancer Metastasis Rev. 2013;**32**:423–48.
17. Bell CC, Gilan O. Principles and mechanisms of non-genetic resistance in cancer. Br J Cancer. 2020;**122**:465–72.
18. Shaffer SM, Emert BL, Reyes Hueros RA, Cote C, Harmange G, Schaff DL, et al. Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors. Cell. 2020;**182**:947–959.e17.
19. Sharma A, Cao EY, Kumar V, Zhang X, Leong HS, Lin AM, et al. Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. Nat Commun. 2018;**9**:4931.
20. Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature. 2008;**453**:544–7.
21. Huang S. The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology? BioEssays. 2012;**34**:149–57.
22. Weng CH, Chen LY, Lin YC, Shih JY, Lin YC, Tseng RY, et al. Epithelial–mesenchymal transition (EMT) beyond EGFR mutations per se is a common mechanism for acquired resistance to EGFR TKI. Oncogene. 2019;**38**:455–68.
23. Morinishi L, Kochanowski K, Levine RL, Wu LF, Altschuler SJ. Loss of TET2 affects proliferation and drug sensitivity through altered dynamics of cell-state transitions. Cell Syst. 2020;**11**:86–94.e5.
24. Karacosta LG, Anchang B, Ignatiadis N, Kimmey SC, Benson JA, Shrager JB, et al. Mapping lung cancer epithelial–mesenchymal transition states and trajectories with single-cell resolution. Nat Commun. 2019;**10**:5587.
25. Saleh T, Bloukh S, Carpenter VJ, Alwohoush E, Bakeer J, Darwish S, et al. Therapy-induced senescence: an 'old' friend becomes the enemy. Cancers. 2020;**12**:822.
26. Paul R, Dorsey JF, Fan Y. Cell plasticity, senescence, and quiescence in cancer stem cells: biological and therapeutic implications. Pharmacol Ther. 2022;**231**:107985.
27. De Angelis ML, Francescangeli F, La Torre F, Zeuner A. Stem cell plasticity and dormancy in the development of cancer therapy resistance. Front Oncol. 2019;**9**.
28. Shibue T, Weinberg RA. EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. Nat Rev Clin Oncol. 2017;**14**:611–29.
29. Khalaf K, Hana D, Chou JT, Singh C, Mackiewicz A, Kaczmarek M. Aspects of the tumor microenvironment involved in immune resistance and drug resistance. Front Immunol. 2021;**12**.
30. Eyler CE, Matsunaga H, Hovestadt V, Vantine SJ, van Galen P, Bernstein BE. Single-cell lineage analysis reveals genetic and epigenetic interplay in glioblastoma drug resistance. Genome Biol. 2020;**21**:174.
31. Karacosta LG. From imaging a single cell to implementing precision medicine: an exciting new era. Emerg Top Life Sci. 2021;**5**:837–47.
32. Bendall SC, Nolan GP, Roederer M, Chattopadhyay PK. A deep profiler's guide to cytometry. Trends Immunol. 2012;**33**:323–32.
33. Todorov H, Saeys Y. Computational approaches for high-throughput single-cell data analysis. FEBS J. 2019;**286**:1451–7.
34. Chen S, Teichmann SA. Completing the cancer jigsaw puzzle with single-cell multiomics. Nat Cancer. 2021;**2**:1260–62.
35. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods. 2015;**12**:519–22.
36. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nat Biotechnol. 2019;**37**:1452–7.
37. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell. 2020;**183**:1103–1116.e20.
38. Todorovic V. Single-cell RNA-seq—now with protein. Nat Methods. 2017;**14**:1028–9.
39. Poplineau M, Platet N, Mazuel A, Hérault L, N'Guyen L, Koide S, et al. Noncanonical EZH2 drives retinoic acid resistance of variant acute promyelocytic leukemias. Blood. 2022;**140**:2358–70.
40. Grosselin K, Durand A, Marsolier J, Poitou A, Marangoni E, Nemati F, et al. High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. Nat Genet. 2019;**51**:1060–66.
41. Cadot S, Valle C, Tosolini M, Pont F, Largeaud L, Laurent C, et al. Longitudinal CITE-Seq profiling of chronic lymphocytic leukemia during ibrutinib treatment: evolution of leukemic and immune cells at relapse. Biomark Res. 2020;**8**:72.
42. Bouzekri A, Esch A, Ornatsky O. Multidimensional profiling of drug-treated cells by Imaging Mass Cytometry. FEBS Open Bio. 2019;**9**:1652–69.
43. Wagner DE, Klein AM. Lineage tracing meets single-cell omics: opportunities and challenges. Nat Rev Genet. 2020;**21**:410–27.
44. Aissa AF, Islam ABMMK, Ariss MM, Go CC, Rader AE, Conrardy RD, et al. Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. Nat Commun. 2021;**12**:1628.

45. Marco E, Karp RL, Guo G, Robson P, Hart AH, Trippa L, et al. Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. Proc Natl Acad Sci U S A. 2014;**111**:E5643–50.
46. Anchang B, Hart TD, Bendall SC, Qiu P, Bjornson Z, Linderman M, et al. Visualization and cellular hierarchy inference of single-cell data using SPADE. Nat Protoc. 2016;**11**:1264–79.
47. Sun Z, Chen L, Xin H, Jiang Y, Huang Q, Cillo AR, et al. A Bayesian mixture model for clustering droplet-based single-cell transcriptomic data from population studies. Nat Commun. 2019;**10**:1649.
48. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Göttgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 2019;**20**:59.
49. Yu L, Cao Y, Yang JYH, Yang P. Benchmarking clustering algorithms on estimating the number of cell types from single-cell RNA-sequencing data. Genome Biol. 2022;**23**:49.
50. Anchang B, Do MT, Zhao X, Plevritis SK. CCAST: a model-based gating strategy to isolate homogeneous subpopulations in a heterogeneous population of single cells. PLOS Comput. Biol. 2014;**10**:e1003664.
51. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;**32**:381–6.
52. Anchang B, Mendez-Giraldez R, Xu X, Archer TK, Chen Q, Hu G, et al. Visualization, benchmarking and characterization of nested single-cell heterogeneity as dynamic forest mixtures. Brief. Bioinform. 2022;**23**:bbac017.
53. Zhao C, Xiu W, Hua Y, Zhang N, Zhang Y. CStreet: a computed Cell State trajectory inference method for time-series single-cell RNA sequencing data. Bioinformatics. 2021;**37**:3774–80.
54. Tran TN, Bader GD. Tempora: cell trajectory inference using time-series single-cell RNA sequencing data. PLOS Comput Biol. 2020;**16**:e1008205.
55. Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. Cell. 2019;**176**:928–943.e22.
56. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;**560**:494–8.
57. Ding J, Condon A, Shah SP. Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. Nat Commun. 2018;**9**:2002.
58. Liu R, Zhang G, Yang Z. Towards rapid prediction of drug-resistant cancer cell phenotypes: single cell mass spectrometry combined with machine learning. Chem Commun. 2019;**55**:616–19.
59. Szubert B, Cole JE, Monaco C, Drozdov I. Structure-preserving visualisation of high dimensional single-cell datasets. Sci Rep. 2019;**9**:8914.
60. Wang D, Gu J. VASC: dimension reduction and visualization of single-cell RNA-seq data by deep variational autoencoder. Genomics Proteomics Bioinformatics. 2018;**16**:320–31.
61. Li S, Zhang B, Zhang N. Network target for screening synergistic drug combinations with application to traditional Chinese medicine. BMC Syst Biol. 2011;**5**:S10.
62. Ryall KA, Tan AC. Systems biology approaches for advancing the discovery of effective drug combinations. J Cheminform. 2015;**7:7**.
63. Baskar R, Fienberg HG, Khair Z, Favaro P, Kimmey S, Green DR, et al. TRAIL-induced variation of cell signaling states provides nonheritable resistance to apoptosis. Life Sci Alliance. 2019;**2**.
64. Hong SP, Chan TE, Lombardo Y, Corleone G, Rotmensz N, Bravaccini S, et al. Single-cell transcriptomics reveals multi-step adaptations to endocrine therapy. Nat Commun. 2019;**10**:3840.
65. Hoffman JA, Papas BN, Trotter KW, Archer TK. Single-cell RNA sequencing reveals a heterogeneous response to glucocorticoids in breast cancer cells. Commun Biol. 2020;**3**:1–11.
66. He L, Bulanova D, Oikkonen J, Häkkinen A, Zhang K, Zheng S, et al. Network-guided identification of cancer-selective combinatorial therapies in ovarian cancer. Brief Bioinform. 2021;**22**:bbab272.
67. Ianevski A, Lahtela J, Javarappa KK, Sergeev P, Ghimire BR, Gautam P, et al. Patient-tailored design for selective co-inhibition of leukemic cell subpopulations. Sci Adv. 2021;**7**:eabe4038.
68. Weiskittel TM, Correia C, Yu GT, Ung CY, Kaufmann SH, Billadeau DD, et al. The trifecta of single-cell, systems-biology, and machine-learning approaches. Genes. 2021;**12**:1098.
69. Liu J, Gao C, Sodicoff J, Kozareva V, Macosko EZ, Welch JD. Jointly defining cell types from multiple single-cell datasets using LIGER. Nat Protoc. 2020;**15**:3632–62.
70. Ma Y, Sun Z, Zeng P, Zhang W, Lin Z. JSNMF enables effective and accurate integrative analysis of single-cell multiomics data. Brief Bioinform. 2022;**23**:bbac105.
71. Dou J, Liang S, Mohanty V, Miao Q, Huang Y, Liang Q, et al. Bi-order multimodal integration of single-cell data. Genome Biol. 2022;**23**:112.
72. Longo SK, Guo MG, Ji AL, Khavari PA. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. Nat Rev Genet. 2021;1–18. doi:10.1038/s41576-021-00370-8.
73. Liu Y, Yang M, Deng Y, Su G, Enninful A, Guo CC, et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. Cell. 2020;**183**:1665–1681.e18.
74. Karacosta, L.G., Pancirer, D., Preiss, J.S. et al. Phenotyping EMT and MET cellular states in lung cancer patient liquid biopsies at a personalized level using mass cytometry. Sci Rep 13, 21781 (2023). https://doi.org/10.1038/s41598-023-46458–5.
75. Allam M, Cai S, Coskun AF. Multiplex bioimaging of single-cell spatial profiles for precision cancer diagnostics and therapeutics. NPJ Precis Oncol. 2020;**4**:1–14.

12

Computational methods to infer lineage decision-making in cancer using single-cell data

Manu Setty

12.1. Introduction

Lineage decisions are a hallmark of the differentiation process, where multi-potent stem cells give rise to the full complement of functionally distinct cell types through a series of cell divisions and lineage decisions to establish and maintain tissue homeostasis. Cells are specified to differentiate to particular lineages following a lineage decision while simultaneously losing the potential to differentiate to other sister lineages. Recent studies have highlighted the role of lineage decision-making process in cancers where fate-committed cells might acquire stem-like properties due to somatic mutations and other perturbations leading to a differentiation process that has gone awry. Single-cell measurement technologies and computational methods have enabled a high-resolution investigation of lineage decisions in biological systems spanning development to cancer.

Classically, lineage decision-making process has been viewed as a discrete process with a series of bifurcations between cell states, where state transitions are punctuated by extensive changes in expression patterns of hundreds of genes. The notion of discrete cell states has also been broadly used in disease such as cancer, where individual cancers have been classified to a discrete number of subtypes based on their transcriptome profiles. However, these ideas largely stemmed from bulk gene expression profiling, measurement technologies that are not equipped to study the lineage decision-making process at a high resolution.

Advances in single-cell RNA-seq (scRNA-seq) technologies can generate transcriptome profiles of individual cells for thousands of cells that have challenged the notion of discrete state transitions and subtypes and have posited that cell-state transitions underlying virtually all biological systems spanning development to disease are continuous in nature [1,2]. While initial single-cell studies suffered from lack of throughput [3], recent advances have led to routine generation of large-scale datasets of tens of thousands of cells. Droplet microfluidics [4,5] was the transformative technology that enabled the profiling of thousands of cells in a single experiment—single cells are encapsulated in oil droplets along with unique cell barcodes and lysis reagents. Cells are lysed within the droplets, and the transcripts of each cell are tagged with the unique cell barcodes contained in the oil droplet. The droplets are then broken, library preparation done in bulk, and the transcriptome and cell barcodes are both recovered by high-throughput sequencing. Split-pool technologies [6] have increased the throughput by another order of magnitude—barcodes are incorporated into cells in 96-well plates, where each cell contains a unique barcode. Cells from all wells are pooled and then randomly redistributed into another barcode contains 96-well plate leading to the incorporation of a second barcode. This process is repeated several times to ensure that each cell is tagged with a unique combination of barcodes.

Consortia such as Human Cell Atlas [7], Human Tumor Atlas Network [8], Tabula Muris [9], and others have leveraged these technological advances to generate datasets containing millions of cells, providing unprecedent scale and resolution to investigate biological systems. A consistent observation across these datasets is that the phenotypic space spanned by single cells is continuous in nature. With a simplifying assumption of each cell representing a unique state, studies have observed that state transitions are gradual and continuous—while the number of genes that change between state transitions can be in tens or hundreds in part due to coordinate gene regulation, the extent of change is incremental unlike the changes over orders of magnitudes observed with bulk data. The continuous nature of cell-state transitions and the resolution of single-cell data has spurred the development of a large number of computational trajectory detection algorithms where the goal is to place the single cells in an order representing their progression and also describe the lineage decision process along the progression [10]. Current approaches for trajectory detection have invariably been developed to study the ordered process such as healthy differentiation and development, regeneration, wound healing, and directed programming. Nonetheless, these approaches have been successfully applied to describe critical cancer-related processes, such as tumour

microenvironment immune cell trajectories [11], metastatic transformation [12], and response to drugs [13].

Lineage plasticity has emerged as a key theme across several different cancers [14]. Differentiated cells demonstrate plasticity by acquiring properties of progenitor cells through de-differentiation or trans-differentiate to alternative lineages, both as a result of somatic mutations and associated perturbations. The applicability of current algorithms that describe lineage decision-making to emergence and progression of highly plastic populations is currently an area of investigation with many efforts underway to develop computational methods that leverage multi-modal single-cell data to describe trajectories in cancer. This chapter describes the state-of-the-art trajectory detection algorithms and their successful application in the context of cancer and concludes with a discussion of efforts to model lineage plasticity.

12.2. Trajectory detection algorithms

Trajectory detection or trajectory inference algorithms are arguably the most studied computational concept with single-cell data. These algorithms provide rich, interpretable views of the data from which hypotheses can be generated for downstream experimental tests and validation. The input to trajectory detection is a single-cell dataset and outputs broadly are an ordering of cells, representing their progression and information about the lineage decision-making process (Figure 12.1). While there are a large number of algorithms available, each with their own assumptions and scenarios in which they are applicable, the following assumptions are generally required for the application of any algorithm: (1) the dataset measures cells that are a part of dynamic biological system; (2) the dataset of interest comprises cells in a continuum, i.e. the application of trajectory detection to a dataset of distinct cell types is not meaningful; (3) there are no missing states in the system including low-frequency transient populations; and (4) virtually all algorithms assume that the flow of information along a trajectory is unidirectional, i.e. cells cannot retrace their path in the reverse order of progression. Most of these assumptions are satisfied in a healthy differentiation and development, but assumption (4) might not be valid in cancers due the extensive observations of lineage plasticity. Further, a number of algorithms require prior biological information to specify the start of the trajectories, but recent approaches can automatically detect the start and terminal states in the system.

Trajectory detection algorithms can be broadly grouped into three classes: (1) graph-based approaches that use nearest neighbour graphs, (2) RNA-velocity-based approaches that take advantage of detection of spliced and unspliced transcripts in scRNA-seq to infer lineage dynamics using splicing dynamics from a population of cells, and (3) approaches that leverage scRNA-seq datasets measured over multiple time points or time-series datasets.

12.2.1. Graph-based trajectory detection algorithms

The first algorithms for trajectory detection, Monocle (1) and Wanderlust (2), were both inspired by concepts from the field of manifold learning. In manifold learning, the goal is to identify the true underlying low-dimensional representation of a dataset when the observations are made in high-dimensional space. This is directly applicable to biological systems since the phenotypic states of the biological system are likely to occupy a substantially lower-dimensional space even though the data is measured in thousands of dimensions representing genes. This is a result of (1) covarying patterns of genes due to tight regulatory mechanisms and (2) the fact that not all genes are expressed in all cell types. Manifold learning also assumes that the observed high-dimensional data is locally Euclidean and thus nearest neighbour graphs are a widely used approach to describe the underlying manifold. The idea behind nearest neighbour graphs is simple: each data point is connected to its k-nearest data points where the number of neighbours, k, is a user-defined parameter. Distances between data points that are not neighbours are measured as steps through the graph rather than directly measuring distances in the observed dimensions. Graph traversals can accurately recover distances of points along the manifold provided that the data and graph construction is not noisy.

Monocle (1) and Wanderlust (2) both use nearest neighbour graphs for trajectory detection. As the first set of algorithms designed for this purpose, they tackled the problem of inferring progression of cells along a single trajectory without branches. Thus, the problem effectively derives a one-dimensional representation of cells, in which the order represents progression of differentiation. Single-cell data is noisy and sparse due to dropouts [15], and as a result the construction of nearest neighbour graphs directly using

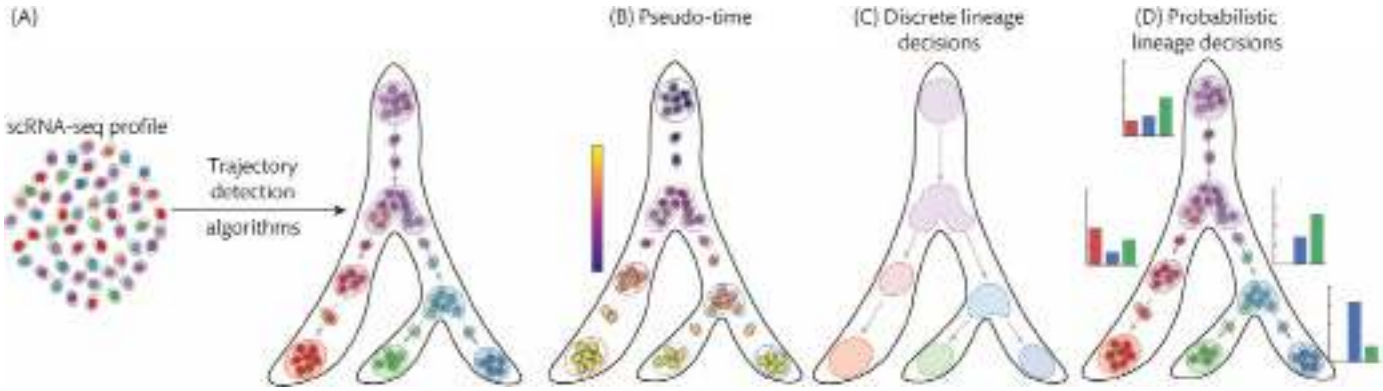


Figure 12.1. Trajectory detection algorithms. Trajectory detection algorithms using single-cell RNA-seq profiles as input (A) to derive a pseudo-temporal ordering of cell representing their progression (B) and lineage decisions either as discrete bifurcations (C) or probabilistic cell-fate choice (D).

the normalized data is highly unreliable. Therefore, most single-cell algorithms use a lower-dimensional representation such as PCA or ICA as a first step. Nearest neighbour graphs are constructed in the reduced dimensional space, following which Wanderlust uses shortest path distances and Monocle uses minimum spanning tree algorithms from a specified start cell to infer the one-dimensional pseudo-temporal ordering. Despite reduced noise in reduced dimensional space, the graphs still contain spurious edges, termed short circuits, that connect unrelated cells leading to loss of accuracy in estimating distances between distant cells. Wanderlust uses landmark points—cells sampled across the manifold—and Monocle uses P–Q graphs to improve the robustness. These algorithms pioneered the idea of trajectory detection from single-cell data, but they have not been extensively used to study cancer.

While the approaches used in Monocle and Wanderlust were powerful in relatively simple trajectories, the increasing complexity of data meant that the nearest neighbour graphs on their own were not sufficiently robust to define trajectories and infer branching or lineage decisions. The use of diffusion maps [16] to describe single-cell phenotypic spaces was a major catalyst in substantially increasing the robustness of trajectory detection algorithms. Diffusion maps can be viewed as a non-linear version of PCA and have proven to be highly effective for single-cell data after their introduction in the Wishbone [17] and diffusion pseudo-time (DPT) [18] algorithms. Diffusion maps also start with nearest neighbour graphs and use the adjacency matrix representation. The distances in the adjacency matrix are transformed into similarities using density-aware Gaussian kernels [19]. Following row normalization of the similarity matrix, each entry in the matrix (i, j) now represents the probability of cell in row i transitioning to column j in one step. It has been demonstrated that the Euclidean distance between any two cells in this representation is equivalent to measuring distances between considering all possible random walks through the graph [16], thus providing a highly reliable metric distance between individual cells. Further, the top eigenvectors of this matrix, termed diffusion maps, are sufficiently representative of the information in the full adjacency matrix. Both Wishbone and DPT leverage diffusion maps to provide robust pseudo-temporal orders and developed heuristics based on triangular inequalities to identify lineage decision-making points in single-cell datasets. Diffusion maps have had a profound impact on single-cell biology and have been extensively used to describe phenotypic process in cancers [12,20].

Graph abstraction is another widely approach to reduce the impact of noise over longer distances on a nearest neighbour graph. Two approaches PAGA [21] and Slingshot [22] use graph abstraction for inferring trajectories and lineage decisions. The first step in both approaches is to cluster the data to identify closely related groups of cells. PAGA than constructs an abstract graph connecting the clusters by enumerating the edges between cells that belong to different clusters. Slingshot uses the clusters as input to construct a minimum spanning tree to identify lineage decisions and then fits a principal curve through each lineage trajectory to infer pseudo-temporal ordering. Both approaches model the biological system as the classical tree or graph structure with discrete bifurcating steps to represent lineage decision-making events.

Single-cell datasets of increasing complexity challenged this classical bifurcating steps for lineage decision-making events, and observations across multiple biological systems led to the hypothesis that the lineage decision-making events are probabilistic in nature given the lack of discrete branch points where a cell fate was decided [23,24]. Palantir [24] was the first algorithm to model lineage decision-making processes as probabilistic events and presented an approach to model single-cell trajectories as a Markov chain. Palantir first used diffusion maps and shortest path distances from a specified start cell in the diffusion space to infer a pseudo-temporal ordering of cells. The pseudo-temporal order was used to derive a heuristic that transformed the undirected nearest neighbour graph to a directed graph such that for each cell the weight of edges going forward in pseudo-temporal order was greater than the backward edges, consistent with the unidirectional flow of trajectories. The directed graph was row normalized to infer a transition matrix or a Markov chain where the entries in the matrix represent transitions between cells representing states. Palantir uses the stationary distribution of the Markov chain to automatically determine the terminal states of the system. The Markov chain is then transformed into an absorbing Markov chain by removing all the outgoing edges from the terminal states, which means analytical solutions can be used to compute the probability of each state reaching the terminal states. Thus, the lineage decision-making process is described probabilistically rather than as discrete bifurcating steps. Further, the branch probabilities for each cell represent a multinomial distribution, and the uncertainty in this distribution can be estimated using entropy, which Palantir nominates to represent the differentiation potential of the cell. Palantir has since been used to successfully describe trajectories in healthy differentiation [24], embryonic development [25,26], immune cell trajectories in tumour microenvironments [11], and metastatic transformation from mouse models of tumour [12] and primary patient samples [27]. The probabilistic view of lineage decision-making has since been experimentally validated using CRISPR-based lineage tracing in a variety of biological systems [28,29]. Some of the other approaches to model probabilistic fate choices include FateID [30], VIA [31], and Grandprix [32].

12.2.2. Lineage dynamics from splicing dynamics

Computational approaches to infer RNA velocity have had a profound impact on inferring dynamics from single-cell data [33]. RNA velocity is a high-dimensional vector that represents the time derivative of gene expression state. scRNA-seq data represents static measurements of transcriptional states. Motivated by the fact that both spliced and unspliced transcripts can be detected from scRNA-seq data, it was proposed that the RNA velocity and hence dynamics of the biological system can be inferred by modelling the splicing dynamics of individual genes [33]. Specifically, the transcription of each gene is modelled as differential equations that describe the evolution of unspliced and spliced counts over time. The unspliced counts are modelled as a function of transcriptional and splicing rates, whereas the spliced counts are modelled as a function of splicing and degradation rates. With the assumption that the spectrum of splicing dynamics for each gene is represented across the cells along with assumptions of time-independent transcriptional and degradation rates and constant splicing rates, a complete solution to rate equation was derived to infer gene-specific transcription and degradation rates [33]. The inferred rates are used to predict the expected state of a cell given its current transcriptional state. The concatenation of these predictions across all genes represents

the RNA velocity vector of the cell. These vectors are projected onto uniform manifold approximation and projections (UMAPs) to visualize the dynamics of the system. The key advantage of RNA velocity approaches is that the amount of prior biological knowledge is minimal—specifically, unlike almost all previous approaches, the start and terminal states are no longer necessary parameters and thus can be utilized to *de novo* identify stem-like cell populations. While the initial efforts to infer RNA velocity made a number of simplifying assumptions and provided only a qualitative description of dynamics, improvements since have led to RNA velocity becoming one of the most powerful tools for inferring trajectories and lineage decision dynamics.

scVelo [34] directly addressed two key limitations of the initial approach to infer RNA velocity from single-cell data: (i) the full dynamics of individuals are captured in the scRNA-seq dataset and (ii) all genes have a shared and constant splicing rate. scVelo introduces a latent time variable that describes the progression of a cell by sharing information with all genes and solves the full gene-wise transcriptional dynamics inferring transcriptional, splicing, and degradation rates for each gene using an expectation–maximization procedure. The transcriptional dynamics are inferred given a latent time, and the latent time is inferred using the rates in the previous iteration. The full dynamical model presented in scVelo substantially improves the accuracy of the cell-state dynamics in a wide variety of biological systems spanning embryonic development [35], differentiation [34], and cancer [36]. The latent time inferred by scVelo represents progression equivalent to pseudo-time inferred by graph-based approaches, but the description of the lineage decision-making process is qualitative. As a solution, scVelo can be used in conjunction with PAGA to derive an abstracted view of lineage decisions.

While RNA velocity approaches leverage biological information to infer dynamics, graph-based approaches are extremely effective in providing a quantitative description of the lineage decision-making process. Inspired by Palantir and scVelo, CellRank [37] models the best of both worlds using RNA velocity estimates and graph-based manifold learning to infer pseudo-time orders of trajectories, lineage probabilities, and differentiation potential, while robustly and automatically inferring the start and terminal states of the system. Starting with the nearest neighbour graph, CellRank weights the edge of each cell as the cosine similarity between the RNA velocity of the cell and the transcriptional change necessary to transition to the neighbour. Therefore, if the neighbourhood transition is consistent with the RNA velocity, the edge weight will be high whereas contradictory directions will result in lower edge weights. This weighting scheme naturally directs the nearest neighbour graph to be consistent with the dynamics inferred using RNA velocity. CellRank then uses clustering approaches to automatically determine the terminal states. The edges are then reversed, and a similar procedure is utilized to determine the start states. The same graph is then transformed to a Markov chain to infer the pseudo-time, branch probabilities, and differentiation potential in a manner similar to Palantir. CellRank has been demonstrated to be highly effective in inferring dynamics of systems where the start of trajectories has not well studied such as regeneration [37]. CellRank has also been employed to infer stem-like or start populations in mouse models of tumour [38] and to model the dynamics of disease progression from single-cell profiles of primary tumours [39].

Despite the enormous impact, there are a number of limitations of RNA velocity estimations that preclude the approach from generalizing to all datasets—time-dependent transcriptional kinetics and lineage-specific dynamics are two of the factors where current approaches fail to accurately infer RNA velocities [40]. scVelo provide diagnostic tools for practitioners to identify whether RNA velocity estimates on their datasets are reliable. Efforts are underway to address these limitations through computational modelling approaches and through the use of additional data modalities such as single-cell epigenomic data.

12.2.3. Lineage decisions in time-series datasets

With the increasing cost-effectiveness of single-cell technologies, time-series experiments, where the system is studied across multiple time points, has become feasible. Time-series scRNA-seq datasets have generated to study a variety of biological systems including embryonic development [6,25,41], *in vitro* directed reprogramming [42], mouse models of tumour [43], and *in vitro* metastatic transformation of cancer cell lines [44]. Based on the resolution of the measured time points, the differences between cells is often a mix of technical batch effects and true biological signal. As a result, the application of typically used batch correction techniques can remove true biological signal [25]. Purpose-built tools such as Harmony-TS [25] can be utilized to connect single-cell disconnected time-series datasets. Harmony-TS augments the nearest neighbour graph with mutually nearest neighbours between cells of successive time points to harmonize time points. The harmonized representation serves as input to graph-based trajectory detection algorithms.

Waddington-OT is a computational algorithm designed to model trajectories using time-series datasets [42]. Waddington-OT uses principles from optimal transport to map cells from each time point to their successive time points. Conversation of mass between time points is a key requirement for the correct application of optimal transport, and Waddington-OT presents extensive work to account for the growth and apoptosis of cells within each time point. Waddington-OT can be used to infer pseudo-temporal trajectories and probabilities quantifying lineage decisions from single-cell time-series data. Waddington-OT has been successfully used to dynamics of *in vitro* reprogramming [42], regeneration [45], and emergence of high-plasticity cell states in mouse tumour models [46]. Recently, optimal transport has also been used to infer network flow models to chart embryonic development collected at successive developmental stages [47].

12.2.4. Multi-modal single-cell datasets

Technologies that can profile multiple modalities per cell simultaneously for thousands of cells are rapidly emerging. These include CITE-seq to measure transcriptomes and proteins [48], paired RNA–ATAC to measure transcriptome and open chromatin [49], CUT&TAG-pro to measure transcriptome and histone modifications [50], and MulTI-TAG [51] to measure multiple histone modifications per cell among others. A number of computational approaches have been developed to learn a joint representation of cells using different modalities such as weighted nearest neighbour graph [52] and MOFA+ [53]. These approaches can better describe the phenotypic space of cells by leveraging both modalities and have demonstrated the differential effectiveness of defining cell states in

diverse contexts. The joint representation from multi-modal data can be used as input for trajectory detection algorithms.

MIRA is a recent algorithm developed for inferring trajectories and regulatory drivers of lineage decisions using paired RNA and ATAC data [54]. MIRA uses topic models to infer a low-dimensional representation for RNA and ATAC modalities separately. The two representations are then concatenated and serve as input for the computation of diffusion maps and trajectory detection using Palantir. In addition to leveraging multiple modalities, these approaches also provide an opportunity to explore the differential impact of modalities on lineage decisions. Modelling lineage decisions using multi-modal datasets is an active area of research with several efforts underway to build mechanistic models of the differentiation process and the dysregulation of such mechanisms in disease.

In addition to the above outlined approaches, trajectory detection algorithms have also been developed using hidden Markov models [55], Bayesian latent variable models [56], and attractor state models [57]. Saelens et al. [58] have developed a framework to comprehensively benchmark different trajectory detection algorithms by devising a number of metrics and biological scenarios. Each algorithm is developed with their own set of assumptions and nuances in which they can be applied to. A careful reading of the assumptions laid out by the algorithm is strongly recommended for practitioners to ensure accurate interpretation of the signal in the data.

12.3. Visualization and downstream applications

UMAPs are the most widely single-cell visualization technique due to their broad applicability to single-cell datasets [59,60]. Results of trajectory detection, such as pseudo-temporal ordering, dynamics from RNA velocity, branch probabilities, and differentiation potential, are visualized using UMAPs. Force-directed layouts have emerged as a powerful alternative to UMAPs for visualizing datasets with trajectories and lineage decisions [61]. Force-directed layouts better preserve the geometry of the data and thus can better highlight the low-frequency transitory populations that are often a characteristic of single-cell data. Tools such as STREAM provide implementations for visualizing linear, branching trajectories and cell densities along them [62]. Results of trajectory detection algorithms that generate lineage probabilities can also be visualized using circular projection plots to summarize the lineage decision-making process [63].

Visualizing of gene expression trends is a powerful technique to interpret the results of trajectories and develop hypothesis (**Figure 12.2**). Generalized additive models provide robust and efficient estimate of gene expression dynamics along pseudo-temporal orders by fitting cubic splines [24,64]. Tools support the visualization of individual genes along multiple lineages simultaneously, which can be effective for understanding key players in lineage decisions or comparative visualization of multiple genes for a particular lineage. Dynamics of multiple genes along a lineage can also be visualized using heat maps to generate a comprehensive view of up- and down-regulation of genes as they shape trajectories [24]. Further, gene expression trends can be clustered using tools, such as Leiden or PhenoGraph, to identify coregulated genes with similar dynamic patterns. Clustering of gene trends are particularly effective for identifying the successive waves of gene expression dynamics that shape lineage decisions. Tools, such as Gene Set Enrichment Analysis [65] or Gene Ontology analysis [66], can be applied for each cluster of gene to infer biological meaning and interpretation.

Tools, such as switchDE [67] and tradeSeq [68], have been developed to identify differential expressed genes and genes with differential dynamics between different lineages. These tools can be effective in identifying the subset of genes that correlate most with the lineage decision-making process. Approaches have also been developed to align trajectories with application to *in vivo–in vitro* trajectories or trajectories between related cells of different species [69].

In the case of trajectories using multi-modal datasets, a comparison of dynamics of the same set of genes across different modalities has led to critical insights into both differentiation and disease. For example, a comparison of gene expression dynamics and dynamics of gene scores derived using ATAC modality has demonstrated the widespread use of open-chromatin priming during lineage decisions [49]. While these approaches are still in their infancy, the sparsity of scATAC-seq data presents a major challenge to model the dynamics of enhancers as they drive lineage decisions. Efforts to summarize single-cell ATAC-seq data into tightly related groups of cells called metacells hold promise of inferring and modelling open-chromatin dynamics as they drive lineage decisions [70].

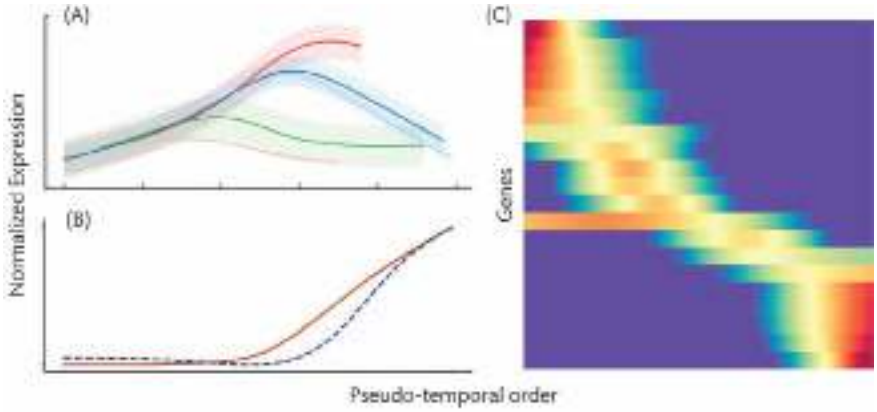


Figure 12.2. Visualization of gene expression trends. (A) Expression dynamics of a given gene can be simultaneously visualized for multiple lineages. (B) Dynamics of multiple genes can be visualized for a lineage. (C) Similar to (B), visualized as a heatmap.

12.4. Lineage plasticity in cancer

Lineage plasticity has emerged as one of the predominant drives of cancer progression and transformation [14]. Terminally differentiated cells gain plasticity either by trans-differentiating different lineages or de-dedifferentiate to acquire stem-like properties as a result of somatic alterations and other factors. Single-cell studies have demonstrated the role of lineage plasticity where cancer transcriptome profiles show patterns related to multiple differentiated cell types or change in the dominant cell-type signal in lung cancer [46,71,72], breast cancer [73], prostate cancer [12], and glioblastoma [74]. Snapshot single-cell profiles of patient samples do not robustly capture the dynamics of emergence and progression of the plastic populations. Therefore, trajectory detection algorithms should only be applied after a careful consideration of the algorithm assumptions and behaviour of the biological system under consideration. Time-series measurements of genetically engineered mouse models, organoid models, and even profiles of the same patient over multiple time points are more amenable to the investigation of cancer lineage decisions through trajectory detection algorithms.

12.5. Conclusions

The study of lineage decisions using trajectory detection algorithms has been one of the lasting impacts of advances in single-cell measurements technologies. While enormous progress has been made in characterizing lineage decision-making in development and differentiation, tools to characterize lineage decisions and plasticity in cancer are under active development. Technological advances, such as lineage tracing [75] and multi-modal measurements, are poised to provide deep insights into lineage plasticity in cancer and are likely to eventually lead to tools and metrics to characterize plasticity from snapshot single-cell measurements of cancer.

REFERENCES

1. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;**32**(4):381–6.
2. Bendall SC, Davis KL, Amir el AD, Tadmor MD, Simonds EF, Chen TJ, et al. Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell. 2014;**157**(3):714–25.
3. Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. Nat Protoc. 2018;**13**(4):599–604.
4. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;**161**(5):1187–201.
5. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015;**161**(5):1202–14.
6. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature. 2019;**566**(7745):496–502.
7. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The human cell atlas. Elife. 2017;**6**.
8. Rozenblatt-Rosen O, Regev A, Oberdoerffer P, Nawy T, Hupalowska A, Rood JE, et al. The human tumor atlas network: charting tumor transitions across space and time at single-cell resolution. Cell. 2020;**181**(2):236–49.
9. Tabula Muris C. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. Nature. 2020;**583**(7817):590–5.
10. Deconinck L, Cannoodt R, Saelens W, Deplancke B, Saeys Y. Recent advances in trajectory inference from single-cell omics data. Curr Opin Syst Biol. 2021;**27**:1000344.
11. Krishna C, DiNatale RG, Kuo F, Srivastava RM, Vuong L, Chowell D, et al. Single-cell sequencing links multiregional immune landscapes and tissue-resident T cells in ccRCC to tumor topology and therapy efficacy. Cancer Cell. 2021;**39**(5):662–77 e6.
12. Chan JM, Zaidi S, Love JR, Zhao JL, Setty M, Wadosky KM, et al. Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. Science. 2022;**377**(6611):1180–91.
13. Chen J, Wang X, Ma A, Wang QE, Liu B, Li L, et al. Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. Nat Commun. 2022;**13**(1):6494.
14. Le Magnen C, Shen MM, Abate-Shen C. Lineage plasticity in cancer progression and treatment. Annu Rev Cancer Biol. 2018;**2**:271–89.
15. Kim TH, Zhou X, Chen M. Demystifying "drop-outs" in single-cell UMI data. Genome Biol. 2020;**21**(1):196.
16. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F, et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. Proc Natl Acad Sci U S A. 2005;**102**(21):7426–31.
17. Setty M, Tadmor MD, Reich-Zeliger S, Angel O, Salame TM, Kathail P, et al. Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat Biotechnol. 2016;**34**(6):637–45.
18. Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics. 2015;**31**(18):2989–98.
19. van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, et al. Recovering gene interactions from single-cell data using data diffusion. Cell. 2018;**174**(3):716–29 e27.
20. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. Cell. 2018;**174**(5):1293–308 e36.
21. Wolf FA, Hamey FK, Plass M, Solana J, Dahlin JS, Gottgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol. 2019;**20**(1):59.
22. Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics. 2018;**19**(1):477.
23. Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM. Fundamental limits on dynamic inference from single-cell snapshots. Proc Natl Acad Sci U S A. 2018;**115**(10):E2467–76.
24. Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D. Characterization of cell fate probabilities in single-cell data with Palantir. Nat Biotechnol. 2019;**37**(4):451–60.
25. Nowotschin S, Setty M, Kuo YY, Liu V, Garg V, Sharma R, et al. The emergent landscape of the mouse gut endoderm at single-cell resolution. Nature. 2019;**569**(7756):361–7.

26. Lotto J, Drissler S, Cullum R, Wei W, Setty M, Bell EM, et al. Single-cell transcriptomics reveals early emergence of liver parenchymal and non-parenchymal cell lineages. Cell. 2020;**183**(3):702–16 e14.
27. Dong R, Yang R, Zhan Y, Lai HD, Ye CJ, Yao XY, et al. Single-cell characterization of malignant phenotypes and developmental trajectories of adrenal neuroblastoma. Cancer Cell. 2020;**38**(5):716–33 e6.
28. Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM. Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science. 2020;**367**(6479).
29. Biddy BA, Kong W, Kamimoto K, Guo C, Waye SE, Sun T, et al. Single-cell mapping of lineage and identity in direct reprogramming. Nature. 2018;**564**(7735):219–24.
30. Herman JS, Sagar, Grun D. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. Nat Methods. 2018;**15**(5):379–86.
31. Stassen SV, Yip GGK, Wong KKY, Ho JWK, Tsia KK. Generalized and scalable trajectory inference in single-cell omics data with VIA. Nat Commun. 2021;**12**(1):5528.
32. Ahmed S, Rattray M, Boukouvalas A. GrandPrix: scaling up the Bayesian GPLVM for single-cell data. Bioinformatics. 2019;**35**(1):47–54.
33. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;**560**(7719):494–8.
34. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol. 2020;**38**(12):1408–14.
35. Amadei G, Handford CE, Qiu C, De Jonghe J, Greenfeld H, Tran M, et al. Embryo model completes gastrulation to neurulation and organogenesis. Nature. 2022;**610**(7930):143–53.
36. Connolly KA, Kuchroo M, Venkat A, Khatun A, Wang J, William I, et al. A reservoir of stem-like CD8(+) T cells in the tumor-draining lymph node preserves the ongoing antitumor immune response. Sci Immunol. 2021;**6**(64):eabg7836.
37. Lange M, Bergen V, Klein M, Setty M, Reuter B, Bakhti M, et al. CellRank for directed single-cell fate mapping. Nat Methods. 2022;**19**(2):159–70.
38. Burdziak C, Alonso-Curbelo D, Walle T, Barriga FM, Reyes J, Xie Y, et al. Epigenetic plasticity cooperates with emergent cell-cell interactions to drive neoplastic tissue remodeling in the pancreas. bioRxiv. 2022:2022.07.26.501417.
39. Guerrero-Juarez CF, Lee GH, Liu Y, Wang S, Karikomi M, Sha Y, et al. Single-cell analysis of human basal cell carcinoma reveals novel regulators of tumor growth and the tumor microenvironment. Sci Adv. 2022;**8**(23):eabm7981.
40. Bergen V, Soldatov RA, Kharchenko PV, Theis FJ. RNA velocity-current challenges and future perspectives. Mol Syst Biol. 2021;**17**(8):e10282.
41. Pijuan-Sala B, Griffiths JA, Guibentif C, Hiscock TW, Jawaid W, Calero-Nieto FJ, et al. A single-cell molecular map of mouse gastrulation and early organogenesis. Nature. 2019;**566**(7745):490–5.
42. Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. Cell. 2019;**176**(4):928–43 e22.
43. Yang D, Jones MG, Naranjo S, Rideout WM, 3rd, Min KHJ, Ho R, et al. Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. Cell. 2022;**185**(11):1905–23 e25.
44. Hodis E, Torlai Triglia E, Kwon JYH, Biancalani T, Zakka LR, Parkar S, et al. Stepwise-edited, human melanoma models reveal mutations' effect on tumor and microenvironment. Science. 2022;**376**(6592):eabi8175.
45. Strunz M, Simon LM, Ansari M, Kathiriya JJ, Angelidis I, Mayr CH, et al. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. Nat Commun. 2020;**11**(1):3559.
46. Marjanovic ND, Hofree M, Chan JE, Canner D, Wu K, Trakala M, et al. Emergence of a high-plasticity cell state during lung cancer evolution. Cancer Cell. 2020;**38**(2):229–46 e13.
47. Mittnenzweig M, Mayshar Y, Cheng S, Ben-Yair R, Hadas R, Rais Y, et al. A single-embryo, single-cell time-resolved model for mouse gastrulation. Cell. 2021;**184**(11):2825–42 e22.
48. Mimitou EP, Cheng A, Montalbano A, Hao S, Stoeckius M, Legut M, et al. Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. Nat Methods. 2019;**16**(5):409–12.
49. Ma S, Zhang B, LaFave LM, Earl AS, Chiang Z, Hu Y, et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell. 2020;**183**(4):1103–16 e20.
50. Zhu C, Zhang Y, Li YE, Lucero J, Behrens MM, Ren B. Joint profiling of histone modifications and transcriptome in single cells from mouse brain. Nat Methods. 2021;**18**(3):283–92.
51. Meers MP, Llagas G, Janssens DH, Codomo CA, Henikoff S. Multifactorial profiling of epigenetic landscapes at single-cell resolution using MulTI-Tag. Nat Biotechnol. 2022;41:708–16.
52. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;**184**(13):3573–87 e29.
53. Argelaguet R, Arnol D, Bredikhin D, Deloro Y, Velten B, Marioni JC, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. Genome Biol. 2020;**21**(1):111.
54. Lynch AW, Theodoris CV, Long HW, Brown M, Liu XS, Meyer CA. MIRA: joint regulatory modeling of multimodal expression and chromatin accessibility in single cells. Nat Methods. 2022;**19**(9):1097–108.
55. Lin C, Bar-Joseph Z. Continuous-state HMMs for modeling time-series single-cell RNA-Seq data. Bioinformatics. 2019;**35**(22):4707–15.
56. Campbell KR, Yau C. A descriptive marker gene approach to single-cell pseudotime inference. Bioinformatics. 2019;**35**(1):28–35.
57. Zhou P, Wang S, Li T, Nie Q. Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics. Nat Commun. 2021;**12**(1):5609.
58. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. Nat Biotechnol. 2019;**37**(5):547–54.
59. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018;**37**:38–44.
60. McInnes L, Healy J, Melville J. UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv e-prints. 2018.
61. Weinreb C, Wolock S, Klein AM. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. Bioinformatics. 2018;**34**(7):1246–8.
62. Chen S, Lake BB, Zhang K. High-throughput sequencing of the transcriptome and chromatin accessibility in the same cell. Nat Biotechnol. 2019;**37**(12):1452–7.

63. Velten L, Haas SF, Raffel S, Blaszkiewicz S, Islam S, Hennig BP, et al. Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol. 2017;**19**(4):271–81.
64. Hastie T, Tibshirani R. Generalized additive-models—some applications. J Am Stat Assoc. 1987;**82**(398):371–86.
65. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;**102**(43):15545–50.
66. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;**25**(1):25–9.
67. Campbell KR, Yau C. switchDE: inference of switch-like differential expression along single-cell trajectories. Bioinformatics. 2017;**33**(8):1241–2.
68. Van den Berge K, Roux de Bezieux H, Street K, Saelens W, Cannoodt R, Saeys Y, et al. Trajectory-based differential expression analysis for single-cell sequencing data. Nat Commun. 2020;**11**(1):1201.
69. Kanton S, Boyle MJ, He Z, Santel M, Weigert A, Sanchis-Calleja F, et al. Organoid single-cell genomic atlas uncovers human-specific features of brain development. Nature. 2019;**574**(7778):418–22.
70. Persad S, Choo Z-N, Dien C, Masilionis I, Chaligné R, Nawy T, et al. SEACells: inference of transcriptional and epigenomic cellular states from single-cell genomics data. bioRxiv. 2022:2022.04.02.486748.
71. Chan JM, Quintanal-Villalonga A, Gao VR, Xie Y, Allaj V, Chaudhary O, et al. Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. Cancer Cell. 2021;**39**(11):1479–96 e18.
72. LaFave LM, Kartha VK, Ma S, Meli K, Del Priore I, Lareau C, et al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. Cancer Cell. 2020;**38**(2):212–28 e13.
73. Rios AC, Capaldo BD, Vaillant F, Pal B, van Ineveld R, Dawson CA, et al. Intraclonal plasticity in mammary tumors revealed through large-scale single-cell resolution 3D imaging. Cancer Cell. 2019;**35**(4):618-32 e6.
74. Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ, et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. Cell. 2019;**178**(4):835–49 e21.
75. Wagner DE, Klein AM. Lineage tracing meets single-cell omics: opportunities and challenges. Nat Rev Genet. 2020;**21**(7):410–27.

13

Analyzing cancer cell-state transition dynamics through live-cell imaging and high-dimensional single-cell trajectory analyses

Jianhua Xing and Weikang Wang

13.1. Hierarchy structure and plasticity of cancer cell population

Heterogeneity of tumour cell is one of the major obstacles in clinical cancer treatment. Increasing accumulative pieces of evidence have revealed that there is subpopulation of cancer stem-like cells (CSCs) in tumour that accounts for recurrence after surgery, chemotherapy, and radiotherapy [1]. CSCs share similarity with normal stem cells, having the ability of self-renewal and differentiation into non-stem cancer cells (NSCCs). Comparing with other cancer cells, CSCs are more resistant to therapy and more aggressive. The concept of CSCs revises our knowledge on cancer cell populations. Previously, researchers used stochastic models or clonal evolution models to describe the dynamics of cancer cell population under treatment. These models assume cancer cells as genetic and phenotypic homogeneous, while intrinsic or extrinsic factors influence different cancer cell dynamics randomly. In the CSC model, cancer cell population are hierarchically organized and CSCs are the drivers of tumour growth and progression. Furthermore, it is reported that CSCs have higher mobility and ability of metastasis [2,3]. The key role of CSCs in tumour development indicates that a complete cure of cancer requires elimination of the CSCs, leading to new strategies suggested in cancer therapy [4,5].

However, several subsequent studies challenged the mechanism of unidirectional conversion from CSCs to NSCCs. Instead, both experiment and theoretical model analyses demonstrate existence of a dynamic equilibrium between CSCs and NSCCs, and the transitions are bidirectional [6,7]. This bidirectional inter-conversions between NSCCs and CSCs, and more generally the phenotypic plasticity of cancer cells, cast doubts on the efficacy of focusing on eradicating CSCs alone (**Figure 13.1**A).

One of the major mechanisms of cell phenotypic plasticity is the epithelial–mesenchymal transition (EMT). EMT is a process that epithelial cells acquire mesenchymal characteristics while losing epithelial features. This process is associated with the invasive properties of CSCs [8]. Through EMT, transcriptional factors including Snail, Slug, Twist, Zeb1, and Zeb2 are activated, marker proteins like E-cadherin are down-regulated and N-cadherin and vimentin are up-regulated. EMT can be induced by a plethora of signals such as TGF-β and Wnt [9–12]. During EMT, the cell morphology switches from a cobbler-stone like shape to a spindle-like shape, accompanied with decreased expression of E-cadherin and increased expression of N-cadherin that reduce the strength of attachment between neighbouring cells as well as that between cells and extra-cellular matrix (ECM) [10,13], which endow the cell with higher motility (**Figure 13.1**B).

EMT is not a simple binary process but a complex transition process that goes through different intermediate states [4,12,14]. The intermediate states, i.e. partial EMT (p-EMT) state, also play important roles in cancer migration and metastasis (**Figure 13.1**B). It is reported that cells that locate on the leading edge of primary tumours show characteristics of p-EMT [15]. The existence of intermediate states implicates a complex spectrum of EMT [9,10].

13.2. Mathematical formulation of cell-state transition dynamics

Mathematically, one uses an array of variables, denoted as $\mathbf{X}$, to specify cell states. Then, different cell types occupy distinct regions in a multi-dimensional phase space defined by $\mathbf{X}$, and cell-state transitions including EMT follow continuous paths that connect the two regions (**Figure 13.1**C) [16].

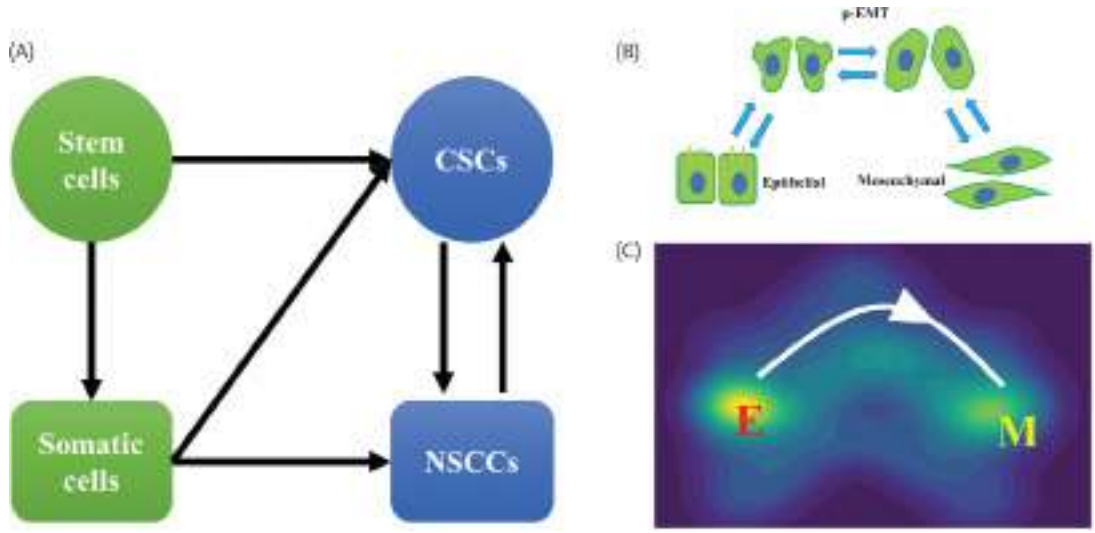


Figure 13.1. Transition between cancer cell state is closely associated with EMT. (A) The bidirectional inter-conversion between CSCs and NSCCs. (B) Illustration of EMT with intermediate partial EMT states. During EMT, cells dissociate with each other and cell morphology varies significantly. (C) Continuum transition path between epithelial and mesenchymal states.

Generically, the single-cell dynamics can be described with a group of Langevin equation as follows:

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(\mathbf{X}, \lambda) + \xi,$$

where $\mathbf{X}$ is the above-defined multi-dimensional state vector of a cell, $\mathbf{F}$ is in general a non-linear vector function representing the interactions or regulation between different components, and ξ is a Gaussian noise term [17]. The control parameters λ specifies extracellular environmental factors or some hidden variables not explicitly specified by $\mathbf{X}$. That is, the underlying regulation relationship between genes may vary when the control parameters change.

The existence of the hidden variables may complicate analyses of systems dynamics from the observed data. Consider the potential system in Figure 13.2A as an example. This is a bistable system coupled to a hidden process. The hidden process represents a control parameter like epigenetic modification for a cellular system, which here we assume that it has a dynamics much slower than the mean transition time between the two potential wells, i.e. the time between observing two consecutive transition events for a population of particles originally residing in the left well and jumping to the right well. Then, we can assume that the value of the hidden slow variable does not change during the transition process. If one can trace individual transition trajectories over time, these trajectories show stepping dynamics characteristic for the bistable system, while the transition positions and the basins vary among different trajectories (Figure 13.2B). On the other hand, when only snapshot data of a population of particles, i.e. the state distribution functions at various time points, is available, information about the temporal correlation of individual cell trajectories is missing, and one cannot deduce the underlying two-state dynamics [18].

13.3. Exploration of cancer state transition with -omic methods

Representing single-cell state $\mathbf{X}$ is one of the key tasks in studying cancer cell-state transitions. In practice, the classification of CSCs and NSCCs has been based on the certain combinations of surface markers, such as CD24, CD44, and CD166 [1]. Similarly, most experimental studies on EMT use a small number of TFs and proteins to classify different EMT phenotypes. However, cell-state transition usually involves global reprogramming of the profiles at the transcriptomic, proteomic, and epigenomic levels. Using a small set of markers sometime leads to inconclusive or spurious cell-state identification. In year 2015, two lineage-tracing studies shook the EMT field by claiming that EMT does not contribute to tumour metastasis [19,20]. However, subsequent studies demonstrate the role of p-EMT on tumour invasion and metastasis, which these two studies missed since they drew their conclusions based on markers for later stage EMT [21]. This debate in the field is an example on the importance of specifying cell phenotypes unambiguously, for which a small number of markers may not be sufficient.

In recent years, the development of single-cell RNA-sequencing techniques has made it possible to study the dynamics of cell phenotype transition in a genome-wide gene expression space. It is now widely used in studying different cell phenotype transition processes such as development and reprogramming. Since the single-cell RNA-sequencing (scRNA-seq) data only provides snapshot distributions of single-cell transcriptomic state, various computational methods have been developed to infer the dynamics from the data. One of the main methods is the pseudo-time method (Figure 13.2C) [22–25]. With this method, each cell is assigned a pseudo-time value based on the similarity of its expression to the predetermined initial and final states, so the cells are time-ordered in the transition processes under study [22,24,26–32]. The pseudo-time methods have been used to infer the transition trajectories or paths in a cell phenotypic transition process and study cancer cell heterogeneity and plasticity [33–38].

A number of studies have been reported on using high-throughput single-cell genomics approaches to analyse the EMT process. Pastushenko et al. used scRNA-seq to compare the expression profile of epithelial Epcam+ and mesenchymal-like Epcam– tumour cells. They identified the intermediate transition states during EMT *in vivo* [39]. Cook and co-worker performed scRNA-seq on

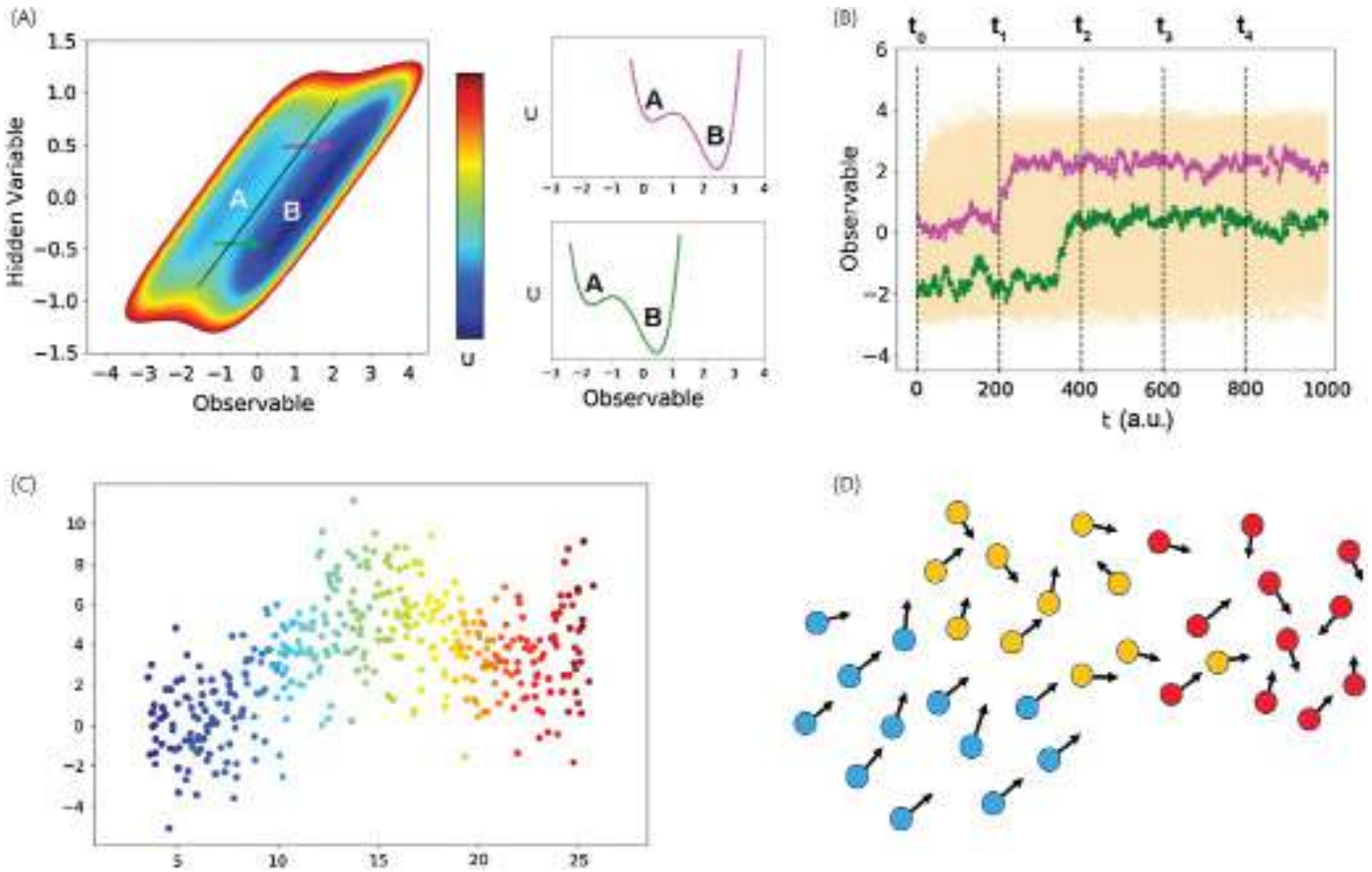


Figure 13.2. Studying cancer cell-state transition dynamics with snapshot data. (A) Example bistable system coupled with a hidden slow variable. Reproduced with permission from [18]. (B) Single-cell trajectory can reveal the dynamics missed by distribution data. Reproduced with permission from [18]. (C) Illustration of pseudo-time analysis on single-cell RNA-sequencing data. Each dot represents a single cell, and colour represents the inferred pseudo-time order value (from blue to red). Axes represent embedding coordinates. (D) Illustration of RNA velocity analysis on single-cell RNA-sequencing data. Each dot represents a single cell (colour represents cell type). Arrow represents cell velocity.

different cancer cell lines treated with TGF-β1, EGF, and TNF for different durations. They found that the EMT dynamics of one cell line under different stimuli was more consistent than that of different cell lines treated with the same stimulus. More importantly, the global expression profiles of cells undergoing EMT are context specific [34], indicating that cells respond to a stimulus with activation of the EMT program and a number of other cellular programs. McFaline-Figueroa et al. compared the scRNA-seq profile of inner and outer cell of colonies of MCF10A cells and found that EMT can take place spontaneously due to spatial location even in the absence of TGF-β [35]. Karacosta et al. analysed multi-dimensional proteomic data with TGF-β-treated lung cancer cell lines using an analysis beyond the pseudo-time method. They divided the cell-state space into discrete regions and modelled transitions among these states as Markovian processes. They found that EMT and the reverse mesenchymal-to-epithelial transition (MET) follow different paths in the expression space of marker genes [40], as expected for a system out of thermodynamic equilibrium. All these studies advanced our understanding on EMT and MET, especially the molecular signatures associated with the transitions.

However, analyses based on single-cell state distributions alone have fundamental limits of inferring dynamics from fixed-cell data [18,30]. Different dynamics can give rise to the same stationary distribution. For instance, oscillations in cell dynamics do not alter the stationary distributions in phase space [30], which cannot be inferred with the pseudo-time method. Mathematically, the governing equation of a stochastic dynamical system can be decomposed into a curl-free part and a divergent-free part, and the stationary distribution is only determined by the curl-free part [41]. While time-series data of the nonstationary distributions can provide additional information, existence of the slow hidden variables discussed above further contributes to heterogeneous single-cell dynamics that complicates the choice of modelling frameworks for analysing the data. To address this limitation of snapshot data, one direction is to include additional dynamical information derived from the data, specifically RNA velocities derived from scRNA-seq data (**Figure 13.2**D) [42]. Qiu et al. developed a computational procedure to learn cellular dynamics from single-cell expression and RNA velocities data [43]. Another direction is to measure single-cell dynamics directly through live-cell imaging, as discussed in the following section.

13.4. Studying cancer cell-state transition with live-cell imaging

With live-cell imaging, one can directly trace single-cell dynamics over time. There are two major technical challenges for extracting information from imaging data. One is single-cell segmentation.

The development of machine learning, especially deep neural networks, makes it possible to perform segmentation on a large batch of imaging data [44–47]. Another one is the limitation of acquiring high-dimensional features from images. While a number of algorithms in image processing enable people to extract information from fixed-cell images [48,49], live-cell imaging has its unique constraints. Traditional live-cell imaging is limited by the maximum number of channels (which is usually five) available for a fluorescence microscope. Photo-toxicity brought by fluorescence microscopy also constrains the duration and frequency of live-cell imaging, especially when performing multiple channels of fluorescence imaging. Fluorescence labelling also affects cell physiology and cellular responses. These limits significantly reduce the application of time lapse imaging on studying single-cell dynamics [50].

Large cell morphological changes typically accompany cell phenotype conversions like EMT [51]. Morphology features have already been widely used in drug selection [52,53] and cell phenotyping, such as quantifying the progress of EMT [54–56]. Wu et al demonstrated that the morphology state of single cell can be used to predict the tumorigenic and metastatic potentials in mouse breast cancer [57]. However, these analyses are based on fixed-cell imaging data and suffer the same limitation as discussed above.

Recently, collective morphology features emerged as a choice alternative to fluorescence-based measurement of gene expression levels in live-cell imaging. Morphology features provide a convenient coordinate system for describing cell phenotype transition. Moreover, it is easy to obtain high-dimension morphological features through label-free imaging that is minimally intrusive and imposes minimal damage to cells. Advance of the field is further catalyzed by recent development of computational approaches of high-quality single-cell segmentation of images for extracting morphology information. Traditional segmentation methods such as watershed and active contour are sensitive to slight variation of imaging condition and are only applicable to simple images. Also, it is time-consuming to segment single cells manually. Developments of machine learning, especially deep convolution neural networks, have improved single-cell segmentation quality and efficiency significantly [44,45,47].

Several studies have attempted to characterize single-cell morphodynamics through live-cell imaging. Gordonov et al. demonstrated that analysis on time series of live single-cell morphology with the hidden Markov model can reveal the heterogeneous dynamics that cannot be captured by fixed-cell imaging, and the morphology dynamics can improve the drug classification or selection accuracy [58]. This result is consistent with the analysis of the example system shown in **Figure 13.2**A, i.e. only time-series modelling can capture the underlying heterogeneous dynamics controlled by hidden variables. Chang et al. explored transition dynamics of mouse embryonic fibroblasts in cytomorphological state space. Through mapping the effective landscape of cell population in this morphology state space, they found the uneven occupation of isogenic cells in the state space, and cell behaviour can be predicted by the reconstructed effective energy landscape [59].

Wang et al. further developed a platform for integrative analyses of live-cell imaging data in the formalism of dynamical systems and applied it on single-cell phenotype transition [18]. They first develop a single-cell segmentation method based on deep convolutional neural networks which combines with the traditional watershed method [47]. This method provides improved single-cell segmentation accuracy comparing with other methods, which is critical for the single-cell tracking. The platform utilized the active shape model to represent single-cell shape [60]. Some ad hoc shape features, such as area and major axis length, were widely used in different research studies. A problem is that the amount of information extracted from the original images with these features cannot be controlled systematically, and it is difficult to reconstruct the original shape with them. Pincus and co-worker compared different shape-representing methods based on three criteria including faithfulness of recording the original morphology, ability of capturing major biological variations, and interpretability of representation [60]. Comparing with other methods like Fourier and Zernike, active shape model together with principal component analysis performs better in processing different types of data [60]. With the active shape model, single-cell morphology is represented as a state vector with hundreds of dimensions.

Wang et al. then applied this platform on TGF-β-induced EMT of A549 lung cancer cells with endogenous vimentin, an intermediate filament protein often used as a mesenchymal marker [18], fused with red fluorescent proteins (**Figure 13.3**A). Cell morphological features were acquired through the differential interference contrast channel, which does not require an additional fluorescence labelling. Therefore, only one fluorescent channel was used for imaging vimentins, which reduced the photo-toxicity to cells and made it possible for long-term live imaging. Through representing single-cell state in the multi-dimensional composite feature space including morphology and vimentin texture features, they acquired single-cell trajectories from live-cell imaging data (**Figure 13.3**B). In the cell morphology state space quantified by the active shape model, the variation mode along the coordinate of the first principal component captures the major variation of cell morphology during EMT that cells stretch into spindle-like shape characteristics of mesenchymal cells. Noting that the subcellular organization of organelle or proteins reflects cell state and cell type [48], they quantified the vimentin texture features with Haralick features [61]. Haralick features are widely used for image profiling in drug screening, phenotype discovery, and classification. They found that the distributions of some Haralick features show significant shifts during EMT.

It has been noted that cell-state transition and chemical reaction share a lot of similarity [62]. In chemical reaction, the term 'transition state' refers to a short-lived intermediate or the bottleneck of the transition dynamics. In a cell-state transition process, intermediate states such as p-EMT states also exist. Wang et al. showed that tools and concepts developed in chemical rate theories, such as transition path theories, can be utilized to analyse single live-cell trajectories in cell-state transition, as discussed below.

After TGF-β treatment for two days, the cell population relaxes from an initial stationary distribution corresponding to the epithelial cell type into a new one in the composite feature space. To understand the dynamics of EMT quantitively, one needs to define the epithelial and mesenchymal states explicitly in the cell-state space. They used a combination of the Gaussian mixture model and label spreading function to define the epithelial state, mesenchymal state, and the intermediate region in the cell-state space. With this definition, one can classify the recorded single-cell trajectories as an ensemble of reactive trajectories that started from the epithelial region and ended in the mesenchymal region with the experimental duration, and as an ensemble of nonreactive trajectories that did not undergo or finish EMT to reach the mesenchymal region.

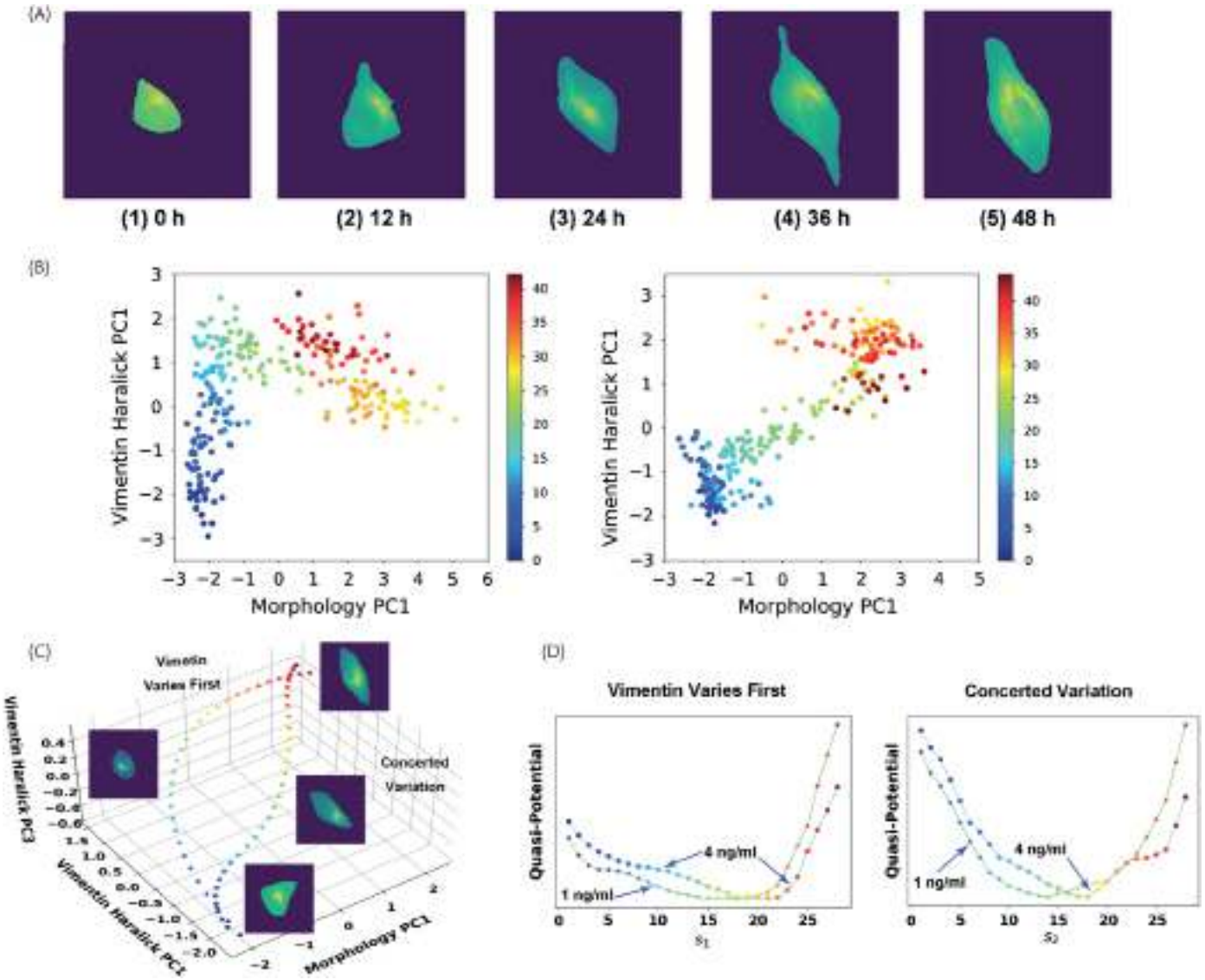


Figure 13.3. Studying cancer cell-state transition dynamics through live-cell imaging. (A) Selected images of an EMT single-cell trajectory at different time points. Reproduced with permission from [18]. (B) Different types of single-cell trajectories in the composite feature space of morphology and vimentin texture feature. Colour represents time. Reproduced with permission from [18]. (C) Reaction coordinates of different transition paths in EMT of A549 lung cancer cell induced by TGF-β. (D) Reconstructed quasi-potential along the reaction coordinates of different paths. Reproduced with permission from [66].

Next, they focused on the ensemble of reactive trajectories undergoing EMT for transition path analyses. Using the dynamics time warping (DTW), they quantified the similarity between different EMT trajectories. In general, DTW is a method that calculates an optimal match between two given sequences. Comparing with an Euclidean distance, DTW can capture the key dynamic feature similarities of two trajectories while neglecting their differences in length and speed. They found that EMT trajectories can be clustered into two categories. One is the vimentin varies first, and the other is morphology and vimentin vary concertedly. The results indicate that EMT can proceed through two parallel paths.

Reaction coordinate (RC) is a one-dimensional (1D) collective variable describing the progression along a continuous reaction path defined in the state space [63–65]. In chemical reaction theories, one way to define the RC is based on sampled transition path ensemble according to the transition path theory. Similarly, Wang et al. generalized an algorithm originally developed in chemical rate theories to reconstruct the two RCs from the measured ensemble of reactive trajectories. With RCs, they separated the state space into different Voronoi cells that can quantify the progression of a single-cell trajectory (Figure 13.3C) [66]. By confining the analysing dynamics of single cell along each RC, the EMT process becomes a 1D convection–diffusion process as

$$\frac{ds}{dt} = F(s) + \xi.$$

For a 1D system even without detailed balance, one can define a quasi-scalar potential φ. The time derivative of s is determined by the quasi-scalar potential,

$$\frac{ds}{dt} = -\frac{d\varphi}{ds} + \xi.$$

As all the single-cell velocities along the RC are directly measured, the derivative of φ with respect to s is obtained by averaging each Voronoi grid:

$$\frac{d\varphi}{ds} = -\left\langle \frac{ds}{dt} \right\rangle_{i_{th}\ grid}$$

By integrating $\frac{d\varphi}{ds}$, the quasi-potential along the RC can be obtained (Figure 13.3D) as

$$\varphi(s) = \varphi(s_0) + \int_{s_0}^{s} \frac{d\varphi}{ds} ds.$$

They also compare the reconstructed potentials of different concentration of TGF-β. In the potential curves of 4 ng/ml concentration, there is a part that is close to a plateau that is like the remnant of the original epithelial attractor. By using a lower concentration of 1 ng/ml of TGF-β, the two paths are nearly kept unchanged. However, the ratio between the numbers of reactive trajectories following the two paths changes. The pseudo-potential for the path with vimentin varying first has a plateau flatter than that of cells treated with 4 ng/ml of TGF-β. On the contrary, the pseudo-potential of concerted variation path does not show such flattening (Figure13. 3D).

To explain the different paths in live-cell imaging experiment result, Wang et al. also proposed a minimal model based on the effective interaction between morphology and vimentin. The metaphorical landscapes in Figure 13.4 provide an intuitive picture of how EMT takes place. Under the control condition, two basins that correspond to epithelial and mesenchymal states are separated by a higher barrier, while the former is stabler. After TGF-β treatment, the epithelial state becomes less and less stable (Figure 13.4A) [18]. Mathematically one can describe the change through bifurcation analyses versus a control parameter, here the concentration of TGF-β. Pitchfork bifurcation and saddle-node bifurcation are two important theoretical mechanisms of critical state transitions in cell-state transitions. Through reconstructing quasi-potentials along RCs of different paths, Wang and co-workers suggested a plausible saddle-node bifurcation mechanism of TGF-β-induced EMT. Mathematically, multiple paths may originate from destabilization of a multi-dimensional epithelial attractor through colliding with multiple saddle points sequentially, which can be understood from the metaphorically landscapes in Fig. 13.4B. With no or low TGF-β, the system resides in the epithelial attractor. Adding TGF-β leads to elevation of the epithelial attractor, and approaching the epithelial attractor and the saddle point connects the epithelial and mesenchymal attractors. The epithelial attractor approaches and collides first with a saddle point to form a barrierless path (concerted variation) to the mesenchymal attractor. Under 1 ng/ml TGF-β concentration, some barrier still exists along vimentin-first

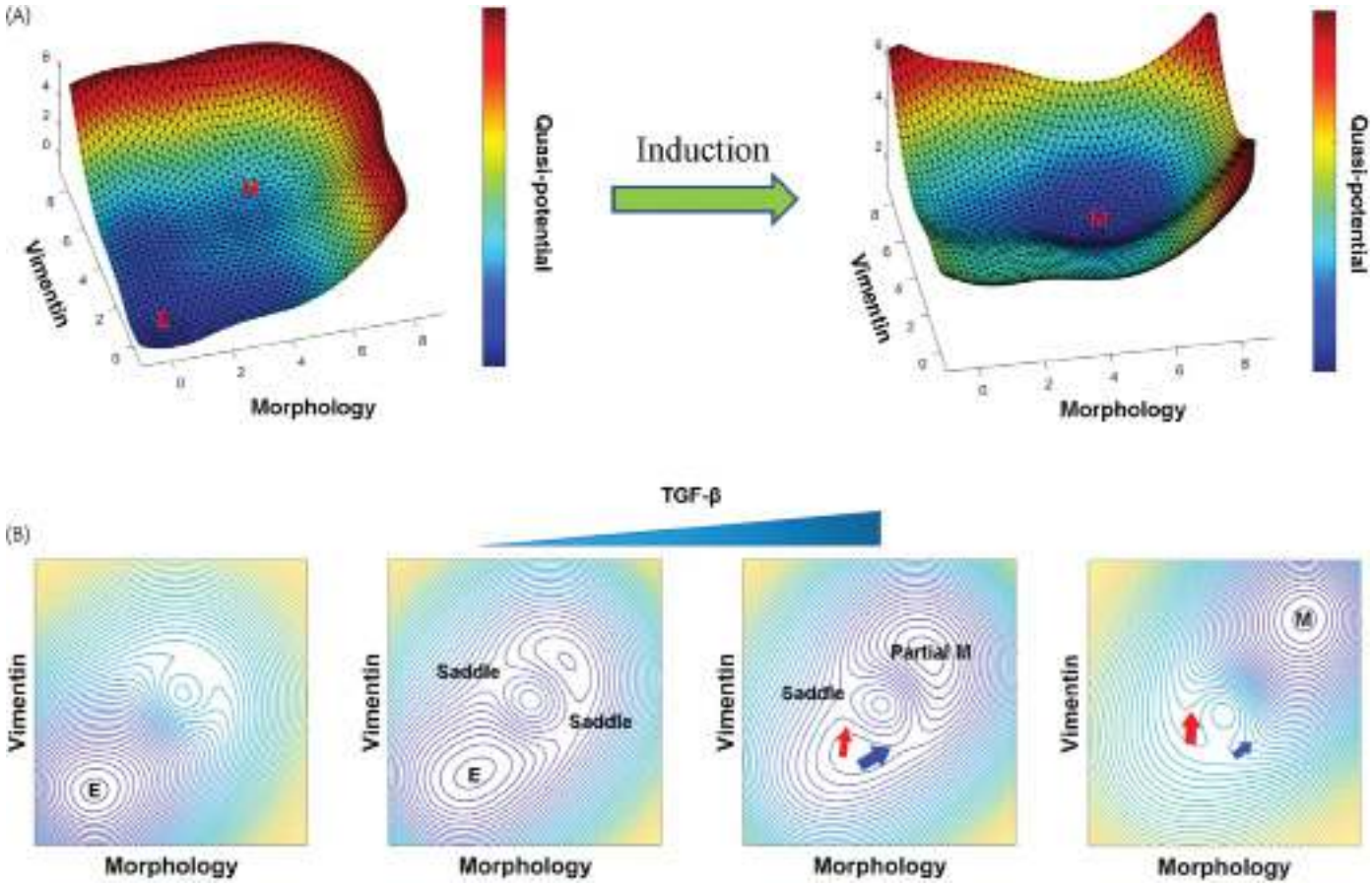


Figure 13.4. Metaphorical landscapes for intuitive understanding of EMT. (A) The effect of TGF-β treatment can be viewed as changing the relative depth of two attractors corresponding to the E and M states. Reproduced with permission from [18]. (B) Contour map view of the sequential change of metaphorical landscapes under TGF-β treatment with different doses. Reproduced with permission from [66].

path. It disappears with a continuous increase in the TGF-β concentration (Figure 13.4B). This process recapitulates the process of saddle-node bifurcation. Further studies are needed to unravel the molecular nature of the two paths.

13.5. Perspectives

To understand development, Waddington suggested a metaphor of a ball sliding downhill along a landscape characterized with valleys and ridges. Recently, some attempts have been made to use different methods to reconstruct the Waddington landscape quantitively [67–69]. For example, Kang and co-worker proposed a network model that incorporates the interaction of metabolism and EMT-related genes and proteins [68]. They analysed the transition paths of EMT and MET and quantified the 2D landscape of this system. Font-Clos et al. proposed a Boolean state model including 72 nodes. By defining a pseudo-Hamiltonian value, they reconstruct the energy landscape of EMT. They found that the intermediate states that separating epithelial and mesenchymal states have higher Hamiltonian value and are unstable upon external perturbation [70]. In the field of scRNA-seq, different methods are also developed to define the landscape of cell-state transition [28,69]. As mentioned above, the usage of these methods is constrained by the fundamental limits of snapshot data.

While 'energy' landscapes provide intuitive pictures, one should be aware that cells function out of thermodynamic equilibrium, and the scalar potential functions only provide partial picture of the cellular dynamics. Only the reconstruction of dynamics on 1D coordinates allows one to follow the procedure of Wang et al. to define a quasi-potential. The term of energy landscape should be used carefully. For nonequilibrium dynamics with more dimensions, the non-gradient force should be taken into consideration [71]. With improved data quality and more sampling, one may reconstruct the vector field that governs the cell-state transition directly.

A method of genome-wide vector field reconstruction called dynamo has been developed in the gene expression state space [43]. The dynamo formalism of Qiu et al. is based on another method called RNA velocity for inferring transition dynamics from single-cell RNA-seq data [42]. Different from the pseudo-time method, RNA velocity utilizes the quantity of spliced and un-spliced RNA [42], or directly measured mRNA turnover dynamics, to estimate the velocity of gene pression. The RNA-velocity method can be used to predict the cell state in next several hours [42,43,72]. Therefore, this method can be used to infer the dynamics missed by the pseudo-time analysis. Dynamo further uses the discrete single-cell expression states and RNA velocities as input to reconstruct a continuous vector field that describes gene regulation relations. Since the RNA velocity suffers from some limitations such that the choice of model and parameters affect the velocity inference, the vector field reconstructed from dynamo also inherits such limitations.

A promising future direction is to combine both live-cell imaging and -omic methods, and exploit the strengths of both methods [73]. A key step is to establish a mapping between a cell state in the composite cell feature space and its correspondence in the expression state space. For such mapping, the dimensionality of single-cell features extracted from live-cell imaging should be increased for improved cell-state resolution. Since the number of fluorescence channel is limited, label-free techniques with deep learning also enable one to extract more features from high-resolution transmitted light images. Developments of 3D imaging techniques like hologram microscope may further increase the resolution and information of transmitted light images.

Further developments of live-cell imaging and single-cell genomics techniques will provide tools for studying heterogenous dynamics, cancer state transition, and cancer treatment [74,75]. The abnormal growth of tumour is one of the major causes of its lethality. Specifically, it has been reported that cell-state transition in development is closely associated with cell cycle. In the field of cancer state transition, the population balance between CSCs and NSCCs indicates that their transitions are coupled with cell cycle. It is of great biomedical significances to study the regulation mechanism that couples the two cellular processes, aiming to redirect the CSC–NSCC transition for more effective biomedical interventions.

REFERENCES

1. Ayob AZ, Ramasamy TS. Cancer stem cells as key drivers of tumour progression. J Biomed. Sci. 2018;**25**(1):20.
2. Shackleton M, Quintana E, Fearon ER, Morrison SJ. Heterogeneity in cancer: cancer stem cells versus clonal evolution. Cell. 2009;**138**(5):822–9.
3. Yabo YA, Niclou SP, Golebiewska A. Cancer cell heterogeneity and plasticity: a paradigm shift in glioblastoma. Neuro-oncology. 2022;**24**(5):669–82.
4. Bocci F, Levine H, Onuchic JN, Jolly MK. Deciphering the dynamics of epithelial–mesenchymal transition and cancer stem cells in tumor progression. Curr Stem Cell Rep. 2019;**5**(1):11–21.
5. Chen J, Li Y, Yu TS, McKay RM, Burns DK, Kernie SG, et al. A restricted cell population propagates glioblastoma growth after chemotherapy. Nature. 2012;**488**(7412):522–6.
6. Yang G, Quan Y, Wang W, Fu Q, Wu J, Mei T, et al. Dynamic equilibrium between cancer stem cells and non-stem cancer cells in human SW620 and MCF-7 cancer cell populations. Br J Cancer. 2012;**106**(9):1512–9.
7. Gupta Piyush B, Fillmore Christine M, Jiang G, Shapira Sagi D, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell. 2011;**146**(4):633–44.
8. Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, et al. The epithelial–mesenchymal transition generates cells with properties of stem cells. Cell. 2008;**133**(4):704–15.
9. Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M, et al. Guidelines and definitions for research on epithelial–mesenchymal transition. Nat Rev Mol Cell Biol. 2020;**21**(6):341–52.
10. Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016. Cell. 2016;**166**(1):21–45.
11. Zhang J, Tian XJ, Zhang H, Teng Y, Li R, Bai F, et al. TGF-β-induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops. Sci Signal. 2014;**7**(345):ra91.
12. Pastushenko I, Blanpain C. EMT transition states during tumor progression and metastasis. Trends Cell Biol. 2019;**29**(3):212–26.
13. Lamouille S, Xu J, Derynck R. Molecular mechanisms of epithelial–mesenchymal transition. Nat Rev Mol Cell Biol. 2014;**15**(3):178–96.

14. Chakraborty P, George JT, Tripathi S, Levine H, Jolly MK. Comparative study of transcriptomics-based scoring metrics for the epithelial-hybrid-mesenchymal spectrum. Front Bioeng Biotechnol. 2020;**8**:220.
15. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell. 2017;**171**(7):1611–24.e24.
16. Burkhardt DB, San Juan BP, Lock JG, Krishnaswamy S, Chaffer CL. Mapping phenotypic plasticity upon the cancer cell state landscape using manifold learning. Cancer Discov. 2022;**12**(8):1847–59.
17. Xing J. Reconstructing data-driven governing equations for cell phenotypic transitions: integration of data science and systems biology. Phys Biol. 2022;**19**(6):061001.
18. Wang W, Douglas D, Zhang J, Kumari S, Enuameh MS, Dai Y, et al. Live-cell imaging and analysis reveal cell phenotypic transition dynamics inherently missing in snapshot data. Sci Adv. 2020;**6**(36):eaba9319.
19. Zheng X, Carstens JL, Kim J, Scheible M, Kaye J, Sugimoto H, et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature. 2015;**527**:525–30.
20. Fischer KR, Durrans A, Lee S, Sheng J, Li F, Wong STC, et al. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature. 2015;**527**:472–6.
21. Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M, et al. Guidelines and definitions for research on epithelial–mesenchymal transition. Nat Rev Mol Cell Biol. 2020;**21**:341–52.
22. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014;**32**:381.
23. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019;**15**(6):e8746.
24. Haghverdi L, Büttner M, Wolf FA, Buettner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. Nat Methods. 2016;**13**(10):845–8.
25. Setty M, Tadmor MD, Reich-Zeliger S, Angel O, Salame TM, Kathail P, et al. Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat Biotechnol. 2016;**34**(6):637–45.
26. Wagner DE, Klein AM. Lineage tracing meets single-cell omics: opportunities and challenges. Nat Rev Genet. 2020;**21**:410–27.
27. Guo M, Bao EL, Wagner M, Whitsett JA, Xu Y. SLICE: determining cell differentiation and lineage based on single cell entropy. Nucleic Acids Res. 2017;**45**(7):e54.
28. Jin S, MacLean AL, Peng T, Nie Q. scEpath: energy landscape-based inference of transition probabilities and cellular trajectories from single-cell transcriptomic data. Bioinformatics. 2018;**34**(12):2077–86.
29. Da Rocha EL, Rowe RG, Lundin V, Malleshaiah M, Jha DK, Rambo CR, et al. Reconstruction of complex single-cell trajectories using CellRouter. Nat Commun. 2018;**9**(1):892.
30. Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM. Fundamental limits on dynamic inference from single-cell snapshots. Proc Natl Acad Sci. 2018;**115**(10):E2467–76.
31. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, et al. Comprehensive integration of single-cell data. Cell. 2019;**177**(7):1888–902.e21.
32. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;**19**(1):15.
33. Dong J, Hu Y, Fan X, Wu X, Mao Y, Hu B, et al. Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. Genome Biol. 2018;**19**(1):31.
34. Cook DP, Vanderhyden BC. Context specificity of the EMT transcriptional response. Nat Commun. 2020;**11**(1):2142.
35. McFaline-Figueroa JL, Hill AJ, Qiu X, Jackson D, Shendure J, Trapnell C. A pooled single-cell genetic screen identifies regulatory checkpoints in the continuum of the epithelial-to-mesenchymal transition. Nat Genet. 2019;**51**(9):1389–98.
36. Bischoff P, Trinks A, Obermayer B, Pett JP, Wiederspahn J, Uhlitz F, et al. Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. Oncogene. 2021;**40**(50):6748–58.
37. Zhang Y, Wang D, Peng M, Tang L, Ouyang J, Xiong F, et al. Single-cell RNA sequencing in cancer research. J Exp Clin Cancer Res. 2021;**40**(1):81.
38. Deshmukh AP, Vasaikar SV, Tomczak K, Tripathi S, den Hollander P, Arslan E, et al. Identification of EMT signaling cross-talk and gene regulatory networks by single-cell RNA sequencing. Proc Natl Acad Sci U S A. 2021;**118**(19): e2102050118.
39. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, et al. Identification of the tumour transition states occurring during EMT. Nature. 2018;**556**(7702):463–8.
40. Karacosta LG, Anchang B, Ignatiadis N, Kimmey SC, Benson JA, Shrager JB, et al. Mapping lung cancer epithelial–mesenchymal transition states and trajectories with single-cell resolution. Nat Commun. 2019;**10**(1):5587.
41. Xing J. Mapping between dissipative and Hamiltonian systems. J Phys A: Math Theor. 2010;**43**:375003.
42. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;**560**(7719):494–8.
43. Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, et al. Mapping transcriptomic vector fields of single cells. Cell. 2022;**185**(4):690–711.
44. Ronneberger O, Fischer P, Brox T, editors. U-Net: convolutional networks for biomedical image segmentation. Medical image computing and computer-assisted intervention—MICCAI. Cham: Springer International Publishing; 2015.
45. Van Valen DA, Kudo T, Lane KM, Macklin DN, Quach NT, DeFelice MM, et al. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. PLoS Comput Biol. [Internet]. 2016;**12**(11):e1005177. Available from: https://doi.org/10.1371/journal.pcbi.1005177.
46. Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. Deep learning for cellular image analysis. Nat Methods. 2019;**16**(12):1233–46.
47. Wang W, Taft DA, Chen YJ, Zhang J, Wallace CT, Xu M, et al. Learn to segment single cells with deep distance estimator and deep cell detector. Comput Biol Med. 2019;**108**:133–41.
48. Gut G, Herrmann MD, Pelkmans L. Multiplexed protein maps link subcellular organization to cellular states. Science (New York, NY). 2018;**361**(6401):eaar7042.
49. Bray MA, Singh S, Han H, Davis CT, Borgeson B, Hartland C, et al. Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. Nat Protoc. 2016;**11**(9):1757–74.
50. Laissue PP, Alghamdi RA, Tomancak P, Reynaud EG, Shroff H. Assessing phototoxicity in live fluorescence imaging. Nat Methods. 2017;**14**(7):657–61.

51. Marklein RA, Lam J, Guvendiren M, Sung KE, Bauer SR. Functionally-relevant morphological profiling: a tool to assess cellular heterogeneity. Trends Biotechnol. 2018;**36**(1):105–18.
52. Caicedo JC, Singh S, Carpenter AE. Applications in image-based profiling of perturbations. Curr Opin Biotechnol. 2016;**39**:134–42.
53. Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, et al. Data-analysis strategies for image-based cell profiling. Nat Methods. 2017;**14**(9):849–63.
54. Leggett SE, Sim JY, Rubins JE, Neronha ZJ, Williams EK, Wong IY. Morphological single cell profiling of the epithelial–mesenchymal transition. Integr Biol. (Cambridge). 2016;**8**(11):1133–44.
55. Haynes J, Srivastava J, Madson N, Wittmann T, Barber DL. Dynamic actin remodeling during epithelial–mesenchymal transition depends on increased moesin expression. Mol Biol Cell. 2011;**22**(24):4750–64.
56. Ren ZX, Yu HB, Li JS, Shen JL, Du WS. Suitable parameter choice on quantitative morphology of A549 cell in epithelial–mesenchymal transition. Biosci Rep. 2015;**35**(3): e00202.
57. Wu PH, Gilkes DM, Phillip JM, Narkar A, Cheng TW, Marchand J, et al. Single-cell morphology encodes metastatic potential. Sci Adv. 2020;**6**(4):eaaw6938.
58. Gordonov S, Hwang MK, Wells A, Gertler FB, Lauffenburger DA, Bathe M. Time series modeling of live-cell shape dynamics for image-based phenotypic profiling. Integr Biol: Quant Biosci Nano Macro. 2016;**8**(1):73–90.
59. Chang AY, Marshall WF. Dynamics of living cells in a cytomorphological state space. Proc Natl Acad Sci U S A. 2019;**116**(43):21556–62.
60. Pincus Z, Theriot JA. Comparison of quantitative methods for cell-shape analysis. J Microsc. 2007;**227**(Pt 2):140–56.
61. Wang J, Zhou X, Bradley PL, Chang SF, Perrimon N, Wong ST. Cellular phenotype recognition for high-content RNA interference genome-wide screening. J Biomol Screen. 2008;**13**(1):29–39.
62. Moris N, Pina C, Arias AM. Transition states and cell fate decisions in epigenetic landscapes. Nat Rev Genet. 2016;**17**(11):693–703.
63. Vanden-Eijnden E, Venturoli M. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. J Chem Phys. 2009;**130**(19):194103.
64. Weinan E, Ren W, Vanden-Eijnden E. Finite temperature string method for the study of rare events. J Phys Chem B. 2005;**109**(14):6688–93.
65. Weinan E, Vanden-Eijnden E. Transition-path theory and path-finding algorithms for the study of rare events. Annu Rev Phys Chem. 2010;**61**:391–420.
66. Wang W, Poe D, Yang Y, Hyatt T, Xing J. Epithelial-to-mesenchymal transition proceeds through directional destabilization of multidimensional attractor. eLife. 2022;**11**:e74866.
67. Li C, Hong T, Nie Q. Quantifying the landscape and kinetic paths for epithelial–mesenchymal transition from a core circuit. Phys Chem Chem Phys. 2016;**18**(27):17949–56.
68. Kang X, Li C. A dimension reduction approach for energy landscape: identifying intermediate states in metabolism-EMT network. Adv Sci. (Weinheim, Baden-Wurttemberg, Germany). 2021;**8**(10):2003133.
69. Wang J, Zhang K, Xu L, Wang E. Quantifying the Waddington landscape and biological paths for development and differentiation. Proc Natl Acad Sci. 2011;**108**(20):8257–62.
70. Font-Clos F, Zapperi S, La Porta CAM. Topography of epithelial-mesenchymal plasticity. Proc Natl Acad Sci U S A. 2018;**115**(23):5902–7.
71. Zhou JX, Aliyu MD, Aurell E, Huang S. Quasi-potential landscape in complex multi-stable systems. J R Soc Interface. 2012;**9**(77):3539–53.
72. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol. 2020;**38**(12):1408–14.
73. Xing J. Reconstructing data-driven governing equations for cell phenotypic transitions: integration of data science and systems biology. Phys Biol. 2022;**19**:061001.
74. Altschuler SJ, Wu LF. Cellular heterogeneity: do differences make a difference? Cell. 2010;**141**(4):559–63.
75. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. Development (Cambridge, England). 2009;**136**(23):3853–62.

14

Emerging single-cell technologies and concepts to trace cancer progression and drug resistance

Syeda Subia Ahmed, Danielle Pi, Nicholas Bodkin*, Vito W. Rebecca, and Yogesh Goyal#*

14.1. Introduction

Recent technological advancements have enabled the measurement of multiscale characteristics of cancer cells at an unprecedented resolution and throughput. In particular, single-cell sequencing techniques have revealed extensive diversity in tumours, often within the same tumour in a given patient. Unsurprisingly, considerable efforts have been devoted to profiling tumour cells in a variety of cancer contexts, such as therapy resistance and metastasis. Typically, such datasets are static in that they only provide disjointed and descriptive snapshots of the process of interest, e.g. separate samples used for measuring molecular signatures before and after drug resistance (**Figure 14.1**). Thus, the eventual phenotypes (e.g. cells *after* becoming resistant or metastatic) are not necessarily predictive of the initial states that drives the very process and vice versa. Inferring state-fate lineage/clonal relationships is further complicated by mounting evidence suggesting that a complex milieu of genetic and non-genetic factors (both intrinsic and extrinsic) can drive cancer cell-fate decisions and evolution, e.g. to metastasize or become resistant [1]. Hence, the lack of longitudinal capabilities precludes us from establishing direct and causal connections between driver cell states and their fate outcomes. Addressing this problem can lead to novel and effective therapeutic opportunities in cancer.

Throughout the text, we conceptualize processes and open questions in terms of cell states and fate outcomes. We define a 'state' as a set of observations that characterize a cell instantaneously where a cell can transition along one or multiple continuums of characteristics. We define 'fate' as the state in which a cell transitions in response to stimuli (e.g. drug exposure). A detailed description of cell 'state', 'fate', and related terms in the context of cancer are provided in a recent review [2]. In this chapter, we cover various experimental (sequencing and imaging) and computational methodologies developed to infer or directly measure longitudinal relationships between cell states and fate outcomes. The technological and conceptual paradigms discussed here are broadly applicable to a variety of cancer contexts, e.g. initiation, evolution, metastasis, and therapy resistance. We highlight key insights gained from such frameworks within these contexts. Lastly, we comment on the computational and technological gaps that can be addressed for delineating clonal state-fate relationships in cancer moving forward.

14.2. Sequencing-based synthetic lineage tracers

Tracing clones by synthetically introducing a foreign sequenceable 'mark' or 'tag' in the DNA of a cell has a rich history (**Figure 14.2**) [3]. Some of the first studies took advantage of randomly integrated retroviral sequences for tracking uniquely 'marked' haematopoietic stem cell lineages using Southern blots in irradiated adult mice [4–6]. Follow-up studies focused on increasing the diversity of the retroviral libraries [7], including CHAPOL, which climbed to a complexity of more than ten million [8]. Tracing cellular clones with sequencing-based DNA tags became particularly popular in the past two decades, in large part owing to rapid advances in next-generation sequencing and genetic engineering technologies. Consequently, the interest and accessibility of studying clonal haematopoiesis paved the way for developing high-throughput sequencing-based lineage-tracing approaches [9–12]. For instance, some studies developed lineage trees of early embryonic development by leveraging genome editing technologies, such as the recently discovered CRISPR/Cas9, to accumulate combinatorial diversity of sequences

* These authors contributed equally.
# Author for correspondence: yogesh.goyal@northwestern.edu

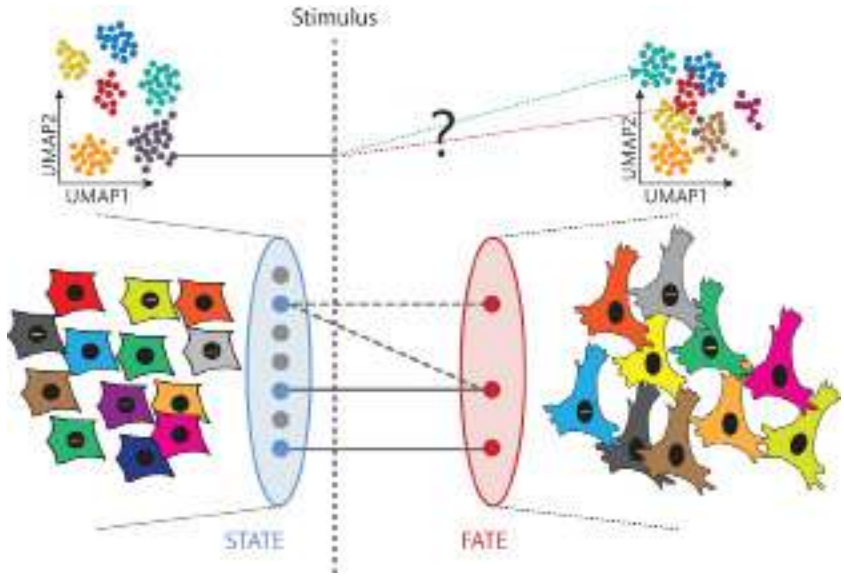


Figure 14.1. A schematic depicting the need to establish a quantitative and direct connection between state and fate of cancer cells.

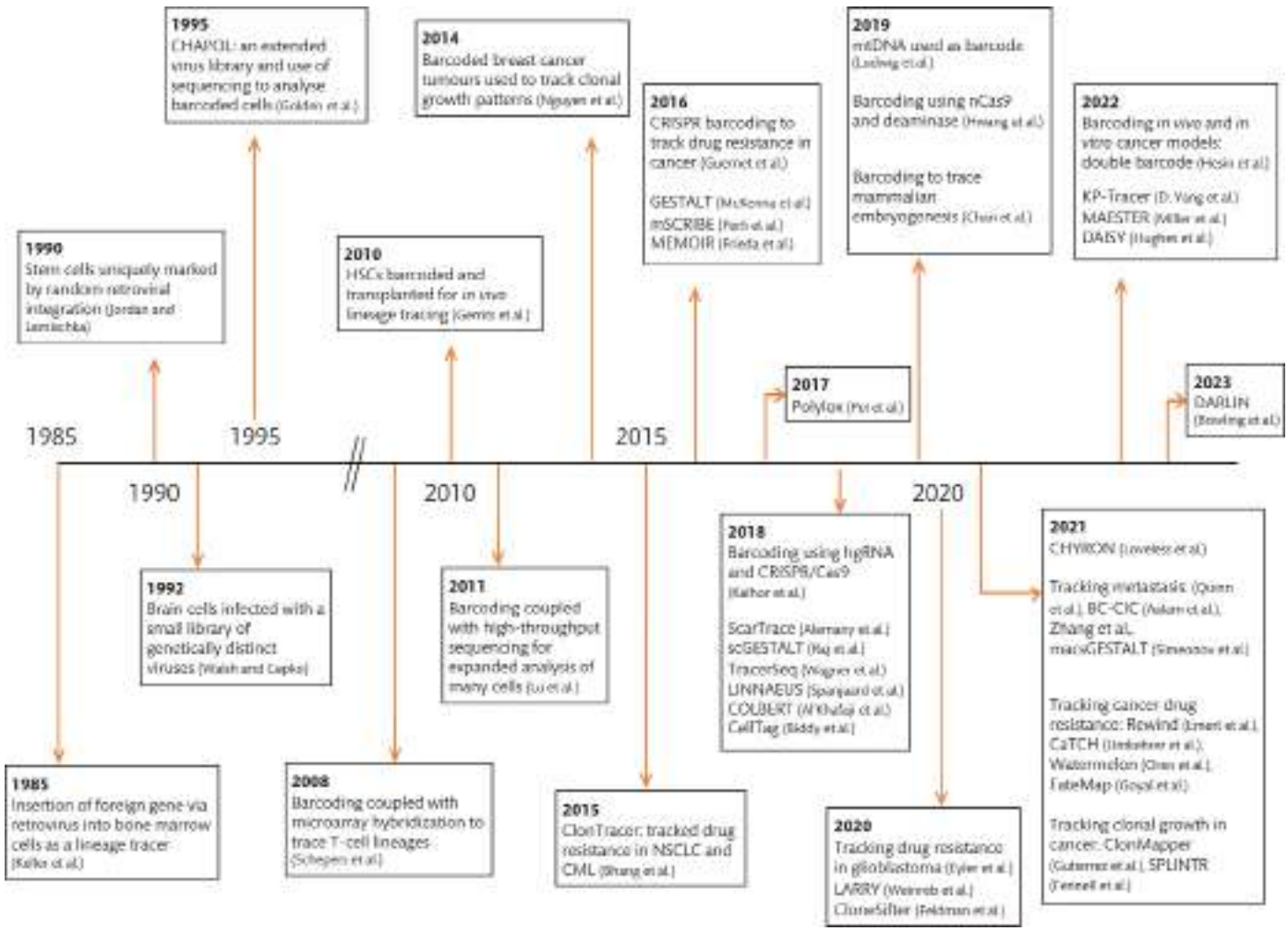


Figure 14.2. Timeline of the development of sequencing-based barcodes and their applications.

resulting in information-rich heritable 'barcodes', such as GESTALT [13–16]. In cancer, the transduction of lentiviral libraries consisting of static pseudorandom DNA sequences was one of the earliest approaches implemented in the context of tracing therapy resistance and metastasis [17–19]. For example, CloneTracer made an important discovery that pre-existing rare subsets of cells could drive resistance as opposed to the acquisition of *de novo* functional mutations during drug treatment [17]. Another study used CRISPR/Cas9 mutable barcodes to trace clonal dynamics in cancer cells exposed to epidermal growth factor receptor (EGFR) inhibitors [20] and compared the nature of responses between lung and breast cancer. While these frameworks pioneered a new wave of high-throughput lineage tracing and provided valuable insights into developmental and cancer biology, they lacked the ability to provide direct single-cell -omic resolution due to the technological limitations of profiling single cells at the time.

The development of single-cell RNA-sequencing technologies initiated a renewed interest in building DNA-barcode lineage tracers that could be coupled with transcriptomic information in single cells [21,22]. As a consequence, several new combined single-cell lineage-tracing and transcriptional profiling frameworks have been developed and implemented in a variety of cancer models [23–32] as well as in other contexts [25,33–45]. Such frameworks have highlighted the contribution of single-cell differences in a variety of cancer contexts, including drug resistance [23,25,26,29,31,46], metastasis [25,47,48], and tumour evolution [24,26,49]. Some other recent approaches have turned to naturally occurring somatic mutations in mitochondrial DNA to capture clonal relationships between transcriptomes of single cells [50–53]. Collectively, these methodologies have enabled us to gain key conceptual insights into cancer, particularly in the context of tracing drug resistance trajectories. As an example, the Obenauf group developed CaTCH [28], a unique CRISPRa-induced barcode expression system, to understand how melanoma clones behave in response to the application of unrelated therapies. They found that clones resistant to targeted therapies were also cross-resistant to immunotherapies [46]. Although this study did not perform the reverse experiment (i.e. immunotherapy as first-line), it stresses the importance of choosing the right order of exposure to unrelated therapies in cancer patients.

In addition, these newer lineage-tracing systems, reported by our groups and others, are well suited for quantifying the timescales and characteristics of non-genetic plasticity, especially as a driver of drug resistance and metastasis across cancers [30,31,54–62]. Indeed, lineage-tracing studies using 'twin' or 'replica' experimental designs enabled our group and others to profile the rare cell states that underlie this plasticity and connect them to their eventual resistant fates [26,27,31,63]. In another example, a modified form of GESTALT [13] was used to simultaneously incorporate both static and mutable barcodes in single cells, identifying a continuum of epithelial-to-mesenchymal transition (EMT) states that underlie metastasis initiated by rare cells in pancreatic cancer [25]. Additionally, another lineage-tracing study in lung cancer identified heritable gene expression states driving metastasis rates and sites [47]. Although single-cell lineage-tracing systems have generated significant interest, they can be technically complex and present computational challenges, as outlined below **in Section 14.4.** Developing clear barcode design guidelines and robust analysis pipelines will be pivotal in overcoming these challenges, leading to a wider and more democratic adoption in basic and clinical cancer studies moving forward.

14.3. Imaging-based synthetic lineage-tracing approaches

Imaging-based lineage-tracing methods date back to the early 20th century, when dyes were used to label amphibian embryos [3,64]. The 1990s marked the beginning of a new era in lineage tracing with the introduction of genetic markers, such as the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* [65,66]. Recent advances in quantitative microscopy, the creation of a spectrally separated suite of photostable fluorescent variants, and the establishment of associated high-throughput analysis now offer complementary approaches to follow the fates of cancer cell clones in cell culture and animal models. In cell culture, many studies have utilized time-lapse imaging dynamics of single or dual colour fluorescent tags to perform fate tracing of single cancer cell clones under selective pressures [67–69]. In fact, such studies were instrumental in establishing that heritable non-genetic differences between single cancer cells in culture alone can drive resistance, such as resistance to apoptosis upon exposure to TRAIL (tumour necrosis factor (TNF)-related apoptosis-inducing ligand) [67]. However, these approaches were unable to uniquely tag clonal populations. More recent studies use a sophisticated combination of fluorescent colours, sequencing technologies, and intracellular localizations to infer clonal dynamics in cancer cells in culture [70,71].

In animal models, early studies used whole-body single colour fluorescent animals to trace the origin of tumour cells through chimerism. For example, by generating chimeras between constitutively active MEK1 animals and GFP-expressing animals that lacked the MEK1 transgene, a study showed that epidermal tumours were polyclonal in origin and that tumour cells could induce oncogenic potential in an otherwise genetically normal cell [72]. While chimerism is a powerful tool, the ability to restrict fluorescent expression to specific cell types greatly expanded the power of lineage tracing and allowed for identification and visualization in a mosaic system. By combining fluorescent reporters with the Cre/lox recombination system, cell tracing could be temporally and spatially induced in cells and tissues of interest. The Cre/lox system has since been refined to enable higher resolution lineage tracing through synthetic constructs such as Brainbow. Brainbow constructs have three possible Cre-based excisions, and labelled cells stochastically express one or more of three fluorescent proteins [73]. This system has since been widely adapted in cancer studies. As an example, a study on colon cancer repurposed the Brainbow system (and aptly renamed it to Crainbow) to fluorescently barcode somatic mutations and directly visualize and compare clonal expansion and oncogenic spread in adult and infant systems [74]. In another study, the Brainbow system was used to visualize keratinocyte clones (dubbed 'Skinbow'), where researchers showed that the cell of origin of basal cell carcinomas appear more frequently near hair follicles on UV-irradiated skin [75]. Further modifications in the initial Brainbow designs, such as simultaneous tagging of nuclear, cytoplasmic, and membranous compartments, have offered enhanced resolution and more robust capabilities in tracking clonal expansions [76]. Initially developed to visualize stem cell dynamics in intestinal

crypt homeostasis, the 'confetti' system has subsequently been used to identify drivers of intestinal adenomas [77] and to trace biased clonal expansion of *KRAS* mutant intestinal crypts leading to colorectal cancer [78]. Confetti systems have also been implemented to study metastasis in various cancers, including pancreatic [79] and breast [80,81] cancers. These studies have elucidated quantitative connections between stages of cancer progression, initiation of metastasis, clonal expansions at metastatic sites, and plasticity mechanisms (such as EMT signatures) underlying metastatic clones. Moreover, a recent study on cancer drug resistance leveraging such confetti systems has demonstrated that chemotherapy specifically targets slow-dividing clones rather than fast-diving ones, which has important clinical implications [82]. Although Brainbow and its derivatives were originally developed as transgenic mouse lines, the tool has been adapted to be used with adeno-associated viral vectors, thus expanding the method to other systems and species [83]. Similarly, lineage-tracing imaging systems specific to a signalling pathway have also been developed to investigate signalling-specific trajectories. For example, a TGF-β-dependent system uncovered the non-genetic heterogeneity that reduced the efficacy of anti-cancer therapies and caused resistance in squamous cell carcinoma [84].

Additionally, it is possible to couple lineage-tracing systems with other genetic manipulations to simultaneously induce perturbation (e.g. gene knockout) and fluorescent labelling in the same cells, such as in the frameworks LeGO and MADM. With mouse models, the MADM framework has been used to reveal a number of insights, such as the tumour cell of origin in glioma and a comparative analysis of tumour progression between pancreatic and lung cancers, among others [85–88]. Similarly, the LeGO system [89] revealed clonal dynamics and heterogeneity underlying metastasis and drug resistance in multiple cancers, including breast cancer [90,91], neuroblastoma [92], glioma [93], and osteosarcoma [94]. Lastly, some other studies have combined visual and DNA barcoding systems for lineage tracing in cancer, for instance, in head and neck squamous cell carcinoma [95] and leukaemia [96]. As molecular sequencing and computational tracking capabilities increase, visual tracers can be integrated with those modalities (**Figure 14.3**) to provide richer state-fate trajectories within single cells in 2D/3D cell culture and animal systems.

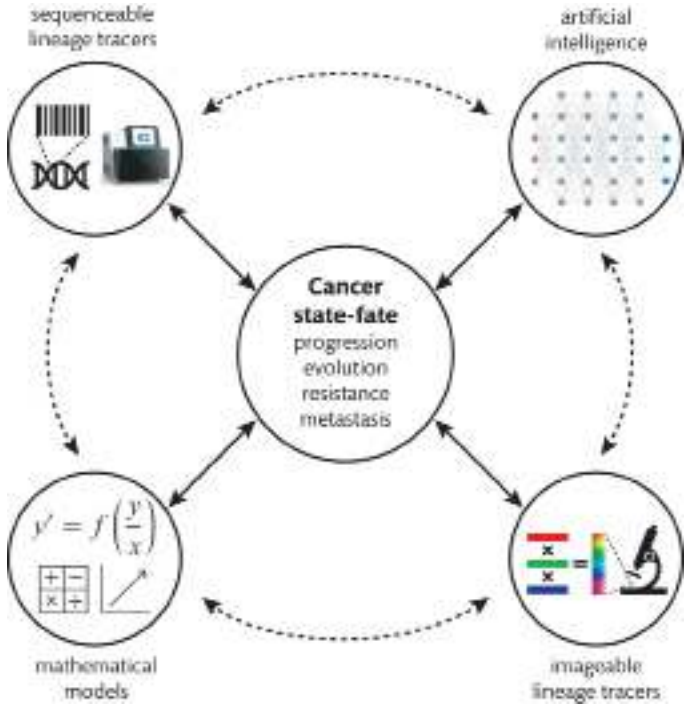


Figure 14.3. A synergistic interplay between various experimental (imaging- and sequencing-based) and computational (mathematical models and artificial intelligence) methodologies is considered critical to elucidate state-fate relationships in a variety of cancer contexts and develop effective therapeutic strategies.

14.4. Computational and theoretical lineage-tracing approaches

Some of the earliest computational efforts geared towards building longitudinal relationships between single cells relied on 'inferring' the time or clonal dimension from snapshots of single-cell sequencing data [97]. One such tool is Monocle [98,99], which orders features of single cells along a pseudo-time by leveraging a cell's asynchronous progression in a specific process, such as differentiation or EMT transitions. Monocle pseudo-time analysis has been widely used in a variety of cancer contexts. For example, a recent study used Monocle 3, the most recent stable version [100] as of 2023, to reveal a 'pseudoEMT' trajectory underlying metastasis in pancreatic cancer [25]. In another example, a recent study has used Monocle 3 to identify pseudo-time co-regulated gene expression trajectories during the treatment of a non-small-cell lung carcinoma cancer cell line with EGFR inhibitor erlotinib [101]. Yet another computational approach that uses static datasets to infer temporal dynamics is RNA velocity [102], which uses intronic and exonic reads from single-cell RNA-seq data to infer temporal trajectories. Similar to Monocle, RNA velocity and other frameworks [103–105] have been implemented to infer temporal paths of initiation, evolution, and drug resistance in multiple cancer systems [106–109]. Other mathematical studies have elucidated theoretical limits on inferences of lineage dynamics from static single-cell datasets [110,111]. Collectively, such inference algorithms have helped circumvent the issue of working with static and disjointed datasets to reconstruct state-fate trajectories in cancer. At the same time, more recent studies have highlighted limitations, underlying assumptions, and potential drawbacks of such inference algorithms and have urged caution when using them [112,113].

Some of the shortcomings of inference algorithms are readily addressed by direct measurements with the lineage-tracing experimental systems discussed above. However, the application of single-cell lineage-tracing approaches (particularly sequencing-based) in cancer is still nascent. The experimental designs and computational pipelines are technically complex and remain lab-specific, and the resulting datasets are prone to misinterpretation. Recognizing this challenge, recent computational studies have attempted to formalize the analysis of integrated single-cell profiling and lineage barcoding datasets [114–118]. Such approaches have extracted additional information by robustly identifying transition or hybrid cell states and potential drivers of drug-resistant fates, not otherwise possible from analysing single-cell sequencing datasets alone. For instance, singletCode, a computational framework leveraging barcoding datasets to identify ground truth singlets within single-cell RNA

sequencing datasets, can discriminate between true hybrid cell states and artefacts resulting from coalesced cells during the implementation of microfluidics-based profiling technologies [117].

Single-cell-identifying information native to a cell, such as genetic variations in nuclear or mitochondrial DNA, which are also used to establish the sequence of cellular events and lineage relationships in single cells, has been largely driven by bioinformatic and statistical advances [119–121]. Some other recent computational tools, such as gene expression memory-based lineage inference (GEMLI), leverage heritable gene expression patterns to infer lineage relationships. Collectively, such frameworks, while still in the early stages, will be critical for extending longitudinal tracking of fate outcomes—from cell culture and animal models (which are more amenable to synthetic tracers) to patient tumour samples collected before and after anti-cancer treatments.

Furthermore, the high-throughput imaging-based studies discussed above come with the inherent challenge of reliably segmenting and tracking a large numbers of cells. Conventional intensity-based and recent machine-learning-based algorithms are critical for reliably segmenting cells in images from time-lapse microscopy [122,123]. The accompanying challenge is to connect single cells from a series of segmented images. Particularly, to make any meaningful measurements in cancer drug resistance studies, the tracking of thousands or more cancer cells simultaneously is needed since only a rare subset escapes the treatment. Accordingly, computational algorithms, such as the Viterbi algorithm applied to time-lapse movies of cancer cells [124,125], have helped reveal important aspects of single-cell biology related to cancer drug resistance, such as the rapid adaptation in rare 'escapee' melanoma cells exposed to the targeted therapy drug dabrafenib [69]. Another effort used a combination of recent deep-learning approaches and analytical tools to reconstruct lineage phylogenies from imaging-based lineage tracers in intact growing tissues, revealing the control principles of growth dynamics from static snapshots [126]. As newer technological advances focus on integrating multiple profiling modalities in cancer cells, such as combining imaging with sequencing data, computational approaches (e.g. [127]) will be key to decoupling technical noise from biological signals driving state-fate relationships in cancer.

At the patient level, another set of frameworks has resulted from advances in artificial intelligence (AI), particularly for predicting anti-cancer therapy treatment outcomes, such as the development of resistance. In fact, some AI-based algorithms developed in a research setting have made it to the clinic, such as a 70-gene signature developed 20 years ago to predict treatment responses for early-stage breast cancer [128]. Having said that, much remains to be done for such algorithms to be robust and reliable in their predictions for patient outcomes. Moving forward, such algorithms are also becoming increasingly sophisticated and are shifting focus to evaluate the potential additive, antagonistic, and synergistic effects of combination therapies, as well as integrating multimodal patient data [129].

14.5. Concluding remarks

Building quantitative and causal relationships between a cell's state and its eventual fate—be it to become cancerous, drug resistant, or to metastasize to distant organs—promises to provide hypotheses for novel treatment strategies. Consider that certain treatment-specific rare cell states are more likely to survive and become resistant. In this case, capturing and profiling these rare cell states that drive clinically relevant behaviours can therefore reveal new therapeutic targets, as opposed to conventional approaches that target the developed behaviour itself. To this end, experimental and computational technological innovations (**Figure 14.3**), such as developing *in situ* techniques to spatially probe clonality directly in patient tumour samples or measure cell–cell communications between clones, will be central to achieving these goals.

Acknowledgements

The authors would like to thank Emanuelle Grody, Cedric Tito Chai, Ian Mellis, and Karun Kiani for their helpful comments. S.A. acknowledges support from the Career Award at the Scientific Interface from BWF (1020614.01) and start-up funds to Y.G. from Northwestern University. D.P. and N.B. acknowledge support from NIH T32 GM144295. V.R. acknowledges support from the K01CA245124-01. Y.G. acknowledges support from the Career Award at the Scientific Interface from BWF (1020614.01) and the Young Investigator Award from the Cancer Research Foundation.

Declaration of interests

Y.G. received consultancy fees from the Schmidt Science Fellows and the Rhodes Trust. All other authors declare no conflict of interest.

REFERENCES

1. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022 Jan;**12**(1):31–46.
2. Sun H, Kumari N, Melzer ME, Backman V, Goyal Y. Navigating nongenetic plasticity in cancer drug resistance. Annu Rev Cancer Biol. 2024;9.
3. Kretzschmar K, Watt FM. Lineage tracing. Cell. 2012 Jan 20;**148**(1-2):33–45.
4. Jordan CT, Lemischka IR. Clonal and systemic analysis of long-term hematopoiesis in the mouse. Genes Dev. 1990 Feb;**4**(2):220–32.
5. Keller G, Paige C, Gilboa E, Wagner EF. Expression of a foreign gene in myeloid and lymphoid cells derived from multipotent haematopoietic precursors. Nature. 1985;**318**(6042):149–54.
6. Lemischka IR, Raulet DH, Mulligan RC. Developmental potential and dynamic behavior of hematopoietic stem cells. Cell. 1986 Jun 20;**45**(6):917–27.
7. Walsh C, Cepko CL. Widespread dispersion of neuronal clones across functional regions of the cerebral cortex. Science. 1992 Jan 24;**255**(5043):434–40.
8. Golden JA, Fields-Berry SC, Cepko CL. Construction and characterization of a highly complex retroviral library for lineage analysis. Proc Natl Acad Sci U S A. 1995 Jun **6**;92(12):5704–8.
9. Lu R, Neff NF, Quake SR, Weissman IL. Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding. Nat Biotechnol. 2011 Oct 2;**29**(10):928–33.

10. Naik SH, Perié L, Swart E, Gerlach C, van Rooij N, de Boer RJ, et al. Diverse and heritable lineage imprinting of early haematopoietic progenitors. Nature. 2013 Apr 11;**496**(7444):229–32.
11. Wu C, Li B, Lu R, Koelle SJ, Yang Y, Jares A, et al. Clonal tracking of rhesus macaque hematopoiesis highlights a distinct lineage origin for natural killer cells. Cell Stem Cell. 2014 Apr 3;**14**(4):486–99.
12. Gerrits A, Dykstra B, Kalmykowa OJ, Klauke K, Verovskaya E, Broekhuis MJC, et al. Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood. 2010 Apr 1;**115**(13):2610–8.
13. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science. 2016 Jul 29;**353**(6298):aaf7907.
14. Kalhor R, Kalhor K, Mejia L, Leeper K, Graveline A, Mali P, et al. Developmental barcoding of whole mouse via homing CRISPR. Science [Internet]. 2018 Aug 31;**361**(6405). Available from: http://dx.doi.org/10.1126/science.aat9804
15. Loveless TB, Grotts JH, Schechter MW, Forouzmand E, Carlson CK, Agahi BS, et al. Lineage tracing and analog recording in mammalian cells by single-site DNA writing. Nat Chem Biol. 2021 Jun;**17**(6):739–47.
16. Perli SD, Cui CH, Lu TK. Continuous genetic recording with self-targeting CRISPR-Cas in human cells. Science [Internet]. 2016 Sep 9;**353**(6304). Available from: http://dx.doi.org/10.1126/science.aag0511
17. Bhang HEC, Ruddy DA, Krishnamurthy Radhakrishna V, Caushi JX, Zhao R, Hims MM, et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. Nat Med. 2015 May;**21**(5):440–8.
18. Nguyen LV, Cox CL, Eirew P, Knapp DJHF, Pellacani D, Kannan N, et al. DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. Nat Commun. 2014 Dec 23;**5**:5871.
19. Wagenblast E, Soto M, Gutiérrez-Ángel S, Hartl CA, Gable AL, Maceli AR, et al. A model of breast cancer heterogeneity reveals vascular mimicry as a driver of metastasis. Nature. 2015 Apr 16;**520**(7547):358–62.
20. Guernet A, Mungamuri SK, Cartier D, Sachidanandam R, Jayaprakash A, Adriouch S, et al. CRISPR-barcoding for intratumor genetic heterogeneity modeling and functional analysis of oncogenic driver mutations. Mol Cell. 2016 Aug 4;**63**(3):526–38.
21. Serrano A, Berthelet J, Naik SH, Merino D. Mastering the use of cellular barcoding to explore cancer heterogeneity. Nat Rev Cancer. 2022 Nov;**22**(11):609–24.
22. Sankaran VG, Weissman JS, Zon LI. Cellular barcoding to decipher clonal dynamics in disease. Science. 2022 Oct 14;**378**(6616):eabm5874.
23. Oren Y, Tsabar M, Cuoco MS, Amir-Zilberstein L, Cabanos HF, Hütter JC, et al. Cycling cancer persister cells arise from lineages with distinct programs. Nature. 2021 Aug;**596**(7873):576–82.
24. Fennell KA, Vassiliadis D, Lam EYN, Martelotto LG, Balic JJ, Hollizeck S, et al. Non-genetic determinants of malignant clonal fitness at single-cell resolution. Nature. 2022 Jan;**601**(7891):125–31.
25. Simeonov KP, Byrns CN, Clark ML, Norgard RJ, Martin B, Stanger BZ, et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. Cancer Cell. 2021 Aug 9;**39**(8):1150–62.e9.
26. Gutierrez C, Al'Khafaji AM, Brenner E, Johnson KE, Gohil SH, Lin Z, et al. Multifunctional barcoding with ClonMapper enables high-resolution study of clonal dynamics during tumor evolution and treatment. Nat Cancer. 2021 Jul 12;**2**(7):758–72.
27. Emert BL, Cote CJ, Torre EA, Dardani IP, Jiang CL, Jain N, et al. Variability within rare cell states enables multiple paths toward drug resistance. Nat Biotechnol. 2021 Jul;**39**(7):865–76.
28. Umkehrer C, Holstein F, Formenti L, Jude J, Froussios K, Neumann T, et al. Isolating live cell clones from barcoded populations using CRISPRa-inducible reporters. Nat Biotechnol. 2021 Feb;**39**(2):174–8.
29. Feldman D, Tsai F, Garrity AJ, O'Rourke R, Brenan L, Ho P, et al. CloneSifter: enrichment of rare clones from heterogeneous cell populations. BMC Biol. 2020 Nov 24;**18**(1):177.
30. Hughes NW, Qu Y, Zhang J, Tang W, Pierce J, Wang C, et al. Machine-learning-optimized Cas12a barcoding enables the recovery of single-cell lineages and transcriptional profiles. Mol Cell. 2022 Aug 18;**82**(16):3103–18.e8.
31. Goyal Y, Dardani IP, Busch GT, Emert B, Fingerman D, Kaur A, et al. Diverse clonal fates emerge upon drug treatment of homogeneous cancer cells. Nature. 2023 July 19;**620**:651–9.
32. Singh I, Fernandez-Perez D, Sanchez PS, Rodriguez-Fraticelli A. Pre-existing stem cell heterogeneity dictates clonal responses to the acquisition of leukemic driver mutations. Cell Stem Cell. 2025 Apr 3;**32**:1–17.
33. Weinreb C, Rodriguez-Fraticelli A, Camargo FD, Klein AM. Lineage tracing on transcriptional landscapes links state to fate during differentiation. Science [Internet]. 2020 Feb 14;**367**(6479). Available from: http://dx.doi.org/10.1126/science.aaw3381
34. Tian L, Tomei S, Schreuder J, Weber TS, Amann-Zalcenstein D, Lin DS, et al. Clonal multi-omics reveals Bcor as a negative regulator of emergency dendritic cell development. Immunity. 2021 Jun 8;**54**(6):1338–51.e9.
35. Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. Nat Biotechnol. 2018 Jun;**36**(5):442–50.
36. Bowling S, Sritharan D, Osorio FG, Nguyen M, Cheung P, Rodriguez-Fraticelli A, et al. An engineered CRISPR-Cas9 mouse line for simultaneous readout of lineage histories and gene expression profiles in single cells. Cell. 2020 Jun 11;**181**(6):1410–22.e27.
37. Li L, Bowling S, Yu Q, McGeary SE, Alcedo K, Lemke B, et al. A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. Cell. 2023 Nov 9;**186**(23):5183–99.
38. Biddy BA, Kong W, Kamimoto K, Guo C, Waye SE, Sun T, et al. Single-cell mapping of lineage and identity in direct reprogramming. Nature. 2018 Dec;**564**(7735):219–24.
39. Frieda KL, Linton JM, Hormoz S, Choi J, Chow KHK, Singer ZS, et al. Synthetic recording and in situ readout of lineage information in single cells. Nature. 2017 Jan 5;**541**(7635):107–11.
40. He Z, Maynard A, Jain A, Gerber T, Petri R, Lin HC, et al. Lineage recording in human cerebral organoids. Nat Methods. 2022 Jan;**19**(1):90–9.
41. Chan MM, Smith ZD, Grosswendt S, Kretzmer H, Norman TM, Adamson B, et al. Molecular recording of mammalian embryogenesis. Nature. 2019 Jun;**570**(7759):77–82.
42. Alemany A, Florescu M, Baron CS, Peterson-Maduro J, van Oudenaarden A. Whole-organism clone tracing using single-cell sequencing. Nature. 2018 Apr 5;**556**(7699):108–12.
43. Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, et al. Simultaneous lineage tracing and cell-type identification

using CRISPR-Cas9-induced genetic scars. Nat Biotechnol. 2018 Jun;**36**(5):469–73.
44. Wagner DE, Weinreb C, Collins ZM, Briggs JA, Megason SG, Klein AM. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science. 2018 Jun 1;**360**(6392):981–7.
45. Jiang CL, Goyal Y, Jain N, Wang Q, Truitt RE, Coté AJ, et al. Cell type determination for cardiac differentiation occurs soon after seeding of human-induced pluripotent stem cells. Genome Biol. 2022 Apr 5;**23**(1):90.
46. Haas L, Elewaut A, Gerard CL, Umkehrer C, Leiendecker L, Pedersen M, et al. Acquired resistance to anti-MAPK targeted therapy confers an immune-evasive tumor microenvironment and cross-resistance to immunotherapy in melanoma. Nat Cancer. 2021 Jul;**2**(7):693–708.
47. Quinn JJ, Jones MG, Okimoto RA, Nanjo S, Chan MM, Yosef N, et al. Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. Science [Internet]. 2021 Feb 26;**371**(6532). Available from: http://dx.doi.org/10.1126/science.abc1944
48. Zhang W, Bado IL, Hu J, Wan YW, Wu L, Wang H, et al. The bone microenvironment invigorates metastatic seeds for further dissemination. Cell. 2021 Apr 29;**184**(9):2471–86.e20.
49. Yang D, Jones MG, Naranjo S, Rideout WM 3rd, Min KHJ, Ho R, et al. Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. Cell [Internet]. 2022 Apr 28; Available from: http://dx.doi.org/10.1016/j.cell.2022.04.015
50. Ludwig LS, Lareau CA, Ulirsch JC, Christian E, Muus C, Li LH, et al. Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. Cell. 2019 Mar 7;**176**(6):1325–39.e22.
51. Miller TE, Lareau CA, Verga JA, DePasquale EAK, Liu V, Ssozi D, et al. Mitochondrial variant enrichment from high-throughput single-cell RNA sequencing resolves clonal populations. Nat Biotechnol [Internet]. 2022 Feb 24; Available from: http://dx.doi.org/10.1038/s41587-022-01210-8
52. Penter L, Ten Hacken E, Southard J, Lareau CA, Ludwig LS, Li S, et al. Mitochondrial DNA mutations as natural barcodes for lineage tracing of murine tumor models. Cancer Res. 2023 Mar 2;**83**(5):667–72.
53. Lareau CA, Liu V, Muus C, Praktiknjo SD, Nitsch L, Kautz P, et al. Mitochondrial single-cell ATAC-seq for high-throughput multi-omic detection of mitochondrial genotypes and chromatin accessibility. Nat Protoc [Internet]. 2023 Feb 15; Available from: http://dx.doi.org/10.1038/s41596-022-00795-3
54. Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell. 2010 Apr 2;**141**(1):69–80.
55. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell. 2011 Aug 19;**146**(4):633–44.
56. Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature. 2017 Jun 15;**546**(7658):431–5.
57. Shaffer SM, Emert BL, Reyes Hueros RA, Cote C, Harmange G, Schaff DL, et al. Memory sequencing reveals heritable single-cell gene expression programs associated with distinct cellular behaviors. Cell. 2020 Aug 20;**182**(4):947–59.e17.
58. Kaur A, Cuenca L, Kiani K, Busch GT, Fingerman D, Dunagin MC, et al. Metastatic potential in clonal melanoma cells is driven by a rare, early-invading subpopulation [Internet]. bioRxiv. 2022 [cited 2022 May 2]. p. 2022.04.17.488591. Available from: https://doi.org/10.1101/2022.04.17.488591
59. Rambow F, Rogiers A, Marin-Bejar O, Aibar S, Femel J, Dewaele M, et al. Toward minimal residual disease-directed therapy in melanoma. Cell. 2018 Aug 9;**174**(4):843–55.e19.
60. Roesch A, Fukunaga-Kalabis M, Schmidt EC, Zabierowski SE, Brafford PA, Vultur A, et al. A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. Cell. 2010 May 14;**141**(4):583–94.
61. Rebecca VW, Herlyn M. Nongenetic mechanisms of drug resistance in melanoma. Annu Rev Cancer Biol. 2020 Mar 9;**4**(1):315–30.
62. Schuh L, Saint-Antoine M, Sanford EM, Emert BL, Singh A, Marr C, et al. Gene networks with transcriptional bursting recapitulate rare transient coordinated high expression states in cancer. Cell Syst. 2020 Apr 22;**10**(4):363–78.e12.
63. Harmange G, Reyes Hueros RA, Schaff D, Emert B, Saint-Antoine M, Kim LC, et al. Disrupting cellular memory to overcome drug resistance. Nat Commun. 2023 Nov 6;14.
64. Vogt W. Gestaltungsanalyse am Amphibienkeim mit Örtlicher Vitalfärbung. Wilhelm Roux Arch Entwickl Mech Org. 1929 Jun 1;**120**(1):384–706.
65. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. Science. 1994 Feb 11;**263**(5148):802–5.
66. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ. Primary structure of the Aequorea victoria green-fluorescent protein. Gene. 1992 Feb 15;**111**(2):229–33.
67. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature. 2009 May 21;**459**(7245):428–32.
68. Roux J, Hafner M, Bandara S, Sims JJ, Hudson H, Chai D, et al. Fractional killing arises from cell-to-cell variability in overcoming a caspase activity threshold. Mol Syst Biol. 2015 May 7;**11**(5):803.
69. Yang C, Tian C, Hoffman TE, Jacobsen NK, Spencer SL. Melanoma subpopulations that rapidly escape MAPK pathway inhibition incur DNA damage and rely on stress signalling. Nat Commun. 2021 Mar 19;**12**(1):1747.
70. Meyer M, Paquet A, Arguel MJ, Peyre L, Gomes-Pereira LC, Lebrigand K, et al. Profiling the non-genetic origins of cancer drug resistance with a single-cell functional genomics approach using predictive cell dynamics. Cell Syst. 2020 Oct 21;**11**(4):367–74.e5.
71. Kaufman T, Nitzan E, Firestein N, Ginzberg MB, Iyengar S, Patel N, et al. Visual barcodes for clonal-multiplexing of live microscopy-based assays. Nat Commun. 2022 May 18;**13**(1):2725.
72. Arwert EN, Lal R, Quist S, Rosewell I, van Rooijen N, Watt FM. Tumor formation initiated by nondividing epidermal cells via an inflammatory infiltrate. Proc Natl Acad Sci U S A. 2010 Nov 16;**107**(46):19903–8.
73. Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature. 2007 Nov 1;**450**(7166):56–62.
74. Boone PG, Rochelle LK, Ginzel JD, Lubkov V, Roberts WL, Nicholls PJ, et al. A cancer rainbow mouse for visualizing the functional genomics of oncogenic clonal expansion. Nat Commun. 2019 Dec 2;**10**(1):5490.
75. Roy E, Wong HY, Villani R, Rouille T, Salik B, Sim SL, et al. Regional variation in epidermal susceptibility to UV-induced

carcinogenesis reflects proliferative activity of epidermal progenitors. Cell Rep. 2020 Jun 2;**31**(9):107702.
76. Snippert HJ, van der Flier LG, Sato T, van Es JH, van den Born M, Kroon-Veenboer C, et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell. 2010 Oct 1;**143**(1):134–44.
77. Schepers AG, Snippert HJ, Stange DE, van den Born M, van Es JH, van de Wetering M, et al. Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. Science. 2012 Aug 10;**337**(6095):730–5.
78. Snippert HJ, Schepers AG, van Es JH, Simons BD, Clevers H. Biased competition between Lgr5 intestinal stem cells driven by oncogenic mutation induces clonal expansion. EMBO Rep. 2014 Jan;**15**(1):62–9.
79. Maddipati R, Stanger BZ. Pancreatic cancer metastases harbor evidence of polyclonality. Cancer Discov. 2015 Oct;**5**(10):1086–97.
80. Cheung KJ, Padmanaban V, Silvestri V, Schipper K, Cohen JD, Fairchild AN, et al. Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. Proc Natl Acad Sci U S A. 2016 Feb 16;**113**(7):E854–63.
81. Rios AC, Capaldo BD, Vaillant F, Pal B, van Ineveld R, Dawson CA, et al. Intraclonal plasticity in mammary tumors revealed through large-scale single-cell resolution 3D imaging. Cancer Cell. 2019 Apr 15;**35**(4):618–32.e6.
82. Tiede S, Kalathur RKR, Lüönd F, von Allmen L, Szczerba BM, Hess M, et al. Multi-color clonal tracking reveals intra-stage proliferative heterogeneity during mammary tumor progression. Oncogene. 2021 Jan;**40**(1):12–27.
83. Cai D, Cohen KB, Luo T, Lichtman JW, Sanes JR. Improved tools for the Brainbow toolbox. Nat Methods. 2013 May 5;**10**(6):540–7.
84. Oshimori N, Oristian D, Fuchs E. TGF-β promotes heterogeneity and drug resistance in squamous cell carcinoma. Cell. 2015 Feb 26;**160**(5):963–76.
85. Muzumdar MD, Dorans KJ, Chung KM, Robbins R, Tammela T, Gocheva V, et al. Clonal dynamics following p53 loss of heterozygosity in Kras-driven cancers. Nat Commun. 2016 Sep 2;**7**:12685.
86. Yao M, Ventura PB, Jiang Y, Rodriguez FJ, Wang L, Perry JSA, et al. Astrocytic trans-differentiation completes a multicellular paracrine feedback loop required for medulloblastoma tumor growth. Cell. 2020 Feb 6;**180**(3):502–20.e19.
87. Zong H, Espinosa JS, Su HH, Muzumdar MD, Luo L. Mosaic analysis with double markers in mice. Cell. 2005 May 6;**121**(3):479–92.
88. Liu C, Sage JC, Miller MR, Verhaak RGW, Hippenmeyer S, Vogel H, et al. Mosaic analysis with double markers reveals tumor cell of origin in glioma. Cell. 2011 Jul 22;**146**(2):209–21.
89. Weber K, Bartsch U, Stocking C, Fehse B. A multicolor panel of novel lentiviral 'gene ontology' (LeGO) vectors for functional gene analysis. Mol Ther. 2008 Apr;**16**(4):698–706.
90. McQuerry JA, Chen J, Chang JT, Bild AH. Tepoxalin increases chemotherapy efficacy in drug-resistant breast cancer cells overexpressing the multidrug transporter gene ABCB1. Transl Oncol. 2021 Oct;**14**(10):101181.
91. Berthelet J, Wimmer VC, Whitfield HJ, Serrano A, Boudier T, Mangiola S, et al. The site of breast cancer metastases dictates their clonal composition and reversible transcriptomic profile. Sci Adv [Internet]. 2021 Jul;**7**(28). Available from: http://dx.doi.org/10.1126/sciadv.abf4408
92. Michaelis M, Selt F, Rothweiler F, Wiese M, Cinatl J Jr. ABCG2 impairs the activity of the aurora kinase inhibitor tozasertib but not of alisertib. BMC Res Notes. 2015 Sep 28;**8**:484.
93. Mohme M, Maire CL, Riecken K, Zapf S, Aranyossy T, Westphal M, et al. Optical barcoding for single-clone tracking to study tumor heterogeneity. Mol Ther. 2017 Mar 1;**25**(3):621–33.
94. Gambera S, Abarrategi A, González-Camacho F, Morales-Molina Á, Roma J, Alfranca A, et al. Clonal dynamics in osteosarcoma defined by RGB marking. Nat Commun. 2018 Sep 28;**9**(1):3994.
95. Roh V, Abramowski P, Hiou-Feige A, Cornils K, Rivals JP, Zougman A, et al. Cellular barcoding identifies clonal substitution as a hallmark of local recurrence in a surgical model of head and neck squamous cell carcinoma. Cell Rep. 2018 Nov 20;**25**(8):2208–22.e7.
96. Cornils K, Thielecke L, Hüser S, Forgber M, Thomaschewski M, Kleist N, et al. Multiplexing clonality: combining RGB marking and genetic barcoding. Nucleic Acids Res. 2014 Apr;**42**(7):e56.
97. Pillai M, Hojel E, Jolly MK, Goyal Y. Unraveling non genetic heterogeneity in cancer with dynamical models and computational tools. Nat Comput Sci. 2023 Apr 24;**3**:301–13.
98. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017 Oct;**14**(10):979–82.
99. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014 Apr;**32**(4):381–6.
100. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature. 2019 Feb;**566**(7745):496–502.
101. Aissa AF, Islam ABMMK, Ariss MM, Go CC, Rader AE, Conrardy RD, et al. Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. Nat Commun. 2021 Mar 12;**12**(1):1628.
102. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018 Aug;**560**(7719):494–8.
103. Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat Biotechnol. 2020 Dec;**38**(12):1408–14.
104. Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, et al. Mapping transcriptomic vector fields of single cells. Cell. 2022 Feb 17;**185**(4):690–711.e45.
105. Wang W, Poe D, Yang Y, Hyatt T, Xing J. Epithelial-to-mesenchymal transition proceeds through directional destabilization of multidimensional attractor. Elife [Internet]. 2022 Feb 21;11. Available from: http://dx.doi.org/10.7554/eLife.74866
106. Fan J, Slowikowski K, Zhang F. Single-cell transcriptomics in cancer: computational challenges and opportunities. Exp Mol Med. 2020 Sep;**52**(9):1452–65.
107. Couturier CP, Ayyadhury S, Le PU, Nadaf J, Monlong J, Riva G, et al. Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. Nat Commun. 2020 Jul 8;**11**(1):3406.
108. Schmidt M, Mortensen LS, Loeffler-Wirth H, Kosnopfel C, Krohn K, Binder H, et al. Single-cell trajectories of melanoma cell resistance to targeted treatment. Cancer Biol Med. 2021 Oct 1;**19**(1):56–73.

109. Zhang Q, He Y, Luo N, Patel SJ, Han Y, Gao R, et al. Landscape and Dynamics of Single Immune Cells in Hepatocellular Carcinoma. Cell. 2019 Oct 31;**179**(4):829–45.e20.
110. Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM. Fundamental limits on dynamic inference from single-cell snapshots. Proc Natl Acad Sci U S A. 2018 Mar 6;**115**(10):E2467–76.
111. Weinreb C, Klein AM. Lineage reconstruction from clonal correlations. Proc Natl Acad Sci U S A. 2020 Jul 21;**117**(29):17041–8.
112. Gorin G, Fang M, Chari T, Pachter L. RNA velocity unraveled. PLoS Comput Biol. 2022;**18**(9): e1010492. https://doi.org/10.1371/journal.pcbi.1010492
113. Bergen V, Soldatov RA, Kharchenko PV, Theis FJ. RNA velocity-current challenges and future perspectives. Mol Syst Biol. 2021 Aug;**17**(8):e10282.
114. Richman LP, Goyal Y, Jiang CL, Raj A. ClonoCluster: a method for using clonal origin to inform transcriptome clustering. Cell Genomics. 2023 Jan 12;100247.
115. Wang SW, Herriges MJ, Hurley K, Kotton DN, Klein AM. CoSpar identifies early cell fate biases from single-cell transcriptomic and lineage information. Nat Biotechnol [Internet]. 2022 Feb 21. Available from: http://dx.doi.org/10.1038/s41587-022-01209-1
116. Leibovich N, Goyal S. Limitations and optimizations of cellular lineages tracking [Internet]. bioRxiv. 2023 [cited 2023 Mar 21]. p. 2023.03.15.532767. Available from: https://www.biorxiv.org/content/10.1101/2023.03.15.532767v1.full
117. Zhang Z, Melzer M, Arun KM, Sun H, Eriksson, CJ, Fabian I, et al. Synthetic DNA barcodes identify singlets in scRNA-seq datasets and evaluate doublet algorithms. Cell Genomics. 2024 July 10;**4**(7).
118. Holze H, Talarmain L, Fennell KA, Lam EY, Dawson MA, and Vassiliadis D. Analysis of synthetic cellular barcodes in the genome and transcriptome with BARtab and bartools. Cell Rep Methods. 2024 Apr 25;**4**(5).
119. Gao R, Bai S, Henderson YC, Lin Y, Schalck A, Yan Y, et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. Nat Biotechnol. 2021 May;**39**(5):599–608.
120. Satas G, Zaccaria S, Mon G, Raphael BJ. SCARLET: single-cell tumor phylogeny inference with copy-number constrained mutation losses. Cell Syst. 2020 Apr 22;**10**(4):323–32.e8.
121. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011 Apr 7;**472**(7341):90–4.
122. Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: a generalist algorithm for cellular segmentation. Nat Methods. 2021 Jan;**18**(1):100–6.
123. Pachitariu M, Stringer C. Cellpose 2.0: how to train your own model. Nat Methods. 2022 Dec;**19**(12):1634–41.
124. Magnusson KEG, Jalden J, Gilbert PM, Blau HM. Global linking of cell tracks using the Viterbi algorithm. IEEE Trans Med Imaging. 2015 Apr;**34**(4):911–29.
125. Tian C, Yang C, Spencer SL. EllipTrack: a global-local cell-tracking pipeline for 2D fluorescence time-lapse microscopy. Cell Rep. 2020 Aug 4;**32**(5):107984.
126. Pi D, Braun J, Dutta S, Patra D, Bougaran P, Mompeon A. et al. Resolving the design principles that control postnatal vascular growth and scaling [Internet]. bioRxiv. 2024 [cited 2025 Mar 24]. p. 2024.12.10.627758; doi: https://doi.org/10.1101/2024.12.10.627758.
127. Bao F, Deng Y, Wan S, Shen SQ, Wang B, Dai Q, et al. Integrative spatial analysis of cell morphologies and transcriptional states with MUSE. Nat Biotechnol [Internet]. 2022 Mar 28. Available from: http://dx.doi.org/10.1038/s41587-022-01251-z
128. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AAM, Mao M, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature. 2002 Jan 31;**415**(6871):530–6.
129. Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D, et al. Artificial intelligence for multimodal data integration in oncology. Cancer Cell. 2022 Oct 10;**40**(10):1095–110.

SECTION 4

Computational approaches to drug development

15

Navigating protein dynamics: Bridging the gap with deep learning and machine intelligence

Supriyo Bhattacharya

15.1. Introduction

Proteins, specialized polymers constituting the core of cellular machinery, perform diverse functions ranging from gene transcription to enzymatic reactions and signalling. The intricate structure of a protein, shaped by its amino acid sequence, chemical environment, and post-translational modifications (PSMs), is a major determinant of cellular function. Under physiological conditions, many proteins fold into stable structures termed the native state, while others persist in partially or fully disordered states. Due to the thermal effect, all proteins exhibit structural variations that range from subtle fluctuations around the native state (rigid proteins) to collective domain motions (partially flexible proteins) and large-scale structural variations (fully flexible disordered proteins). The distribution of structures or conformations that proteins adopt constitutes the structural ensemble, and the temporal trajectories associated with the exchange among these conformations constitute protein dynamics.

Natural variation in protein structure is biologically relevant in many cellular processes, including enzyme function, interaction with partner proteins, muscle contraction, and actin cytoskeleton formation, to name a few. The timescale of transition between protein conformations also has an important impact on cellular function. For example, such transition timings influence the kinetics of enzymatic processes (e.g. phosphorylation and dephosphorylation cycles) [1] and the binding and unbinding of partner proteins and transcription factors [2], thus regulating the timescales of key cellular events such as cell-state transitions. Dysregulation of protein dynamics through mutations or large-scale sequence alterations has been linked to disease phenotypes, including cancer and neurodegenerative ailments [3].

Over the past several decades, a diverse array of *in silico* methods has emerged to mechanistically link protein dynamical characteristics with physicochemical properties (e.g. amino acid sequence and PSM) and the extrinsic environment (e.g. temperature, pH, binding of small molecule ligands, and partner proteins). Molecular dynamics (MD), a primary approach, simulate protein dynamics at the atomistic resolution by modelling the protein as a collection of atoms, surrounded by ions and solvent molecules. Forces among these atoms are modelled using analytical functions of distance and chemical properties (force fields [4]), and the resulting motions are modelled using classical mechanics (Newton's laws of motion) [5]. The spatiotemporal trajectory of the entire system is then obtained by integrating a set of coupled differential equations over time. MD is a physics-based interpretable method that simulates protein dynamics in its physiological environment while providing key mechanistic insights into the temporal evolution of protein trajectories, such as the role of specific inter-residue contacts.

Despite the strengths of MD, key challenges in scalability and simulating longer timescales persist [6]. Computational time scales with the square of the number of atoms and the integration time step in MD, typically in the femtosecond (10^{-15} s) range, faces limitations when exploring biologically relevant timescales extending to milliseconds or even seconds [5]. Proteins experience thermal motions ranging from high-frequency vibrations to slower domain movements (**Figure 15.1**). The integration timestep in MD is therefore constrained by the highest frequency vibrational modes. Given the current computational resources available to the community, MD-based methods can explore protein dynamics typically up to several microseconds and, for smaller proteins, milliseconds [7]. Thus, there exists a substantial gap between timescales explored by MD and biologically relevant timescales, necessitating novel methods.

Addressing these challenges, machine learning methods come into play, aiming to extract key features and their intricate correlations from existing data. Generative AI, a subset of deep learning (DL) or artificial intelligence (AI), efficiently creates data resembling training datasets based on learned features [8]. Given their ability to comprehend complex co-dependencies, generative AI methods have found applications in diverse domains, including image and voice synthesis, speech recognition, and bioinformatic data imputation [9–12]. Transformer and recurrent neural-network-based

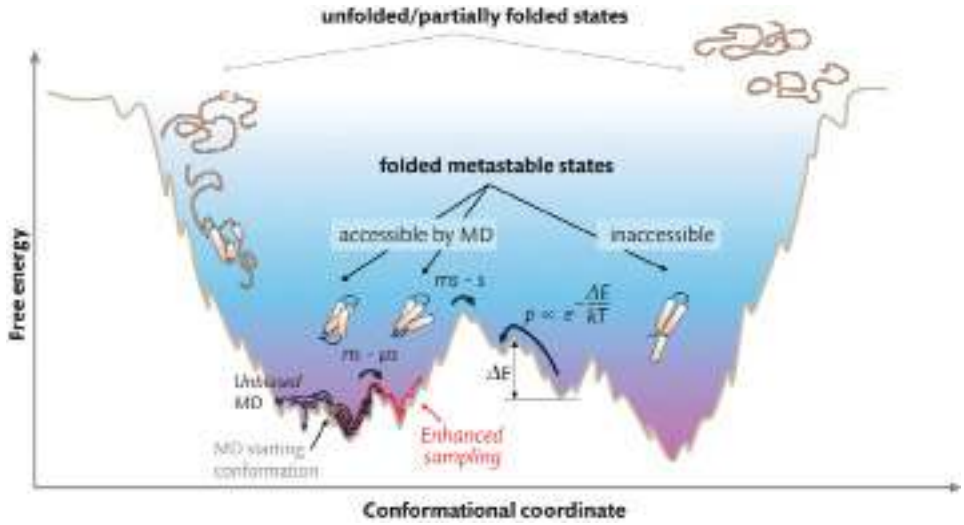


Figure 15.1. Schematic representing the two-dimensional projection of a multidimensional protein free energy landscape showing local minima and maxima representing metastable states and energy barriers, respectively. The probability p of transitioning into a neighbouring metastable state scales as the inverse exponential of barrier height $\Delta E\left(p \propto exp\left[-\frac{\Delta E}{kT}\right]\right)$, where k is the Boltzmann constant and T is the temperature; consequently, the timescale of barrier crossing ($1/p$) scales as the exponential of barrier height.

approaches, adept at discerning hidden patterns within sequences, offer promise in predicting temporal data, including those derived from MD trajectories.

While the application of AI in protein dynamics is still emerging, notable progress is evident from the increasing number of publications in this field. This chapter seeks to provide a concise overview of the synergies between AI and protein dynamics, shedding light on the potential of these innovative methodologies in advancing our understanding of cellular processes at the molecular level. Notably, a few excellent, detailed reviews on this subject already exist, and the reader is strongly encouraged to consult them for further information [13,14]. Given the rapidly evolving nature of DL, we discuss the main challenges associated with classical MD methods and how AI can be used to address them. We augment these discussions with a few chosen examples, focusing on the latest literature in a non-technical manner, appealing to a broader audience.

15.2. Classical approaches for improving MD-based conformational sampling

While, in theory, conventional MD is capable of reproducing realistic kinetics of protein systems, in reality, it is faced with serious challenges limiting its applicability to large macromolecular structures. In a thermodynamic ensemble (distribution of all possible configurations of a system under a given thermodynamic condition), each state is associated with an energy value, contributed by the interactions among its components. The probability that the system exists in a given state at a constant temperature follows the Boltzmann distribution, scaling inverse exponentially with energy [15]. Due to numerous interatomic/solvent interactions and a crowded cellular environment, energies associated with protein conformations tend to be highly variable, leading to a 'rugged' energy landscape with local wells separated by barriers [16] (Figure 15.1). During MD, the average timescale of escaping from a local well scales exponentially with the barrier height (the energy difference between the highest and lowest points within the well) (Figure 15.1). This leads to frequent scenarios where an MD simulation will get 'stuck' in a metastable state and not sample the rest of the conformational space, leading to incorrect statistics [17]. Enhanced sampling methods address these issues by combining MD simulations with clever statistical schemes [18]. These methods can be categorized into two broad classes: (1) trajectory ensemble based and (2) biased energy landscape based.

In trajectory ensemble-based methods, multiple MD trajectories are generated for the same system by varying parameters such as temperature or random seed (adaptive seeding). Utilizing cluster computing to conduct parallel MD simulations enables efficient exploration of a broader conformational space, surpassing the capabilities of a single extended simulation. Replica exchange and weighted ensemble methods fall into this category. In replica exchange methods, parallel independent (replica) simulations are run by varying parameters such as temperature (temperature replica exchange) or force field (Hamiltonian replica exchange), with frequent exchange of these parameters among the trajectories. These exchanges facilitate low-temperature replicas to overcome energy barriers, improving the overall efficiency of exploring the conformational landscape [19]. The weighted ensemble method generates multiple MD trajectories of the same system and weights these trajectories based on their progress towards a desired conformational state (e.g. the folded state of a protein). Trajectories with higher weights that sample the desirable part of the energy landscape are replicated, while those that sample non-desirable conformations are terminated. Weights of the trajectories are readjusted at every step to preserve the statistics of the conformational ensemble. In cases where an energy barrier (rare event) separates the desired state from the starting state, weighted ensemble methods can increase the probability of the system to sample rare events and overcome energy barriers [20,21].

Biased energy landscape-based methods seek to increase the efficiency of sampling the energy landscape within a single trajectory by introducing bias in the intermolecular forces. The bias can be targeted towards reaching a desired state (e.g. steered MD [22]) or enhance the sampling along a reaction coordinate (RC; one-dimensional parameter that best describes the desired conformational change over time) defined using a few chosen collective variables (CVs; low-dimensional functions of molecular coordinates that capture key aspects of coordinated protein movements) by discouraging the system from sampling already explored states (e.g. metadynamics [23]). Non-targeted bias is also possible, such as in the case of accelerated MD, where the force bias is introduced to artificially lower energy barriers separating local wells [24]. This facilitates a wider coverage of the conformational space at the expense of sampling many irrelevant conformations. Proper reweighing of the trajectories must be performed post-simulation to recover accurate statistics, and this can be quite challenging, especially for larger systems [25].

Despite proven success in many cases, enhanced sampling MD methods are not without their challenges. Replica exchange methods can be computationally expensive, due to the need to simulate many replicas in parallel, often ranging in the hundreds. This requires large computing clusters with fast information exchange among the processors. The efficiency of weighted ensemble methods relies on optimized binning along the RCs and appropriate pruning and merging strategies for low- and high-weight trajectories [21]. Biased energy landscape methods, such as metadynamics, are sensitive to the choice of relevant CVs [26]. To address some of these challenges, hybrid methods have been suggested that combine multiple enhanced sampling strategies to harness the added advantages of each method, e.g. replica exchange with collective-variable tempering (metadynamics) [27]. Although effective in small systems, these methods suffer in efficiency while sampling larger systems, such as multi-protein complexes. The performance of current MD-based approaches is limited by the integration timestep (1–4 fs), which is determined by the highest frequency motions in the system, namely the bond vibrations. Therefore, new innovative methods are needed to predict the time evolution of molecular systems that circumvent sequential integration along the time axis using small timesteps. In recent decades, advancement in AI using DL has provided such an opportunity.

15.3. Overview of deep learning

AI is a field of computer algorithms whose goal is to perform complex tasks that require human intelligence but in a scaled-up automated manner that surpasses human capacity. To learn how to perform specific tasks, AI systems require large volumes of training data. DL is a subfield within AI that employs multiple layers of artificial neural networks (ANNs) to learn complex patterns and interdependencies within training datasets [28]. DL has found widespread applications in fields, such as image and speech recognition and natural language processing.

The architecture of ANNs is inspired by the organization of neurons within biological brains [29]. ANNs are computational models consisting of layers of interconnected nodes (neurons), each with its own parameters, i.e. weights and biases (these parameters are optimized during the training process using existing data). The input data is transmitted across the ANN layers via the interconnections while undergoing mathematical transformations at each node according to pre-defined (typically non-linear) activation functions. Each layer learns to extract features from the data, and the subsequent layers learn to combine these features to make increasingly abstract representations. The output from DL is a set of predictions based on the input data.

During prediction or classification, DL methods capture the intricate variability within input data, often characterized by numerous dimensions, by distilling them into a concise set of key features. These features are then utilized to forecast outcomes. Thanks to their hierarchical architecture and use of non-linear functions, DL approaches bring increased flexibility to predictive modelling compared to conventional approaches of dimensionality reduction such as principal component analysis or matrix factorization (using linear functions) that are typically employed in analysing MD data. Moreover, the key features learned through DL can be leveraged to generate new data (generative AI), preserving the statistical relationships and patterns found in the training dataset. Generative AI has demonstrated its capability for generating authentic content such as artificial human faces, voices, and written text. Notably, it exhibits potential in constructing extended timescale MD trajectories based on existing simulation data, a topic we will delve into shortly. However, prior to that discussion, let's briefly explore the DL methodologies frequently applied within the MD domain.

15.3.1. Autoencoders and VAEs

Autoencoders are constructed using two ANNs, an encoder and a decoder [30]. The encoder ANN processes the input data to express it in terms of a limited number of features (information bottleneck), also known as latent variables. The decoder ANN then uses the latent variables to reconstruct the input data as accurately as possible. Running the input data through an information bottleneck forces the autoencoder to learn the fundamental features that contribute to the diversity and codependence within the input data and reject statistical noise [31]. Variational autoencoders (VAEs) retain the autoencoder architecture while incorporating variational inference for training purposes [32]. Instead of directly optimizing the model parameters, it optimizes a lower bound on the likelihood of the data while constraining the latent variables to follow a parametric statistical function such as the Gaussian distribution. This introduces a probabilistic element to the latent space, where each point in the latent space corresponds to a probability distribution rather than a single point. By sampling in the latent variable space and processing through the decoder ANN, new data can be generated that retains the statistical dependencies and patterns found in the input data. For example, VAEs can be trained using MD data to represent and explore protein conformational landscapes in the latent variable space. By doing so, certain problems associated with conventional MD (e.g. difficulty in overcoming energy barriers) can be potentially resolved. Protein conformations generated by VAE, applied in the conventional way, do not incorporate any temporal relationship among the conformations (there are specialized autoencoders that can learn temporal relationships; see time-lagged VAE [33] later in the discussion). However, certain generative AI methods, such as transformers, can learn temporal sequences as discussed later.

15.3.2. Generative adversarial network

Generative adversarial networks (GANs) are generative AI frameworks comprised of two neural networks, the generator and the discriminator [34]. The generator creates new data similar to the training data, while the discriminator tries to determine whether the generated data is real (from the training dataset) or fake (generated). Both the generator and the discriminator are refined through adversarial training such that, ultimately, the data produced by the generator becomes indistinguishable from real data. GANs have been successfully applied in generating realistic content, such as images, text, music, and speech. Trained using MD-derived ensembles, GANs have the potential to generate realistic protein structures.

15.3.3. Transformers

Transformers are DL methods used in learning sequential data such as written language and temporal variations [35]. Transformers use a mechanism called self-attention that assigns certain parts of the input sequence (e.g. certain words in a sentence) higher weights (attention scores) than others based on their importance in predicting future sequences. In addition, transformers analyse different lengths of input sequences in parallel (multi-head attention), deriving multiple weights for the same word and capturing both short- and long-range context-dependent relevance. By pooling weights obtained from multi-head attention mechanisms, transformers can efficiently learn the association and contextual relevance of each word in a sentence, even when those associations are only apparent within long stretches of input sequences. Although the previous description used written language as an example, the data used by transformers need not only be sequences of words but they can also be, e.g., sequences of protein conformations observed in a MD trajectory. Transformers can learn the relationships among the conformations that follow the temporal sequence and predict future time trajectories [36]. Besides temporal data, transformers have also been used to learn amino acid sequence–structure relationships to predict a single protein structure (e.g. AlphaFold2 [37]) or structural ensembles (e.g. idpGAN [38]), when trained using MD data from multiple proteins.

15.4. Deep learning in MD

In recent years, machine learning and especially DL have been applied to various aspects of MD-driven exploration of protein conformations, ranging from quantum-mechanical force-field calculation [39], transition path [40,41] and rare event sampling (e.g. in conjunction with quantum computing [42]), developing coarse-grain force fields [43], to generating MD trajectories [44] and improving the efficiency of enhanced sampling algorithms [45]. There are still many challenges in applying DL methods to MD. In brief, some of the major tasks involved in implementing DL in the MD field are as follows:

(1) Transforming chemical and 3D molecular coordinates into machine-learnable features.
(2) Encoding kinetic relationships using spatiotemporal features from MD trajectories.
(3) Forecasting future system states based on learned past trends using available (sometimes limited) MD data.

Each of these steps involves many adjustable parameters that need to be determined using system knowledge or through optimization operations. The model parameters and, in some cases, the choice of the model itself (e.g. the ANN topology and feature engineering approach) could depend on the specific protein system under study.

15.4.1. Encoding of machine-learnable features from 3D molecular coordinates

The fundamental information obtained from MD simulations is the three-dimensional (Cartesian) coordinates of protein and solvent atoms for a finite number of temporal snapshots. While Cartesian coordinates are generally not deemed as an optimal input format for DL [33], exceptions exist [38,41,46]. Instead, spatial relationships among protein atoms (e.g. interatomic distances and angles among atom triplets) and internal coordinates (dihedral angles formed by four consecutive protein atoms) derived from Cartesian coordinates proved to be more suitable for training DL models since such descriptions are invariant to centre-of-mass rotation and translation [47,48]. For example, interatomic distance matrices derived from multiple MD frames can be converted into images for training generative AI models based on convolutional neural networks (CNNs) [49] (Figure 15.2A). Alternatively, the protein can be modelled as a graph topology (individual protein atoms as nodes and bonds as graph edges) and used in training graph neural networks (GNNs) [43].

The resultant DL output derived from the transformed input features can be converted back into Cartesian coordinates using suitable force fields or protein chemistry-based restraints (e.g. dihedral restraints based on Ramachandran diagrams of individual amino acids). Some approaches amalgamate the aforementioned principles. For example, leveraging the inter-residue connectivity, CNNs can be trained by performing convolution along the protein sequence using a one-dimensional kernel [41]. Here, the convolution is executed over pre-defined windows along the amino acid chain (utilizing Cartesian coordinates of protein atoms within the local window), thereby encoding the molecular topology in the training process. Notably, in this specific instance, the authors integrated a physics-based force field into the training, generating physically plausible conformations from undersampled regions in the protein landscape that were not present in the training data. The idpGAN method incorporated amino acid sequence information in the training process using a transformer architecture within a GAN framework [38]. Trained using (coarse-grained) MD-derived protein structures, idpGAN was shown to generate structural ensembles of intrinsically disordered proteins, exhibiting amino acid sequence-specific characteristics.

DL agents trained using structural ensembles can be effective in sampling the flexible degrees of freedom in a protein molecule. Such approaches have practical application in tasks, such as protein–protein docking or ensemble docking of small molecules, and have been shown in certain cases to outperform classical methods of conformation generation, e.g. sampling along the top principal components [46]. The DL approaches discussed so far are intended to generate protein ensembles, where the structures need not follow any temporal sequence. For DL agents to predict the temporal sequence of protein conformations, kinetic information must be integrated into the training process, as discussed next.

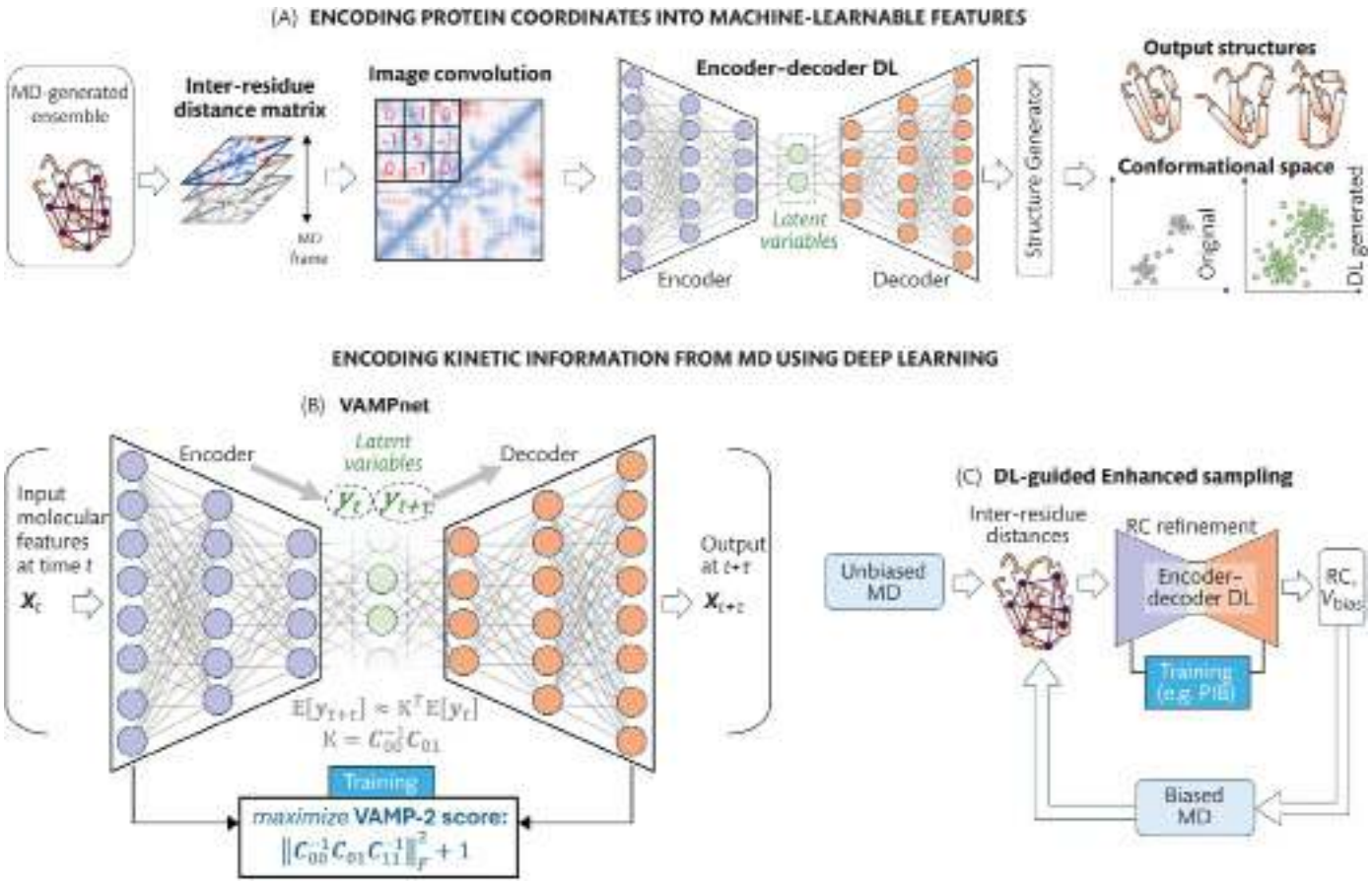


Figure 15.2. Description of key aspects of modelling MD data using DL. (A) Example pipeline for encoding protein coordinates from MD into machine-learnable features for training a convolutional neural-network-based autoencoder. (B) and (C) Examples of encoding kinetic relationships from MD using DL. (B) VAMPnet: Using the input molecular features (e.g. inter-residue distances) at time t, encoder network generates latent variables y_t, which are related to $y_{t+\tau}$ at time delay τ through a linear transformation (Markov approximation from Koopman theory). Using transformed variables $y_{t+\tau}$, decoder then generates the molecular features at time $t+\tau$. The encoder–decoder DL is trained by maximizing the VAMP-2 score. C_{00} and C_{11} are mean variances of y_t and $y_{t+\tau}$, respectively, while C_{01} is the mean covariance between y_t and $y_{t+\tau}$. (C) General framework for DL-guided enhanced sampling. Iterations between DL and biased MD guide the simulations along a preferred reaction coordinate (RC) that facilitates escape from metastable states. The enhancer–decoder DL network learns the kinetic relationships in the input data using a predetermined statistical training objective, such as the past–future information bottleneck (PIB) (55). With each iteration, a new improved RC and corresponding bias potential V_{bias} are derived, which guide the next round of MD. The process converges when the derived RC does not change further.

15.4.2. Encoding kinetic relationships using MD data

Mathematically, MD simulations behave as dynamical systems where the system state (coordinates of thousands of protein and solvent atoms and their velocities) at time $t + \Delta t$ (Δt: integration timestep, 1–4 fs) is a non-linear function of the state at time t. The non-linearity arises due to the physics-based force-field functions used in computing pairwise interatomic interactions. Thus, the computation time scales as the square of the number of atoms. Secondly, the highest frequency motions (i.e. bond vibrations) suggest an integration timestep to be limited to a few femtoseconds. Together, these two factors are responsible for inefficient computation, leading to the challenge of exploring biologically relevant timescales, as discussed in the introduction. These challenges have been addressed recently by treating MD as a stochastic process combined with DL-based transformation of the system states to more manageable latent variable spaces. Exciting developments have emerged from these efforts, as discussed later.

Protein dynamics can also be described as stochastic processes where the probability of future states depends on past states. A simplification of this principle results in the Markov model, where the probability of observing a given conformation at time $t + \tau$ (τ is referred to as lag time) depends only on the conformation at time t and not on any other times in history. Due to the non-linearity mentioned earlier, casting protein dynamics as a Markov process is inherently challenging. The existing Markov state model (MSM) approach approximates the protein conformation space by clustering an ensemble of MD trajectories into kinetically related microstates (a heuristic yet practical approach) [50]. The term 'kinetically related' means that protein structures within a microstate are likely to experience frequent interconversions over time, while the transition rate between microstates depends on the adjoining energy barrier. The MSM is then constructed by estimating transition rates among the microstates over a suitable lag time that captures slow meaningful conformational changes, filtering out fast transitions that are likely to be thermal noise. In theory, once all relevant microstates and the rules governing their temporal exchanges

have been enumerated, future time evolution can be predicted. In practice, however, MSM suffers from insufficient statistics and the challenge to identify relevant microstates, especially the ones representing rare events. Recent machine-learning-based approaches have shown how to address these challenges, leading to improved Markov models of MD-based systems.

It is well known that, using an appropriate transformation of the original state variables into a latent space, any non-linear system can be recast such that the system state (in the latent space) at time $t + \tau$ is obtainable through a linear transformation of the latent variables at time t [51]. Although this formulation is exact when the dimension of the latent space approaches infinity, an approximate solution is achievable using a limited number of dominant latent variables and a sufficiently long lag time τ [44]. This leads to the variational principle for Markov process (VAMP) [52,53] and its DL-based implementation to predict protein dynamics, VAMPnet (**Figure 15.2**B) [44]. VAMPnet-based approaches seek to find an optimal latent space transformation of the protein coordinates (representing the best approximation of an MD system as a Markov model) by training neural networks against available MD data. The quality of the transformation can be assessed by evaluating the VAMP score [44,54]. The VAMP-based approach was shown to be superior to MSM (albeit in simulations of simple systems such as alanine dipeptide) in approximating protein dynamics as Markov models, although their practicality in handling large macromolecular systems is yet to be evaluated [44].

The encoder–decoder DL architecture (such as the one used in VAMPnet) has emerged as a suitable choice for learning kinetic relationships in protein dynamics since they challenge DL to recreate input data patterns using a limited number of latent variables (information bottleneck). This ensures that the key (typically non-linear) relationships among the input variables responsible for broader kinetics are captured rather than thermal fluctuations and noise. Examples of DL methods employing the encoder–decoder architecture are time-lagged autoencoder (TAE) [33] and the past–future information bottleneck (PIB)-based approach by Tiwary and coworkers [55] (discussed in the Enhanced Sampling section). The TAE encodes the past–future relationship using an autoencoder, where the encoder neural network converts the input molecular coordinates at time t into latent variables. Then, instead of reconstructing the input, the decoder uses the latent variables to predict molecular coordinates at a future instant $t + \Delta t$ (hence the acronym 'time-lagged'). In essence, TAE derives CVs that are non-linear functions (implemented in DL) of the input coordinates. This places TAE in the general category of methods seeking to find optimal CVs, including traditional approaches such as time-lagged-independent component analysis where the CVs are modelled as linear rather than non-linear functions [33].

15.4.3. Scaling up to larger systems

DL methods such as VAMPnets, which express global dynamics using latent variables, are limited by the complexity of the conformational space and are challenging to scale up to large biomolecular systems. For proteins with intrinsically disordered domains or multi-protein assemblies, the number of relevant microstates can be prohibitively large to model using latent variables.

One potential workaround is to decompose a large system into separate components or sub-systems and learn their individual kinetics using neural nets. The global dynamics is then learned in terms of weighted contributions from the individual sub-systems [56]. This approach has led to iVAMPnet (independent Markov decomposition VAMPnet), where the protein is decomposed into semi-independent domains, whose individual kinetics can be learned using separate VAMPnets [54]. The global dynamics is then modelled by combining the dynamics from individual sub-systems (i.e. as the Kronecker product of all component states) into a single Koopman framework. During training, the optimal sub-system decomposition is achieved by maximizing kinetic independence across sub-systems while tuning the individual VAMPnets to reproduce slow kinetics within each sub-system. The positive aspect of this method is that the user only needs to specify the number of components that the system should be decomposed into. The identity of these components and their relaxation timescales are determined during the training process.

The application of the iVAMPnet method to the synaptotagmin C2A domain (whose kinetics could not be reasonably modelled using a single VAMPnet) correctly identified the sub-systems that were previously known to exhibit uncoupled dynamics [54,56]. Additionally, the identified sub-systems are biologically interpretable; in the case of synaptotagmin, they represent the calcium binding region and the loops on the opposite side of a beta sheet (refer to Figure 5 of ref. [54]). While the iVAMPnet approach is suitable for systems that can be decomposed into a discrete number of minimally coupled sub-systems, different approaches may be needed to address highly dynamic proteins (e.g. IDPs that are almost entirely disordered) without clearly independent sub-systems or where the individual sub-systems are simply too large and complicated for a single VAMPnet to model.

15.4.4. Enhanced conformational sampling using deep learning

As intricate, connected systems, protein free energy landscapes/surfaces (FESs) exhibit ruggedness, characterized by numerous metastable states. To navigate these landscapes effectively, a class of MD-based methods called adaptive sampling has emerged. These methods aim to enhance the exploration of FESs by selectively restarting multiple MD simulations from undersampled states or biasing the MD trajectory along specific RCs to increase the probability of sampling rare events (**Figure 15.2**C) [18,47).

Recent advancements have demonstrated the integration of machine learning with adaptive sampling MD simulations as a means of further enhancing the exploration of protein FESs [47,55–58]. Iterative coupling of MD with DL can streamline the identification of optimal CVs or guide the selection of microstates to restart new MD simulations from. The maximum entropy VAMPnet method (Maxent VAMPnet) combines DL with the classical approach of clustering MD conformations into microstates to guide future simulations [57]. At each step of the iterative process, the uncertainty (in the form of Shannon's entropy) of MD conformations to be assigned to one of the microstates is estimated using a VAMPnet. New MD simulations are then started from conformations with high Shannon entropy. These conformations are likely to reside in high energy regions outside of the already sampled microstates and are expected to facilitate exploration of new microstates.

The efficiency of adaptive sampling methods relies on the choice of appropriate CVs in defining an RC that captures slow principal

motions, rare events, and associated kinetics [18]. In most scenarios, CVs are system dependent and selected using human intuition rather than data-centric or algorithmic approaches. The PIB principle [55] seeks to derive CVs that are maximally predictive of the future while using as little information about the past as possible [59–61]. The method works by iterating between rounds of MD and DL, where an RC is learned at each iteration by training an autoencoder using the PIB criterion. The next round of MD is then biased along the newly derived RC. Using a few biomolecular systems (e.g. alanine dipeptide and small molecule dissociation from a protein cavity), it was shown that successive iterations improved the estimated RC in its capacity to represent the collective motion and kinetics of the system. For the small molecule (ligand) dissociation dynamics, the authors used a limited set of inter-residue and protein–ligand distances as input coordinates. So, it is conceivable that the quality of the retrieved dynamics and FES will depend on the choice of input coordinates. However, DL–MD iterative methods like the above can improve the sampling of metastable states in the FES, especially the ones representing rare events that are often associated with high kinetic barriers.

15.5. Discussion

Advancements in recent decades have seen machine learning, especially DL, being applied to nearly every domain of science and technology. Leveraging the capacity of neural networks to model arbitrarily complex mathematical functions and functionals [62,63], DL holds the potential to supplant classical statistical methods, albeit at the expense of interpretability. Numerous recent studies have showcased DL's effectiveness in analysing and enhancing MD simulations. However, like any other machine learning approaches, DL's performance hinges on the quality and quantity of training data.

Unlike many fields where training data is obtained from real-life observations or experiments, MD data is generated through simulations, often constrained by their limited accessible timescales. Given the nature of this training data, the efficacy of DL methods in forecasting long timescale protein dynamics, especially for larger systems, remains uncertain. Presently, at the nascent stage of development, DL approaches are demonstrated in small peptide or protein systems. It is unclear how these methods will scale up to larger systems, such as protein complexes, since important questions regarding the optimal number of latent variables or the required length of MD simulations for training are unresolved. However, DL presents a promising framework when coupled with enhanced sampling MD in traversing protein FESs. For instance, adaptive RCs derived from iterations between MD and DL can facilitate more efficient sampling of long timescales or rare events compared to unbiased MD of similar durations [55], or suggest new regions in the FES for extended sampling, even in large systems (e.g. the SARS-Cov2 Spike protein) [64].

Beyond temporal trajectory prediction (where successive conformations are kinetically related), generative AI techniques can derive protein ensembles by learning structural constraints from MD data. AI-generated ensembles offer significant speed advantages over traditional methods, such as Monte Carlo sampling or MD simulations, especially when accelerated using specialized hardware. These ensembles can be further refined using experimental measurements, such as small-angle X-ray scattering and nuclear magnetic resonance [65], and rapidly deployed in applications involving small molecule drug/PROTAC discovery and the design of antibodies and CAR-T.

In summary, the practical challenges of applying DL in MD remain unresolved, although recent progress suggests that the field is moving in a promising direction. Future advancements, including faster computing technologies like quantum computing, may bridge the gap between DL's modelling capabilities and physiologically relevant protein complexes, while novel algorithms could facilitate the forecasting of long timescale dynamics using limited MD data. The effective integration of machine learning with physics-based simulations holds the potential to develop highly accurate multiscale methods [66] that seamlessly traverse biological length and timescales.

REFERENCES

1. Mylona A, Theillet FX, Foster C, Cheng TM, Miralles F, Bates PA, et al. Opposing effects of Elk-1 multisite phosphorylation shape its response to ERK activation. Science. 2016;**354**(6309):233–7.
2. Stanley N, Esteban-Martín S, De Fabritiis G. Kinetic modulation of a disordered protein domain by phosphorylation. Nat Commun. 2014;**5**(1):5272.
3. Mittag T, Kay LE, Forman-Kay JD. Protein dynamics and conformational disorder in molecular recognition. J Mol Recognit. 2010;**23**(2):105–16.
4. Vanommeslaeghe K, Guvench O, MacKerell AD, Jr. Molecular mechanics. Curr Pharm Des. 2014;**20**(20):3281–92.
5. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. Neuron. 2018;**99**(6):1129–43.
6. Hartmann C, Klus S, Schütte C. Overcoming the timescale barrier in molecular dynamics: Transfer operators, variational principles and machine learning. Acta Numerica. 2023;**32**:517–673.
7. Shaw DE, Dror RO, Salmon JK, Grossman JP, Mackenzie KM, Bank JA, et al. Millisecond-scale molecular dynamics simulations on Anton. Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis. Portland, Oregon: Association for Computing Machinery; 2009. p. Article 65.
8. Aithani L, Alcaide E, Bartunov S, Cooper CDO, Doré AS, Lane TJ, et al. Advancing structural biology through breakthroughs in AI. Curr Opin Struct Biol. 2023;**80**:102601.
9. Zhan F, Yu Y, Wu R, Zhang J, Lu S, Liu L, et al. Multimodal image synthesis and editing: the generative AI era. IEEE Trans Pattern Anal Mach Intell. 2023;**45**(12):15098–119.
10. Mira R, Vougioukas K, Ma P, Petridis S, Schuller BW, Pantic M. End-to-end video-to-speech synthesis using generative adversarial networks. IEEE Trans Cybernetics. 2022.
11. Chen C, Hu Y, Yang C-HH, Siniscalchi SM, Chen P-Y, Chng ES. Hyporadise: an open baseline for generative speech recognition with large language models. arXiv preprint arXiv:230915701. 2023.
12. Camino RD, Hammerschmidt CA, State R. Improving missing data imputation with deep generative models. ArXiv. 2019;abs/1902.10666.
13. Noé F, Tkatchenko A, Müller KR, Clementi C. Machine learning for molecular simulation. Annu Rev Phys Chem. 2020;**71**:361–90.
14. Glielmo A, Husic BE, Rodriguez A, Clementi C, Noé F, Laio A. Unsupervised learning methods for molecular simulation data. Chem Rev. 2021;**121**(16):9722–58.

15. Bais FA, Farmer JD. The physics of information. In: Adriaans P, van Benthem J, editors. Philosophy of Information. Amsterdam: North-Holland; 2008. p. 609–83.
16. Milanesi L, Waltho JP, Hunter CA, Shaw DJ, Beddard GS, Reid GD, et al. Measurement of energy landscape roughness of folded and unfolded proteins. Proc Natl Acad Sci U S A. 2012;**109**(48):19563–8.
17. Wiehe K, Schmidler SC, editors. Monitoring convergence of molecular simulations in the presence of kinetic trapping. 2011. Accessed from the site: http://www2.stat.duke.edu/~scs/Papers/ConvergeDiagnos_JCTC.pdf
18. Hénin J, Lelièvre T, Shirts MR, Valsson O, Delemotte L. Enhanced sampling methods for molecular dynamics simulations. 2022 February 01;[arXiv:2202.04164 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2022arXiv220204164H.
19. Rhee YM, Pande VS. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. Biophys J. 2003;**84**(2):775–86.
20. Aristoff D, Copperman J, Simpson G, Webber RJ, Zuckerman DM. Weighted ensemble: Recent mathematical developments. J Chem Phys. 2023;**158**(1):014108.
21. Zuckerman DM, Chong LT. Weighted ensemble simulation: review of methodology, applications, and software. Annu Rev Biophys. 2017;**46**:43–57.
22. Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, et al., editors. Steered molecular dynamics. Berlin, Heidelberg: Springer; 1999.
23. Laio A, Gervasio FL. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep Prog Phys. 2008;**71**(12):126601.
24. Voter AF. Hyperdynamics: accelerated molecular dynamics of infrequent events. Phys Rev Lett. 1997;**78**(20):3908–11.
25. Miao Y, Feher VA, McCammon JA. Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. J Chem Theory Comput. 2015;**11**(8):3584–95.
26. Blumer O, Reuveni S, Hirshberg B. Stochastic resetting for enhanced sampling. J Phys Chem Lett. 2022;**13**(48):11230–6.
27. Gil-Ley A, Bussi G. Enhanced conformational sampling using replica exchange with collective-variable tempering. J Chem Theory Comput. 2015;**11**(3):1077–85.
28. Sarker IH. Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci. 2021;**2**(6):420.
29. Walczak S, Cerpa N. Artificial neural networks. In: Meyers RA, editor. Encyclopedia of physical science and technology. 3rd ed. New York: Academic Press; 2003. p. 631–45.
30. Kramer MA. Nonlinear principal component analysis using autoassociative neural networks. AIChE J. 1991;**37**(2):233–43.
31. Tapia NI, Estévez PA. On the information plane of autoencoders. In2020 International Joint Conference on Neural Networks (IJCNN) 2020 Jul 19 (pp. 1–8). IEEE.
32. Pinheiro Cinelli L, Araújo Marins M, Barros da Silva EA, Lima Netto S. Variational autoencoder. In: Variational methods for machine learning with applications to deep networks. Cham: Springer International Publishing; 2021. p. 111–49.
33. Wehmeyer C, Noé F. Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. J Chem Phys. 2018;**148**(24):241703.
34. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: an overview. IEEE Signal Process Mag. 2018;**35**(1):53–65.
35. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. Advances in neural information processing systems. 2017;30.
36. Tsai S-T, Kuo E-J, Tiwary P. Learning molecular dynamics with simple language model built upon long short-term memory neural network. Nat Commun. 2020;**11**(1):5115.
37. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;**596**(7873):583–9.
38. Janson G, Valdes-Garcia G, Heo L, Feig M. Direct generation of protein conformational ensembles via machine learning. Nat Commun. 2023;**14**(1):774.
39. Unke OT, Chmiela S, Sauceda HE, Gastegger M, Poltavsky I, Schütt KT, et al. Machine learning force fields. Chem Rev. 2021;**121**(16):10142–86.
40. Ray D. Path sampling and machine learning approaches to biomolecular simulation. UC Irvine; 2022.
41. Ramaswamy VK, Musson SC, Willcocks CG, Degiacomi MT. Deep learning protein conformational space with convolutions and latent interpolations. Phys Rev X. 2021;**11**(1):011052.
42. Ghamari D, Hauke P, Covino R, Faccioli P. Sampling rare conformational transitions with a quantum computer. Sci Rep. 2022;**12**(1):16336.
43. Husic BE, Charron NE, Lemm D, Wang J, Pérez A, Majewski M, et al. Coarse graining molecular dynamics with graph neural networks. J Chem Phys. 2020;**153**(19):194101.
44. Mardt A, Pasquali L, Wu H, Noe F. VAMPnets for deep learning of molecular kinetics. Nat Commun. 2018;**9**(1):5.
45. Shamsi Z, Cheng KJ, Shukla D. Reinforcement learning based adaptive sampling: REAPing rewards by exploring protein conformational landscapes. J Phys Chem B. 2018;**122**(35):8386–95.
46. Degiacomi MT. Coupling molecular dynamics and deep learning to mine protein conformational space. Structure. 2019;**27**(6):1034–40 e3.
47. Salawu EO. DESP: deep enhanced sampling of proteins' conformation spaces using AI-inspired biasing forces. Front Mol Biosci. 2021;**8**:587151.
48. Hoffmann M, Noé F. Generating valid Euclidean distance matrices. arXiv preprint arXiv:191003131. 2019.
49. Eguchi RR, Choe CA, Huang P-S. Ig-VAE: generative modeling of protein structure by direct 3D coordinate generation. PLOS Comput Biol. 2022;**18**(6):e1010271.
50. Pande VS, Beauchamp K, Bowman GR. Everything you wanted to know about Markov state models but were afraid to ask. Methods. 2010;**52**(1):99–105.
51. Koopman BO. Hamiltonian systems and transformation in Hilbert space. Proc Natl Acad Sci U S A. 1931;**17**(5):315–8.
52. Noé F, Nüske F. A variational approach to modeling slow processes in stochastic dynamical systems. Multiscale Model Simul. 2013;**11**(2):635–55.
53. Wu H, Noé F. Variational approach for learning Markov processes from time series data. J Nonlinear Sci. 2020;**30**(1):23–66.
54. Mardt A, Hempel T, Clementi C, Noé F. Deep learning to decompose macromolecules into independent Markovian domains. Nat Commun. 2022;**13**(1):7101.
55. Wang Y, Ribeiro JML, Tiwary P. Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. Nat Commun. 2019;**10**(1):3573.

56. Hempel T, Del Razo MJ, Lee CT, Taylor BC, Amaro RE, Noé F. Independent Markov decomposition: Toward modeling kinetics of biomolecular complexes. Proc Natl Acad Sci U S A. 2021;**118**(31):e2105230118.
57. Kleiman DE, Shukla D. Active learning of the conformational ensemble of proteins using maximum entropy VAMPNets. J Chem Theory Comput. 2023;**19**(14):4377–88.
58. Ribeiro JML, Bravo P, Wang Y, Tiwary P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). J Chem Phys. 2018;**149**(7):072301.
59. Tishby N, Pereira FC, Bialek W. The information bottleneck method. 2000 April 01;[physics/0004057 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2000physics...4057T.
60. Tishby N, Zaslavsky N. Deep learning and the information bottleneck principle. 2015 March 01;[arXiv:1503.02406 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2015arXiv150302406T.
61. Creutzig F, Globerson A, Tishby N. Past–future information bottleneck in dynamical systems. Phys Rev E. 2009;**79**(4):041925.
62. Cybenko G. Approximation by superpositions of a sigmoidal function. Math Control Signals Syst. 1989;**2**(4):303–14.
63. Hornik K, Stinchcombe M, White H. Multilayer feedforward networks are universal approximators. Neural Networks. 1989;**2**(5):359–66.
64. Casalino L, Dommer A, Gaieb Z, Barros EP, Sztain T, Ahn SH, et al. AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics. bioRxiv. 2020.
65. Eriksson Lidbrink S. Determining protein conformational ensembles by combining machine learning and SAXS. 2023. Available from: https://www.diva-portal.org/smash/get/diva2:1766567/FULLTEXT01.pdf
66. Dada JO, Mendes P. Multi-scale modelling and simulation in systems biology. Integr Biol. 2011;**3**(2):86–96.

16

Cancer-related intrinsically disordered proteins: Functional insights from energy landscape analysis

Vitor B.P. Leite, Murilo N. Sanches, and Rafael G. Viegas

16.1. Introduction

Cancer is one of the toughest health challenges for humanity. It is a multifaceted subject, and a complete solution to the problem is unthinkable without understanding its molecular mechanisms, particularly cancer-related proteins. Enormous progress has been made in comprehending healthy functioning biochemical processes, and this knowledge is the cornerstone to identifying and characterizing all sorts of pathologies, including cancer. The specificity of protein mechanisms laid the foundation of structural biology: the statement that the unique three-dimensional (3D) structure of a protein is intrinsically related to its function, such as in a lock-and-key system [1]. Along with this idea and through Anfinsen's findings, the understanding that the amino-acid sequence encodes the essential information for proteins to reach their functional conformation was realized [2]. The puzzle of how a protein can attain its native state was known as the ' protein folding problem', which has been successfully explained by the energy landscape theory (ELT), at least for proteins with well-defined native structures. However, a significant portion of proteins do not have a stable structure, at least when not complexed to ligands or binding partners [3 –5]. Such unstructured proteins populate a diverse set of interconverting conformations to achieve their function and are known as intrinsically disordered proteins (IDPs). They are often related to cancers, but the lack of reference conformational states for comparison presents significant conceptual and methodological challenges. In this chapter, we discuss the ELT , IDP's features, computational challenges, methods to address them, and finally, recent insights obtained from these approaches.

16.2. Protein folding and the energy landscape

More than 50 years have passed since Anfinsen experimentally showed that the bovine ribonuclease A can refold to its functional form after denaturation without the help of any biological machinery or catalysts [6]. In this way, the Anfinsen's experiment suggested that all the information for building a functional protein was encoded in its amino-acid sequence. These facts led Anfinsen to formulate the 'thermodynamic hypothesis', which states that the 3D native structure corresponds to a minimum in the Gibbs free energy of the system [2]. Since then, finding the physical –chemical mechanism underlying the folding process has become a crucial problem in protein science. Cyrus Levinthal soon realized that, due to the high number of degrees of freedom, there would not be enough time for a protein to sample all possible conformations in the search for the most stable one [7,8]. This impossibility was known as Levinthal's paradox and led Levinthal to propose that the folding should proceed by fast special kinetic pathways.

The ELT, which is based on principles of statistical mechanics, has reconciled the experimental observations with the apparent Levinthal's paradox. The ELT uses a framework of spin glass, a magnetic system in which the spins randomly interact with each other. These random interactions are associated with amino-acid conformations of random heteropolymers, which create a rugged potential energy surface due to many competing interactions. The fact the energy of such interactions cannot be simultaneously minimized is called 'frustration'. A naturally occurring protein follows the principle of minimum frustration, in which frustrated interactions are minimized at their native conformations, and this state is associated with the global energy minimum of the system [9]. The energy of other conformations correlates with the structural similarity to the global minimum and yields a low-energy basin of states. The overall energy landscape is known as a protein folding funnel [10].

In this representation, the conformation similarity with the native structure (Q) serves as a reaction coordinate of the folding process, and its free energy ($F(Q)$) acts as the effective one-dimensional potential of the system. With this simplified representation, one can bridge theory and experiments, extracting folding energy barriers,

diffusion coefficient along Q ($D(Q)$), folding rates, and other experimental variables, such as Φ-values [11].

16.3. Intrinsically disordered proteins

The notion that a functional protein requires a well-defined 3D structure has been defied in the past few decades by the discovery of flexible, unstructured proteins that remain disordered in physiological conditions, at least *in vitro* [3,4]. They are now widely known as intrinsically disordered proteins . In addition, there are partially disordered proteins, which are composed of both ordered and intrinsically disordered regions (IDRs). While initially considered the exception, it is now acknowledged that a significant fraction of proteins lack unique 3D structures. In fact, it has been reported that some eukaryotic proteomes comprise 12% of fully disordered proteins and 43% of proteins contain large disordered regions (equal or larger than 41 residues) [12]. In spite of being disordered, these proteins play crucial roles in transcription, cell signalling pathways, cell cycle regulation, and other functions [13]. On the other hand, over expression and aggregation of IDPs are the cause of many chronic human diseases [14], cancer [15], and neurodegenerative diseases [16].

Different from folded globular proteins, IDPs are highly flexible and better described by interconverting ensembles. Although some IDPs may behave like random coils, it has been shown that IDPs may also present some kind of structural order, represented by preferential ensembles, which are characterized by small regions of the conformation space separated by energies barriers, resulting in a heterogeneous quasi-continuum of structures [17,18].

Although it is easy to depict the interaction mechanism of ordered proteins through the lock-and-key model [1], binding mechanisms of disordered binding regions (DBRs) and IDPs are still not well understood. In this sense, it has been proposed that these interactions occur via a 'coupled folding and binding '. Two mechanisms were initially proposed: (i) the ensemble selection mechanism, which states that the IDP samples a number of different ensembles so that the binding partner chooses the most fitted one; and (ii) the induced fit mechanism, in which the protein folds after binding to its target ligand, i.e. the binding process induces a structure not sampled in the free state [19]. Both mechanisms characterize a disorder-to-order transition, resulting in a stable structure for the bound state. Nevertheless, this is not always the case, and some IDPs may retain some degree of disorder or even remain fully disordered upon binding, such as in polyelectrolyte complexes [20]. IDPs may also engage in fuzzy interactions in which their bond state is context dependent, varying with the binding partner or cellular conditions [21]. For instance, disordered regions may fold to different structures according to the binding partner.

Due to this large repertoire of binding modes, IDPs play an important role in protein interaction networks (PINs). PINs are represented by graphs in which proteins occupy nodes (edges) connected to their binding partners. In these networks, some proteins occupy a hub position, connected to many partners, whereas the great majority of proteins are connected to few partners or to only one, as in the case of end proteins [22]. Studies on PINs from different eukaryotic interactomes have demonstrated that most hub proteins are IDPs [13]. In this case, the intrinsic disorder endows the IDP with the ability to bind to many partners with similar affinities. In addition, by engaging in promiscuous binding, IDPs can participate in multiple and unrelated signalling networks that could be regulated by different pathways [22].

These facts led to the formulation of the MRK hypothesis [23] (named after Mahmoudabadi, Rangarajan , and Kulkarni) almost a decade ago. In this hypothesis, the authors postulated that, due to their rapid conformation dynamics and promiscuous interaction, IDPs would cause conformational noise. In biological systems, the term noise is used to express the variability encountered in the system due to stochastic processes. One example of such variability is the well-known transcriptional noise, which stands for the heterogeneity in gene expression in an isogenic cell population. In this sense, conformational noise refers to the variability originated from the rapid conformational dynamics and promiscuous non-functional interactions of IDPs. This unbalanced activity may lead to a rewiring of PINs, affecting the cellular fate. In this way, the conformational noise may be responsible for a cell phenotypic switch, in which, for instance, a cell can change from a normal to a malignant state. In conclusion, the MRK model provides an insightful non-genetic mechanism to generate heterogeneity that may have important implications in cancer progression and therapeutics [24].

To fully understand the implications of conformation heterogeneity in biological function, it is necessary to characterize the energy landscape of IDPs and be able to extract insightful features from it. Whereas a random coil would present a flat landscape where all states have almost the same energy, a protein that folds in an all-or-none fashion would present a single deep funnelled landscape. The energy landscape of IDPs lays between these two extremes and may be almost flat for highly disordered proteins, or present several local minima, being multi-funnelled and highly frustrated [17,25,26]. However, the determination of the conformation ensemble of IDPs is still challenging for both theoretical and experimental methods. Currently, this problem has been tackled by integrative approaches combining both molecular dynamics (MD) with improved force fields and experimental techniques, such as nuclear magnetic resonance (NMR), small-angle X-ray scattering (SAXS), and Förster resonance energy transfer [27,28]. In the next section, methods developed to extract biophysically relevant features from conformation ensembles are presented.

16.4. Computational approaches

Theoretical–computational approaches use molecular modelling to mimic the behaviour of biomolecules, and MD is the most used method to simulate the time evolution of these systems. The atoms or a group of atoms can be treated as the smallest individual units, depending on the coarse-grained level of representation. These units are considered particles connected by spring-like potentials, interacting with the other particles in their range based on their specific potentials. The evolution of the system over time is calculated using Newton's equations that describe their motions, with the forces between the particles and their potential energies often calculated using interatomic potentials and molecular mechanics force fields [29].

The force field consists of a set of parameters and energy functions that are used to calculate the potential energy of a system. These

parameters are usually derived from experiments and attempt to reproduce a realistic behaviour. Several approaches have been used to develop new force fields or parameterize existing ones to improve the sampling for IDPs [30]. The CHARMM36m is one example which, using all-atom representation, presents coherence with NMR spectroscopy and SAXS measurements of small IDPs [31]. A comparison between some of these force fields when applied to a group of IDPs is provided in [32].

Obtaining reliable force fields for IDPs is more difficult because the error in energy estimates linearly increases with disorder [33], so results are highly susceptible to small force field variations. One reasonable approach is to take a coarse-grained model, where amino-acids are represented only by some of their atoms, aiming to understand some essential features of the system. This is the case for the Associative memory, Water-mediated, Structure and Energy Model (AWSEM) [34], which is based on ELT and the principle of minimum frustration, and restricts the representation of each amino acid to C_α, C_β, and O atoms. Along with physics-based potentials, AWSEM implements bioinformatically motivated biasing potentials, such as the fragment memory term, which provides local structural biasing based on the information of locally similar sequence fragments. The AWSEM-IDP, a branch specifically developed for IDP simulation, presents re-weighted parameters and integration with experimental data, with the addition of a radius of gyration potential and fragment memories both based on structural ensembles from either NMR ensemble or atomistic simulations [33]. Another possibility to fine-tuning the force field is using the available experimental data as simulation constraints, also known as MDFIT [35].

16.5. Reaction coordinates and dimensionality reduction

Investigation methods explore molecular systems using reductionist approaches, in which one seeks to understand the molecular processes in terms of a few reaction coordinates or order parameters. These effective variables try to convey the essence of such mechanisms. In the simplest computational representations, each amino acid is the smallest individual unit, and they do not consider the individual atoms and solvent molecules. For a protein with N amino acids, one has the order of $3N$ degrees of freedom, which is a high -dimensional system of difficult visualization and insightful representation, even in such a coarse-grained model.

A possible solution is to use *a priori* reaction coordinates that are usually meaningful variables, such as radius of gyration, end-to-end distances, or similarity degree with a reference structure Q (as discussed above). There are even computationally costly methods that try to identify possible reaction coordinates [36]. However, such procedures have the potential to shroud the richness of the energy landscape and its dynamics.

MD trajectories can also be analysed using dimensionality reduction (DR) techniques, which aim to provide visualization of the complex manifold in two or three dimensions, retaining the relevant information about the system dynamics. In the past decades, several DR methods have been proposed using both linear and non-linear approaches and making different assumptions about the original manifold to be mapped [37, 38]. Combined with effective sampling from molecular simulations, these techniques may provide insightful visualization of the conformation space of proteins, enabling the identification of collective motions, intermediate ensembles, meta-stable states, and molecular mechanisms that otherwise would not be accessible from the raw trajectory data. Next, we discuss a few well-known DR techniques.

16.5.1. Principal component analysis

The search for an adequate low-dimensional space for describing protein collective motions began with a seminal paper by García [39]. In this work, the author derived a set of equations to find, from a minimization procedure, the most relevant directions to describe the collective motions of the protein crambin. The derived equations are equivalent to the so-called principal component analysis (PCA), a linear DR method that since then has been widely used in the analysis of MD trajectories. Essentially, the PCA analysis consists of the following steps: (i) align all the conformations to a single frame in order to remove translation and rotation; (ii) generate a matrix of coordinates of the system, e.g. Cartesian coordinates of C_α atom or dihedral angles; (iii) calculate a covariance matrix of the coordinates; and (iv) diagonalize the covariance matrix to find its eigenvectors and eigenvalues. It is possible to show that the eigenvalue is equal to the variance of the data along the direction of its corresponding eigenvector. The original set of atomic coordinates may then be projected onto the directions of these eigenvalues, revealing the motions in each of these principal directions. The dimensionality of the system can be drastically reduced if a small number of these eigenvalues account for a great fraction of the total variance. In this case, these directions are known as essential coordinates, while the low-dimensional space formed by them is known as essential space (ES). The motions as observed in the ES are referred to as essential dynamics and are related to functional motions of proteins [40]. For further examples and discussion about the method, we refer to the recent review [41].

16.5.2. Multidimensional scaling

Classical multidimensional scaling (MDS) [42] is a technique widely used to reduce the dimensionality and visualize similarity or dissimilarity in high-dimensional data. In MDS, the coordinates of each object in the high-dimensional space are not required. Instead, only a matrix of pairwise proximity of points is needed. Initially, MDS techniques were developed in the field of psychophysics and sensory analysis, where the objects to be mapped were stimuli [43]. Although classical MDS considers that the dissimilarity can be represented by the Euclidean distance in the original space, it may be calculated using any distance metric such as geodesic or cosine distance. Thus, considering n objects populating a high-dimensional space, the input data is an $n \times n$ matrix in which each entry contains the dissimilarity, δ_{ij}, between the object pair (i, j). Then, the method aims to place a set of n points into a low-dimensional space with the constraint that the pairwise Euclidean distances are approximately equal to dissimilarities. Mathematically, this constraint is expressed through a loss (or cost) function given by

$$L = \sum_{i<j} \left(\delta_{ij} - || \, y_i - y_j \, || \right)^2 , \tag{16.1}$$

where y_i is the position vector of the projected point i and $|| \, y_i - y_j \, ||$ is the Euclidean distance between points representing objects i and j

. Broadly, in classical MDS, a Gram matrix of inner products can be found from the double-centered dissimilarity matrix. The eigenvalue problem can be solved to find the principal coordinates. Although the classical MDS may have an analytical solution, other methods, such as nonmetric MDS, were developed to project data in which dissimilarities do not satisfy the metric requirements, for instance, qualitative dissimilarity in psychophysics. In this case, the projected points may preserve a rank order instead of distances like in classical MDS. To achieve such projections, nonmetric MDS usually minimizes a stress function through iterative minimization procedures. The set of MDS techniques may be used to minimize different stress functions through different minimization algorithms [43]. For mathematical details about this procedure, we refer to the review [44] and references therein.

One great advantage of MDS is that only a dissimilarity matrix is needed as input to achieve the matrix of coordinates on the plane. To describe how similarity evolves along simulation trajectory, the most used strategy is to remove translation and rotation from all frames, by aligning them to a reference structure and then calculating the coordinate root-mean-square deviation (RMSD). In the case of MDS, the construction of the dissimilarity matrix does not require a reference structure but still requires pairwise alignment, which may not be the best approach if sampled structures are too diverse to be aligned [38]. Another approach that does not require alignment involves dealing with internal distances. For instance, for a protein consisting of N C_α atoms, there will be $N(N-1)/2$ internal distance terms that can be used to calculate a distance root-mean-square deviation (dRMSD) [45]. Like RMSD, the dRMSD can also be calculated from a reference structure or in a pairwise way. Several other dissimilarity metrics have been proposed such as contact map distances, power distances, and dihedral distances based on the Ramachandran angles [45,46]. It should be noted that different similarity measures may highlight different features of the conformational space. Nevertheless, the combination of these measures with DR techniques may circumvent such limitations and unravel global and/or local features of the system.

One interesting example of how MDS can be applied to MD trajectories was provided by Pisani et al. [47]. The authors used classical MDS and RMSD-based dissimilarity to unveil meta-stable states and characterize the conformational ensemble of the cancer-related cyclin-dependent kinase CDK2. Specifically, they used 255 structures obtained from crystallographic data from the Protein Data Bank to project an effective low-dimensional conformation space. In the resulting projected space, which accounted for 80% of the data variance, it was possible to discriminate five clusters corresponding to different known states (e.g. active, inactive, open, etc.). The low-dimensional conformation landscape was further populated by running regular and accelerated MD simulations for CDK2 in apo state and for CDK2 complexed with fluorophore 8-anilino-1-naphtalenesulfonic acid (ANS) molecules. The new generated conformations were added to the initial space through an out-of-sample embedding procedure (see details in the original work [47]). The authors also constructed the energy landscape using order parameters that had already been used to describe kinase ensembles, such as the RMSD of specific secondary motifs. As a result, the space generated by MDS was able to discriminate new meta-stable states not seen in the landscape generated by the usual order parameters. In addition, the landscape for the ANS -bound state showed a populational shift towards inactive states. Thus, the analysis of the low-dimensional landscape shed light on molecular mechanisms, revealing meta -stable structures that may be potential candidate structures for the designing of allosteric inhibitors.

16.5.3. Non-linear dimensionality reduction

Despite being robust techniques, it has been acknowledged that both PCA and classical MDS only give good results when the data lies on a manifold that is isomorphic to a hyperplane. When data is sampled from complex manifolds, e.g. isomorphic to a torus, the points can not be simply mapped onto a hyperplane while preserving all the distances between them [38]. Several non-linear DR techniques have been proposed to address this issue, such as Isomap [48], diffusion maps [49], t-SNE [50], UMAP [51], and Sketch-map [52]. Additionally, deep machine learning techniques have also been applied to manifold learning and visualization, like the EncoderMap [53]. For a recent review of these methods and their applications in the context of molecular simulation, we refer to [38,52,54]. By choosing different loss functions, optimization procedures, and dissimilarity metrics, different features of the original manifold can be accurately depicted in a low-dimensional space. For example, the method may choose to preserve mainly local or global features.

16.6. Energy landscape visualization method

A new approach to generate an intuitive visualization of high-dimensional data is the energy landscape visualization method (ELViM) [55,56]. Similar to MDS methods, the ELViM projection relies on two fundamental steps: (i) the generation of a dissimilarity matrix based on a specific proximity measurement and (ii) the minimization of a loss function through an iterative procedure. The proximity measure used by ELViM is based on an order parameter that was first devised to study spin glasses and then successfully applied as a reaction coordinate or a structural similarity measure [57–59]. The similarity between conformations i and j is given by

$$Q_w^{i,j} = \frac{1}{N_p}\sum_{m,n} exp\left[-\frac{\left(r_{m,n}^i - r_{m,n}^j\right)^2}{2\sigma_{m,n}^2}\right], \qquad (16.2)$$

where $r_{m,n}^{i(j)}$ is the Euclidean distance between atoms m and n from conformation i or j, and N_p is the total number of atom pairs. The $\sigma_{m,n}$ is a weighting parameter given, in angstroms, by $\sigma_{m,n} = \sigma_0 |m-n|^\varepsilon$, where $\varepsilon = 0.15$ and $\sigma_0 = 1.0$ Å [58]. The pairwise dissimilarity is defined as $\delta_{ij} = 1 - Q_w^{i,j}$, so it is 0 for identical structures and tends to 1 for very different structures. As an advantage, this metric only relies on internal pairwise distances (C_α or heavy atoms) that can be promptly obtained from MD simulations.

To generate a low-dimensional visualization, ELViM seeks to minimize a loss function (Equation (16.1)) so that $\delta_{ij} \approx || y_i - y_j ||$, meaning that the Euclidean distances in the plane fit the dissimilarities as well as possible. A force-based technique called force scheme is adopted to achieve such results [60]. In this technique, in each

iteration, each point is chosen as a reference and all other points are slightly perturbed, being attracted to or repelled from the reference in order to approximate the distance to the dissimilarity. A minimization procedure ensures the best configuration and usually converges up to a number of iteration equal to the square root of the number of conformations. After reaching equilibrium, conformations that are close on the multidimensional energy landscape are attracted to the same basin in the final projection. From a mathematical viewpoint, this DR can be considered an ill-posed problem; there is no unique solution, and the results depend on the ELViM-adjusted parameters. In our experience, however, if the system's complexity is not too high, results are robust, stable, and can give new insights into molecular mechanisms.

Further analysis can be performed by using heat maps to colour the projection's points according to any relevant biophysical quantity, such as the radius of gyration (Rg), the RMSD , or any reaction coordinate values. Comparing how these quantities vary over the projected space can bring new insights into the dynamics and function of proteins that can not be extracted by any other means from the abundant data contained in the simulation trajectory. Moreover, several ensembles of conformations, obtained under different physical– chemical conditions, can be analysed using the same conformational phase space. This allows for a differential analysis between different systems. Next, we show one example of how this methodology can be applied to elucidate the conformation diversity of an IDP, the disordered prostate-associated gene 4 protein (PAGE4), and its phosphoforms.

16.7. Functional insights from the PAGE4 energy landscape

PAGE4 is an IDP that belongs to the family of cancer/testis antigens. It is highly expressed in the human foetal prostate, trophoblasts, and placenta. During prostate cancer, its expression is elevated, which plays an important role both in the benign and malignant diseases [61,62]. PAGE4 is a transcriptional coactivator and a stress-response factor that potentiates the transactivation by c-Jun, which heterodimerizes with c-Fos to form the activator protein-1 (AP-1) [63]. It has been experimentally shown that the phosphorylation of PAGE4, mainly at Threonine 51, by the homeodomain interacting protein kinase 1 (HIPK1) results in an ensemble of more compact structures that can bind to AP-1 and potentiate the c-Jun activity [64]. On the other hand, PAGE4 can also be hyperphosphorylated by the CDC-like kinase 2 (CLK2), resulting in an ensemble of more expanded structures that has reduced affinity to AP-1 and attenuates the c-Jun activity [63].

Atomistic details about the system were provided by Oliveira et al. [26], which applied the ELViM to analyse a coarse-grained AWSEM trajectory of the PAGE 4 and its phosphoforms. **Figure 16.1**a shows the two-dimensional (2D) ELViM projection as a function of the radius of gyration, surrounded by typical structures that exemplify each region of the phase space. In this projection, each point corresponds to a different sampled structure. The axes have no particular meaning, and the important aspect is the relative distance between conformations. This method allows for the clustering of conformations to obtain a continuous projection, with the most compact structures on one side and the most extended on the other. Another advantage of ELViM is the ability to analyse each ensemble individually. Since the position of each configuration is assigned based on their dissimilarity value, it is possible to compare different ensembles simultaneously. This is demonstrated using the 2D density of states of each phosphoform shown in **Figure 16.1**b. The density of state demonstrates that the wild type (WT) and the HIPK1 PAGE4 structures populate the bottom right of the projection, while the CLK2-PAGE4 structures are more spread out. Combining this result with the radius of gyration analysis, one can note that the WT and the HIPK1-PAGE4 conformations correspond to the more compact ones, while most of the CLK2-PAGE4 are near the extended and more disordered region. These results are in agreement with the experimental finding previously discussed. The analyses of the main basins in the projection revealed that for the wild-type PAGE4

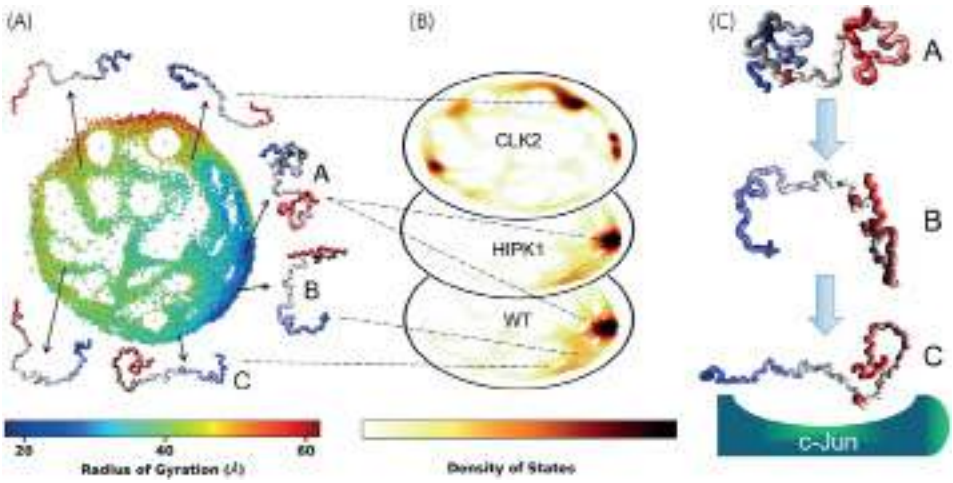


Figure 16.1. (A) ELViM projection as a function of the radius of gyration. The structures around the projection illustrate some ensemble regions of the conformation phase space, with the N-terminals shown in red and C-terminals in blue. (B) Density of state for each phosphoform. Different phosphoform cases have their density of states (σ) normalized by their highest value, with the highest density occurring for HIPK1. Their relative maximum values are $\sigma^{max}_{WT} / \sigma^{max}_{HIPK1} = 0.57$ and $\sigma^{max}_{CLK2} / \sigma^{max}_{HIPK1} = 0.29$. (C) Suggested fly-casting mechanism of WT, with the Thr51 highlighted in red to show the conformational change that the C-terminal undergoes from region A to B to C.

a loop is formed in the N-terminal region, and the C-terminal region presents stretched conformations, which exposes the region where c-Jun binds, facilitating c-Jun docking to occur [26]. This behaviour is in line with the fly-casting mechanism, in which the protein weakly binds at a large distance and folds as it approaches the binding site [65]. This mechanism is frequently observed in the conformational dynamics of IDPs as depicted in **Figure 16.1**c, in which the C-terminal is shown in blue and undergoes a conformational extension with the Thr51 highlighted in red. This structural change allows PAGE4 to dock to the c-Jun and can be traced back at the ELViM projection as the cluster examples A, B, and C.

The work of Oliveira and co-workers also makes it possible to visualize the conformational changes induced in HIPK1-PAGE4 ensembles, obstructing the binding site and lowering the PAGE4 affinity to c-Jun, when compared to the WT , as found in experiments [64]. This lower affinity results in the dissociation of PAGE4 and c-Jun, which promotes the interaction of c-Jun to other partners and potentiates its transactivation [64]. Analogous studies have addressed other pathology-related IDPs, the amyloid-β [66], and other complex folding systems [67], which have also elucidated unprecedented details of these energy landscapes.

16.8. Conclusions

The discovery of IDPs and IDRs in the past decades has shaken the protein science field, mainly regarding the well-established structure–function paradigm. The fact that IDPs occupy hub positions in PINs and are related to many pathologies has inspired multiple studies addressing how biological functions arise from conformational dynamics and structural disorder. The determination of the conformation ensembles of IDPs has been a challenge for both experimental and theoretical techniques. Whereas most biophysical experiments used to study IDPs result in ensemble averages, computational modelling aims to provide an atomistic description of these ensembles. However, computational models also present many limitations due to difficulties inherent in setting an appropriate force field and ensuring proper sampling. Nevertheless, a lot of progress has been made, and many strategies have focused on integrative approaches, bridging theory and experiment. The ELT has provided a powerful resource for tackling complex systems. We have shown that methods such as the ELViM can provide invaluable insights into IDP mechanisms, particularly cancer-related ones.

REFERENCES

1. Fischer E. Einfluss der configuration auf die Wirkung der enzyme. Ber Dtsch Chem Ges. 1894;**27**:2985–93.
2. Anfinsen CB. Principles that govern the folding of protein chains. Science. 1973;**181**:223–30.
3. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol. 1999;**293**:321–31.
4. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, et al. Intrinsically disordered protein. J Mol Graph Model. 2001;**19**:26–59.
5. Uversky VN, Kulkarni P. Intrinsically disordered proteins: chronology of a discovery. Biophys Chem. 2021;**279**:106694.
6. Anfinsen CB, Haber E, Sela M, White FH. The Kinetics of formation of native ribonuclease during oxidation of the reduced popypeptide chain. Proc Natl Acad Sci. 1961;**47**:1309–14.
7. Levinthal C. Are there pathways for protein folding? J Chem Phys. 1968;**65**:44–5.
8. Levinthal C. How to fold graciously. Mossbauer Spectroscopy in Biological Systems. 1969;**67**:22–4.
9. Bryngelson JD, Wolynes PG. Spin glasses and the statistical mechanics of protein folding. Proc Natl Acad Sci. 1987;**84**:7524–28.
10. Leopold PE, Montal M, Onuchic JN. Protein folding funnels: a kinetic approach to the sequence-structure relationship. Proc Natl Acad Sci. 1992; 89:8721–25.
11. Fersht A, Sato S. Φ-Value analysis and the nature of protein-folding transition states. Proc Natl Acad Sci. 2004; **101**:7976–81.
12. Bogatyreva NS, Finkelstein AV, Galzitskaya OV. Trend of amino acid composition of proteins of different taxa. J Bioinform Comput Biol. 2006; **04**:597–608.
13. Kulkarni P, Bhattacharya S, Achuthan S, Behal A, Jolly MK, Kotnala S, et al. Intrinsically disordered proteins: critical components of the wetware. Chem Rev. 2022; **122**:6614–33.
14. Kulkarni P, Uversky VN. Intrinsically disordered proteins in chronic diseases. Biomolecules. 2019;**9**(4):147.
15. Iakoucheva LM, Brown CJ, Lawson JD, Obradović Z, Dunker AK. Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol. 2002; **323**:573–84.
16. Uversky VN. Intrinsic disorder in proteins associated with neurodegenerative diseases. Front Biosci (Landmark Ed). 2009;**14**(14):5188–238.
17. Choi U, Sanabria H, Smirnova T, Bowen M, Weninger K. Spontaneous switching among conformational ensembles in intrinsically disordered proteins. Biomolecules. 2019;**9**:114.
18. Kulkarni P, Leite VBP, Roy S, Bhattacharyya S, Mohanty A, Achuthan S, et al. Intrinsically disordered proteins: Ensembles at the limits of Anfinsen's dogma. Biophys Rev. 2022;**3**(1):011306.
19. Boehr DD, Nussinov R, Wright PE. The role of dynamic conformational ensembles in biomolecular recognition. Nat Chem Biol. 2009;**5**(11):789–96.
20. Schuler B, Borgia A, Borgia MB, Heidarsson PO, Holmstrom ED, Nettels D, et al. Binding without folding—the biomolecular function of disordered polyelectrolyte complexes. Curr Opin Struct Biol. 2020;**60**:66–76.
21. Fuxreiter M. Classifying the binding modes of disordered proteins. Int J Mol Sci. 2020;**21**(22):8615.
22. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J. 2005; **272**(20):5129–48.
23. Mahmoudabadi G, Rajagopalan K, Getzenberg RH, Hannenhalli S, Rangarajan G, Kulkarni P. Intrinsically disordered proteins and conformational noise: implications in cancer. Cell Cycle. 2013;**12**(1):26–31.
24. Saxena K, Subbalakshmi AR, Kulkarni P, Jolly MK. Cancer: more than a geneticist's Pandora's box. J Biosci. 2022;**47**:21.
25. Chebaro Y, Ballard AJ, Chakraborty D, Wales DJ. Intrinsically disordered energy landscapes. Sci Rep. 2015;**5**:10386.
26. Oliveira Junior AB, Lin X, Kulkarni P, Onuchic JN, Roy S, Leite VBP. Exploring energy landscapes of intrinsically disordered proteins: insights into functional mechanisms. J Chem Theory Comput. 2021;**17**(5):3178–87.

27. Thomasen FE, Lindorff-Larsen K. Conformational ensembles of intrinsically disordered proteins and flexible multidomain proteins. Biochem Soc Trans. 2022;**50**:541–54.
28. Chan-Yao-Chong M, Durand D, Ha-Duong T. Molecular dynamics simulations combined with nuclear magnetic resonance and/or small-angle X-ray scattering data for characterizing intrinsically disordered protein conformational ensembles. J Chem Inf Model. 2019;**59**(5):1743–58.
29. Frenkel D, Smit B. Understanding molecular simulation: from algorithms to applications. 2nd ed. San Diego: Academic Press; 2001.
30. Mu J, Liu H, Zhang J, Luo R, Chen HF. Recent force field strategies for intrinsically disordered proteins. J Chem Inf Model. 2021;**61**(3):1037–47.
31. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat Methods. 2017;**14**(1):71–3.
32. Rahman MU, Rehman AU, Liu H, Chen HF. Comparison and evaluation of force fields for intrinsically disordered proteins. J Chem Inf Model. 2020;**60**(10):4912–23.
33. Wu H, Wolynes PG, Papoian GA. AWSEM-IDP: a coarse-grained force field for intrinsically disordered proteins. J Phys Chem B. 2018;**122**(49):11115–125.
34. Davtyan A, Schafer NP, Zheng W, Clementi C, Wolynes PG, Papoian GA. AWSEM-MD: protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. J Phys Chem B. 2012;**116**(29):8494–503.
35. Ratje AH, Loerke J, Mikolajka A, Brünner M, Hildebrand PW, Starosta AL, et al. Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. Nature. 2010; **468**(7324):713–16.
36. Best RB, Hummer G. Microscopic interpretation of folding s translocation by transition path ensemble. Proc Natl Acad Sci U S A. 2016;**113**(12):3263–68.
37. Tribello GA, Gasparotto P. Using dimensionality reduction to analyze protein trajectories. Front Mol Biosci. 2019;**6**:46.
38. Glielmo A, Husic BE, Rodriguez A, Clementi C, Noé F, Laio A. Unsupervised learning methods for molecular simulation data. Chem Rev. 2021;**121**(16):9722–58.
39. García AE. Large-amplitude nonlinear motions in proteins. Phys Rev Lett. 1992; **68**(17):2696–99.
40. Amadei A, Linssen ABM, Berendsen HJC. Essential dynamics of proteins. Proteins: Struct Funct Bioinform. 1993;**17**(4):412–25.
41. Palma J, Pierdominici-Sottile G. On the uses of PCA to characterise molecular dynamics simulations of biological macromolecules: basics and tips for an effective use. Chemphyschem. 2023;**24**(2):e202200491.
42. Torgerson WS. Multidimensional scaling: I. Theory and method. Psychometrika. 1952;**17**(4):401–19.
43. Mead AL. Review of the development of multidimensional scaling methods. J R Stat Soc: Ser D (Statistician). 1992;**41**(1):27–39.
44. France SL, Carroll JD. Two-way multidimensional scaling: a review. IEEE Trans Syst Man Cybern Pt C (Appl Rev). 2011;**41**(5):644–61.
45. Wallin S, Farwer J, Bastolla U. Testing similarity measures with continuous and discrete protein models. Proteins. 2003;**50**(1):144–57.
46. Cossio P, Laio A, Pietrucci F. Which similarity measure is better for analyzing protein structures in a molecular dynamics trajectory? Phys Chem Chem Phys. 2011;**13**(22):10421–5.
47. Pisani P, Caporuscio F, Carlino L, Rastelli G. Molecular dynamics simulations and classical multidimensional scaling unveil new metastable states in the conformational landscape of CDK2. PLoS ONE. 2016;**11**(4):e0154066.
48. Tenenbaum JB, de Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. Science. 2000;**290**:2319–23.
49. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F, et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. Proc Natl Acad Sci. 2005;**102**(21):7426–7431.
50. Van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;**9**:2579–605.
51. McInnes L, Healy J, Saul N, Großberger L. UMAP: uniform manifold approximation and projection. J Open Source Softw. 2018;**3**(29):861.
52. Ceriotti M, Tribello GA, Parrinello M. Simplifying the representation of complex free-energy landscapes using sketch-map. Proc Natl Acad Sci. 2011; **108**(32):13023–8.
53. Lemke T, Peter C. EncoderMap: dimensionality reduction and generation of molecule conformations. J Chem Theory Comput. 2019;**15**(2):1209–15.
54. Trozzi F, Wang X, Tao P. UMAP as a dimensionality reduction tool for molecular dynamics simulations of biomacromolecules: a comparison study. J Phys Chem B. 2021;**125**(19):5022–34.
55. Oliveira AB Jr, Fatore FM, Paulovich FV, Oliveira ON Jr, Leite VBP. Visualization of protein folding funnels in lattice models. PLoS ONE. 2014; **9**(7):e100861.
56. Oliveira AB Jr, Yang H, Whitford PC, Leite VBP. Distinguishing biomolecular pathways and metastable states. J Chem Theory Comput. 2019;**15**(11):6482–90.
57. Hardin C, Eastwood MP, Prentiss MC, Luthey-Schulten Z, Wolynes PG. Associative memory Hamiltonians for structure prediction without homology: alpha/beta proteins. Proc Natl Acad Sci U S A. 2003; **100**(4):1679–84.
58. Cho SS, Levy Y, Wolynes PG. P versus Q: structural reaction coordinates capture protein folding on smooth landscapes. Proc Natl Acad Sci. 2006; **103**(3):586–91.
59. Potoyan DA, Papoian GA. Regulation of the H4 tail binding and folding landscapes via Lys-16 acetylation. Proc Natl Acad Sci. 2012;**109**(44):17857–62.
60. Tejada E, Minghim R, Nonato LG. On improved projection techniques to support visual exploration of multi-dimensional data sets. Inf Vis. 2003;**2**:218–31.
61. Zeng Y, He Y, Yang F, Mooney SM, Getzenberg RH, Orban J, et al. The cancer/testis antigen prostate-associated gene 4 (PAGE4) is a highly intrinsically disordered protein. J Biol Chem. 2011;**286**(16):13985–94.
62. Lin X, Roy S, Jolly MK, Bocci F, Schafer NP, Tsai MY, et al. PAGE4 and conformational switching: insights from molecular dynamics simulations and implications for prostate cancer. J Mol Biol. 2018;**430**(16):2422–38.
63. Kulkarni P, Jolly MK, Jia D, Mooney SM, Bhargava A, Kagohara LT, et al. Phosphorylation-induced conformational dynamics in an intrinsically disordered protein and potential role in phenotypic heterogeneity. Proc Natl Acad Sci. 2017;**114**(13):E2644–53.
64. Mooney SM, Qiu R, Kim JJ, Sacho EJ, Rajagopalan K, Johng D, et al. Cancer/testis antigen PAGE4, a regulator of c-Jun transactivation, is phosphorylated by homeodomain-interacting

protein kinase 1, a component of the stress-response pathway. Biochemistry. 2014;**53**(10):1670–1679.
65. Shoemaker BA, Portman JJ, Wolynes PG. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci. 2000; **97**(16):8868–73.
66. Sanches MN, Knapp K, Oliveira AB Jr, Wolynes PG, Onuchic JN, Leite VBP. Examining the ensembles of Amyloid-β monomer variants and their propensities to form fibers using an energy landscape visualization method. J Phys Chem B. 2022;**126**(1):93–9.
67. Sanches MN, Parra RG, Viegas RG, Oliveira AB, Wolynes PG, Ferreiro DU, et al. Resolving the fine structure in the energy landscapes of repeat proteins. QRB Discov. 2022;**3**:e7.

17

Targeting RAS

Priyanka Prakash

17.1. Introduction

RAS enzymes are molecular switches that cycle between an active 'on' and an inactive 'off' state [1–3]. Active RAS is GTP-bound, while inactive RAS is GDP-bound [4]. Given that the intrinsic exchange rate of inactive to active transition is slow, RAS interacts with guanine exchange factors (GEFs) that catalyze the exchange of GDP to GTP [5–8]. GTPase-activating proteins (GAPs), on the other hand, catalyze the hydrolysis of GTP in order to make the enzyme inactive to cycle it back to the GDP-bound state [5,6,9,10]. There are three human isoforms of RAS: H-, N-, and KRAS [11]. The sequence and structural similarity among the three isoforms are well known. From structural standpoint, RAS is made up of a catalytic G-domain that contains the active site or nucleotide-binding site where GDP/GTP binds a hypervariable region (HVR) that attaches itself to the cell membrane (**Figure 17.1**). The G-domain contains two lobes [12]: the effector lobe or lobe 1 that contains the switch region that is the interaction site of all the binding partners of RAS, such as GEFs and GAPs. The other lobe is referred to as the allosteric lobe or lobe 2. Allosteric lobe has sequence-based differences among different isoforms [11,13]. These differences are located on the three helices of the allosteric lobe, i.e. α3, α4, and α5 (Fig 17.1). Attached to the allosteric lobe is the hypervariable or HVR region that exhibits the most prominent differences among the isoforms [14,15]. HVR varies among the three isoforms primarily at two levels: (1) sequence and (2) the number of lipidations [11,15,16]. KRAS has only a single lipid anchor (farnesylation); HRAS has two palmitoylation and one farnesylation sites making a total of three lipid attachment points, while NRAS has a single palmitoylation and a single farnesylation site making a total of two lipid attachment points. Despite sharing structural and sequence similarities, different RAS isoforms show differential functional output under normal conditions [3,17]. Among other factors, isoform-specific differences in the HVR of RAS is among the key regions that govern these observed differences in the functional output [18].

Mutations primarily at the three hotspot residues, i.e. the positions, 12, 13, and 61, has been found associated with deadly cancer types, such as pancreatic, colorectal, kidney, gallbladder, melanoma, and others [11]. These mutations disturb the fine balance in the RAS cycle leading to an aberrant signalling of MAPK or PI3K [19] pathways, and there are several excellent reviews that the interested readers are recommended to refer for more details [3,11,20,21]. The isoform-specific differences in the functional output is well reported, such as pancreatic and colorectal are predominantly associated with mutant KRAS, melanoma is largely associated with mutant NRAS, and kidney and gallbladder cancer are primarily associated with mutant HRAS [22]. The functional output of different RAS mutants is also different for an isoform (see, for instance, [11]). Owing to their involvement in a variety of deadly cancer types, drugs against mutant RAS are essential to abrogate their abnormal signalling. Worldwide a multi-pronged research focused on a variety of areas, such as small molecules, immunotherapy, peptidomimetics, etc. are underway to address these challenges [23,24]. As an example, one such effort from immunotherapy to treat pancreatic cancer containing KRAS mutations is based upon KRAS peptide vaccine that is currently under clinical trial. While too early to conclude anything and challenging at the same time, however, if successful, this will provide us with a promising alternative approach of targeting KRAS-driven cancers. In addition, mRNA vaccines for pancreatic cancers are also under clinical trial which are first of its kind.

In this chapter, I first discuss the small-molecule drug design focused on RAS that has gained a fast pace starting 2012–2013 which led to the successful identification of the first FDA-approved anti-KRAS drug that is currently in the clinic and targets a specific mutant of KRAS(G12C). There are several excellent studies in this field, and readers are referred to many excellent articles and reviews in this area [20,23,25–28]. Here, I primarily focus upon those initial studies that paved the way towards the successful identification of mutant KRAS-specific drugs and focus at the same time on the binding pockets [29–32]. Next, I highlight selected molecular simulation-based studies. My goal here is primarily to highlight simulation studies as applicable in the fast-paced industry setting instead of focusing on the investigation of conformational dynamics of RAS's G-domain or membrane-bound RAS; here, I highlight the power of probe-based simulations or mixed-probe MD simulations as applied to RAS proteins and the role of water dynamics in a drug discovery process. There are excellent review and research articles that discuss the conformational dynamics of RAS, RAS–ligand complexes, RAS nanoclusters, dimerization, and others, and interested readers are suggested to refer refs. [2,13–15,33–40]. At the end, I briefly hover

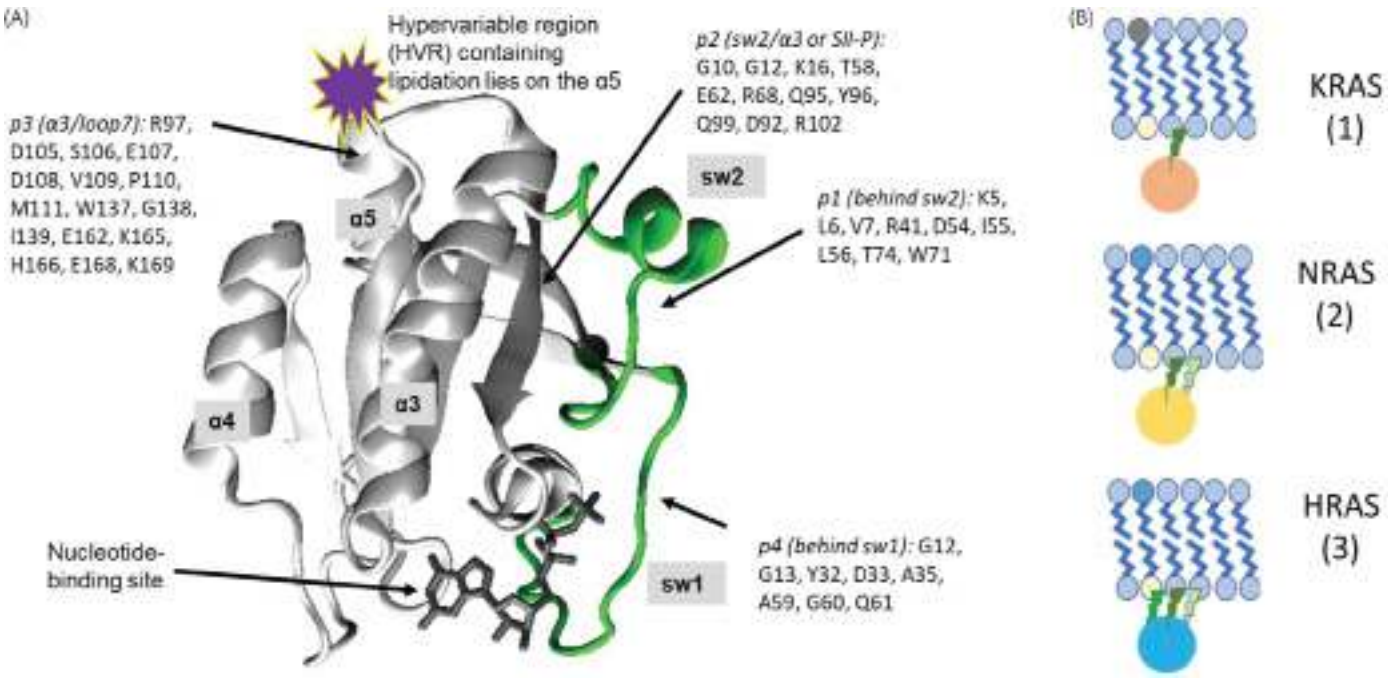


Figure 17.1. (A) Structural elements of RAS. The switch region and the nucleotide-binding site form the effector-binding lobe (lobe 1) and the three helices, helix3, helix4, and helix5, form the allosteric lobe (lobe 2). helix5 is attached to the hypervariable region (HVR) and contains multiple lipidation sites. RAS has four allosteric pocket, and the key residues lining these pockets are shown. (B) KRAS has a single lipidation (farnesylation); NRAS two lipid attachment points (one palmitoylation and one farnesylation), and HRAS three lipid attachment points (two palmitoylation and one farnesylation). Within brackets are the number of lipidation.

over a yet underexplored area and a potential role of membrane dynamics as an alternative strategy to target challenging drug targets, such as RAS and RAF kinases.

17.2. RAS targeting: binding pockets

For the direct targeting of RAS, there are studies in the literature that reports the presence of four allosteric pockets on the surface of this single-domain protein (Figure 1 of [41] and **Figure 17.2**). These four pockets are as follows: (A) p1 or sw1/sw2 pocket that lies behind the sw-2 and occupies a hydrophobic pocket that opens as a result of flipping of Tyr71 and loss of interaction between R41 and D54 [29,31,32]. Compounds that bind to this pocket may extend towards the switch-2 or switch-1 in the upwards direction (**Figure 17.2**, blue box) or towards the outside, e.g. BI-2852 [42] (**Figure 17.2**). (B) p2 or SII-P pocket, which is the most explored pocket till date. This is due to the successful approach of covalent targeting of G12C mutant of KRAS [30]. (C) p3 pocket in which the metal–cyclens bind near helices 4 and 5 that are proximal to the membrane anchor region [43,44]. Recent studies with compounds binding to p3 pockets have further strengthened this yet underexplored pocket/region in RAS, but this pocket comes with its own challenges of being proximal to the membrane anchor (see Section 17.5 for a detailed discussion on compounds binding p3). Since no crystal structures are yet available for these p3 binders and they are obtained from utilizing computational approaches, therefore, I discuss them separately. (D) pocket p4 that lies behind sw1 (not shown in **Figure 17.2**). Currently, as evidence of p4 binders, there are not many reports but andrographolide, which is a natural product and its analogues are predicted using computational approaches to bind p4 pocket, and an NMR-based study shows the binding of metal–cyclens in this region, as discussed below. However, thus far no X-ray structure is available for compounds bound to p4 pocket [45,46]. All these four allosteric pockets in RAS were predicted using computational approaches in 2011 by Grant et al. [45], and these pieces of evidence exist for the two of them, p1 and p2, at this time.

Owing to the initial failures in drugging RAS and thus marked as undruggable for several decades, the revolutionary studies in the year 2012 onwards acted as a game changer in the design of anticancer RAS-targeting drugs and [29–32] guided the field towards the FDA-approved drugs, thereby marking the beginning of an era with immense potential for the future drug-discovery efforts in the field [26]. A variety of inhibitors of RAS, such as cyclic compounds, natural derivatives, and biologics, have been reported in the literature [47,48], and here I focus only on a few selected articles. In 2010, Kalbitzer's group identified previously unknown binding sites in RAS using metal–cyclens (**Figure 17.2**) [43,44,46]. GppNHp-bound wild-type RAS showed two binding regions. One was located near the γ-phosphate where Cu^{2}-cyclen could directly interact with the γ-phosphate and the amines of the cyclen hydrogen bond with Gly12, Asp33, Ala35 (T35A mutant), and Ala59. NMR showed the perturbation of the following residues: Gly13, Tyr32, Ala59, Gly60, and Gln61. The second was the binding site 2 that was found distant from the nucleotide-binding site and localized near the loop7 and helix5 with the chemical shift perturbations (CSPs) most predominant for Asp105, Ser106, Asp107, Asp108, Val109, and Met111 residues and those near the C-terminus (or helix5), i.e. Glu162, Gln165, and His166. X-ray structure was solved with Zn^{2+}–cyclen bound to HRAS-wt-GppNHp (wt: wild type), and the binding region 2 was

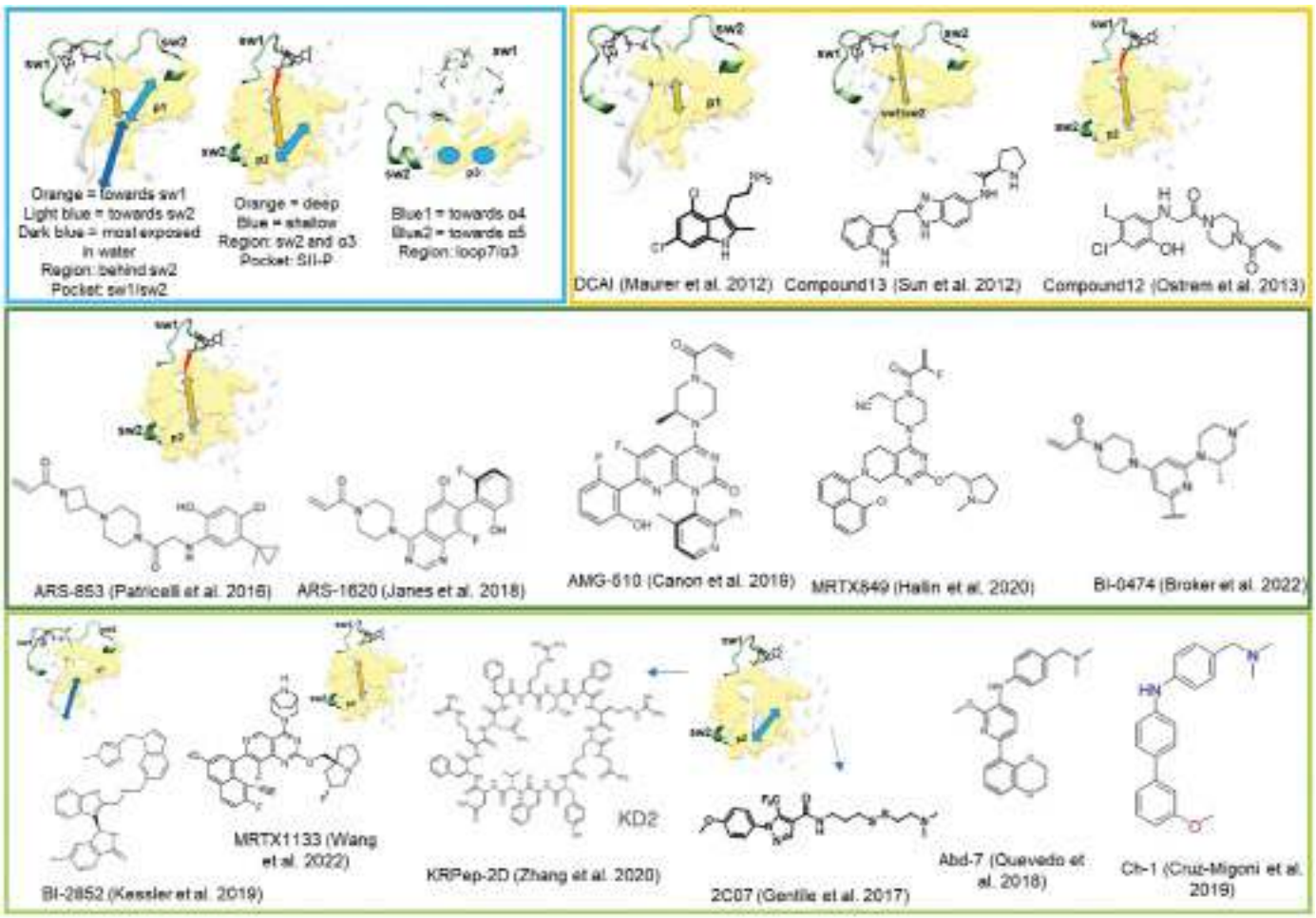


Figure 17.2. Blue box: Allosteric binding pockets on the surface of RAS. p1 or sw1/sw2 pocket. Left: The position where arrows meet is a common hydrophobic binding pocket, and then the compound can either be extended towards sw1 or sw2 (orange and blue arrows) or points towards the outside, e.g. BI-2852 compounds (dark blue arrow). Middle: pocket p2 or SII-P pocket as described by Shokat (Ostrem et al. [30]); the orange arrow shows the deeper pocket that is observed in the majority of compounds that bind in this region, such as MRTX894 (covalent) or MRTX1184 (non-covalent), as well as a shallow pocket that is nucleotide-specific and is shallower than that observed in 2C07 compounds (blue arrow). Red rugged arrow shows covalent linkage. Right: The pocket p3 that lies near the loop7 and helix3/helix4/helix5 of the region where the membrane anchor is positioned. Metal–cyclens are shown to bind this pocket. Orange box: The four initial studies that provided promising evidence towards the sw1/sw2 and SII-P binding pockets and showed abrogation of either RAS–SOS or RAS–effector interactions as a consequence of the compound/fragment binding. Dark green box: Compounds targeting the SII-P pocket with attached covalent warhead. Light green box: Compounds, peptides (KRpep-2D), cross-over compounds (Ch-1), and biologics-derived compounds (Abd-7) for non-G12C cases. See the text for discussion and Figure 17.1 for structural element description.

observed while binding site 1 was not. Metal–cyclens stabilized the weak-effector binding state, and isothermal titration calorimetry showed that the RAS-binding domain of RAF kinase to RAS–GppNHp-wt (wt: wild type) was perturbed in the presence of either Zn^{2+} or Cu^{2+} cyclen [43]. These studies served as initial evidence for pockets p1 and p3 (Figure 17.2).

In 2012 and 2013, three independent studies reported allosteric binding pockets on the surface of RAS for the first time with desired functional consequences (Figure 17.2). Fesik's group carried out an NMR-based fragment screening containing 11,000 fragments and identified about 140 fragments that bind the inactive GDP-bound state of KRAS [32]. A proprietary library obtained from ChemBridge, ChemDiv, and other sources was utilized. The fragment binding was in the range of 1–2 mM and observed for G12D-KRAS, G12V-KRAS, and wild-type KRAS and HRAS [32]. Chemotypes, such as indoles, phenols, and sulphonamides, were identified from the screening. The fragment library was constructed utilizing the rule of three with molecular weight ≤300, *c* log *P* ≤ 3.0, ≤3 hydrogen bond donors, and ≤4 rotatable bonds. Library was filtered for including only the soluble compounds and fragments from known drug molecules but not including the reactive and poorly soluble fragments. The presence of a hydrogen bond donor such as –NH on the indole or the –OH on the phenol was reported important for binding (Figure 17.2). Substituents that occluded this hydrogen bond formation resulted in reduced binding affinity [32]. The binding pocket that was invisible in the apo enzyme became prominent in the presence of selected fragments. It was shown that the triad-interacting partners of Tyr71 (D54, R41, and S39] were perturbed and Met67 underwent a shift, thereby giving rise to slight movement (or changed conformation) of switch-2 and making the binding pocket more 'open' for binding

fragments. Primarily, it is the flip of Tyr71 away that opens the ligand binding site. Overlay of the X-ray structure of RAS bound to the compound on the RAS–SOS complex suggested that the compound will likely inhibit their interaction. The functional relevance was reported from the SOS-mediated nucleotide exchange assay where unlabelled GDP was exchanged by BODIPY-GTP, and a decrease in the fluorescence was observed in the presence of selected analogues of the fragments [32]. Another study that appeared around the same time in 2012 was from Genentech in which Maurer et al. [29] performed an NMR-fragment screening using a 3,300-compound library where compounds were obtained from commercially available sources. The criteria for their selection were ≤16 heavy atoms, *c* log $P \leq 3$, and no reactive functional groups. The CSPs of 25 compounds out of the total 240 primary hits mapped to a contiguous site, and these 25 fragments were selected as confirmed hits. These hits contained aromatic heterocycles with fused rings, directly linked or linked via one or two linker atoms. The screenings were performed using the KRAS-G12D mutant (referred in the chapter as $KRas_m$) with either GDP-bound or containing a mixture of GDP or GMppCp. All 25 hits showed CSPs for V8, T74/G75 and L56/D57, and additional regions around residues, such as K5, L6, V7, I55, L56, and T74. One of the bound fragments that was larger in size, DCAI, showed the perturbation of the downstream signalling and inhibited the SOS-mediated nucleotide exchange for RAS [29]. It blocked nucleotide exchange and released reactions with IC_{50} of 342 and 155 μM, respectively. DCAI was found to block the first phase of exchange reaction by preventing nucleotide release from KRAS. Mechanistic exploration revealed that DCAI interferes with SOS^{cat} (catalytic domain of SOS) binding with KRAS, thereby preventing nucleotide exchange, and that the compounds did not affect the RAS–effector interactions (though their inhibitory action in *in vivo* assays could not be ruled out). DCAI was tested both in the *in vitro* and *in vivo* assays [29]. The fragment-induced expansion of the binding pocket was observed, and the breaking of the salt bridge between R41 and D54 was the key for blocking the RAS–SOS interaction. The bound site of DCAI on RAS did not overlay with the effector binding region and thus expected to not affect the RAS–effector interaction. In 2013, Shima et al. [31] identified a set of Kobe compounds and its analogues that were specific for MRAS-GTP with P40D mutation (MRAS with P40 renders it HRAS type) and also showed activity for KRAS-G12V-GppNHp bound. These compounds were identified using an *in silico* screening of a library containing 40,882 compounds against a high-resolution crystal structure of $MRAS^{P40D}$–GppNHp. A total of ~100 compounds were tested in the *in vitro* assays for their potential to inhibit the RAS–RAF interaction (HRAS–GTP and $MRAS^{P40D}$–GppNHp). Two compounds, Kobe0065 and Kobe2602, showed potent activity to competitively inhibit the RAS–effector interaction for HRAS-G12V mutant. Similar inhibition was observed for KRAS-G12V as well. The compound did not engage D54 that likely serves as the reason for observed differences in the inhibitory activities of Kobe compounds and Maurer's DCAI compounds (**Figure 17.2**) [31].

17.3. Ras targeting: covalent inhibitors

Ostrem et al. showed the presence of an allosteric pocket that lie between switch-2 and helix3 and that was specific to the GDP-bound G12C mutant of KRAS [30]. This was a breakthrough study that utilized the tethering approach [25]. Nucleophilic Cys at the 12th position in the G12C mutant of KRAS was harnessed to make irreversible covalent modification first via chemical modification using the disulphide library followed by the attachment of other carbon-based electrophiles, such as acrylamides and vinyl sulphonamides. Initially, a disulfide fragment-based screening was performed with a library containing 480 compounds against KRAS-G12C with GDP-bound state. Both vinyl sulphonamides and acrylamides performed better with former than latter owing to its high reactivity. Different compounds induced shifts in the position of either switch-2 and switch-1 (e.g. compounds 6 and 8 from Ostrem et al. [30]). This clearly showed that different fragments had potential to induce different conformational changes in different regions of RAS. Compounds 8 and 12 were shown to shift the equilibrium towards the inactive GDP-bound KRAS-G12C and to abrogate RAS–RAF interaction in the G12C-specific cell lines [30]. This and several follow-up studies are among those hallmark studies that ignited the hope of drugging the undruggable RAS back again, and since that time we have indeed taken a leap forward [28,49–51]. Several efforts were initiated to improve the first-generation compounds that were developed by Shokat's group in 2013. For instance, to obtain binders with higher binding affinities and better cellular efficacy, such as from the first-generation compounds, the hydrophobic region near Met72 and the electrophile linker were optimized to achieve a relatively superior analogue from the previously reported compound 12 from Ostrem's 2013 article [30] (**Figure 17.2**); in the biochemical assays, ARS-853 engaged KRAS(G12C) and showed an improvement of 600-fold over compound 12 from Ostrem's original article (**Figure 17.2**). ARS-853 blocked the SOS-catalyzed nucleotide exchange and did not bind the active state of RAS as the compound occupied the binding pocket where the γ-phosphate of GTP is positioned [50]. In 2018, Janes et al. [49] published ARS-1620 as a further improvement over ARS-853, and the authors proved that their compound works under *in vivo* conditions by targeting the mutant KRAS [49] (**Figure 17.2**). Then, Amgen and Mirati's G12C-specific compounds were developed, i.e. AMG510 [52] and MRTX849 [53] (**Figure 17.2**). Another interesting strategy was proposed where crossover compounds were obtained by merging PPI binders (that are inactive) and biologics-derived compounds (that are active) to create new and potent RAS inhibitors, for instance, Ch-1, as shown in **Figure 17.2** [54].

While the majority of compounds were selective for inactive G12C mutant of KRAS, there are several other mutant KRAS forms that are involved in deadly cancer types, such as pancreatic and colorectal cancers [11,28]. Mirati developed MRTX1133 that is specific for G12D and does not rely on the covalent modifications [55], and it has recently received IND clearance by US FDA enabling phase 1 initiation for first-in-class oral KRAS-G12D-selective inhibitor. In addition, there are a number of studies with a similar pocket region (though not identical, e.g. see [56]) which could also be open in the active state of RAS, for instance, in G12D—though there are differences in the pocket region, i.e. inactive state G12C pockets are much deeper with a pronounced hydrophobic environment via V9 and M72 while those in the case of GTP-bound state have been reported to be more on the surface (**Figure 17.2**). 2C07 was the first report of a small-molecule binding to an active RAS state that inhibited the downstream signalling via the PI3K pathway [56]. A cyclic peptide KRAS-pep2D was reported to bind similar regions spanning

switch-2 and helix-3 [56–59]. KS-58 was reported later to be more potent [60]. The exploration of the SII-P pocket of KRAS for other reversible inhibitors has been recently reported by Shokat [24], and another recent study presented a fragment optimization approach for targeting the SII-P pocket to obtain reversible inhibitors [61]. The BI-2852 compound was reported as a high nanomolar binder and a tool compound for the sw1/sw2 pocket region as well as a viable option for targeting other mutants [42]. The progress on the covalent inhibitor design and mechanism has been clearly summarized by the leaders in the group in a number of excellent articles [27,28].

Studies worth mentioning here are those that were reported in late 1990s with collaborative efforts of Schering-Plough Research Institute and Agouron Pharmaceuticals [62,63]. Utilizing the NMR spectroscopy, spectroscopy-derived NOE restraints, molecular modelling, and Monte Carlo based refinements, Taveras et al. reported an allosteric pocket that extended in the switch-2 pocket of G12V GDP-bound HRAS [62]. The compound SCH-54292 showed greater solubility over others with an IC_{50} of 0.7 μM. Though the attempts made by using X-ray crystallography were failed, but the structural model obtained using Monte Carlo refinements showed an allosteric pocket that did not overlap with the nucleotide-binding site but rather extended below to lie between switch-2 and helix3 (similar to the SII-P or p2 pocket; Figure 17.2). The cellular effects were tested on NGF-induced neurite outgrowth of PC12 cells that have been linked to the RAS nucleotide exchange. One of the analogues of SCH-54292 showed inhibition of NGF-stimulated neurite outgrowth in the 10–20 μM range. However, SCH-54292 did not show cellular effects at this concentration, but other concentrations were not tested [62]. The study found that the binding of this inhibitor was enhanced at low Mg^{2+} concentrations. The answer to this question was addressed in a follow-up study where it was found that the low Mg^{2+} concentration stabilizes a new conformational state of the RAS [63]. Upon addition of excess Mg^{2+}, however, the chemical shifts reverted back to the regular conformational state. Therefore, the study showed that there are two conformational states of RAS and the interconversion occurs between them. The regions that showed major differences between the two states lie at the switch regions and near the nucleotide-binding site. The binding pocket of SCH-54292 was reported to be formed by similar residues as the covalent and non-covalent binders that bind the SII-P pocket (Figure 17.2). The reported residues were Gln99, Ile100, Val108, Met72, Tyr96, Ala11, Lys16, Gly12, Gly60, and Gln61. The naphthyl of the SCH-54292 was found to lie in a hydrophobic pocket formed by Met72, Gln99, Ile100, and Val103, and the phenol lie near Tyr96, Lys16, Gly60, and Gln61. Lys16 was proposed to be engaging the carbonyl of the compound.

17.4. RAS targeting: antibodies/monobodies

Antibody designs or mini-proteins are another area of research in RAS drug design and development, and a lot of effort was devoted by several different groups in this direction, for instance, [64–70] (Table 17.1). Recently reported R15 monobody trapped RAS in its apo state and effectively inhibited the interaction with the downstream binding partners [65]. In addition, the binding with monobody trapped multiple oncogenic mutants—suggesting the immense therapeutic potential for targeting the apo state of RAS. The mechanism of inhibition was shown to be a competition of R15 with GEFs. Using *in vitro* and *in vivo* studies, another pan-Ras iMab, inRas37, was identified that blocks the RAS–effector interaction and inhibits the growth of mutant form of RAS [71]. In this case, antibodies do not work in isolation, but combinations with other known inhibitors were tested and shown to work effectively. For instance, RAS mutants with concurrent PI3K mutations could work better by utilizing hybrid therapy, i.e. combining inRas37 and PI3K inhibitors; another protein YAP1 was found elevated in mutant-RAS-dependent colorectal tumours where combined therapy with inRas37 and Yap1 inhibitors alleviated the problem [71,72]. 12VC1 monobody was reported to be selective for the mutant RAS at the 12th position with binding affinity for G12C mutant being the most effective [70]. The X-ray structure was obtained for 12VC1 with HRAS-G12C mutant bound to GTPγS. The binding pocket was shallow near the nucleotide-binding site and engaged D33 from switch-1 and nucleotide, but the overall binding interface between 12VC1 and RAS was extended in the entire switch-1 and switch-2 region going downwards towards the β-strands. Proteasomal degradation is an emerging promising area for degrading oncogenic RAS [73]. 12VC1 monobody and its variants were used as a warhead and fused with the E3 ubiquitin ligase VHL and shown to degrade KRAS(G12C) and KRAS (G12V) [70]. Affimers are another promising biologics given the area that they

Table 17.1. RAS-antibody-based selected studies.

Antibody design	Reference
Intracellular antibody capture technology	[68,69]
NS1 monobody bound to the α4–α5 interface, binding region known	[67]
RT11-1 (first-generation) and Ras37 (second-generation) pan RAS monobody, block RAS–effector interactions, and binding site is not reported	[71,72]
12VC1 monobody, specific for active state of mutant RAS (12th position mutants such as G12C/G12V), binding site is known in complex with 12VC1+HRAS(G12C)GTPγS, and is a shallow pocket while K6 binds in the sw1/sw2 pocket	[70]
K3 and K6 affimers, binding pocket is known for both, and K3 affimer binds in the sw2/α3 pocket while k6 binds in the sw1/sw2 pocket	[64]
Jam-20 is a pan-RAS monobody that binds in the RAS switch region and blocks RAS–effector interactions	[75]
R15 monobody, selective for apo state of RAS, competes with GEF for binding RAS, and could either bind directly at the switch region or at an allostreric site	[65]

cover the surface of RAS [64]. Two affimers K3 and K6 were reported by a group in the United Kingdom [64]. K6 binds near the switch-1 and switch-2 interface, and K3 affimer binds between switch-2 and helix-3. K3 affimer demonstrated isoform specificity with the most effective inhibition of KRAS (144 ± 94 nM) followed by HRAS (2585 ± 335 nM), but IC_{50} is not obtainable for NRAS. NS1 monobody was found to bind at the allosteric lobe of RAS, i.e. the interface of helix-4 and helix-5 [67]. A group in the United Kingdom developed an effective proteolytic affinity-directed protein missile (AdPROM) that led to efficient degradation of endogenous target proteins [74]. They attached AdPROM to RAS-specific NS1 monobody and observed that RAS can undergo protein degradation using NS1 monobody as the warhead attached to VHL showed that RAS could be degraded. Another antibody is identified as JAM20 that is a pan-RAS monobody and targets the sw1/sw2 pocket [75]. Terrence Rabbits group identified the anti-RAS antibody that binds near the switch region and prevented the interaction of RAS with its downstream effectors [69] using the antibody-derived compound technology. The crystal structure showed that the antibody covered the same interface where RAS downstream effectors bind [69]. The pharmacophore generated from this antibody, a small-molecule Abd-7, was shown to inhibit the downstream signalling of RAS [76] (see Figure 17.2).

17.5. RAS targeting: molecular simulations to identify allosteric pockets and investigate water dynamics for structure-based drug discovery

Computational studies have played a pivotal role and span from predicting allosteric binding pockets in KRAS to understanding the conformational dynamics of its ligand-bound forms (such as AMG510 and ARS-893) and predicting kinetics. In 2011, a combined molecular dynamics simulation and bioinformatic approach predicted novel allosteric sites on the surface of RAS enzyme [45]. In this study, the conformational flexibility of the protein was taken into consideration by the generation of MD-derived ensembles and conformationally distinct X-ray structures. Next, a consensus obtained from multiple pocket identification methods (fragment-based: FTMAP; grid-based: AutoLigand; ligand-based: BlindDock) revealed a total of four novel allosteric pockets identified on the surface of RAS. No single method predicted all four pockets, thereby further emphasizing the importance of a consensus approach derived by using a variety of methodologies. One of the pockets, however, was identified by all three methods and this pocket was located near the membrane-proximal C-terminus (termed as pocket p3). The p3 pocket is formed by loop7, loop9, and helix5. These pockets were subjected to virtual screening against the ZINC and NCI compound libraries, and the cellular assays showed that the chosen binders inhibit the downstream signalling of the RAS pathway. Interestingly, the majority of the chosen experimental targets were predicted to bind the p3 pocket. Utilizing a combined approach of MD/computational methods and cell-based assays, this study showed the presence of four druggable allosteric pocket on the surface of RAS [45] and suggested that the p3 pocket could be subdivided into two regions, p3a and p3b.

There are several excellent studies reported in the literature that have applied mixed-solvent or probe-based MD on challenging drug targets, and here I focus on the one applied on RAS [77–80]. A computational equivalent of multi-solvent crystallography termed as probe-based molecular dynamics (pMD) has been applied to KRAS [41]. pMD is advantageous because the probe binding occurs when the protein is in motion. It is a grid-based approach where probe occupancies are calculated for the cubic grids referred to as voxels. The time-averaged number density per voxel is used to calculate the grid-binding free energy (ΔG_{grid}). Several careful filtering criteria are applied to ΔG_{grid} for every voxel, and the resulting interaction spots are further clustered based upon defined rules. These clustered interaction spots are utilized to estimate the binding affinities (K_d) and the druggability of distinct binding pockets. pMD predicted five druggable sites and three sub-sites on the surface of KRAS with binding affinities ranging from high micromolar to low millimolar [41]. These druggable sites were found to be in excellent agreement with the previously reported allosteric ligand-binding sites in RAS proteins, pockets p1, p2, p3, and p4, as described in Section 17.2. An excellent agreement between the pMD-derived and experimentally observed K_d values were observed as shown in Figure 17.3A. pMD

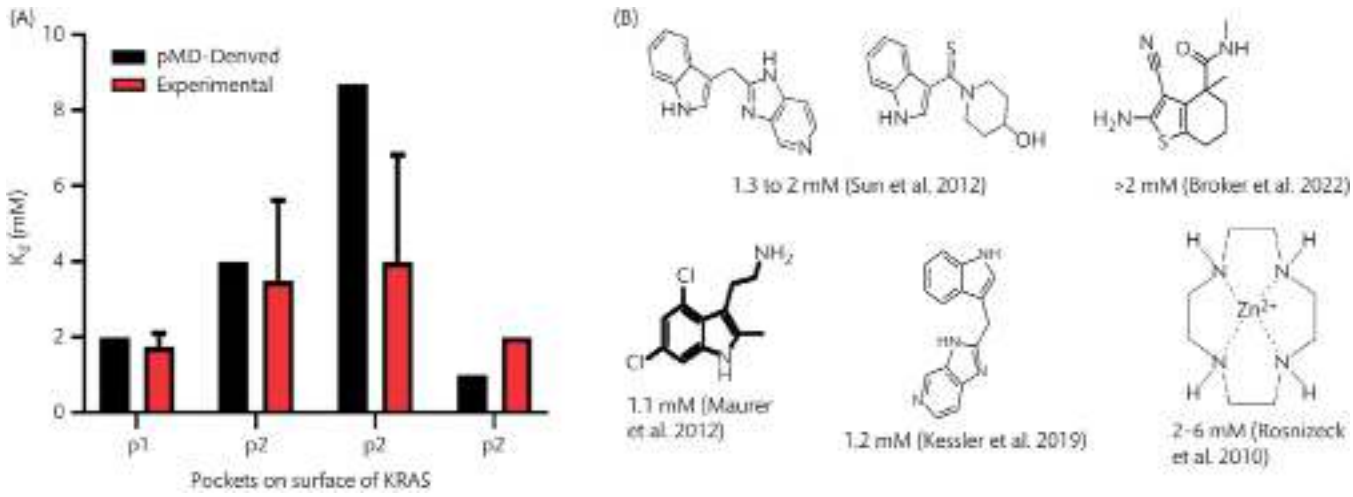


Figure 17.3. (A) The comparison of pMD-derived (probe-based molecular dynamics simulations) K_d values with experimentally determined ranges of fragments binding to RAS. (B) Selected fragments with their experimental K_d values.

estimates are more comparable with the fragment-bound RAS; therefore, I focused on a few selected fragment-bound studies of KRAS to convey the readers that pMD is an efficient technique to estimate K_d values. Some fragments are highlighted in Figure 17.3B. The p1 pocket appears to be in an excellent agreement between the MD-derived and experimentally observed K_d values (Figure 17.3). Note that while p2 is the most studied pocket currently, this is largely true in the context of Cys12 or covalently attached compounds to the Cys12 of KRAS-G12C mutant. In contrast, p1 pockets bound to several fragments/compounds are all non-covalent. pMD-derived results pinpointed a druggable site and a small sub-site comprising the p1 pocket; while the sub-site had an estimated K_d value of 55 mM, the druggable site estimated K_d was 2 mM. While averaging the two resulted in a combined K_d of 0.12 mM, but in Figure 17.3A a predicted K_d of 2 mM was used for the druggable binding pocket. This is because the fragment-based studies reported a deeper hydrophobic cavity in the p1 pocket combined with a shallow cleft formed by adjacent residues in switch1, E37, and D38 [29,31,32], and a K_d of 2 mM is estimated for the hydrophobic cavity of p1. p2 pocket's pMD-derived K_d and its comparison with experiments could not be performed by Prakash et al. [41], but with the recently reported reversible covalent inhibitor design approach from the Fesik group this could be achieved now. This is because, in addition to the S39-site modified fragment screen, Broker et al. [61] also reported fragments in G12V-KRAS without any other modifications in the switch region (see Figure 17.3B). While earlier studies that supported the p3 binding pockets were based upon metal–cyclens [43,44], rigorous computational work including molecular dynamics simulations combined with virtual screening led to the identification of two compounds that bind the p3 pocket. Extensive experimental characterization for these compounds is reported in the published studies, as discussed below, and one patent that appears to be the first patent targeting this novel pocket. These compounds are shown to interfere either with the calmodulin binding with KRAS-G12V [81] or are shown to impair the interaction of KRAS-G12D with BRAF, thereby preventing further downstream signalling [82]. Feng et al. performed a virtual screen against the previously reported p3 pocket and found 77 promising candidates. Several biochemical screening approaches, such as microscale thermophoresis, thermal-shift assay, NMR line broadening, and HSQC NMR spectroscopy, were applied to the promising candidates and KAL-21404358 was found as the most favourable compound showing positive results across all biochemical assays. Abuasaker et al. reported compound P14B to increase the apoptosis in DLD-1 cells that harbour an oncogenic allele, thereby making these compounds and this binding site important. The binding affinities of these p3 binders were relatively low as compared to the p2/p1 binders reported in the literature which is interesting given this pocket is proximal to the membrane. KAL-21404358 was shown to have a variable binding affinity for different nucleotide states that ranged from 88 to 146 μM, whereas P14B was reported to be even lower, i.e. 32.8 μM. Differences in the interacting residues forming the binding pocket may be responsible for the observed differences, for instance, while KAL's binding pocket is formed by R97, D105, S106, E107, D108, V109, P110, M111, W137, G138, I139, E162, K165, and H166, the P14B was situated a bit more proximal to the HVR with additional contributions from the residues K169 and E168. Additionally, latter study by Abusaker et al. showed two different binding modes of the same compound with significant differences in the pocket environment. Similar differences in the p3 pocket have been reported previously by Grant et al. [45] where the p3 pocket was predicted to occur as p3a and p3b. Prior to 2019, the evidence for the p3 pocket was metal–cyclens alone (Prakash et al, 2015), but since they will be much closer to the fragments as compared to the full compounds (KAL-21404358 and P14), the cyclens were used to compare the pMD-derived K_d with the experimentally reported values (Figure 17.3A). The p4 pocket is the one where metal–cyclen binding has been shown and no crystal structure is available, but the K_d value was obtained from the NMR study as shown in Figure 17.3A [46]. The pMD study also showed immense potential in predicting the protein–protein interaction surfaces as discussed by Prakash et al.

In any drug-discovery campaign, displacing unfavourable water molecules is a standard technique used for enhancing binding affinity. Given the water dynamics, specifically if the binding pocket is lined by flexible region such as in the case of RAS, these predictions become very challenging. For KRAS Q61H mutant, a belt of six water molecules with high residence times obtained from an extensive molecular dynamics simulation approach has been reported that likely connects the distant nucleotide-binding pocket region with the membrane-proximal allosteric lobe [83]. This study showed that the perturbation of key water molecules has potential to modulate the conformational state of RAS. Of all, perturbation in W4 led to major shifts in the conformational state of the protein with W4 acting more like a part of the protein playing functional role. W3, on the other hand, was observed to be less affected by the changes in the dynamics of even switch-2 to which it directly associates (W3 is associated with Thr58 and Gly10). A computational study published afterwards determined a number of conserved water sites on RAS, and their water # 307, 302, and 306 were identical to W1, W3, and W6 (Figure 17.4). Quite interestingly, a recent study showed that water coordinating Gly10 and Thr58 (similar to W3 from Prakash et al. [83] and W302 from Kearney et al. [84]) stays conserved in the majority of X-ray structures of G12C-KRAS bound with non-covalent compounds in the p2 pocket (or SII-P pocket), and whenever it is replaced by compounds, the observed effect is suboptimal [85]. It is interesting to observe that this water remains intact though the Watermap predictions indicated otherwise [85]. Khrenova et al. combined the molecular dynamics and QM/MM approach and demonstrated an excellent agreement between their predicted K_{inact} and K_i values and the experimental values for the KRAS-G12C

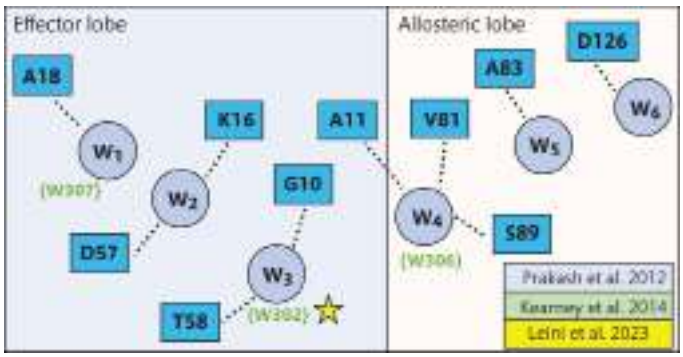


Figure 17.4. A network of conserved and high-residence water molecules connecting different regions of RAS as predicted from computational studies.

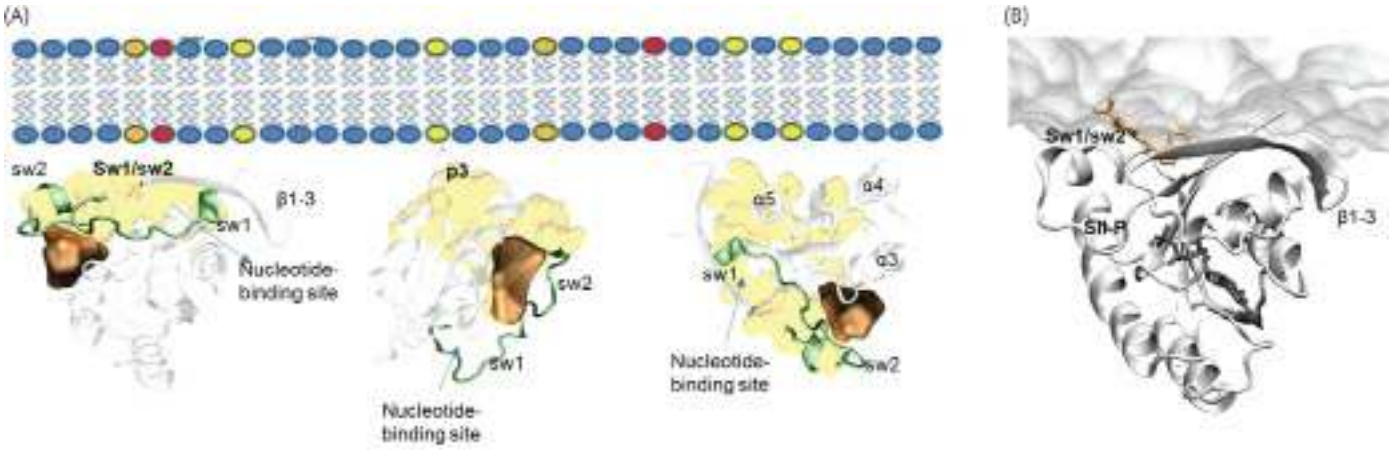


Figure 17.5. Literature reported orientations of RAS near the membrane surface. (A) Left: The occluded orientation where the sw1/sw2 pocket directly faces the membrane. Middle: G-domain orientation where the G-domain is largely solvent exposed; in this orientation, the p3 pocket or the membrane-proximal pocket faces towards the membrane. Right: G-domain orientation where the allosteric-lobe residues make a direct contact with the membrane. As an example, NS1 monobody is shown to bind this RAS interface. (B) RAS in complex with compound 2 as reported in Fang et al. (compound 2 is in orange).

complex with ARS-853. Their study is a first of its kind that utilizes the computational simulations for the calculation of K_{inact} and K_i values for covalent inhibitor bound protein complexes as applied to KRAS.

17.6. Conclusions

Computational approaches, such as molecular dynamics simulations, virtual screening approaches, molecular docking, and others, have immense power in exploring mechanisms, generating multiple substates, predicting kinetics, to name a few, and hence are useful contributors in the drug-discovery processes [13,86–89]. MD studies with covalently bound AMG510 with KRAS and ARS-853 provided insights into the mechanisms of inhibitor interaction and the conformational dynamics of the switch regions in the presence of covalent inhibitors explaining the dependence of the rate constant of covalent complex formation on the inhibitor's concentration, and showed that different oxidation states of Cys12 stabilize KRAS-G12C in G12D mutant-specific conformation, among others [86,90,91]. MD simulations provide information on atomic details of protein–drug interactions, predict cryptic allosteric pockets, provide information on the kinetics, and may provide mechanistic explanation for observations to progress the drug design process. The role of membrane in RAS biology is well known, and there have been several computational and experimental studies that are available in the literature for interested readers on this topic [13,34,37,40,87,88,92–102]. Studies have reported the K_d of RAS–membrane association, for instance, using surface plasmon resonance and neutron reflectometry. Van et al. [101] showed that KRAS preferential association with anionic model membranes (a 70:30 POPC:POPS composition) had a dissociation constant of $K_d \approx 1.5$ μM. The pMD approach predicted the druggable binding pockets on the surface of RAS, and the predicted K_d values were found to be in good agreement with the known allosteric pockets. The pMD approach was further extended to pMD membrane where the probe-based MD was carried out in the presence of membrane for a membrane-bound RAS complex. Although the determination of K_d values, specifically for pockets proximal to membrane, could be a bit challenging, it indicated significant differences in binding pocket dynamics when compared with solution pMD [103]. MD studies have shown a direct and transient interaction of its G domain with the membrane that may have potential for novel drug-discovery approaches (Figure 17.5). The membrane–RAS interface for drug design purposes has been explored by Ikura's group to understand the mechanism of action of compounds that were found to be active in the presence of phospholipids developed by Novartis [104,105]. In addition, MD studies have indicated that the direct and transient interaction of downstream players of RAS, such as RAF kinase with the membrane, may hold functional relevance during RAF activation that is likely generalizable to a number of kinases [36,99]. Due to multiple complexities that underlie the membrane interactions of RAS and the membrane–RAS interface, it requires further exploration and is an active area of research by several eminent scientists. Year 2010 onwards there have been an immense progress in the field of RAS drug discovery that eventually led to the identification of the first FDA-approved drug in 2021. Combined contributions from a variety of computational and experimental approaches have encouraged towards this success and will continue to do so in the future.

REFERENCES

1. Vetter IR, Wittinghofer A. The guanine nucleotide-binding switch in three dimensions. Science. 2001;**294**(5545):1299–304.
2. Wittinghofer A, Vetter IR. Structure-function relationships of the G domain, a canonical switch motif. Annu Rev Biochem. 2011;**80**:943–71.
3. Bourne HR, Sanders DA, McCormick F. The GTPase superfamily: a conserved switch for diverse cell functions. Nature. 1990;**348**(6297):125–32.
4. Wittinghofer A, Waldmann H. Ras—a molecular switch involved in tumor formation. Angew Chem Int Ed Engl. 2000;**39**(23):4192–214.

5. Cherfils J, Zeghouf M. Regulation of small GTPases by GEFs, GAPs, and GDIs. Physiol Rev. 2013;**93**(1):269–309.
6. Bos JL, Rehmann H, Wittinghofer A. GEFs and GAPs: critical elements in the control of small G proteins. Cell. 2007;**129**(5):865–77.
7. Hunter JC, Manandhar A, Carrasco MA, Gurbani D, Gondi S, Westover KD. Biochemical and structural analysis of common cancer-associated KRAS mutations. Mol Cancer Res. 2015;**13**(9):1325–35.
8. Boriack-Sjodin PA, Margarit SM, Bar-Sagi D, Kuriyan J. The structural basis of the activation of Ras by Sos. Nature. 1998;**394**(6691):337–43.
9. Scheffzek K, Ahmadian MR, Kabsch W, Wiesmuller L, Lautwein A, Schmitz F, et al. The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. Science. 1997;**277**(5324):333–8.
10. Ahmadian MR, Stege P, Scheffzek K, Wittinghofer A. Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. Nat Struct Biol. 1997;**4**(9):686–9.
11. Cox AD, Der CJ. Ras history: the saga continues. Small GTPases. 2010;**1**(1):2–27.
12. Gorfe AA, Grant BJ, McCammon JA. Mapping the nucleotide and isoform-dependent structural and dynamical features of Ras proteins. Structure. 2008;**16**(6):885–96.
13. Prakash P, Gorfe AA. Lessons from computer simulations of Ras proteins in solution and in membrane. Biochim Biophys Acta. 2013;**1830**(11):5211–8.
14. Gorfe AA. Mechanisms of allostery and membrane attachment in Ras GTPases: implications for anti-cancer drug discovery. Curr Med Chem. 2010;**17**(1):1–9.
15. Hancock JF. Lipid rafts: contentious only from simplistic standpoints. Nat Rev Mol Cell Biol. 2006;**7**(6):456–62.
16. Hancock JF, Magee AI, Childs JE, Marshall CJ. All Ras proteins are polyisoprenylated but only some are palmitoylated. Cell. 1989;**57**(7):1167–77.
17. Castellano E, Santos E. Functional specificity of ras isoforms: so similar but so different. Genes Cancer. 2011;**2**(3):216–31.
18. Inder K, Harding A, Plowman SJ, Philips MR, Parton RG, Hancock JF. Activation of the MAPK module from different spatial locations generates distinct system outputs. Mol Biol Cell. 2008;**19**(11):4776–84.
19. Yan J, Roy S, Apolloni A, Lane A, Hancock JF. Ras isoforms vary in their ability to activate Raf-1 and phosphoinositide 3-kinase. J Biol Chem. 1998;**273**(37):24052–6.
20. Downward J. Targeting RAS signalling pathways in cancer therapy. Nat Rev Cancer. 2003;**3**(1):11–22.
21. Lee S, Rauch J, Kolch W. Targeting MAPK signaling in cancer: mechanisms of drug resistance and sensitivity. Int J Mol Sci. 2020;**21**(3):1102.
22. Haigis KM, Kendall KR, Wang Y, Cheung A, Haigis MC, Glickman JN, et al. Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. Nat Genet. 2008;**40**(5):600–8.
23. Ostrem JM, Shokat KM. Direct small-molecule inhibitors of KRAS: from structural insights to mechanism-based design. Nat Rev Drug Discov. 2016;**15**(11):771–85.
24. Vasta JD, Peacock DM, Zheng QH, Walker JA, Zhang ZY, Zimprich CA, et al. KRAS is vulnerable to reversible switch-II pocket engagement in cells. Nat Chem Biol. 2022;**18**(6):596–604.
25. Erlanson DA, Braisted AC, Raphael DR, Randal M, Stroud RM, Gordon EM, et al. Site-directed ligand discovery. Proc Natl Acad Sci USA. 2000;**97**(17):9367–72.
26. Stephen AG, Esposito D, Bagni RK, McCormick F. Dragging Ras back in the ring. Cancer Cell. 2014;**25**(3):272–81.
27. Westover KD, Janne PA, Gray NS. Progress on covalent inhibition of KRAS(G12C). Cancer Discov. 2016;**6**(3):233–4.
28. Zheng QH, Peacock DM, Shokat KM. Drugging the next undruggable KRAS Allele-Gly12Asp. J Med Chem. 2022;**65**(4):3119–22.
29. Maurer T, Garrenton LS, Oh A, Pitts K, Anderson DJ, Skelton NJ, et al. Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. Proc Natl Acad Sci U S A. 2012;**109**(14):5299–304.
30. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. Nature. 2013;**503**(7477):548–551.
31. Shima F, Yoshikawa Y, Ye M, Araki M, Matsumoto S, Liao JL, et al. In silico discovery of small-molecule Ras inhibitors that display antitumor activity by blocking the Ras-effector interaction. Proc Natl Acad Sci USA. 2013;**110**(20):8182–7.
32. Sun Q, Burke JP, Phan J, Burns MC, Olejniczak ET, Waterson AG, et al. Discovery of small molecules that bind to K-Ras and inhibit SOS-mediated activation. Angew Chem Int Edit. 2012;**51**(25):6140–3.
33. Gorfe AA, Cho KJ. Approaches to inhibiting oncogenic K-Ras. Small GTPases. 2021;**12**(2):96–105.
34. Hancock JF. Ras proteins: different signals from different locations. Nat Rev Mol Cell Biol. 2003;**4**(5):373–84.
35. Ingolfsson HI, Neale C, Carpenter TS, Shrestha R, Lopez CA, Tran TH, et al. Machine learning-driven multiscale modeling reveals lipid-dependent dynamics of RAS signaling proteins. Proc Natl Acad Sci U S A. 2022;**119**(1):e2113297119.
36. Prakash P. A regulatory role of membrane by direct modulation of the catalytic kinase domain. Small GTPases. 2021;**12**(4):246–56.
37. Simanshu DK, Philips MR, Hancock JF. Consensus on the RAS dimerization hypothesis: strong evidence for lipid-mediated clustering but not for G-domain-mediated interactions. Mol Cell. 2023;**83**(8):1210–5.
38. Van QN, Prakash P, Shrestha R, Balius TE, Turbyville TJ, Stephen AG. RAS nanoclusters: dynamic signaling platforms amenable to therapeutic intervention. Biomolecules. 2021;**11**(3):377.
39. Weise K, Kapoor S, Denter C, Nikolaus J, Opitz N, Koch S, et al. Membrane-mediated induction and sorting of K-Ras microdomain signaling platforms. J Am Chem Soc. 2011;**133**(4):880–7.
40. Zhou Y, Prakash P, Gorfe AA, Hancock JF. Ras and the plasma membrane: a complicated relationship. Cold Spring Harb Perspect Med. 2018;**8**(10):a031831.
41. Prakash P, Hancock JF, Gorfe AA. Binding hotspots on K-ras: Consensus ligand binding sites and other reactive regions from probe-based molecular dynamics analysis. Proteins. 2015;**83**(5):898–909.
42. Kessler D, Gmachl M, Mantoulidis A, Martin LJ, Zoephel A, Mayer M, et al. Drugging an undruggable pocket on KRAS. Proc Natl Acad Sci U S A. 2019;**116**(32):15823–9.
43. Rosnizeck IC, Graf T, Spoerner M, Trankle J, Filchtinski D, Herrmann C, et al. Stabilizing a weak binding state for effectors in the human ras protein by cyclen complexes. Angew Chem Int Ed Engl. 2010;**49**(22):3830–3.

44. Rosnizeck IC, Spoerner M, Harsch T, Kreitner S, Filchtinski D, Herrmann C, et al. Metal-bis(2-picolyl)amine complexes as state 1(T) inhibitors of activated Ras protein. Angew Chem Int Ed Engl. 2012;**51**(42):10647–51.
45. Grant BJ, Lukman S, Hocker HJ, Sayyah J, Brown JH, McCammon JA, et al. Novel allosteric sites on Ras for lead generation. PLoS ONE. 2011;**6**(10):e25711.
46. Kalbitzer HR, Spoerner M. State 1(T) inhibitors of activated Ras. Enzymes. 2013;**33 Pt A**:69–94.
47. Buyanova M, Cai S, Cooper J, Rhodes C, Salim H, Sahni A, et al. Discovery of a bicyclic peptidyl pan-Ras inhibitor. J Med Chem. 2021;**64**(17):13038–53.
48. Hocker HJ, Cho KJ, Chen CY, Rambahal N, Sagineedu SR, Shaari K, et al. Andrographolide derivatives inhibit guanine nucleotide exchange and abrogate oncogenic Ras function. Proc Natl Acad Sci U S A. 2013;**110**(25):10201–6.
49. Janes MR, Zhang J, Li LS, Hansen R, Peters U, Guo X, et al. Targeting KRAS mutant cancers with a covalent G12C-specific inhibitor. Cell. 2018;**172**(3):578–89 e17.
50. Patricelli MP, Janes MR, Li LS, Hansen R, Peters U, Kessler LV, et al. Selective inhibition of oncogenic KRAS output with small molecules targeting the inactive state. Cancer Discov. 2016;**6**(3):316–29.
51. FDA Approves First KRAS Inhibitor: Sotorasib. Cancer Discov. 2021;**11**(8):OF4.
52. Canon J, Rex K, Saiki AY, Mohr C, Cooke K, Bagal D, et al. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. Nature. 2019;**575**(7781):217–23.
53. Hallin J, Engstrom LD, Hargis L, Calinisan A, Aranda R, Briere DM, et al. The KRAS(G12C) inhibitor MRTX849 provides insight toward therapeutic susceptibility of KRAS-mutant cancers in mouse models and patients. Cancer Discov. 2020;**10**(1):54–71.
54. Cruz-Migoni A, Canning P, Quevedo CE, Bataille CJR, Bery N, Miller A, et al. Structure-based development of new RAS-effector inhibitors from a combination of active and inactive RAS-binding compounds. P Natl Acad Sci USA. 2019;**116**(7):2545–50.
55. Wang X, Allen S, Blake JF, Bowcut V, Briere DM, Calinisan A, et al. Identification of MRTX1133, a noncovalent, potent, and selective KRAS(G12D) Inhibitor. J Med Chem. 2022;**65**(4):3123–33.
56. Gentile DR, Rathinaswamy MK, Jenkins ML, Moss SM, Siempelkamp BD, Renslo AR, et al. Ras binder induces a modified switch-II pocket in GTP and GDP states. Cell Chem Biol. 2017;**24**(12):1455–66 e14.
57. Niida A, Sasaki S, Yonemori K, Sameshima T, Yaguchi M, Asami T, et al. Investigation of the structural requirements of K-Ras(G12D) selective inhibitory peptide KRpep-2d using alanine scans and cysteine bridging. Bioorg Med Chem Lett. 2017;**27**(12):2757–61.
58. Sogabe S, Kamada Y, Miwa M, Niida A, Sameshima T, Kamaura M, et al. Crystal structure of a human K-Ras G12D mutant in complex with GDP and the cyclic inhibitory peptide KRpep-2d. ACS Med Chem Lett. 2017;**8**(7):732–6.
59. Sakamoto K, Kamada Y, Sameshima T, Yaguchi M, Niida A, Sasaki S, et al. K-Ras(G12D)-selective inhibitory peptides generated by random peptide T7 phage display technology. Biochem Biophys Res Commun. 2017;**484**(3):605–11.
60. Sakamoto K, Masutani T, Hirokawa T. Generation of KS-58 as the first K-Ras(G12D)-inhibitory peptide presenting anti-cancer activity in vivo. Sci Rep-UK. 2020;**10**(1):21671.
61. Broker J, Waterson AG, Smethurst C, Kessler D, Bottcher J, Mayer M, et al. Fragment optimization of reversible binding to the switch II pocket on KRAS leads to a potent, in vivo active KRASG12C inhibitor. J Med Chem. 2022 Nov 10;**65**(21):14614–29.
62. Taveras AG, Remiszewski SW, Doll RJ, Cesarz D, Huang EC, Kirschmeier P, et al. Pas oncoprotein inhibitors: The discovery of potent, ras nucleotide exchange inhibitors and the structural determination of a drug–protein complex. Bioorgan Med Chem. 1997;**5**(1):125–33.
63. Ganguly AK, Wang YS, Pramanik BN, Doll RJ, Snow ME, Taveras AG, et al. Interaction of a novel GDP exchange inhibitor with the Ras protein. Biochemistry-US. 1998;**37**(45):15631–7.
64. Haza KZ, Martin HL, Rao A, Turner AL, Saunders SE, Petersen B, et al. RAS-inhibiting biologics identify and probe druggable pockets including an SII-alpha3 allosteric site. Nat Commun. 2021;**12**(1):4045.
65. Khan I, Koide A, Zuberi M, Ketavarapu G, Denbaum E, Teng KW, et al. Identification of the nucleotide-free state as a therapeutic vulnerability for inhibition of selected oncogenic RAS mutants. Cell Rep. 2022;**38**(6):110322.
66. McGee JH, Shim SY, Lee SJ, Swanson PK, Jiang SY, Durney MA, et al. Exceptionally high-affinity Ras binders that remodel its effector domain. J Biol Chem. 2018;**293**(9):3265–80.
67. Spencer-Smith R, Koide A, Zhou Y, Eguchi RR, She F, Gajwani P, et al. Inhibition of RAS function through targeting an allosteric regulatory site. Nat Chem Biol. 2017;**13**(1):62–8.
68. Tanaka T, Rabbitts TH. Intrabodies based on intracellular capture frameworks that bind the RAS protein with high affinity and impair oncogenic transformation. Embo J. 2003;**22**(5):1025–35.
69. Tanaka T, Williams RL, Rabbitts TH. Tumour prevention by a single antibody domain targeting the interaction of signal transduction proteins with RAS. EMBO J. 2007;**26**(13):3250–9.
70. Teng KW, Tsai ST, Hattori T, Fedele C, Koide A, Yang C, et al. Selective and noncovalent targeting of RAS mutants for inhibition and degradation. Nat Commun. 2021;**12**(1):2656.
71. Shin SM, Kim JS, Park SW, Jun SY, Kweon HJ, Choi DK, et al. Direct targeting of oncogenic RAS mutants with a tumor-specific cytosol-penetrating antibody inhibits RAS mutant-driven tumor growth. Sci Adv. 2020;**6**(3):eaay2174.
72. Shin SM, Choi DK, Jung K, Bae J, Kim JS, Park SW, et al. Antibody targeting intracellular oncogenic Ras mutants exerts anti-tumour effects after systemic administration. Nat Commun2017 May 10;**8**:15090.
73. Bekes M, Langley DR, Crews CM. PROTAC targeted protein degraders: the past is prologue. Nat Rev Drug Discov. 2022;**21**(3):181–200.
74. Roth S, Macartney TJ, Konopacka A, Chan KH, Zhou HJ, Queisser MA, et al. Targeting endogenous K-RAS for degradation through the affinity-directed protein missile system. Cell Chem Biol. 2020;**27**(9):1151–1163.e6.
75. Wallon L, Khan I, Teng KW, Koide A, Zuberi M, Li J, et al. Inhibition of RAS-driven signaling and tumorigenesis with a pan-RAS monobody targeting the Switch I/II pocket. Proc Natl Acad Sci U S A. 2022;**119**(43):e2204481119.
76. Quevedo CE, Cruz-Migoni A, Bery N, Miller A, Tanaka T, Petch D, et al. Small molecule inhibitors of RAS-effector protein interactions derived using an intracellular antibody fragment. Nat Commun. 2018;**9**(1):3169.
77. Seco J, Luque FJ, Barril X. Binding site detection and druggability index from first principles. J Med Chem. 2009;**52**(8):2363–71.

78. Bakan A, Nevins N, Lakdawala AS, Bahar I. Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. J Chem Theory Comput. 2012;**8**(7):2435–47.
79. Guvench O, MacKerell AD, Jr. Computational fragment-based binding site identification by ligand competitive saturation. PLoS Comput Biol. 2009;**5**(7):e1000435.
80. Raman EP, Yu W, Guvench O, Mackerell AD. Reproducing crystal binding modes of ligand functional groups using site-identification by ligand competitive saturation (SILCS) simulations. J Chem Inf Model. 2011;**51**(4):877–96.
81. Abuasaker B, Garrido E, Vilaplana M, Gomez-Zepeda JD, Brun S, Garcia-Cajide M, et al. α4-α5 Helices on surface of KRAS can accommodate small compounds that increase KRAS signaling while inducing CRC cell death. Int J Mol Sci. 2023;**24**:18.
82. Feng HZ, Zhang Y, Bos PH, Chambers JM, Dupont MM, Stockwell BR. K-Ras(G12D) has a potential allosteric small molecule binding site. Biochemistry-US. 2019;**58**(21):2542–54.
83. Prakash P, Sayyed-Ahmad A, Gorfe AA. The role of conserved waters in conformational transitions of Q61H K-ras. PLoS Comput Biol. 2012;**8**(2):e1002394.
84. Kearney BN, Johnson CW, Roberts DM, Swartz P, Mattos C. DRoP: a water analysis program identifies Ras-GTP-specific pathway of communication between membrane-interacting regions and the active site. J Mol Biol. 2014;**426**(3):611–29.
85. Leini R, Pantsar T. In silico evaluation of the Thr58-associated conserved water with KRAS switch-II pocket binders. J Chem Inf Model. 2023;**63**(5):1490–505.
86. Pantsar T. KRAS(G12C)-AMG 510 interaction dynamics revealed by all-atom molecular dynamics simulations. Sci Rep-UK. 2020;**10**(1):11992.
87. Prakash P, Litwin D, Liang H, Sarkar-Banerjee S, Dolino D, Zhou Y, et al. Dynamics of membrane-bound G12V-KRAS from simulations and single-molecule FRET in native nanodiscs. Biophys J. 2019;**116**(2):179–83.
88. Prakash P, Gorfe AA. Probing the conformational and energy landscapes of KRAS membrane orientation. J Phys Chem B. 2019;**123**(41):8644–52.
89. Lu SY, Ni D, Wang CX, He XH, Lin HW, Wang Z, et al. Deactivation pathway of Ras GTPase underlies conformational substates as targets for drug design. ACS Catal. 2019;**9**(8):7188–96.
90. Khrenova MG, Kulakova AM, Nemukhin AV. Proof of concept for poor inhibitor binding and efficient formation of covalent adducts of KRAS(G12C) and ARS compounds. Org Biomol Chem. 2020;**18**(16):3069–81.
91. Huynh MV, Parsonage D, Forshaw TE, Chirasani VR, Hobbs GA, Wu H, et al. Oncogenic KRAS G12C: kinetic and redox characterization of covalent inhibition. J Biol Chem. 2022;**298**(8):102186.
92. Abankwa D, Gorfe AA, Inder K, Hancock JF. Ras membrane orientation and nanodomain localization generate isoform diversity. Proc Natl Acad Sci U S A. 2010;**107**(3):1130–5.
93. Abankwa D, Hanzal-Bayer M, Ariotti N, Plowman SJ, Gorfe AA, Parton RG, et al. A novel switch region regulates H-ras membrane orientation and signal output. EMBO J. 2008;**27**(5):727–35.
94. Cao S, Chung S, Kim S, Li Z, Manor D, Buck M. K-Ras G-domain binding with signaling lipid phosphatidylinositol (4,5)-phosphate (PIP2): membrane association, protein orientation, and function. J Biol Chem. 2019;**294**(17):7068–84.
95. Kapoor S, Triola G, Vetter IR, Erlkamp M, Waldmann H, Winter R. Revealing conformational substates of lipidated N-Ras protein by pressure modulation. Proc Natl Acad Sci USA. 2012;**109**(2):460–5.
96. Lee KY, Fang Z, Enomoto M, Gasmi-Seabrook G, Zheng L, Koide S, et al. Two distinct structures of membrane-associated homodimers of GTP- and GDP-bound KRAS4B revealed by paramagnetic relaxation enhancement. Angew Chem Int Ed Engl. 2020;**59**(27):11037–45.
97. Lopez CA, Agarwal A, Van QN, Stephen AG, Gnanakaran S. Unveiling the dynamics of KRAS4b on lipid model membranes. J Membrane Biol. 2021;**254**(2):201–16.
98. Ngo VA, Garcia AE. Millisecond molecular dynamics simulations of KRas-dimer formation and interfaces. Biophys J. 2022;**121**(19):3730–44.
99. Prakash P, Hancock JF, Gorfe AA. Three distinct regions of cRaf kinase domain interact with membrane. Sci Rep-UK. 2019 Feb 14;**9**(1):2057.
100. Prakash P, Zhou Y, Liang H, Hancock JF, Gorfe AA. Oncogenic K-Ras binds to an anionic membrane in two distinct orientations: a molecular dynamics analysis. Biophys J. 2016;**110**(5):1125–38.
101. Van QN, Lopez CA, Tonelli M, Taylor T, Niu B, Stanley CB, et al. Uncovering a membrane-distal conformation of KRAS available to recruit RAF to the plasma membrane. Proc Natl Acad Sci USA. 2020;**117**(39):24258–68.
102. Zhou Y, Gorfe AA, Hancock JF. RAS nanoclusters selectively sort distinct lipid headgroups and acyl chains. Front Mol Biosci. 2021;**8**:686338.
103. Prakash P, Sayyed-Ahmad A, Gorfe AA. pMD-Membrane: a method for ligand binding site identification in membrane-bound proteins. PLoS Comput Biol. 2015 Oct 27;**11**(10):e1004469.
104. Fang ZH, Marshal CB, Nishikawa T, Gossert AD, Jansen JM, Jahnke W, et al. Inhibition of K-RAS4B by a unique mechanism of action: stabilizing membrane-dependent occlusion of the effector-binding site. Cell Chem Biol. 2018 Nov 15;**25**(11):1327–1336.e4.
105. Jansen JM, Wartchow C, Jahnke W, Fong SS, Tsang T, Pfister K, et al. Inhibition of prenylated KRAS in a lipid environment. PLoS ONE. 2017 Apr 6;**12**(4):e0174706.

SECTION 5

Statistical methods and data mining, machine learning, artificial intelligence, and cloud computing

18

The power of connection—enabling collaborative, multimodal data analysis at petabyte scale to advance understanding of oncology

Brandi N. Davis-Dusenbery, Cera R. Fisher, Rowan Beck, and Zelia F. Worman

18.1. Introduction

Rapid access to large quantities of diverse and high-quality biomedical data is transforming the research enterprise. This evolution has been particularly notable in the field of oncology and has been fuelled by the success of translating new discoveries into improvements in patient care. Today, researchers are routinely accessing and analysing multimodal datasets collected from tens of thousands of patients and representing petabytes of data. This scale is made possible by cloud-based, collaborative environments that enable the engagement of interdisciplinary groups including bench biologists, computer scientists, mathematicians, clinicians, and others to reveal novel insights from large, complex datasets.

Some of the most visible examples of the increased scale of oncology research are the large consortia studies that bring together people, data, and ideas across geographies and disciplines. For example, the landmark paper highlighting the findings of the International Cancer Genome Consortium (ICGC) included more than 1,300 authors from research institutions throughout the world [1]. The foundation of this work was the unified analysis of whole genome sequencing from 2,658 tumour-normal pairs across 38 tumour types. More than 100 people were involved in the orchestration of three computational workflows across multiple cloud environments. This harmonization required more than a year to complete but provided the foundation for international scientists to make practice-changing discoveries. Despite the breadth of this study, approximately 5% of the analysed samples did not contain a known driver mutation [1], suggesting that additional large studies are required to understand the full repertoire of cancer drivers. Fortunately, these future studies can take advantage of the significant advances in scalability and interoperability that would enable the completion of a similar project in just a few weeks.

18.2. Diverse data sources to support biomedical discovery

Biomedical data is generated from a wide range of sources and each imposes unique properties that shape how the data can be effectively used to drive discovery. For the purposes of this discussion, we define three major sources based on the origination of the data: Clinical Data (also known as Real World Data or RWD), Investigator-Initiated Studies, and Consortia-Initiated Studies. We have defined these primarily based on the origination of the data; however, they are neither exhaustive nor exclusive. For example, some Investigator-Initiated Studies may include Clinical Data and/or aggregations of Consortia Initiated-Studies. Each data source has a unique profile that shapes how the data can be effectively used to drive discovery (see Table 18.1). In the following sections, we dive more deeply into each data source and highlight opportunities for researchers to access it.

18.2.1. Real World Data

Data generated during the course of clinical care, also called Real World Data, has attracted significant interest over the past years due to the massive volume of data and the potential to reveal insights into vast populations of individuals. RWD includes the data derived from electronic medical records, billing claims, registries, patient surveys, or wearable devices [2]. The use of RWD is particularly exciting as it relates to addressing clinical and policy questions that may not be feasible to address using traditional clinical trials approaches. For example, significant progress has been made in identifying analytical techniques whereby trial treatment arms may be compared to synthetic control arms composed of RWD from a comparative population [3]. This can reduce the number of participants in a trial thus reducing costs and increasing speed. While the cost

Table 18.1. Biomedical data streams and characteristics.

	Clinical Data (RWD)	Investigator Initiated	Consortia Initiated
Volume[a]	High	Moderate	Low
Velocity[b]	Very high	High/moderate	Low
Cost to generate[c]	Low	Moderate	High
Standardization and QC[d]	Low	Moderate	High
Specificity[e]	Low	Variable	Low/moderate
FAIR-ness[f]	Low	Variable	Very high

[a]Volume of data generation: how much total data is generated per unit time.
[b]Velocity of data generation: how quickly is new data added.
[c]Cost to generate: what is the investment required to create the data.
[d]Standardization and QC: how uniform is the data.
[e]Specificity: was the data generated to test or refute a specific hypothesis.
[f]FAIR-ness: a compound measure evaluating Findability, Accessibility, Interoperability, and Reusability.

to generate RWD is low because it is being generated as part of care, the lack of standardization and specificity make deriving insights from RWD challenging. Furthermore, the highly personal nature of clinical data and the evolving landscape of patient consent mean that researcher access to that data is frequently extremely limited. Advances in tokenization, aggregation, and anonymization are beginning to remove these barriers, but further implementations are still needed to enable research [4]. Furthermore, RWD is increasingly being incorporated into Consortia-Initiated studies, which currently represent the most accessible way for most researchers to compliantly work with that data.

18.2.2. Investigator-Initiated Studies

A second data stream, referred to here as Investigator-Initiated studies, represents data collected through the efforts of a single or small number of collaborating laboratories. That data may include profiling or perturbation studies and is frequently oriented to a specific hypothesis. While generated at a lower velocity and volume than RWD, these studies collectively represent petabytes of data that is readily available to researchers for reuse. Continued efforts towards encouraging data sharing have improved not only the accessibility of that data but also the quality of submissions. For example, the revised US National Institute of Health (NIH) Data Management and Sharing Policy [5], as well as many scientific journals, mandate sharing of high-throughput data generated in the course of an analysis. There remains high variability across studies; however, continued attention on valuing data sharing aligned with the principles of Findable, Accessible, Interoperable, and Reusable (FAIR) data has increased the potential for secondary data use.

Increasingly, data repositories serve to convene researchers with a shared research focus across multiple high-throughput data sources, such as genomics, proteomics, imaging, and sensor data. The primary repository for sharing non-human, non-identifiable, or aggregated sequencing data is the Sequence Read Archive that as of 2020 catalogued more than 9 million records and 12 petabytes of data [6,7].

18.2.3. Consortia-Initiated Data

This leads us to the final primary data stream, Consortia-Initiated data. Adoption of robust data sharing and collaboration practices represents a relatively recent phenomenon in the biological sciences particularly in contrast to fields, such as physics, astronomy, and atmospheric sciences. While international collaborations like ICGC (discussed in Section 18.1) have been successful in bringing together existing datasets, generating new data under uniform methods can be costly and requires concerted funding, typically from governments or not-for-profit organizations [8]. These investments have proven transformative in our understanding of health and disease. For example, the analysis of longitudinal clinical data from participants in the Framingham Heart Study established obesity, increased blood cholesterol, and blood pressure as major contributors to cardiovascular disease [9]. Managing these risk factors is now a central component of standard of care, and over the past 50 years, the rate of heart disease has decreased nearly 70% in both men and women [10].

Just as the Framingham Heart Study provided the spark for long-term epidemiological studies, The Cancer Genome Atlas (TCGA) demonstrated the power of combining multiple high-throughput data modalities with uniform clinical data elements. TCGA was originally conceived and launched in 2005, just two years after the completion of the Human Genome Project. Over the next 12 years, samples from 11,000 patients with 33 distinct tumour types would be analysed across multiple modalities as part of the TCGA program. In all, more than 2.5 petabytes of data would be generated as part of TCGA. The primary analysis of that data transformed our understanding of molecular classification, oncogenic processes, and signalling pathways in cancer [11–13]. Importantly, that data contributed to a dramatic shift in clinical cancer care with molecular profiling increasingly serving a central role in determining treatment avenues, particularly in some cancer types such as glioma [14,15].

The impact of TCGA and other Consortia-Initiated studies extends far beyond the primary landmark papers describing the data generation. Indeed, because these datasets are typically generated with the express purpose of reuse, they represent a rich resource for quickly developing and testing new hypotheses. Despite the conclusion of TCGA in 2018, the number of papers including references to TCGA in either title or abstract has continued to increase years later, see **Figure 18.1**. These nearly 27,000 manuscripts represent just a fraction of the studies that have been informed by this incredible resource, and novel approaches powered by cloud computing are critical to ensuring that all researchers can effectively access and analyse resources like TCGA.

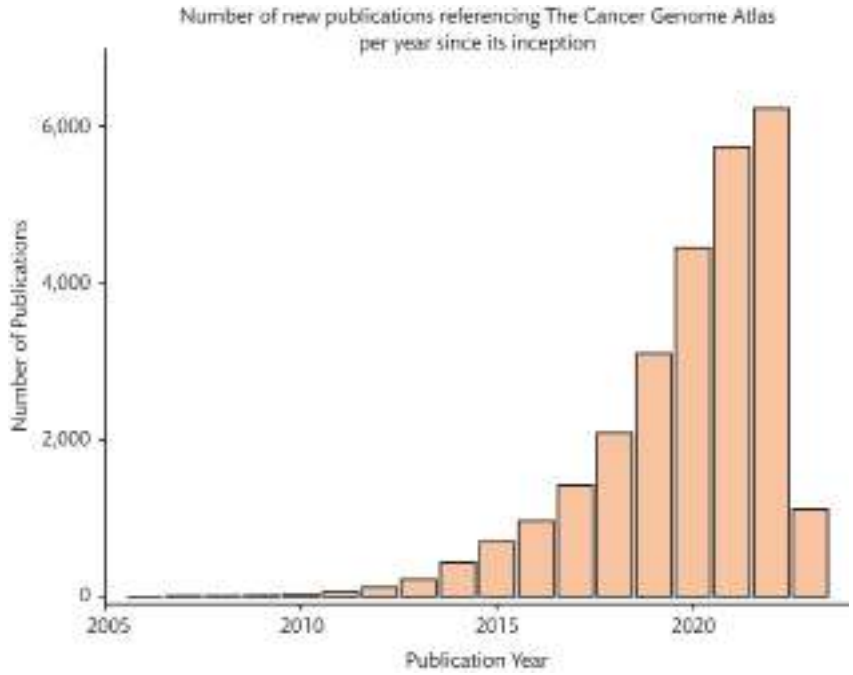


Figure 18.1. Publications enabled by The Cancer Genome Atlas (TCGA) dataset. TCGA was initiated in 2005, completed the bulk of its data production by 2010, and officially came to an end in 2018. Since it was first conceived of, a total of 26,686 publications referencing it have been indexed in PubMed. The bar plot shows the increase in publications referencing TCGA since 2005. The most recent completed year at the time of writing, 2022, observed 7,734 publications. The steep rate of increase even after 2018 demonstrates that this resource continues to provide benefits to researchers and the global health community. *Data Source*: PubMed (https://pubmed.ncbi.nlm.nih.gov/) using the following advanced search pattern: ((TCGA data*[Title/Abstract]) OR (Cancer Genome Atlas[Title/Abstract])) AND (("2004/01/01"[Date - Publication]: "3000"[Date - Publication]))(TCGA data*[Title/Abstract]) OR (Cancer Genome Atlas[Title/Abstract]).

The successes of TCGA and other large-scale studies have driven massive investment into coordinating large research efforts that ultimately generate large volumes of high impact data. These investments accelerated in response to the coronavirus disease 2019 crisis with more than $1 Billion invested in projects including Data Coordinating Centers (DCCs) in 2020 by the National Institutes of Health alone. As shown in Figure 18.2, while the total funding for DCCs has slightly decreased since 2020, the number of DCCs

Figure 18.2. Increase in DCC projects compared to funding since 1984. The number of Data Coordination projects approved for funding by all NIH agencies since 1984 are represented by the blue bar plots and the right-hand *y*-axis. The black line segment plot tracks the aggregated funding for these projects against the left-hand *y*-axis in US dollars. The number of projects has steadily increased since 2007 as an increased awareness of large data-driven projects was recognized across the NIH. The rate of funding has approximately tracked with the total number of projects with a significant peak in 2020 likely related to SARS-CoV-

is steadily growing, illustrating the continued investment in these activities by the NIH.

18.3. Key characteristics of platforms that can accelerate effective data analysis

As academic, government, not-for-profit, and private data generating programs increase in size and complexity, it can become more difficult for data to be effectively accessed by the research community or for it to be combined with other data. Without dedicated focus, rich data sources can become data silos that significantly reduce their potential. Fortunately, emergent themes are helping to reduce data silos and accelerate effective data analysis.

18.3.1. Cloud-based data

First, the increased adoption of cloud computing allows a single copy of data to be accessed and analysed by researchers across the world [16–20]. This dramatically reduces data storage costs and enables researchers from all institutions to equitably access data generated through all of the data streams discussed above. For example, the complete repository of TCGA represents more than 2.5 petabytes of data, which at 2022 list prices would cost approximately $750,000 annually to maintain in cloud storage [21]. Based on conservative estimates, streamlining the availability of the data via a single cloud-based copy provides more than 100-fold cost savings in aggregate (see Table 18.2). Of course, many researchers may use local storage or high-performance clusters at their university. Because the costs of purchasing and maintaining these resources are typically covered in indirect fees, the true cost is difficult to determine though undoubtedly a single cloud-based copy provides dramatic cost savings in aggregate.

18.3.2. Interoperability and APIs

A second emergent theme supporting the reduction of data silos is the increased focus on interoperability standards and Application Programming Interfaces (APIs). The leading organization advancing this focus is the Global Alliance for Genomics and Health (GA4GH). Composed of more than 650 organizations from 56 different countries, the GA4GH community collaborates to develop frameworks and standards to support the responsible, voluntary, and secure sharing of genomic and health-related data [22]. A critical enabler of progress by the GA4GH community has been the focus on 'driver projects' that have defined clear use cases and requirements for interoperability challenges. This focus has ensured that the community stays oriented towards addressing the problems faced by researchers today rather than creating complex standards that address hypothetical future possibilities. Particularly, important and well-adopted standards for data interoperability include the Data Repository Service (DRS) and GA4GH Passports and Visas [23,24]. DRS allows data consumers to access authorized datasets regardless of the repository in which they are stored or managed while passports and visas provide a programmatic approach to support robust data access policies and governance systems. The NIH Cloud Platform Interoperability (NCPI) effort has successfully leveraged these APIs and other standards to support federated data exchange across the National Cancer Institute's Cancer Research Data Commons, the National Heart Lung and Blood Institute's BioData Catalyst, the National Human Genome Research Institute's Anvil Platform, and the Common Fund's Kids First Data Resource Center. Similar to GA4GH, the NCPI has focused on using real-world scientific use cases as a driver to prioritize development to support interoperability.

Table 18.2. Evaluation of the cost savings of accessing a single copy instead of individual investigators independently storing the data.

	Individual investigators store a fraction of data for one year	Centralized data storage for three years
Cost of 1 copy of TCGA/yr	$750,000	$750,000
Fractional copies	26,000[a] × 10%[b] = 2,600	1
Years storage	1.5[c] × 30%[d] = 0.45	10[e]
Total cost of storage	**$877M**	**$7.5M**

Notes: We conservatively estimate that each of the 26,000 + individual manuscripts published to date using TCGA data[a] required access to at least 10%[b] of the data and spanned 18 months[c] from initiation to publication. Cost savvy investigators may download, process, and then delete data throughout the period of manuscript preparation, so we estimate the required data storage time of 30%[d]. In contrast, a single copy would need to be available over the entire period of investigation for all authors, so we use an estimate of 10 years to accommodate the majority of published papers to date[e].

18.3.3. Data security

In addition to supporting access and interoperability, it is of course absolutely critical that data is housed in highly secure and compliant environments that enforce patient consent [25–27]. As cloud infrastructures have grown in maturity, there is a growing agreement that cloud-based storage represents a more secure storage solution than previous paradigms in which individual investigators would download copies of the data for use locally. By maintaining a single copy of the primary data, institutions are able to readily trace data access and prevent access by unapproved individuals. Audit trails allow further understanding of researcher utilization as well as the detection of any anomalous events. In many cases, the same technology that provides deep auditability of data use can also facilitate reproducibility of computational analyses.

18.3.4. Analytical reproducibility

Challenges in reproducing and replicating the outcomes of biomedical research have plagued the biomedical field. Indeed, a 2016 survey revealed that more than 70% of researchers have tried and failed to reproduce other scientists' results and more than 50% could not reproduce their own experiments [28]. While the complexity of biological systems can introduce unknown variables into clinical and wet-lab experiments, most researchers agree that the analysis of the same data with the same computational tools should yield the same results. However, without focus or tooling, this seemingly simple expectation has proved quite challenging [29]. Tool versions and reference data are evolving rapidly, and most computational biology algorithms will have tens of different parameters that are set at runtime and typically not well reported in published studies.

When working with biomedical data, researchers are typically faced with two primary types of tasks, each with differing requirements and reproducibility considerations. Batch analyses are

typically computationally intensive, multi-step processes that reduce primary data into harmonized information. For example, quality control, read alignment, and somatic variant call for matched tumour-normal whole genome sequencing experiments. In contrast, interactive analysis typically uses scripting, statistical, and other iterative methods to transform information into insights. Ultimately, combining these insights leads to an impact on patient care through new practice guidelines, new therapies, or preventative measures. Considering this ultimate goal, it is critical that computational analyses are performed in a way that makes reproducing them easy.

The development of Docker and other containerization strategies has significantly streamlined the distribution of compiled algorithms that can run in a reproducible manner across diverse computational environments [30,31]. Containerization enables all code dependencies to be packaged together and executed in an isolated manner on top of the operating system's kernel. This means that an algorithm can run exactly the same way across different cloud providers or on-prem hardware. It also avoids the frequently frustrating and error-prone process of compiling computational tools.

In addition to containerization, standardized frameworks for describing computational executions are necessary to support the reproducibility of analysis. There are numerous frameworks available for the description and automation of computational tasks, and the bioinformatics community has developed or tailored these to address the specific use cases faced by researchers [32]. Common workflow language (CWL) [33], workflow description language (WDL) [34], and Nexflow [35] currently represent the most commonly used frameworks. Unsurprisingly, given the diversity of the bioinformatics community, these frameworks provide different user experiences and features that influence their application under different circumstances. Both WDL and Nextflow represent domain-specific languages with a single primary open-source implementation. Both languages enable a rapid development of complex bioinformatics workflows, particularly for sophisticated users. In contrast, CWL represents a community-defined specification with multiple execution engines. This approach enables continued innovation and ensures portability across compliant engines. In particular, CWL enables a detailed description of tool inputs, parameters, as well as computational requirements that can be dynamically determined based on the specific run. This allows tool developers to create highly composable and computationally optimized tools that can then be reused and remixed by others in the community, including through the use of visual interfaces Figure 18.3. The comprehensive, human- and machine-readable description provided by CWL is utilized by the BioCompute Object standard that is being developed as a means to efficiently submit descriptions of fully reproducible experiments to regulatory agencies like the FDA [36,37].

In contrast to batch analysis, interactive analysis is, by definition, a process of exploration and discovery. Typically, researchers use a variety of methods to clean, transform, analyse, and visualize information in order to identify and present scientific insights. Maintaining a clear record of these transformations allows others to understand and reproduce identified results. Additionally, because in many cases analysts will work with the same dataset over many months, having a record of analysis promotes efficiency and enables collaboration. Jupyter [38] and RStudio [39] notebooks provide a rich analysis experience specifically tailored to promoting reproducibility by enabling researchers to weave together executable code chunks with markdown-based text and figures. Cloud environments that enable these tools to be run directly on data with expandable computational resources further support secure collaboration on large datasets.

18.4. Multimodal data for cancer research

As discussed above, a variety of approaches have been developed to support data aggregation and analysis. The National Cancer Institute (NCI) served as an important trailblazer in the space of making large-scale datasets available to the community via TCGA. Faced with the challenge of democratizing access to TCGA and other NCI datasets, the NCI launched the Cloud Pilot program in 2014 with the objective of developing a cloud-based ecosystem that brought together investigator- and consortia-initiated data within secure, scalable, collaborative, and reproducible environments [40–43]. These pilot projects have since evolved into the NCI Cloud Resources which connect to multiple application-specific 'data nodes'. This approach enables specific tooling, data harmonization, and access

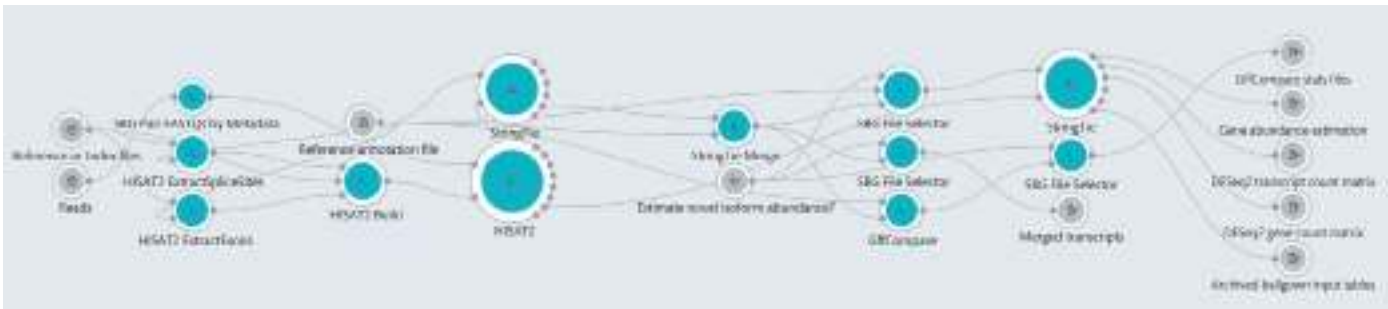


Figure 18.3. Composable workflows enable the democratization of bioinformatics. Workflows described in CWL can be visualized and interactively edited by researchers with minimal coding expertise that facilitates the use of data by biologists and clinicians. For example, the HISAT2-StringTie Workflow can be used to perform a gene abundance estimation of RNA-Seq data (i.e. quantification) of a unified set of genes common for all samples in the analysis. The workflow is based on the HISAT2 Nature protocol paper (62) with the last step (differential expression testing) omitted. This allows a researcher to plug in alternative differential testing approaches, also described in CWL. In this workflow diagram, grey circles represent data—both input files and outputs—while teal circles represent software tools wrapped in CWL for cloud deployment. Connections between teal circles represent the flow of outputs from one tool becoming the inputs for the next tool.

approaches to be applied to diverse data types while also enabling the cross analysis and computational integration across, e.g. genomics, proteomics, and imaging data. The connected data nodes are described in the following subsections. Section 18.5 highlights a number of examples of research that has been enabled by coupling a rich cancer dataset with a powerful analysis ecosystem with the characteristics described in Section 18.3.

18.4.1. Genomics, transcriptomics, and other molecular data

Established in 2014, the Genomic Data Commons (GDC) enables search and discovery of molecular data including genomic, transcriptomic, and epigenomic sequencing modalities [44]. The 36.0 data release from December 2022 included data from 86,513 cases spanning 74 projects and representing 67 primary tissue sites [45]. Lung, bone marrow, breast, and colorectal sites are most highly represented, each with more than 8,000 cases included in the repository. In addition to providing a rich search interface, the GDC performs harmonization and standardization of molecular, biospecimen, and clinical data. While researchers still need to be aware of experimental batch effects, the standardization of computational processing workflows enables the combination of data from multiple projects. Data from the GDC falls into two categories: open access and controlled access. Open access data includes aggregated information that does not include individually identifiable information such as gene expression levels. Any researcher can access that data either by downloading or connecting to a cloud resource. Controlled access data does include individual-level information and requires authorization, in most cases from the NIH database of Genotypes and Phenotypes (dbGaP) [46]. Depending on the specific dataset, authorization may require engaging an Institutional Review Board, and/or certifying the data will only be used for specific purposes.

18.4.2. Proteomics data

The Proteomic Data Commons (PDC) also provides highly curated and standardized biospecimen, clinical, and proteomic data. Reflecting the broad range of proteomic analysis, the PDC houses data representing diverse analytical fractions, including acetylome, glycoproteome, phosphoproteome, proteome, and ubiquitylome derived from multiple experimental technologies. More than 375 million spectra were derived from 123 studies, and approximately 4,000 cases were available in the December 2022 data release [47]. Similar to the GDC, the PDC houses both controlled and open access data, and controlled access data is authorized via dbGaP.

18.4.3. Imaging data

Imaging represents a wide range of applications from clinical and pre-clinical imaging and radiological images, such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), digital pathology, and multispectral microscopy. Just as in the case of molecular data, raw imaging data must be processed, annotated, and modelled to support cross-comparison and derivation of insights. Reflecting the complexity and uniqueness of imaging data, the NCI has also established the Imaging Data Commons (IDC) [48]. The October 2022 data release included 128 collections representing more than 470,000 image series from 63,316 cases. Unlike the PDC and GDC, all data available on the IDC has been certified to be deidentified and is therefore open access.

18.4.4. Multi-species data

The fourth data node of the Cancer Research Data Commons (CRDC) is the Integrated Canine Data Commons (ICDC) [49]. Although at first glance it may seem curious that the NCI would invest in a data repository focused on canine data, the analysis of canine cancer progression can provide important insights that impact human disease. For many, their dog is a bona fide family member, and as a result, dogs are frequently exposed to similar environmental conditions as humans. They additionally receive healthcare and participate in clinical trials. Finally, their accelerated ageing process and breed-specific disease predisposition provide an interesting backdrop to study human disease. As of January 2023, the ICDC provides access to data from nearly 600 cases representing more than 75 different breeds. Studies include the PRE-medical Cancer Immunotherapy Network Canine Trials (PRECINCT) and the Comparative Oncology Program. All data is open access.

18.4.5. Supporting the long tail of data modalities and analyses

The type of data generated in the course of biomedical research is diverse and wide ranging. Additionally, even when the type of data (for example whole genome sequencing) is the same as the focus of a CRDC node, it may not be appropriate or feasible to perform the standardized processing required as part of a CRDC data node. To accommodate these situations and to support researcher's compliance with data-sharing policies, the NCI developed the Cancer Data Service (CDS) [50]. This solution provides a flexible and responsive approach for researchers to quickly and securely share data. As of February 2023, numerous datasets from the Childhood Cancer Data Initiative as well as the Human Tissue Atlas Network are available through this repository. As data standards continue to develop, data may be transitioned from CDS to dedicated data nodes.

18.5. Combining cloud-based data and analysis environments to accelerate research

As data volume and complexity increases, it becomes critical that the secure, collaborative, and reproducible environments discussed above are able to interface with the rich cloud-based data hosted by data repositories. In the context of the CRDC, this is accomplished through the development of three Cloud Resources by Seven Bridges [41], Institute of Systems Biology, [43], and the Broad Institute [42]. In the following sections, we provide examples of how researchers have used the Seven Bridges resource in the hope that these examples provide further inspiration to the type of analysis that these environments can accelerate.

18.5.1. Multi-omic analysis to understand retroposon activity in cancer

Retrotransposons are genomic DNA elements that are copied and inserted into different genomic locations. In particular, the Long Interspersed Element-1 (LINE-1) retroposon is the only active and autonomous retroposon present in the human genome. Although

derepression of LINE-1 is frequently observed in human cancer, its role in tumorigenesis is incompletely understood. In order to comprehensively measure LINE-1 activity across different tumour types, Dr McKerrow et al. at New York University, Langone Health, combined genomic, transcriptomic, and proteomic data from TCGA and CPTAC studies [51]. Overall, LINE-1 activity was evaluated through LINE-1 RNA levels, ORF1 protein and phosphorylation levels, as well as somatic insertions. This analysis showed that ORF1p could serve as a simple and high-quality measure of overall LINE-1 activity in large, multi-omic datasets. The investigators went on to show that LINE-1 activity level is correlated with p53 mutation and copy number alterations in multiple cancer types as well as alterations in cell cycle progression markers [51]. This work provides a deeper understanding of the role of LINE-1 in oncogenesis which could lead to novel treatment or prevention strategies. The ability to deploy novel computational tools and rapidly access large volumes of multimodal data was critical for the completion of this work.

18.5.2. Pancancer analysis of gene fusions

Gene fusions are known to drive some cancer types. Perhaps most infamously, a fusion between the BCL and ABL genes underlies deregulated cellular signalling in most cases of chronic myelogenous leukaemia. Accurate detection of fusions can be complicated by the heterogeneity of tumour specimens that reduce signal and the introduction of false positives during the sequencing process. To address these challenges, Dr Dehghannasiri et al. at Stanford University developed a novel algorithm termed Data-Enriched Efficient PrEcise STatistical (DEEPEST) fusion detection that uses statistical modelling to reduce false positives and increase sensitivity of detection [52]. After optimizing this tool to run in the cloud, the investigators analysed RNA-sequencing data from nearly 10,000 tumour samples spanning 33 different cancer types within TCGA on the CGC. The tool identified 31,007 fusions, 30% more than other methods while reducing false positives 10-fold. Kinase domains, anaerobic metabolism, and DNA-binding proteins were significantly enriched in identified fusions. Importantly, in addition to creating a knowledge base of novel fusions and reporting frequency across cancer types, the authors also provided the optimized DEEPEST workflows that allow other researchers to easily extend this work to novel samples [53].

18.5.3. Data harmonization, sharing, and analysis of Patient-derived Xenograft data

Patient-derived Xenograft (PDX) models are cancer models in which patient tissue, often tumour, is transplanted from a human patient into a mouse model for use in supporting personalized medicine research and pre-clinical and co-clinical trials. The PDX Network (PDXNet) leveraged the CGC to provide a secure environment to facilitate collaboration, promote data harmonization, and centralize access to NCI-funded PDXNet consortium resources [54]. As of December 2022, more than 330 models, spanning 33 different cancer types had been catalogued. Molecular data from these models was harmonized in the cloud using workflows that were benchmarked and optimized by the consortia. That data was then used by the consortia to reveal insights into tumorigenesis. For example, Guillen et al. identified that matched PDXs and PDX-derived organoids could be used for drug screening as these samples were representative of endocrine-resistant, treatment-refractory, and metastatic breast cancers [55]. Both whole exome and RNA sequencing were processed using the Cancer Genomics Cloud in accordance with PDXNet-approved pipelines. Using this method, researchers identified a Food and Drug Administration (FDA)-approved drug with high efficacy against the models. Treatment with this therapy resulted in a complete response for the individual and a progression-free survival period more than three times longer than their previous therapies.

18.6. Conclusions and outlook

The ability for researchers and clinicians from a wide variety of disciplines and backgrounds to access and analyse large volumes of multidimensional cancer data is creating unprecedented opportunities to advance our understanding of tumorigenesis and develop new strategies for effective therapeutics. As the volume, velocity, and complexity of data continues to increase, it is particularly important that collaboration across domains is facilitated in a secure and reproducible manner. For example, the incorporation of mathematical and systems biology approaches can allow the identification of signals that, when contextualized by clinicians or cancer biologists, can have a significant impact on patient care.

The past 10 years have been marked by a dramatic increase in the adoption of scalable and secure approaches for data distribution and analysis of investigator- and consortia-derived data [56–60]. Radical transformations in cancer patient care, notably the incorporation of molecular markers to define therapeutic pathways and the use of liquid biopsy to detect and monitor cancer, have occurred. We foresee that the next decade will increasingly focus on supporting incorporation of individual-level clinical data into multidimensional machine learning and AI models that will not only allow unique insights into health and disease but will also dramatically improve each patient's care journey. Numerous advances in policy, algorithms, and infrastructure are necessary to realize this future to create positive outcomes for patients and avoid unintended drawbacks.

REFERENCES

1. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. Nature. 2020 Feb;**578**(7793):82–93.
2. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence—what is it and what can it tell us? New Engl J Med. 2016 Dec 8;**375**(23):2293–7.
3. Booth CM, Karim S, Mackillop WJ. Real-world data: towards achieving the achievable in cancer care. Nat Rev Clin Oncol. 2019 May;**16**(5):312–25.
4. Kiernan D, Carton T, Toh S, Phua J, Zirkle M, Louzao D, et al. Establishing a framework for privacy-preserving record linkage among electronic health record and administrative claims databases within PCORnet®, the National Patient-Centered Clinical Research Network. BMC Res Notes. 2022 Oct 31;**15**(1):337.
5. 2023 NIH Data Management and Sharing Policy [Internet]. [cited 2022 Dec 21]. Available from: https://oir.nih.gov/sourcebook/intramural-program-oversight/intramural-data-sharing/2023-nih-data-management-sharing-policy

6. Kodama Y, Shumway M, Leinonen R, International Nucleotide Sequence Database Collaboration. The sequence read archive: explosive growth of sequencing data. Nucleic Acids Res. 2012 Jan;**40**(Database issue):D54–6.
7. DbGaP summary statistics [Internet]. [cited 2022 Dec 21]. Available from: https://www.ncbi.nlm.nih.gov/projects/gap/summaries/cgi-bin/summariesInteractive.cgi
8. Repositories for Sharing Scientific Data [Internet]. [cited 2022 Dec 21]. Available from: https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/repositories-for-sharing-scientific-data
9. Hajar R. Framingham contribution to cardiovascular disease. Heart Views. 2016 Apr–Jun;**17**(2):78–81.
10. Weir HK, Anderson RN, Coleman King SM, Soman A, Thompson TD, Hong Y, et al. Heart disease and cancer deaths—trends and projections in the United States, 1969–2020. Prev Chronic Dis. 2016 Nov 17;**13**:E157.
11. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell. 2018 Apr 5;**173**(2):291–304.e6.
12. Ding L, Bailey MH, Porta-Pardo E, Thorsson V, Colaprico A, Bertrand D, et al. Perspective on oncogenic processes at the end of the beginning of cancer genomics. Cell. 2018 Apr 5;**173**(2):305–20.e10.
13. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic signaling pathways in the cancer genome atlas. Cell. 2018 Apr 5;**173**(2):321–37.e10.
14. Killock D. CNS cancer: molecular classification of glioma. Nat Rev Clin Oncol. 2015 Sep;**12**(9):502.
15. Jones DTW, Bandopadhayay P, Jabado N. The power of human cancer genetics as revealed by low-grade gliomas. Annu Rev Genet. 2019 Dec 3;**53**:483–503.
16. SPARK • Health data integration platform [Internet]. [cited 2022 Dec 28]. Available from: https://spark.sbgenomics.com/
17. Pishvaian MJ, Blais EM, Brody JR, Lyons E, DeArbeloa P, Hendifar A, et al. Overall survival in patients with pancreatic cancer receiving matched therapies following molecular profiling: a retrospective analysis of the Know Your Tumor registry trial. Lancet Oncol. 2020 Apr;**21**(4):508–18.
18. PanCAN's Precision Promise Adaptive Clinical Trial Platform—[Internet]. Pancreatic Cancer Action Network. PanCAN; 2016 [cited 2023 Jan 12]. Available from: https://pancan.org/research/precision-promise/
19. Break Through Cancer [Internet]. Break Through Cancer. 2017 [cited 2022 Dec 28]. Available from: https://breakthroughcancer.org/
20. Grossman RL, Abel B, Angiuoli S, Barrett JC, Bassett D, Bramlett K, et al. Collaborating to compete: Blood Profiling Atlas in cancer (BloodPAC) consortium. Clin Pharmacol Ther. 2017 May;**101**(5):589–92.
21. AWS Pricing Calculator [Internet]. [cited 2022 Dec 28]. Available from: https://calculator.aws/#/
22. Rehm HL, Page AJH, Smith L, Adams JB, Alterovitz G, Babb LJ, et al. GA4GH: International policies and standards for data sharing across genomic research and healthcare. Cell Genom [Internet]. 2021 Nov 10;**1**(2). Available from: http://dx.doi.org/10.1016/j.xgen.2021.100029
23. data-repository-service-schemas: A repository for the schemas used for the Data Repository Service [Internet]. Github; [cited 2022 Dec 28]. Available from: https://github.com/ga4gh/data-repository-service-schemas
24. researcher_ids at master · ga4gh-duri/ga4gh-duri.github.io [Internet]. Github; [cited 2022 Dec 28]. Available from: https://github.com/ga4gh-duri/ga4gh-duri.github.io
25. ISO/IEC 27001 and related standards [Internet]. ISO. 2022 [cited 2023 Jan 12]. Available from: https://www.iso.org/isoiec-27001-information-security.html
26. Program Basics [Internet]. [cited 2023 Jan 12]. Available from: https://www.fedramp.gov/program-basics/
27. Joint Task Force Transformation Initiative. Security and Privacy Controls for Federal Information Systems and Organizations [Internet]. National Institute of Standards and Technology; 2015 Jan [cited 2023 Jan 12]. Report No.: NIST Special Publication (SP) 800-53 Rev. 4 (Withdrawn). Available from: https://csrc.nist.gov/publications/detail/sp/800-53/rev-4/archive/2015-01-22
28. Baker M. 1,500 scientists lift the lid on reproducibility [Internet]. Nature Publishing Group UK. 2016 [cited 2023 Feb 21]. Available from: http://dx.doi.org/10.1038/533452a
29. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. PLoS Comput Biol. 2013 Oct;**9**(10):e1003285.
30. Merkel D. Docker: lightweight Linux containers for consistent development and deployment. Linux J. 2014 Mar 1;**2014**(239):2.
31. Overview [Internet]. Docker Documentation. 2023 [cited 2023 Feb 22]. Available from: https://docs.docker.com/get-started/
32. Leipzig J. A review of bioinformatic pipeline frameworks. Brief Bioinform. 2017 May 1;**18**(3):530–6.
33. common-workflow-language. common-workflow-language/common-workflow-language [Internet]. GitHub. [cited 2015 Feb 22]. Available from: https://github.com/common-workflow-language/common-workflow-language
34. OpenWDL [Internet]. [cited 2023 Feb 23]. Available from: https://openwdl.org/
35. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol. 2017 Apr 11;**35**(4):316–9.
36. Xiao N, Koc S, Roberson D, Brooks P, Ray M, Dean D. BCO App: tools for generating BioCompute Objects from next-generation sequencing workflows and computations. F1000Res. 2020 Sep 16;**9**:1144.
37. Simonyan V, Goecks J, Mazumder R. BioCompute objects—a step towards evaluation and validation of biomedical scientific computations. PDA J Pharm Sci Technol. 2017 Mar–Apr;**71**(2):136–46.
38. Kluyver T, Ragan-Kelley B, Pérez F, Granger B, Bussonnier M, Frederic J, et al. Jupyter Notebooks—a publishing format for reproducible computational workflows. In: Loizides F, Scmidt B, editors. Positioning and power in academic publishing: players, agents and agendas. IOS Press; 2016. p. 87–90.
39. Allaire J. RStudio: integrated development environment for R. Boston, MA. 2012;**770**(394):165–71.
40. Kibbe W, Klemm J, Quackenbush J. Cancer informatics: new tools for a data-driven age in cancer research. Cancer Res. 2017 Nov 1;**77**(21):e1–2.
41. Lau JW, Lehnert E, Sethi A, Malhotra R, Kaushik G, Onder Z, et al. The cancer genomics cloud: collaborative, reproducible, and democratized—a new paradigm in large-scale computational research. Cancer Res. 2017 Oct 31;**77**(21):e3–6.
42. Birger C, Hanna M, Salinas E, Neff J, Saksena G, Livitz D, et al. FireCloud, a scalable cloud-based platform for collaborative genome analysis: Strategies for reducing and controlling costs [Internet]. bioRxiv. 2017 [cited 2023 Feb 23]. p. 209494. Available from: https://www.biorxiv.org/content/10.1101/209494v1

43. Reynolds SM, Miller M, Lee P, Leinonen K, Paquette SM, Rodebaugh Z, et al. The ISB cancer genomics cloud: a flexible cloud-based platform for cancer genomics research. Cancer Res. 2017 Nov 1;**77**(21):e7–10.
44. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. New Engl J Med. 2016 Sep 22;**375**(12):1109–12.
45. GDC [Internet]. [cited 2023 Feb 24]. Available from: https://portal.gdc.cancer.gov/
46. Tryka KA, Hao L, Sturcke A, Jin Y, Wang ZY, Ziyabari L, et al. NCBI's database of genotypes and phenotypes: dbGaP. Nucleic Acids Res. 2014 Jan;**42**(Database issue):D975–9.
47. Proteomic Data Commons [Internet]. [cited 2023 Feb 24]. Available from: https://proteomic.datacommons.cancer.gov/pdc/
48. Fedorov A, Longabaugh WJR, Pot D, Clunie DA, Pieper S, Aerts HJWL, et al. NCI Imaging Data Commons. Cancer Res. 2021 Aug 15;**81**(16):4188–93.
49. ICDC [Internet]. [cited 2023 Feb 27]. Available from: https://caninecommons.cancer.gov/#/
50. Cancer Data Service [Internet]. [cited 2023 Feb 27]. Available from: https://datacommons.cancer.gov/repository/cancer-data-service
51. McKerrow W, Wang X, Mendez-Dorantes C, Mita P, Cao S, Grivainis M, et al. LINE-1 expression in cancer correlates with p53 mutation, copy number alteration, and S phase checkpoint. Proc Natl Acad Sci U S A [Internet]. 2022 Feb 22;**119**(8). Available from: http://dx.doi.org/10.1073/pnas.2115999119
52. Dehghannasiri R, Freeman DE, Jordanski M, Hsieh GL, Damljanovic A, Lehnert E, et al. Improved detection of gene fusions by applying statistical methods reveals oncogenic RNA cancer drivers. Proc Natl Acad Sci U S A. 2019 Jul 30;**116**(31):15524–33.
53. Zhao Y, Fang LT, Shen TW, Choudhari S, Talsania K, Chen X, et al. Whole genome and exome sequencing reference datasets from a multi-center and cross-platform benchmark study. Sci Data. 2021 Nov 9;**8**(1):296.
54. Evrard YA, Srivastava A, Randjelovic J, Doroshow JH, Dean DA, 2nd, Morris JS, et al. Systematic establishment of robustness and standards in patient-derived xenograft experiments and analysis. Cancer Res. 2020 Jun 1;**80**(11):2286–97.
55. Guillen KP, Fujita M, Butterfield AJ, Scherer SD, Bailey MH, Chu Z, et al. A human breast cancer-derived xenograft and organoid platform for drug discovery and precision oncology. Nat Cancer. 2022 Feb;**3**(2):232–50.
56. Thistlethwaite LR, Petrosyan V, Li X, Miller MJ, Elsea SH, Milosavljevic A. CTD: an information-theoretic algorithm to interpret sets of metabolomic and transcriptomic perturbations in the context of graphical models. PLoS Comput Biol. 2021 Jan;**17**(1):e1008550.
57. White BS, Woo XY, Koc S, Sheridan T, Neuhauser SB, Wang S, et al. A pan-cancer PDX histology image repository with genomic and pathological annotations for deep learning analysis [Internet]. bioRxiv. 2022 [cited 2023 Mar 25]. p. 2022.10.26.512745. Available from: https://www.biorxiv.org/content/biorxiv/early/2022/10/27/2022.10.26.512745
58. Chatrath A, Kiran M, Kumar P, Ratan A, Dutta A. The germline variants rs61757955 and rs34988193 are predictive of survival in lower grade glioma patients germline variants are predictive of survival in LGG patients. Mol Cancer Res. 2019;**17**(5):1075–86.
59. Chatrath A, Przanowska R, Kiran S, Su Z, Saha S, Wilson B, et al. The pan-cancer landscape of prognostic germline variants in 10,582 patients. Genome Med. 2020 Feb 17;**12**(1):15.
60. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, et al. Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature. 2020 Mar;**579**(7800):567–74.
61. RePORT ⟩ RePORTER [Internet]. [cited 2023 Mar 30]. Available from: https://reporter.nih.gov/
62. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016 Sep;**11**(9):1650–67.

19

Interpretation of machine learning models in cancer: The role of model-agnostic explainable artificial intelligence

Colton Ladbury and Arya Amini

19.1. Introduction

Statistical models typically used in oncology include linear regression, logistic regression, and Cox-proportional hazards regression. These yield odds ratios, hazards ratios, coefficients, and *p*-values that are simple to read and utilize. Oncology is a data-driven field with nuanced clinical questions, which necessitates more complicated models in some cases, such as modelling non-linear and/or distribution assumption-free relationships, interaction effects, or image analysis, to better utilize data and make more informed and accurate decisions. Machine learning (ML) algorithms have the potential to develop better and more therapeutically useful models as processing capacity has increased [1]. However, when models become more complicated in comparison to ordinary regression models, they become more difficult for end users (including clinicians, patients, administrators, stakeholders, and others) to grasp, resembling a metaphorical 'black box' [2]. End users are interested not only in the accuracy of a model's output, but also in how that output is produced and how it may be changed [3]. That is a fundamental challenge to using ML algorithms in the oncology clinic [4].

The trade-off between interpretability and complexity has resulted in the development of so-called 'explainable artificial intelligence (XAI)' techniques, which seek to produce interpretable visualizations and sets of rules to represent the inner workings of the ML 'black box' [5, 6]. This is accomplished by both local explanations of single data points, which illustrate how the model arrives at an individual prediction, and all local explanations, which permit a more global understanding of how the algorithm works (**Figure 19.1**). Local Interpretable Model-agnostic Explanations (LIME) [7] and SHapley Additive exPlanations (SHAP) [8] are the most commonly used XAI frameworks and have been used in fields, such as insurance, finance, and healthcare. Examples of the plots generated using SHAP are shown in **Figure 19.2**. These frameworks have gained significant popularity in part because they are 'model-agnostic', meaning they can be applied to ML algorithms irrespective of how they function. Details of XAI use cases for SHAP and LIME can be found in **Table 19.1**. Other non-model-agnostic XAI frameworks do exist [9] but are not discussed in this chapter. In oncology specifically, ML models and model-agnostic XAI have been used to investigate clinical oncologic questions. This chapter summarizes the use of model-agnostic XAI in the oncology literature, separated by discipline [10].

19.2. Application of Artificial Intelligence in the Diagnosis and Workup of Cancer

19.2.1. Prognostication

Perhaps the area of oncology that has been most extensively explored using XAI is prognostication, which is of no surprise given the abundance of large datasets (including the National Cancer Database and the Surveillance, Epidemiology, and End Results) with outcomes data. These are of great interest to both clinicians and patients, given that this approach not only can help counsel patients on prognosis and planning but might also inform on potential interventions that can lead to improvements in outcomes.

A clinical example that lends itself nicely to the utility of XAI is modelling of interactions between disease characteristics and prostate cancer and their impact on survival. Li et al. modelled the impact of prostate-specific antigen (PSA), percent positive cores (PPC), and Gleason score on survival using the extreme gradient boosted (XGB) tree algorithm [11]. Specifically, they sought to examine non-linear relationships and interactions, which facilitate identification of prognostic thresholds. Since LIME primarily looks only at individual predictions, SHAP tends to be the main framework to perform such analyses. Although the impact of the specified factors on survival was not controversial, modelling with XAI revealed nuances that contradict modern risk stratification. Visualization of such interactions is only possible by using a more complicated model than standard regressions and then explaining

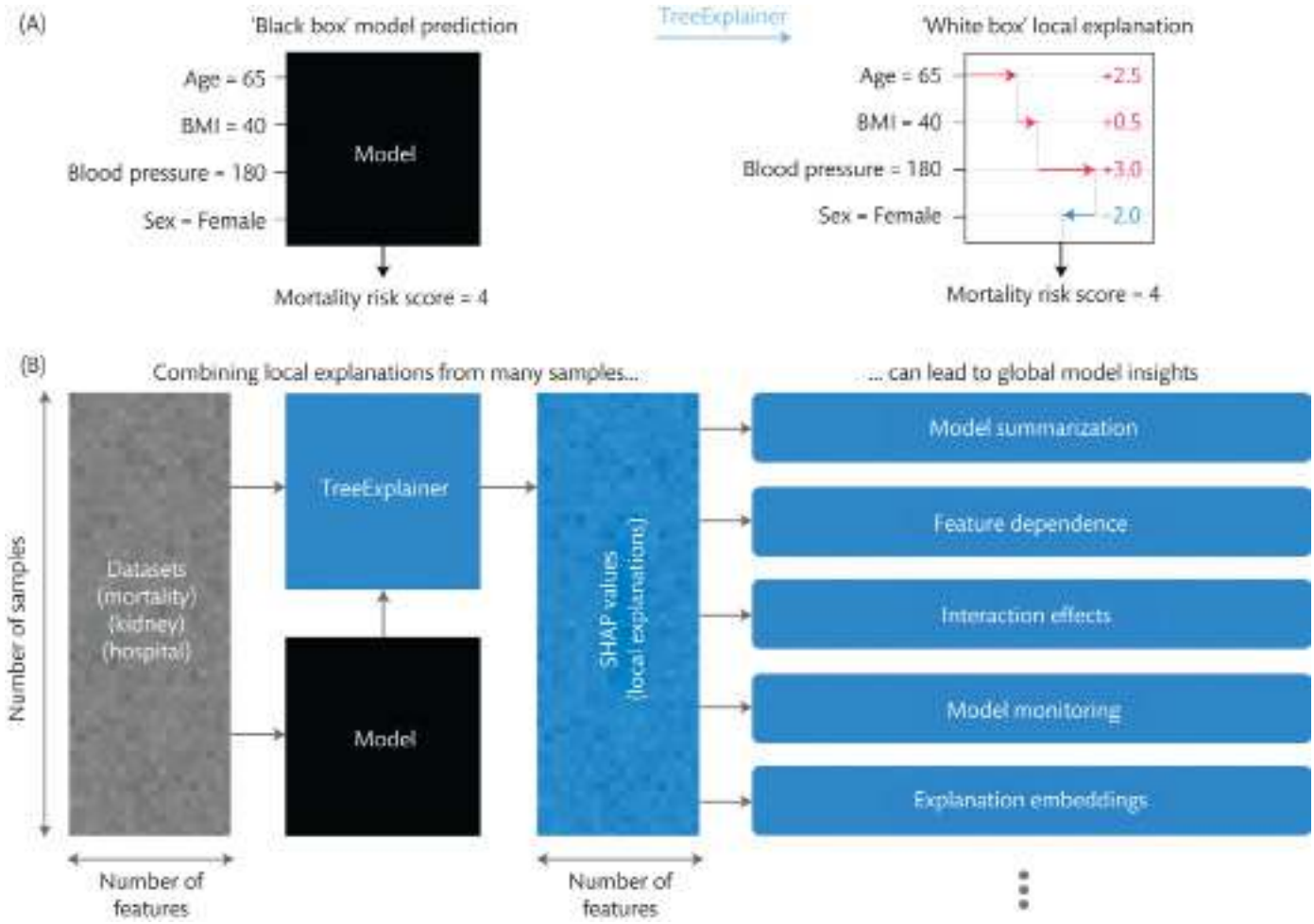


Figure 19.1. Use of explainable artificial intelligence in visualizing the inside of the 'black box' via local explanations (A) and combined local explanations to generate global insights (B). *Source:* Reproduced with permission from [6].

that model. For example, when examining the interaction between percentage of positive cores and Gleason score, the SHAP dependence plots revealed that PPC is largely irrelevant if Gleason score is 7 or less. Additionally, when Gleason score is 8 or higher, a PPC threshold of 0.7 best illustrates differences in outcomes, as opposed to the threshold of 0.5 that is used in modern practice to differentiate between favourable and unfavourable intermediate risk prostate cancer. The plots are also able to illustrate phenomena such as patients with high Gleason score and exceedingly low PSA counterintuitively having inferior survival compared to patients with more intermediate PSA levels. These example plots can be found in **Figure 19.3**. In these interaction plots, the combined effect of the two variables are plotted, with a value of zero representing an overall neutral effect.

Bertsimas et al. also used XAI to visualize non-linear relationships, thereby identifying prognostic thresholds. The authors used an XGB model and the SHAP framework to explore the impact of lymph node ratio on survival compared to lymph node count in pancreatic cancer, which is used in the current American Joint Committee on Cancer (AJCC) staging system [12]. Their model using lymph node ratio outperformed the AJCC schema at predicting one-year overall survival with an area under the curve (AUC) of 0.638 versus 0.586 (though not validated via DeLong's test or confidence intervals). Using the SHAP dependence plots, the authors, via inspection, identified thresholds for lymph node ratio and tumour size. The authors' identification of relevant lymph node ratio thresholds was feasible by using the graphical interpretation of dependence plots generated by the SHAP framework, and given the interaction is non-linear, identification of thresholds with linear methods, such as logistic regression or Cox regression, would have required some trial and error and would not have been as precise.

The aforementioned studies primarily examined XAI plots that summarize all predictions within the datasets. XAI also has been used to examine prognostication in individual predictions (i.e. patients). Both SHAP and LIME have this functionality implemented. In a study by Jansen et al., both SHAP and LIME are used to explain an XGB model of 10-year overall survival in breast cancer patients [13]. In this study, the authors use LIME to model individual patients and predictions of overall survival, which explains how individual patients' characteristics yield their prediction for 10-year overall survival. They did the same with SHAP, but as detailed in **Table 19.1**, SHAP has improved functionality in illustrating global impact of features in all patients, so they also presented a summary plot. Lastly, they compared all individual patient explanations produced by LIME and SHAP, demonstrating agreement in 87.8%–99.9% (95.4% overall) of cases depending on the feature examined.

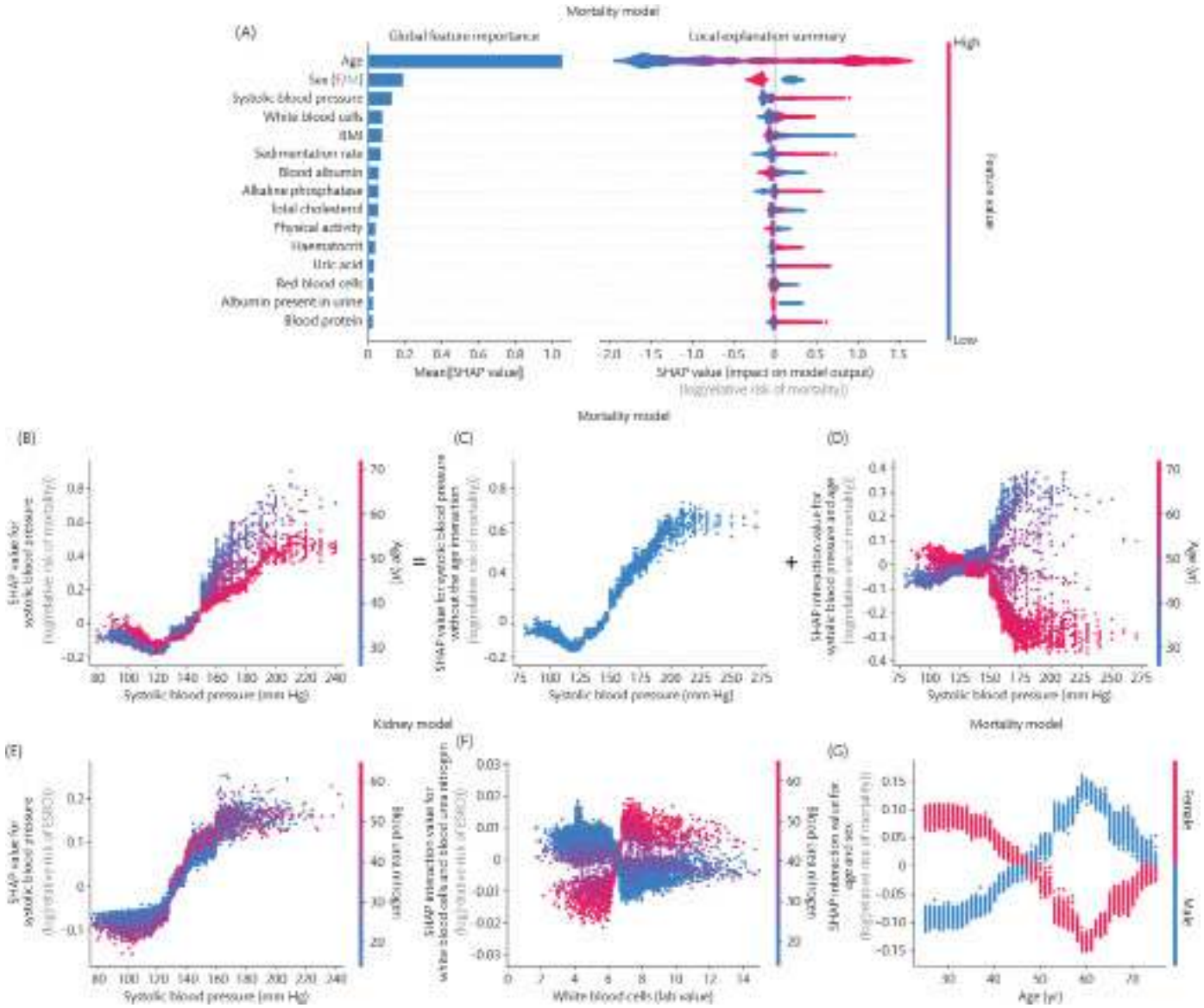


Figure 19.2. Examples of types of plots generated using explainable artificial intelligence including feature importance (A), dependence (B, C, E), and interaction (D, F, G) plots. *Source:* Reproduced with permission from [6].

This study highlights key distinctions between LIME and SHAP; both approaches can yield consistent local results, but SHAP is able to examine global trends within models. The same approach, examining individual patients and overall trends, has been performed on models predicting survival with nasopharyngeal cancer tumour burden [14].

Lastly, perhaps the way XAI was used most commonly for prognostication and oncology in general was to delineate global feature

Table 19.1. Overview of LIME and SHAP use cases.

Use case	Suitable frameworks	Overview	Example plots
Local explanation	LIME, SHAP	-Tells you, in a local sense, what is the most important attribute around the data point of interest -Computationally efficient	-Waterfall plot -Force plot
Global model explanation	SHAP and LIME (only through pooled explanations)	-Calculates the average marginal contribution of a feature value over all possible combinations of predictions and sets of inputs -Decomposes the final prediction into the contribution of each feature -Computationally expensive (aside from tree-based algorithms)	-Feature importance plot -Dependence plot -Interaction plot

Abbreviations: LIME: Local Interpretable Model-agnostic Explanations; SHAP: SHapley Additive exPlanations.

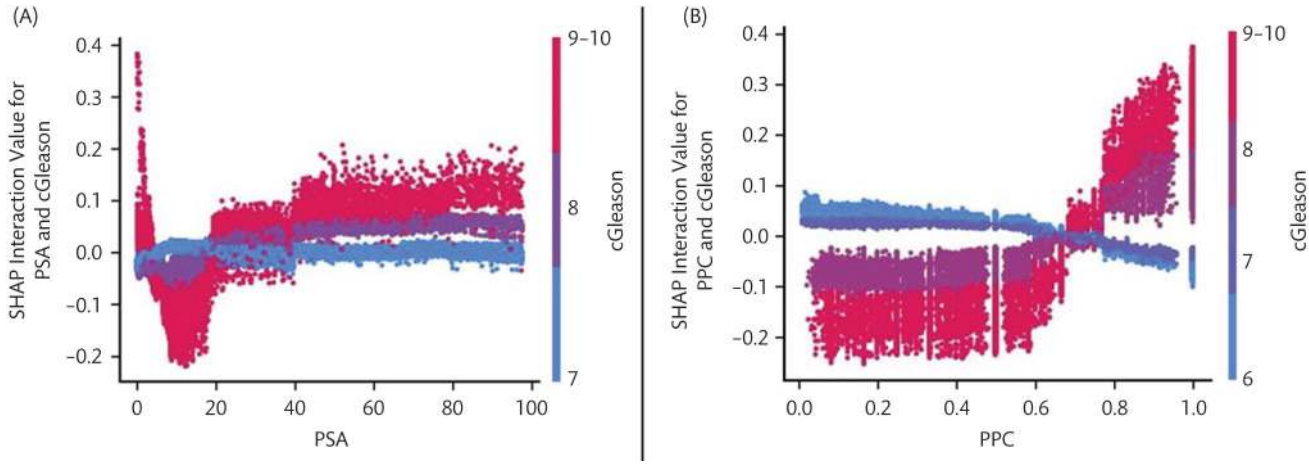


Figure 19.3. SHAP plots visualizing non-linear interactions between prognostic features in prostate cancer, including interaction between Gleason score and PSA (A) and between percent positive cores and Gleason score (B). *Source:* Adapted with permission from [11].

importance by summarizing all model predictions and outputting each feature's mean impact across all predictions. This permits a general understanding that how the overall model functions, similar to the summary statistics generated by regression models. Moncada-Torres et al. used XAI to explain a general model of overall survival in breast cancer [15]. Importantly, models such as Cox regression are routinely used for such applications, with outputs being readily interpretable by most end users. ML algorithms, namely the XGB tree algorithm, significantly outperformed Cox regression. However, the standard output of such an algorithm is not as readily understandable in terms of how its output is computed. In this example, the authors used the SHAP framework to identify feature importance within the model, which can be compared to intuition to build trust in the model. By using the SHAP framework, the study was able to illustrate the prognostic significance of multiple commonly used variables, thereby facilitating adoption of algorithms such as XGB.

Several other studies have used similar approaches to the aforementioned study to create powerful prognostic models to illustrate nuanced interactions that influence prognosis, with XAI used to explain their ML model on a global level. Additional examples include predicting 30-day mortality following colorectal cancer surgery [16], five-year survival and oesophageal cancer [17], characterizing the influence of ethnicity on outcomes and multiple myeloma [18], predicting hospital length of stay [19], and risk of skeletal-related events following discontinuation of denosumab among patients with bone metastases [20].

19.2.2. Diagnosis

An additional area of research interest is ways to improve diagnosis of cancer, where ML models have proven to be valuable tools, but given the high stakes of a cancer diagnosis, it is important that the predictions of such models are both accurate and easy to explain to clinicians and patients, representing a great opportunity for XAI.

In one such application, Suh et al. explored a model predicting prostate cancer in general as well as clinically significant prostate cancer prior to prostate biopsy [21]. In this study, the authors used XAI frameworks for one of the reasons detailed in the previous section (understanding global feature importance) as well as a new one (feature selection/construction for building clinically relevant tools). When explained using SHAP summary visualizations, the authors identified important features predicting prostate cancer and clinically significant prostate cancer, which included known predictive factors such as PSA and Gleason score, and how changing these factors individually influenced risk. These visualizations facilitated translation to a risk calculator via aiding selection of salient features that could be deployed as a data-driven risk estimator that would be generally applicable to the oncology clinic.

In another study, Kwong et al. reported on a model that predicted side-specific extraprostatic extension and pre-prostatectomy patients [22], again using SHAP summaries to gauge global feature importance. Furthermore, the authors examined the non-linear relationships between relevant factors and probability of site-specific extraprostatic extension with dependence plots, which is of clinical relevance given that features such as percent positive cores were relatively noncontributory until reaching approximately 75% based on inspection of dependence plots. Though these conclusions are qualitative in nature, they inform quantitative hypotheses that can inform other statistical approaches and overall clinical intuition. Thus, the XAI complemented the ML model by not only providing information to surgeons on risk of extraprostatic extension but also providing explanations of how that risk was calculated for given patient, where identification of non-linear interactions is valuable.

19.2.3. Radiomics

A primary field of oncology that has benefited greatly from XAI is radiomics, given that it has the reputation of being a 'black box' or 'fishing expedition' [23, 24]. Radiomics is defined as a process designed to extract quantitative features from imaging, which can subsequently be used for hypothesis generation, testing, or both [25]. In oncology, this might mean using imaging characteristics to predict tumour molecular features, behaviour, and prognosis. In doing so, a patient's image might be input into a radiomic algorithm and predict risk of relapse. However accurate such a prediction might be, it is also of vital importance that providers be able to understand why the imaging produces such a prediction.

Radiomics specifically represents a field where LIME might be preferred in certain scenarios, given that radiomics is highly interested in individual predictions and commonly uses non-tree-based algorithms, meaning SHAP becomes computationally intensive. Several radiomics studies focus on using XAI and individual predictions.

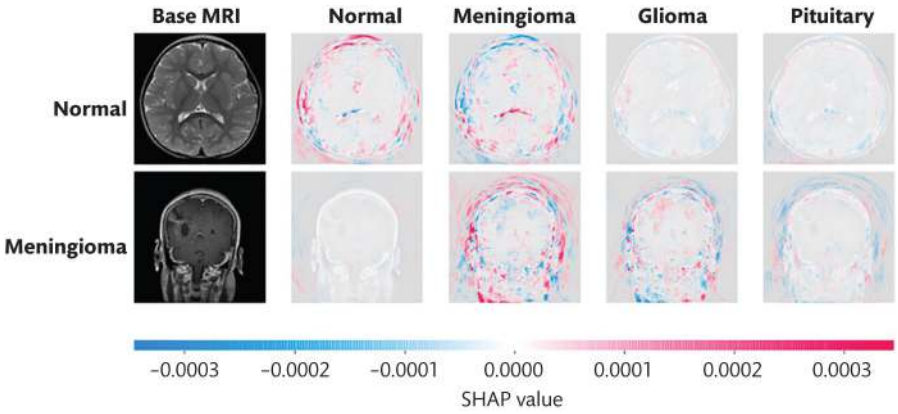


Figure 19.4. SHAP plots visualizing neural network identification of a normal brain and meningioma. Source: Adapted with permission from [28].

One example that has been explored in the literature is molecular classification of gliomas. In gliomas, molecular subtypes related to isocitrate dehydrogenase (IDH) mutations and 1p/19q co-deletion are of vital relevance to prognosis and might also inform treatment options [26]. Although confirmation of such abnormalities occurs via pathologic analysis, it can be useful to identify patients with such abnormalities prior to surgery or if pathologic conformation is not possible. Manikis et al. reported on radiomic patterns that can predict IDH mutational status with an accuracy of 73.6% and an associated sensitivity and specificity of 0.6 and 0.736, respectively [27]. Following model development, results were explained using both the SHAP and LIME frameworks, which were able to identify important radiomic features that lead to an image being classified as IDH-mutant, which the authors were then able to correlate with biological behaviour of IDH-mutant gliomas.

XAI can also be used to describe how a given imaging voxel contributes to a model's output. Gaur et al. used a deep learning model to identify brain tumour subtypes, with an accuracy of 94.64% [28]. To explain how their model made its predictions, they provide examples using the SHAP framework where SHAP values are superimposed on imaging voxels, and in doing so, illustrate graphically and intuitively how example images are classified as normal or meningioma, as illustrated in Figure 19.4. In these images, the probability of a given classification is increased with red pixels and decreased with blue pixels. As can be seen, for the normal magnetic resonance imaging (MRI), the normal classification has the greatest amount of red pixels. The same idea holds true for the meningioma classification. The authors use both SHAP and LIME again because this classification problem is highly interested in individual cases. Another similar published application is classification of ultrasound imaging of lymph nodes to predict nodal metastasis and early breast cancer (achieving an accuracy of 81.05%, which used LIME to graphically identify regions of interest in individual images) [29].

In addition to explaining the radiomic features, XAI can also aid in development of the models. A major problem in ML is overfitting training data, which can be a result of feeding the excessive features in the algorithm, including features that are largely irrelevant or strongly correlated with the causative features. This is particularly relevant in radiomics, where massive amounts of data points might be generated from a given image set. A challenge XAI helps address is identifying only the most important features predictive of the outcome, which improves final model feature selection, thus eliminating the irrelevant ones, leading to a better model that is more robust and less prone to overfitting. Kha et al. have reported on identification of a radiomic signature predictive of 1p/19q co-deletion [30]. During model development, the SHAP framework was used to identify the most important features to include in their model, leading to improved model performance when less important features, based on average SHAP values, were removed, with an AUC of 0.710 before feature selection and 0.753 after.

Other applications of combination ML and XAI in radiomics simply use the resulting visualizations to identify global feature importance, to better understand and build trust in their respective models. These include prediction of early progression of nasopharyngeal cancer following intensity-modulated radiation therapy using MRI (which used the SHAP framework to demonstrate that select radiomic features were more predictively powerful than staging) [31] and classification of breast cancer molecular subtypes using non-contract computed tomography (achieving an accuracy of 71.3%, using SHAP to identify important radiomic features) [32]. In total, these explanations of radiomic models allow providers to better understand what the model is looking at, which eases application to the clinic by presenting them as more than abstract 'black boxes'.

19.2.4. Pathology

Pathology is another diagnostic specialty in oncology where XAI has improved predictive ML models. Like radiomics, pathology is also a field highly interested in individual predictions, where both SHAP and LIME can often be applicable. In a study similar to the one performed by Gaur et al. in radiomics, Palatnik de Sousa et al. used XAI to depict how a neural network identified tumours in histology samples of lymph node metastases [33]. Given that this problem involved a convolution neural network and was primarily interested in individual patients, this was a prime example where LIME was a good choice, where SHAP would be more computationally inefficient. Using this method, the model output can overlay histology slides, and end users can see locations of interest identified by the model. These overlays permitted the authors to biologically validate results by comparing highlighted areas of interest with clinical intuition. This technique has also been applied to diagnosing leukaemia (accuracy of 98.38%) [34]. Examination of individual patients has also been used on structured data to help classify primary and metastatic cancers using origin-based DNA methylation profiles [35].

Next, pathology studies have harnessed plots of non-linear inter-

change. Chakraborty et al. examined the influence of the tumour microenvironment on prognosis in breast cancer [36]. In this study, not only did XAI facilitate identification of prognostic factors via global feature importance, but it also illustrated non-linear interactions that permitted identification prognostic thresholds for each factor, which could help tailor identification of treatments that could manipulate the tumour microenvironment and improve prognosis. In this study, the authors used SHAP dependence plots to identify inflection points to correspond to potentially clinically relevant thresholds, generally corresponding to where SHAP values crossed zero. Similarly, a paper examines the impact of synoptic reporting on survival in patients with prostate cancer using an XGB model explained by SHAP, as reported by Janssen et al. [37]. As this study has the possibility to benefit pathology using XAI summary plots, the authors could conclude that synoptic reporting, specifically reporting pathologic data in a structured manner, is the second most important factor after age. In this example, these plots allow interpretation of an ML model that could lead to real policy changes, such as standard implementation of synoptic reporting, which would be difficult to illustrate to decision-makers without XAI.

Lastly, as is the case in other disciplines, XAI has been used to explain overall models to better understand how they generally function. Meena and co-worker developed a model for diagnosis of squamous cell carcinoma (SCC) based on genetic signature [38]. In a simple example of how SHAP values can be powerful, when distinguishing between healthy tissue and SCC, only a single gene (HNRNPM) produced a significant impact on the model, which was interestingly not among the 20 most important genes in actinic keratosis based on SHAP values. Although accurate diagnosis of SCC is useful, and the model was 92.86% accurate, these signatures made available by XAI are highly valuable to end users evaluating reports without access to sophisticated models, permitting powerful qualitative conclusions. Similar approaches have been applied to haematopoietic cancer subtype classification (with an accuracy of 97.01%) [39].

19.3. Treatment selection

Following diagnosis, an area where XAI has a significant opportunity to influence the clinic is in treatment selection. This might include identification of optimal treatment options and predicting outcomes of a given intervention. This is of great interest to the oncology clinic, as any additional information to determine optimal patient care is appreciated, but understanding such information is equally important in order to be able to explain recommendations to patients, their family, and other members of the care team.

Dependence plots generated by XAI can be useful for identification of predictive thresholds. Ladbury et al. [40] and Zarinshenas et al. [41] examined the prognostic and predictive values of nodal burden in endometrial cancer and locally advanced non-small-cell lung cancer (NSCLC), respectively, with associated XAI plots aiding in addressing controversies in the field. In endometrial cancer, via qualitative inspection of SHAP plots, XAI facilitated identification of a threshold of four or more positive nodes where treatment with adjuvant chemoradiation achieved optimal outcomes, while chemotherapy alone had a neutral effect and radiation alone had a deleterious effect. This finding adds insight following publication of PORTEC-3 and GOG 258, wherein optimal adjuvant therapy and sequencing remains unclear [42, 43]. In locally advanced NSCLC, again via qualitative inspection of SHAP plots, XAI enabled identification of nodal thresholds including three or more positive lymph nodes or a lymph node ration of 0.34 or greater as possible scenarios where addition of postoperative radiotherapy might improve outcomes, which is an area of controversy following publication of the LungART and PORT-C trials suggesting no benefit with postoperative radiotherapy [44, 45]. These conclusions were only possible because XAI enabled graphical depiction of interactions between nodal burden and treatments in the models, allowing for identification of predictive thresholds. The associated plots illustrating these interactions are found in **Figure 19.5**. Using a value of zero as neutral, the aforementioned thresholds can be identified, where patients who do not receive postoperative radiotherapy, represented in blue, have increased risk above those thresholds.

Beyond predicting optimal treatments, XAI has also demonstrated ability to help globally explain models that predict outcomes of treatments. Namely, Laios et al. [46] and Bang et al. [47] explored models that predicted for complete cytoreduction in ovarian cancer and curative resection in early undifferentiated gastric cancer. In the case of ovarian cancer, the model predicted R0 resection with an AUC of 0.866, using SHAP to identify important features. Additionally, SHAP identified non-linear interactions via inspection of dependence plots, such as significant decreases in R0 resection rates with peritoneal carcinomatosis indices greater than five and no significant change in R0 resection rates in years from 2017 onwards. In the case of gastric cancer, the model predicted curative resection with an accuracy of 89.8% and used SHAP to identify important factors. In combination, these two studies using XAI provide end users with useful information that can help counsel patients not only on odds of curative resection but also explain how these estimates were determined.

19.4. Epidemiology

An additional area involves improvements in epidemiology analyses using XAI, which benefit from examining global interactions and individual predictions. Ahmed et al. used ML to explore spatial variability of lung and bronchus cancer mortality rates across the contiguous United States [48]. The authors used dependence plots, break down plots, and maps to visually and geographically represent how key factors were interrelated and might affect specific geographic locations rather than simply describing the United States population as a whole. For example, they use XAI not only to show that Union County, Florida, has an almost 13-fold higher risk of mortality than Summit County, Utah, but also show that elevation is the largest protective factor in Summit County, while smoking is the largest risk factor in Union County. These conclusions were made possible by the authors using waterfall plots for corresponding individual explanations to visualize how model behaviour can vary drastically based on the given example. This information permits the model to be both predictive and able to inform possible interventions. In a study by Kobylińska et al., XAI was used to investigate the influence of factors on lung cancer screening [49]. The authors used the SHAP framework to produce summary plots, to overall illustrate lung cancer risk in lung cancer screening populations, dependence plots to show how changing individual variables influenced risk, and plots of individual predictions, which can be used how individual

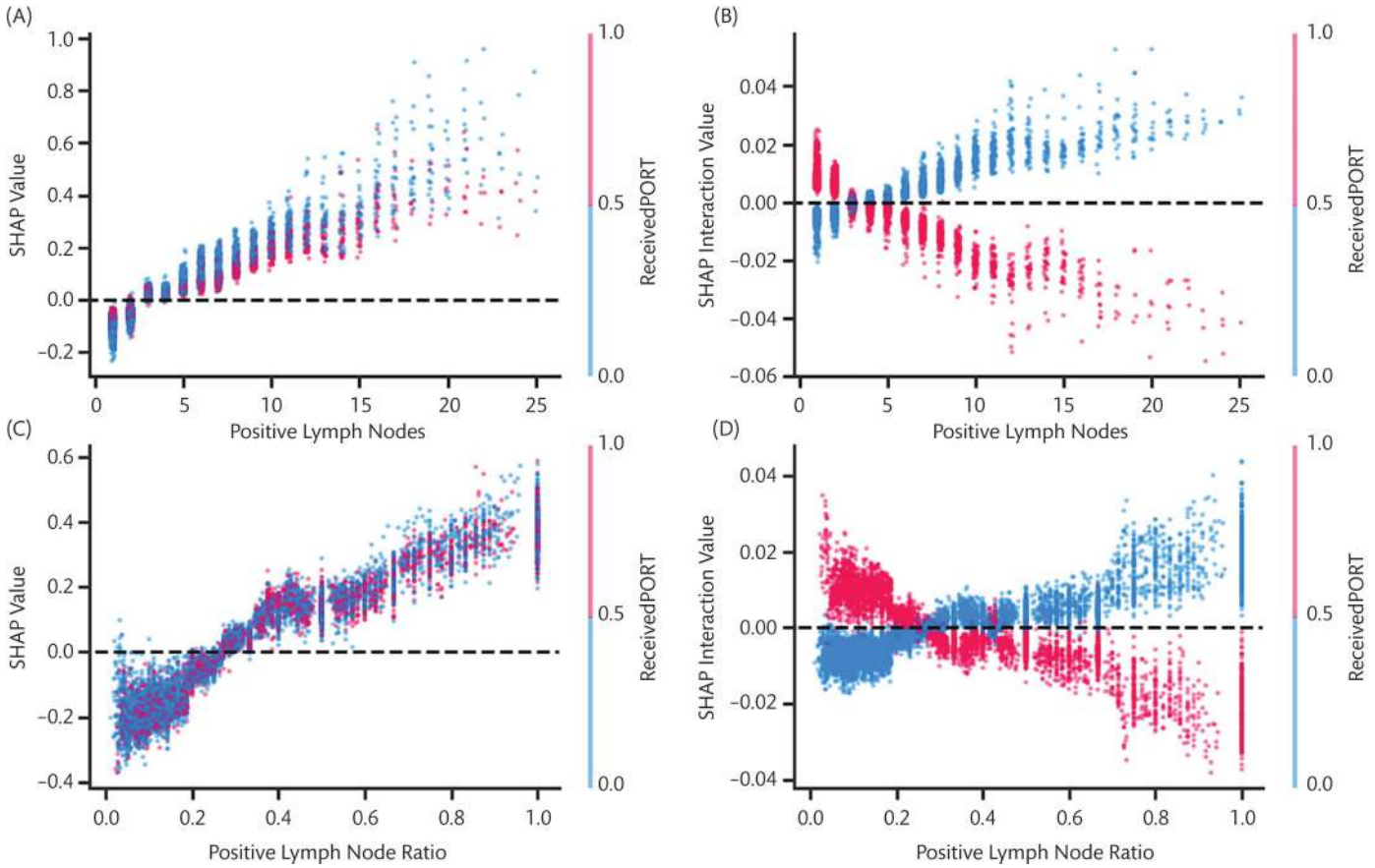


Figure 19.5. SHAP dependence plots (A, C) and interaction plots (B, D) illustrating thresholds for lymph node burden predictive of benefit of PORT in completely resected N2 NSCLC. *Source:* Reproduced with permission from [41].

patient risk is calculated. This information is useful, as it can not only be broadly used to counsel patients on decreasing lung cancer risk but can also inform specific patients of the best way to decrease their risk by identifying the most impactful factors to them.

Similar to other disciplines, studies in epidemiology have also benefited from XAI aiding with feature selection/construction. In a study by Richter and co-worker, electronic health record (EHR) data was utilized to predict risk of developing melanoma [50]. In the study, the use of XAI not only depicted important factors that are associated with risk of developing melanoma, but it also facilitated improvements in model efficiency that would facilitate the use of ML algorithms on EHR data by helping to select only the most relevant and impactful features, which would otherwise require prohibitive computational resources due to large data size and data missingness.

19.5. Radiation treatment workflow

One more niche area where XAI has been evaluated is in the radiation oncology clinic aiding in workflow for treatment planning. The first scenario that was evaluated by Siciarz et al. was clinical decision support systems for radiation plan evaluation in brain tumours [51]. The authors used XAI to look at the model globally, to examine interactions within features, and to evaluate individual predictions. This model classified treatment plans based on whether the target volume planning objective was met or whether the target volume planning objective was met or not due to a priority trade-off due to organ at risk constraints. Such a model uses knowledge from previous radiation treatment plans to inform on when trade-offs might be necessary. SHAP was used on the model to be able to determine relevant dosimetric factors that would inform whether a plan was acceptable, which can provide useful feedback to medical physicist and radiation oncologists when determining ways to further optimize plans that are not deemed acceptable. In this case, XAI was helpful in that it provides information on generally what leads to objectives being met and also providing a breakdown of specific cases. Next, once the plan is approved, quality assurance is required to ensure patient's safety and avoid clinically significant errors such as delivery of the desired dose. This is a process that can be somewhat automated with ML models. Again, it is important that a model predicts that a plan can be safely delivered and for end users to understand why this conclusion was made to facilitate improvements. Chen et al. explored a model that evaluated whether a plan would be deemed acceptable and used XAI to identify features that led to such a prediction, both globally and individually, which aids in further automating the quality assurance process [52].

19.6. Discussion

The studies discussed above, including how model-agnostic XAI was used by the authors, is summarized in Table 19.2. We found that

Table 19.2. Summary of identified model-agnostic XAI studies in oncology.

			Purpose of XAI				
Study	**Cancer type**	**XAI framework**	**Delineation of global feature importance**	**Characterization of individual prediction feature importance**	**Visualization of non-linear relationships and interactions**	**Identification of prognostic and/or predictive thresholds**	**Feature selection**
Prognostication							
Li et al. [11]	Prostate	SHAP			✓	✓	
Bertsimas et al. [12]	Pancreatic	SHAP			✓	✓	
Jansen et al. [13]	Breast	SHAP, LIME	✓	✓	✓		
Chen et al. [14]	Nasopharyngeal	SHAP	✓	✓	✓		
Moncada-Torres et al. [15]	Breast	SHAP	✓				
Van Den Bosch et al. [16]	Colorectal	SHAP	✓				
Gong et al. [17]	Oesophageal	SHAP	✓		✓		
Farswan et al. [18]	Multiple myeloma	SHAP	✓		✓		
Alsinglawi et al. [19]	Lung	SHAP	✓				
Jacobson et al. [20]	Bone metastases	SHAP	✓		✓		
Diagnosis							
Suh et al. [21]	Prostate	SHAP	✓				✓
Kwong et al. [22]	Prostate	SHAP	✓		✓		
Radiomics							
Manikis et al. [27]	Glioma	SHAP, LIME	✓	✓			
Gaur et al. [28]	Brain tumour	SHAP, LIME		✓			
Lee et al. [29]	Breast	LIME		✓			
Kha et al. [30]	Glioma	SHAP	✓				✓
Du et al. [31]	Nasopharyngeal	SHAP	✓		✓		
Wang et al. [32]	Breast	SHAP	✓				
Pathology							
Palatnik de Sousa et al. [33]	General	LIME		✓			
Abir et al. [34]	Leukaemia	LIME		✓			
Modhukar et al. [35]	Breast	LIME		✓			
Chakraborty et al. [36]	Breast	SHAP, LIME	✓		✓	✓	
Janssen et al. [37]	Prostate	SHAP	✓		✓		
Meena et al. [38]	Skin	SHAP	✓				
Park et al. [39]	Leukaemia	SHAP	✓				
Treatment selection							
Ladbury et al. [40]	Endometrial	SHAP			✓	✓	
Zarinshenas et al. [41]	Lung	SHAP			✓	✓	
Laios et al. [46]	Ovarian	SHAP	✓	✓	✓		
Bang et al. [47]	Gastric	SHAP	✓				
Epidemiology							
Ahmed et al. [48]	Lung	SHAP		✓	✓		
Kobylińska et al. [49]	Lung	SHAP	✓	✓	✓		
Richter et al. [50]	Melanoma	SHAP	✓				✓
Radiation treatment workflow							
Siciarz et al. [51]	Brain tumour	SHAP	✓	✓	✓		
Chen et al. [52]	Prostate	SHAP, LIME	✓	✓			

Abbreviations: XAI: explainable artificial intelligence; SHAP: SHapley Additive exPlanations; LIME: Local Interpretable Model-

the use of XAI could be divided into five categories: delineation of global feature importance, characterization of individual prediction feature importance, visualization of non-linear relationships and interactions, identification of prognostic and/or predictive thresholds, and feature selection/construction. These permit additional conclusions to be drawn from already ML algorithms, which are key to implementation in the oncology clinic, given that they improve understandability and therefore confidence.

In general, XAI is a vibrant area of ML research, with hundreds of papers produced in the field over the past decade. While there is a wide adoption of interpretability 'add-ons' (to the 'classical' ML classifiers and estimators), such as SHAP and LIME, there is yet no unifying XAI framework in the broad ML field, let alone biomedical ML. This situation is likely to change, given the heightened interest in XAI within both theoretical and applied ML research communities. Adoption of the cutting-edge advances and methods in XAI ML research in oncology applications, preferably biomedical data-specific, is the next frontier. That being said, there are already existing methods and algorithms in the ML/AI toolkit that are more explainable and interpretable by design, such as probabilistic causal modelling, various Bayesian methodologies, and fast/oblique decision trees. Adaptation of such techniques to the oncology spaces is an emerging trend [53-57]. Other areas of active research with direct ties to oncology are computation of hazard ratios from XAI output [58] and using XAI as an explicit feature selection/construction mechanism [59].

Although XAI is a powerful tool for improving the interpretability and thereby adoption of ML models in the oncology clinic, it is not without limitations, which do also apply to oncology research [60]. First, as discussed previously, an active area of research is implementation of models that are both accurate and interpretable, which potentially bypassed the need for XAI entirely. Therefore, the notion that ML inherently is a black box that must be explained must not always be true. Nevertheless, less interpretable models continue to be popular in oncology, as detailed by this review, meaning XAI still has a niche to fill as long as the gap between using accurate and interpretable models is being bridged. Next, there are inherent inaccuracies in explanations of certain ML models; many 'explanations' are actually approximations or probabilistic measures, as creating exact representations is computationally prohibitive [7, 8]. Furthermore, attempts to distil models may necessitate simplifications in order to be interpretable, which can raise concerns for applicability as well as human error when the explanations are used to make modifications for individual cases. Of course, this is a major hindrance to adoption in a high-stakes field such as oncology. When addressing this problem, a critical question for researchers is model selection; 'simpler' models such as LASSO regression may yield superior generalization results (thanks to decreased overfitting) relative to feature selection facilitated with XAI and a complex ML model. Notably, XAI complements feature selection (or includes feature selection, in the explainability-embedded ML methods), not 'competes' with it. A potential analysis framework might include variable selection, then XAI, and then further variable selection guided by XAI if necessary.

Future work should also seek to characterize 'significance' of resulting model explanations, which might be accomplished with a permutation test such as a permutation LASSO for variable selection [61] or aforementioned efforts to generate hazard ratios and *p*-values from explanations [58]. Lastly, the explanations can only be as good as their models, so it is important that models undergo robust testing with suitable performance metrics and are empirically sound before being explained, yielding improved efficacy and safety. Notably, a major limitation of most of the included studies in this review is a lack of external validation of model results as well as cross-validation of explanations. Future work should ensure that models are extensively quality controlled, tested, and validated, optimally via a schema, such as Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) [62], minimum information about clinical artificial intelligence modelling (MI-CLAIM) [63], or Radiomics Quality Score [64], of course acknowledging that due to availability of suitable datasets may limit external validation. Despite these limitations, it is not our recommendation that XAI have no place in oncology. On the contrary, as discussed above, it is our belief that it is a valuable tool, but it does need to be implemented responsibly and assessed critically before being used to influence patient care.

19.7. Conclusions

ML undoubtedly has the potential to improve the cancer clinic. In addition to delivering accurate predictive models, it is critical that models be interpretable by the practitioners who will use them; otherwise, adoption in the clinic would be limited due to their intricate and sometimes indecipherable character. XAI methodologies, such as LIME and SHAP, can generate powerful visualizations that illustrate the inner workings of ML algorithms used in a variety of oncologic fields, making them easier for average end users to understand and, in some cases, providing actionable information that end users can use to improve patient outcomes. Further, XAI facilitates feature selection/construction, identification of prognostic and/or predictive thresholds, and overall confidence in the models, among other benefits. To ensure that ML oncologic research achieves its maximal benefit and reach, future studies should consider utilization of XAI frameworks, which can make the models more understandable to end users without technical acumen that would otherwise be needed to interpret ML literature.

Acknowledgements

Adapted with permission from Ladbury C, et al. Utilization of model-agnostic explainable artificial intelligence frameworks in oncology: a narrative review. *Transl Cancer Res.* 2022;11(10):3853–3868.

REFERENCES

1. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. New Engl J Med. 2019;**380**(14):1347–58.
2. Papernot N, McDaniel P, Goodfellow I, Jha S, Celik ZB, Swami A, editors. Practical black-box attacks against machine learning. Proceedings of the 2017 ACM on Asia conference on computer and communications security; 2017.
3. Diprose WK, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. J Am Med Inform Assoc. 2020;**27**(4):592–600.

4. Price WN. Big data and black-box medical algorithms. Sci Transl Med. 2018;**10**(471):eaao5333.
5. Gunning D, Stefik M, Choi J, Miller T, Stumpf S, Yang G-Z. XAI—explainable artificial intelligence. Sci Robot. 2019;**4**(37):eaay7120.
6. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nat Mach Intell. 2020;**2**(1):56–67.
7. Ribeiro MT, Singh S, Guestrin C, editors. 'Why should I trust you?' Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016.
8. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;**30**:4768–77.
9. van der Velden BHM, Kuijf HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Med Image Anal. 2022;**79**:102470.
10. Ladbury C, Zarinshenas R, Semwal H, Tam A, Vaidehi N, Rodin AS, et al. Utilization of model-agnostic explainable artificial intelligence frameworks in oncology: a narrative review. Transl Cancer Res. 2022;**11**(10):3853–68.
11. Li R, Shinde A, Liu A, Glaser S, Lyou Y, Yuh B, et al. Machine learning-based interpretation and visualization of nonlinear interactions in prostate cancer survival. JCO Clin Cancer Inform. **2020**(4):637–46.
12. Bertsimas D, Margonis GA, Huang Y, Andreatos N, Wiberg H, Ma Y, et al. Toward an optimized staging system for pancreatic ductal adenocarcinoma: a clinically interpretable, artificial intelligence-based model. JCO Clin Cancer Inform. **2021**(5):1220–31.
13. Jansen T, Geleijnse G, Van Maaren M, Hendriks MP, Ten Teije A, Moncada-Torres A. Machine learning explainability in breast cancer survival. Digital Personalized Health and Medicine: IOS Press; 2020. p. 307–11.
14. Chen X, Li Y, Li X, Cao X, Xiang Y, Xia W, et al. An interpretable machine learning prognostic system for locoregionally advanced nasopharyngeal carcinoma based on tumor burden features. Oral Oncol. 2021;**118**:105335.
15. Moncada-Torres A, Van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Explainable machine learning can outperform Cox regression predictions and provide insights in breast cancer survival. Sci Rep. 2021;**11**:6968.
16. Van Den Bosch T, Warps A-LK, De Nerée Tot Babberich MPM, Stamm C, Geerts BF, Vermeulen L, et al. Predictors of 30-day mortality among Dutch patients undergoing colorectal cancer surgery, 2011-2016. JAMA Netw Open. 2021;**4**(4):e217737.
17. Gong X, Zheng B, Xu G, Chen H, Chen C. Application of machine learning approaches to predict the 5-year survival status of patients with esophageal cancer. J Thorac Dis. 2021;**13**(11):6240–51.
18. Farswan A, Gupta A, Sriram K, Sharma A, Kumar L, Gupta R. Does ethnicity matter in multiple myeloma risk prediction in the era of genomics and novel agents? Evidence from real-world data. Front Oncol. 2021;**11**:720932.
19. Alsinglawi B, Alshari O, Alorjani M, Mubin O, Alnajjar F, Novoa M, et al. An explainable machine learning framework for lung cancer hospital length of stay prediction. Sci Rep. 2022;**12**:607.
20. Jacobson D, Cadieux B, Higano CS, Henry DH, Bachmann BA, Rehn M, et al. Risk factors associated with skeletal-related events following discontinuation of denosumab treatment among patients with bone metastases from solid tumors: a real-world machine learning approach. J Bone Oncol. 2022;**34**:100423.
21. Suh J, Yoo S, Park J, Cho SY, Cho MC, Son H, et al. Development and validation of an explainable artificial intelligence-based decision-supporting tool for prostate biopsy. BJU Int. 2020;**126**(6):694–703.
22. Kwong JCC, Khondker A, Tran C, Evans E, Cozma AI, Javidan A, et al. Explainable artificial intelligence to predict the risk of side-specific extraprostatic extension in pre-prostatectomy patients. Can Urol Assoc J. 2022;**16**(6):213–21.
23. Incoronato M, Aiello M, Infante T, Cavaliere C, Grimaldi A, Mirabelli P, et al. Radiogenomic analysis of oncological data: a technical survey. Int J Mol Sci. 2017;**18**(4):805.
24. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. Theranostics. 2019;**9**(5):1303–22.
25. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;**278**(2):563–77.
26. Ludwig K, Kornblum HI. Molecular markers in glioma. J Neuro-Oncol. 2017;**134**(3):505–12.
27. Manikis GC, Ioannidis GS, Siakallis L, Nikiforaki K, Iv M, Vozlic D, et al. Multicenter DSC–MRI-based radiomics predict IDH mutation in gliomas. Cancers. 2021;**13**(16):3965.
28. Gaur L, Bhandari M, Razdan T, Mallik S, Zhao Z. Explanation-driven deep learning model for prediction of brain tumour status using MRI image data. Front Genet. 2022;**13**:822666.
29. Lee YW, Huang CS, Shih CC, Chang RF. Axillary lymph node metastasis status prediction of early-stage breast cancer using convolutional neural networks. Comput Biol Med. 2021;**130**:104206.
30. Kha Q-H, Le V-H, Hung TNK, Le NQK. Development and validation of an efficient MRI radiomics signature for improving the predictive performance of 1p/19q co-deletion in lower-grade gliomas. Cancers. 2021;**13**(21):5398.
31. Du R, Lee VH, Yuan H, Lam K-O, Pang HH, Chen Y, et al. Radiomics model to predict early progression of nonmetastatic nasopharyngeal carcinoma after intensity modulation radiation therapy: a multicenter study. Radiol: Artif Intell. 2019;**1**(4):e180075.
32. Wang F, Wang D, Xu Y, Jiang H, Liu Y, Zhang J. Potential of the non-contrast-enhanced chest CT radiomics to distinguish molecular subtypes of breast cancer: a retrospective study. Front Oncol. 2022;**12**:848726.
33. Palatnik de Sousa I, Maria Bernardes Rebuzzi Vellasco M, Costa da Silva E. Local interpretable model-agnostic explanations for classification of lymph node metastases. Sensors (Basel). 2019 Jul 5;**19**(13):2969.
34. Abir WH, Uddin MF, Khanam FR, Tazin T, Khan MM, Masud M, et al. Explainable AI in diagnosing and anticipating leukemia using transfer learning method. Comput Intell Neurosci. 2022;**2022**:5140148.
35. Modhukur V, Sharma S, Mondal M, Lawarde A, Kask K, Sharma R, et al. Machine learning approaches to classify primary and metastatic cancers using tissue of origin-based DNA methylation profiles. Cancers (Basel). 2021 Jul 27;**13**(15):3768.
36. Chakraborty D, Ivan C, Amero P, Khan M, Rodriguez-Aguayo C, Başağaoğlu H, et al. Explainable artificial intelligence reveals novel insight into tumor microenvironment conditions linked with better prognosis in patients with breast cancer. Cancers. 2021;**13**(14):3450.
37. Janssen FM, Aben KK, Heesterman BL, Voorham QJ, Seegers PA, Moncada-Torres A. Using explainable machine learning to explore the impact of synoptic reporting on prostate cancer. Algorithms. 2022;**15**(2):49.

38. Meena J, Hasija Y. Application of explainable artificial intelligence in the identification of squamous cell carcinoma biomarkers. Comput Biol Med. 2022;**146**:105505.
39. Park KH, Batbaatar E, Piao Y, Theera-Umpon N, Ryu KH. Deep learning feature extraction approach for hematopoietic cancer subtype classification. Int J Environ Res Public Health. 2021 Feb 23;**18**(4):2197.
40. Ladbury C, Li R, Shiao J, Liu J, Cristea M, Han E, et al. Characterizing impact of positive lymph node number in endometrial cancer using machine-learning: A better prognostic indicator than FIGO staging? Gynecol Oncol. 2022;**164**(1):39–45.
41. Zarinshenas R, Ladbury C, McGee H, Raz D, Erhunmwunsee L, Pathak R, et al. Machine learning to refine prognostic and predictive nodal burden thresholds for post-operative radiotherapy in completely resected stage III-N2 non-small cell lung cancer. Radiother Oncol. 2022 Aug;**173**:10–18.
42. Matei D, Filiaci V, Randall ME, Mutch D, Steinhoff MM, Disilvestro PA, et al. Adjuvant Chemotherapy plus Radiation for Locally Advanced Endometrial Cancer. New Engl J Med. 2019;**380**(24):2317–26.
43. De Boer SM, Powell ME, Mileshkin L, Katsaros D, Bessette P, Haie-Meder C, et al. Adjuvant chemoradiotherapy versus radiotherapy alone for women with high-risk endometrial cancer (PORTEC-3): final results of an international, open-label, multicentre, randomised, phase 3 trial. Lancet Oncol. 2018;**19**(3):295–309.
44. Le Pechoux C, Pourel N, Barlesi F, Lerouge D, Antoni D, Lamezec B, et al. Postoperative radiotherapy versus no postoperative radiotherapy in patients with completely resected non-small-cell lung cancer and proven mediastinal N2 involvement (Lung ART): an open-label, randomised, phase 3 trial. Lancet Oncol. 2021.
45. Hui Z, Men Y, Hu C, Kang J, Sun X, Bi N, et al. Effect of postoperative radiotherapy for patients with pIIIA-N2 non–small cell lung cancer after complete resection and adjuvant chemotherapy. JAMA Oncol. 2021;**7**(8):1178.
46. Laios A, Kalampokis E, Johnson R, Thangavelu A, Tarabanis C, Nugent D, et al. Explainable artificial intelligence for prediction of complete surgical cytoreduction in advanced-stage epithelial ovarian cancer. J Pers Med. 2022 Apr 10;12(4):607.
47. Bang CS, Ahn JY, Kim JH, Kim YI, Choi IJ, Shin WG. Establishing machine learning models to predict curative resection in early gastric cancer with undifferentiated histology: development and usability study. J Med Internet Res. 2021;**23**(4):e25053.
48. Ahmed ZU, Sun K, Shelly M, Mu L. Explainable artificial intelligence (XAI) for exploring spatial variability of lung and bronchus cancer (LBC) mortality rates in the contiguous USA. Sci Rep. 2021 Dec 16;**11**(1):24090.
49. Kobylińska K, Orłowski T, Adamek M, Biecek P. Explainable machine learning for lung cancer screening models. Appl Sci. 2022;**12**(4):1926.
50. Richter AN, Khoshgoftaar TM. Efficient learning from big data for cancer risk modeling: a case study with melanoma. Comput Biol Med. 2019;**110**:29–39.
51. Siciarz P, Alfaifi S, Uytven EV, Rathod S, Koul R, Mccurdy B. Machine learning for dose-volume histogram based clinical decision-making support system in radiation therapy plans for brain tumors. Clin Transl Radiat Oncol. 2021;**31**:50–7.
52. Chen Y, Aleman DM, Purdie TG, McIntosh C. Understanding machine learning classifier decisions in automated radiotherapy quality assurance. Phys Med Biol. 2022;**67**(2). doi: 10.1088/1361-6560/ac3e0e.
53. Kalet AM, Doctor JN, Gennari JH, Phillips MH. Developing Bayesian networks from a dependency-layered ontology: a proof-of-concept in radiation oncology. Med Phys. 2017;**44**(8):4350–9.
54. Trilla-Fuertes L, Gámez-Pozo A, Arevalillo JM, López-Vacas R, López-Camacho E, Prado-Vázquez G, et al. Bayesian networks established functional differences between breast cancer subtypes. PLoS ONE. 2020;**15**(6):e0234752.
55. Park SB, Hwang K-T, Chung CK, Roy D, Yoo C. Causal Bayesian gene networks associated with bone, brain and lung metastasis of breast cancer. Clin Exp Metastasis. 2020;**37**(6):657–74.
56. Wang X, Branciamore S, Gogoshin G, Ding S, Rodin AS. New analysis framework incorporating mixed mutual information and scalable Bayesian networks for multimodal high dimensional genomic and epigenomic cancer data. Front Genet. 2020;**11**.
57. Djulbegovic B, Hozo I, Dale W. Transforming clinical practice guidelines and clinical pathways into fast-and-frugal decision trees to improve clinical care strategies. J Eval Clin Pract. 2018;**24**(5):1247–54.
58. Sundrani S, Lu J. Computing the hazard ratios associated with explanatory variables using machine learning models of survival data. JCO Clin Cancer Inform. 2021;**5**:364–78.
59. Marcílio WE, Eler DM, editors. From explanations to feature selection: assessing SHAP values as feature selection mechanism. 2020 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI); 2020: IEEE.
60. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;**1**(5):206–15.
61. Yang S, Wen J, Eckert ST, Wang Y, Liu DJ, Wu R, et al. Prioritizing genetic variants in GWAS with lasso using permutation-assisted tuning. Bioinformatics. 2020;**36**(12):3811–7.
62. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;**162**(1):W1–73.
63. Norgeot B, Quer G, Beaulieu-Jones BK, Torkamani A, Dias R, Gianfrancesco M, et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. Nat Med. 2020;**26**(9):1320–4.
64. Sanduleanu S, Woodruff HC, De Jong EE, Van Timmeren JE, Jochems A, Dubois L, et al. Tracking tumor biology with radiomics: a systematic review utilizing a radiomics quality score. Radiother Oncol. 2018;**127**(3):349–60.

20

Applying cloud computing and informatics in cancer

Jay G. Ronquillo

20.1. Biomedical informatics

Biomedical informatics is a growing field of medicine that leverages the increased availability of technology applicable in healthcare [1]. At one level, it is the integration and synthesis of numerous fields, such as medicine, engineering, computer science, public health, biostatistics, and data science. In addition, it is also the broad intersection of healthcare and technology and the practical application of technology (in its many forms) to positively impact the health of both individuals and populations. The American Medical Informatics Association (AMIA) Board published a white paper that provides a good description of the key components of the field of biomedical informatics. More specifically, there are several important subcategories under the umbrella of 'biomedical informatics', including (1) bioinformatics and imaging informatics, (2) translational bioinformatics and clinical research informatics, and (3) health informatics [2]. These subcategories of informatics represent the entire spectrum of healthcare, from the microscopic (i.e. molecules, cells, and tissues) to the macroscopic (i.e. patients, communities, and populations), as well as the translational elements in between [2].

Health informatics is an especially important subcategory which can be divided, respectively, into its individual and population level components of (1) clinical informatics and (2) public health informatics [2,3]. Clinical informatics can be further classified according to specific clinical areas of focus, such as medical informatics (focused on healthcare providers in general), nursing informatics, and dental informatics [2]. These classifications and subclassifications may overlap and are not mutually exclusive.

Individuals who work or practice in the field are called 'informaticists' or 'informaticians' and can be certified in either clinical informatics or health informatics [3–5]. Board certification in clinical informatics commonly requires (1) a medical degree, (2) medical license, (3) primary board certification, (4) passing the Clinical Informatics Board Exam, and either (5a) a demonstrated number of years practicing clinical informatics or graduate degree in biomedical informatics (Practice Pathway) or (5b) completion of an Accreditation Council for Graduate Medical Education (ACGME)-accredited fellowship in clinical informatics (Fellowship Pathway) [5,6]. Alternatively, certification in health informatics requires (1) a graduate degree in health informatics or a related field, (2) a demonstrated number of years of qualifying health informatics work experience, and (3) passing the AMIA Health Informatics Certification Exam [7].

Separate practice analyses were performed to better understand the domains, tasks, and knowledge commonly leveraged for the successful practice of both clinical informatics and health informatics [8,9]. For clinical informatics, there are currently five major domains of practice: (1) fundamental knowledge and skills, (2) improving care delivery and outcomes, (3) enterprise information systems, (4) data governance and data analytics, and (5) leadership and professionalism [9,10]. Similarly, for health informatics, the five key domains of practice are (1) foundational knowledge; (2) enhancing health decision-making, processes, and outcomes; (3) health information systems; (4) data governance, management, and analytics; and (5) leadership, professionalism, strategy, and transformation [8,11]. While the high-level domains appear similar at first glance, the practice analysis breakdowns of common tasks, knowledge, and skills highlight different areas of emphasis. Clinical informatics, for example, has the largest number of knowledge/skills focused on enterprise information systems and a strong emphasis on improving care delivery and outcomes. Health informatics, on the other hand, has the largest number of knowledge/skills targeting health information systems, with a heavy emphasis on data governance, management, and analytics. These differences likely reflect the common medical settings of clinicians practicing clinical informatics, in contrast to the diverse and more data-intensive backgrounds of professionals practicing health informatics [8–12]. Both clinical and health informaticists engage in broad aspects of cancer research, as their 'toolbox' of versatile abilities and skills allow them to solve complex problems across disciplines, industries, and sectors [13–19].

20.2. Informatics for big data research

20.2.1. Biomedical big data

An important aspect of cancer research recently enabled by informatics is the use of biomedical 'big data', defined as the growing volume, variety, and velocity of data describing the many dimensions of human health [20–24]. This 'real-world evidence' or 'real-world data' has opened up new avenues of research and allowed large population-based studies of diverse patient communities to be performed at an unprecedented scale and scope [25–29]. However, working with large healthcare databases differs from traditional clinical trials where study design, patient recruitment, and data collection procedures are clearly defined and well structured [28]. Indeed, many real-world healthcare databases are complex aggregations of heterogeneous sources, encoded through a fragmented collection of data standards, or even composed of substantial volumes of unstructured data [28–30]. As a result, these real-world studies often require significant data processing or 'wrangling' before any comprehensive analysis can be performed, making the application of informatics and data science critical for success [28–30].

Traditional approaches for clinical trials generally involve designing a study, recruiting patients, collecting data, and analysing results [28]. The complex and manual processes involved with clinical trials often require large teams of specially trained staff, deep investment of resources, and multiple years of effort to find, recruit, and enroll eligible patients [16,28,31]. However, for studies that primarily focus on publicly available datasets and real-world data, the research workflow is often much faster, requires fewer resources, and enables 'recruitment' of hundreds of thousands (and possibly millions) of eligible patients. In general, the workflow for data-driven studies can be summarized through the design, development, and implementation of an 'informatics pipeline' composed of three phases: (1) identification of appropriate datasets; (2) extraction, integration, and harmonization of data ('data wrangling'); and (3) analysis and visualization of results (Figure 20.1).

20.2.2. Dataset identification (informatics pipeline I)

Finding, evaluating, and selecting robust datasets is a critical first step for big data research [27]. Once candidate datasets have been identified, investigators should assess their scale and scope by understanding the population of interest (e.g. patients, participants, samples, files, etc.), extent of standardization, degree of missingness, and availability of other fields (e.g. metadata) describing the core entities of interest [32,33]. Good datasets often have a data dictionary or other documentation that provide an overview of the metadata fields available and their possible values. In general, data that has been structured, standardized, or harmonized into a common data model will be easier to analyse and require less preprocessing than raw unstructured datasets [27]. In addition, investigators integrating multiple datasets together (e.g. to benefit from large sample sizes) must decide which dataset will serve as the study's core content and which one(s) will play a more supplementary role used for annotation. For example, a researcher may use de-identified patient data from institutional Electronic Health Records (EHRs) as the core dataset for their study but enhance it with population data from the US Centers for Disease Control and Prevention (CDC) or US Census Bureau. Finally, it will also be important to assess the credibility of each source of data, as datasets generated by different sectors (e.g. government, industry, and nonprofit) will vary in terms of scale, scope, completeness, update frequency, access restrictions, degree of standardization, and even potential quality [32,33].

From a practical perspective, one approach to assessing the quality of a dataset involves using the 'FAIR' guiding principles for scientific data: Findable, Accessible, Interoperable, and Reusable [34]. Firstly, according to Wilkinson et al., datasets with high 'findability' are characterized as having unique and persistent identifiers, being annotated with comprehensive metadata, and accessible through an easily searchable repository [34]. Secondly, datasets with high 'accessibility' possess robust and clearly defined authentication, authorization, and retrieval protocols revolving around persistent identifiers and metadata [34]. Thirdly, highly 'interoperable' datasets leverage robust data models and vocabularies that make scalable integration with other datasets possible [34]. Finally, 'reusability' describes datasets with clearly defined data usage rules and provenance to enable repeated, consistent use across different research domains [34].

Once datasets have been selected for a study, it will also be important to understand specific access requirements. While some datasets may be open access and publicly available, others (e.g. those with sensitive patient or patient health information) may have more rules and restrictions regarding access. Organizations sharing robust research datasets usually have their privacy and access policies clearly described, and may require that investigators (or institutions) enter into specific data use agreements, obtain institutional review board approval, complete required research/ethics training, or submit short research proposals for review.

20.2.3. Data wrangling (informatics pipeline II)

Once datasets are selected, the technical aspects of accessing, integrating, and transforming data must be addressed. This could be done by writing and running queries using a database language like Structured Query Language (SQL) to extract needed data and relevant fields. Alternatively, many real-world datasets are accessible through an Application Programming Interface (API), which allows scripts (e.g. in programming languages such as Python or R) to access the data over the internet [24,35]. For researchers using cloud-based resources, this may also involve transferring datasets into a cloud workspace or environment where they can be further transformed into an analysis-ready format.

Data integration and harmonization can be a complex and non-trivial task, especially when dealing with large heterogeneous datasets. In general, the harmonization process involves making sure all key data and metadata fields are in a structured and standardized format

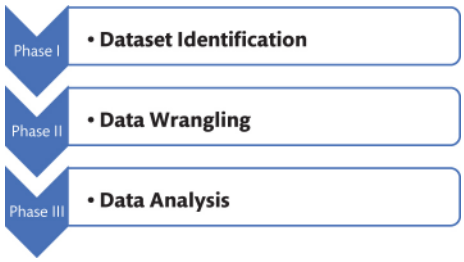


Figure 20.1. Informatics pipeline for biomedical big data research.

[36,37]. This requires assessing a dataset's level of standardization, which often falls into one of the three categories: (1) raw unstructured data that needs to be structured and standardized, (2) structured data that is not standardized or only partially standardized, and (3) structured data that has been fully standardized [36,37]. Raw unstructured data can be converted into a structured format using a variety of tools and techniques, from simple (but time-consuming) approaches, such as manual review and extraction by a subject matter expert, to automated approaches, such as natural language processing [38–40]. Structured data can be further transformed and standardized through the mapping of relevant fields to any number of biomedical terminologies, ontologies, nomenclatures, classifications, or vocabularies [41,42]. Some examples include mapping clinical diagnoses to International Classification of Diseases (ICD) codes, laboratory tests to Logical Observation Identifiers Names and Codes (LOINC) codes, medical procedures to Current Procedural Terminology (CPT) codes, or genes as represented in the Human Genome Organization (HUGO) Gene Nomenclature Committee (HGNC) database [37,41,42]. The final step in the harmonization process involves integrating structured and standardized datasets so that common data elements in different datasets follow the same biomedical data standards. Informaticians with appropriate subject matter expertise and a strong understanding of biomedical terminologies are capable of effectively performing these steps. The endgame of this process should be a single, clean, integrated dataset that is ready for robust analyses.

20.2.4. Data analysis (informatics pipeline III)

While the options can be broad, the analysis of large real-world datasets usually start with understanding fundamental characteristics of the population being studied, often described in the first table ('Table 1') of most published studies. Depending on whether data are categorical (e.g. nominal and ordinal) or numerical (e.g. discrete and continuous), descriptive statistics can include frequencies and proportions (or percentages), measures of central tendency (e.g. mean, median, and mode), as well as measures of variability (e.g. standard deviation, range, and interquartile range) [43–45].

In general, biostatistical approaches for hypothesis testing involve parametric tests when assumptions of normality, equal variances, and independence are satisfied [43,44]. Common parametric tests include the following: two-sample unpaired *t*-test when assessing mean differences in two different groups, analysis of variance (ANOVA) for comparing differences in means between three or more groups, and the Pearson correlation coefficient to assess the relationship between two different sets of data [43,44]. Nonparametric tests are used when the assumptions for parametric tests are not satisfied, often deal with frequencies (i.e. counts), and commonly include the following: chi-square test, Fisher's exact test, Mann–Whitney *U* or Wilcoxon rank sum test (analogous to the two-sample unpaired *t*-test), Wilcoxon signed-rank test (the nonparametric version of the paired *t*-test), Kruskal–Wallis (the nonparametric ANOVA), and Spearman's rho (roughly equivalent to a non-parametric Pearson correlation coefficient) [43,44].

Additional analysis and visualization of integrated data can be performed using several approaches. There are many options ranging from commercial off-the-shelf software to open-source tools to home-grown applications. Many solutions are standalone technologies designed specifically to perform data analysis and visualization. However, many packages and libraries have also been developed for popular programming languages like Python (e.g. pandas, Matplotlib, and Seaborn) and R (e.g. dplyr, ggplot2, and Plotly), and which can be leveraged through interactive computing notebooks (such as Jupyter Notebooks) [35].

20.3. Informatics and cancer research

While informatics has been leveraged in many areas of biomedical research, there are subtle nuances when applying informatics to oncology. For example, while some data standards contain broadly defined content that must be adapted for cancer research, others have content designed specifically for cancer. For example, the ICD is a standardized vocabulary to describe many different conditions. Current versions in existing US health information technology (IT) systems include International Classification of Diseases, Ninth Revision (ICD9) and International Classification of Diseases, Tenth Revision (ICD10), which contain a rich collection of diagnoses ranging from hypertension to diabetes to Alzheimer's disease, along with many different types of cancers [46]. However, while ICD9 and ICD10 codes can describe general cancer sites (e.g. lung cancer, colorectal cancer, etc.), International Classification of Diseases for Oncology, Third Edition codes were created to more specifically and accurately describe cancer as a combination of topography, morphology, and behaviour codes [47]. Firstly, topography describes the general location or site of the cancer (e.g. lung cancer, colorectal, breast, prostate, etc.). Secondly, morphology corresponds to the specific histology associated with the cancer (e.g. adenocarcinoma, squamous cell, intraductal, small cell, etc.). Finally, the behaviour code distinguishes whether the tumour is a benign neoplasm, *in situ* neoplasm, primary malignant neoplasm, or secondary malignant neoplasm [47]. In practice, identifying a cancer cohort in real-world databases is complex and often involves a combination of different structured data elements (e.g. cancer-specific diagnoses, treatments, procedures, visits, etc.) as well as diverse data sources (e.g. EHRs, claims, patient surveys, etc.) [48–50].

20.4. Cloud computing

The onset of cloud computing has presented new challenges and opportunities for biomedical research in general and cancer research in particular. There are active discussions about the overall benefits of performing biomedical research through cloud computing compared to 'on-premise' resources [35,51]. Broadly speaking, researchers who use 'cloud computing' will access and interact with data almost exclusively through the Internet since all data-related hardware and software are remotely provided by a third-party vendor with cloud expertise (e.g. Google, Amazon, and Microsoft) [51]. From the researcher perspective, this effectively means that all relevant data are stored in the cloud, and all analyses are performed with software and analytics tools residing in the cloud as well. There are minimal upfront costs by the researcher's home institution since cloud providers are responsible for all updates and maintenance, giving users high service availability with relatively little downtime [51]. In addition, many cloud vendors offer investigators a variety of cloud 'services' (e.g. 'software as a service', 'infrastructure as a

service', and 'platform as a service') through an on-demand, pay-as-you-go basis, allowing for researchers to add or remove services to match their specific needs [51].

In contrast, 'on-premise(s)' computing involves computational hardware and software that is licensed, installed, and maintained by the researcher's home institution. Specifically, all research data are stored on servers controlled and maintained within the researcher's organization, along with all relevant software and subsequent analyses. This approach requires larger upfront institutional investments in both technical infrastructure and staff for its maintenance and support (e.g. software and security updates, hardware replacement, or upgrades). However, it also enables the home institution to provide a very consistent, stable, and controlled technical environment for its research investigators.

One way to better understand these differences is through the analogy of 'buying a home' ('on-premise') vs. 'renting an apartment' ('cloud'). A prospective homeowner must cover substantial upfront costs to purchase the home as well as assume responsibility for addressing all issues with its repair and upkeep. One major advantage is that the homeowner has full control over everything they want to do with the home (e.g. remodelling, renovating, landscaping, etc.). In contrast, someone renting an apartment cannot meaningfully alter the apartment (e.g. remove walls, change appliances, etc.), but only worries about paying for services they use (e.g. heat, water, and electricity) rather than fixing things that break, which is handled by the landlord. Apartment renters also have the flexibility to move to larger or smaller apartments (or entirely different locations) based on their needs and finances. In a nutshell, researchers using on-premise technologies (the 'homeowner') have not only full control of resources but also full responsibility for its maintenance, while cloud computing (the 'renter') offloads both control and responsibility to cloud vendors (the 'landlord') in exchange for dramatic scalability and convenience. Just as individuals have different housing needs and preferences, specific investigators have different research goals that could be addressed through cloud technologies.

20.5. Cloud costs

Understanding how to monitor and manage costs involved with using the cloud remains an important barrier to wider adoption of cloud resources, and there are several rules of thumb for aspiring cloud researchers [52,53]. Broadly speaking, cloud technologies provide three important functions for biomedical big data: (1) cloud storage, (2) cloud computation, and (3) cloud services [54]. Firstly, there is a direct relationship between costs and data storage volume, with costs increasing as data storage requirements increase [53,55]. Also, data storage retrieval time ('latency') has an inverse relationship with costs, meaning cloud costs become more expensive when researchers need faster data retrieval times and less expensive if slower times are acceptable [53,55]. Secondly, costs associated with cloud computation have similar relationships, with costs higher for research projects that require (1) more highly configured virtual machines instead of ones with fewer features, (2) longer instead of shorter data analysis and processing times, or (3) continuous, on-demand availability of compute resources instead of flexible or off-peak usage [52,53,56].

Finally, cloud 'services' have the most variable range of cloud costs as their pricing structures depend on the specific technology being used, the cloud provider who developed it, and the usage needs of the researcher. Cloud-based healthcare natural language processing services, for example, take large volumes of raw unstructured text, extract relevant medical terms, and assign each term a specific code from selected biomedical vocabularies [53,57,58]. However, the pricing for this service depends on many factors, including the total characters processed, type of analysis (e.g. entity analysis, sentiment analysis, syntax analysis), how much de-identification of sensitive data was performed, which biomedical data standards were mapped, and even where in the cloud documents were stored or analysed [53,57,58]. In summary, many factors can influence cloud costs, but informed researchers can configure their cloud environments to minimize unneeded costs and maximize research productivity.

20.6. Cancer research and the cloud

Practically speaking, there are two main reasons an investigator would want to use the cloud for cancer research. Firstly, the cloud will likely become necessary to perform 'big data' studies of unprecedented scale and scope. The Genomic Data Commons, for example, is one of the many large cloud-based repositories of cancer data maintained by the National Cancer Institute as part of the Cancer Moonshot initiative [59]. At the time of this writing, the Genomic Data Commons alone contains nearly 3 petabytes of data or roughly 3,000 times the maximum capacity of a top-of-the-line laptop [59]. For investigators planning to perform research involving millions of patients and billions of data points, big data storage and computation needs will far surpass the capacity of a single computer (or even institution) to physically handle this workload, and the cloud becomes a clear, scalable solution. Secondly, cloud vendors with deep technical expertise have created many robust big data analytics tools, including services expected to play a key role in cancer research: natural language processing; search, image/video processing; and even artificial intelligence, machine learning, or deep learning [52,60–64]. For cancer researchers who cannot find useful open-source solutions or do not want to build analytics tools from scratch, these proprietary fee-based services provide a powerful option.

20.7. Future directions

20.7.1. NCI Cancer Research Data Commons

The National Cancer Institute (NCI) Cancer Research Data Commons (CRDC) is a cloud-based cancer research data ecosystem created to accelerate cancer research through big data [22,61,65–67]. In general, the CRDC is made up of two major components: (1) large cancer datasets centralized in focused repositories called 'data commons' and (2) technical infrastructure in the form of cloud-based platforms or 'cloud resources'. Datasets in the CRDC with similar modalities are grouped into individual data commons. For example, cancer-related genomic, proteomic, and imaging data are stored in the CRDC's Genomic Data Commons, Proteomic Data Commons, and Imaging Data Commons, respectively [22,65]. Other databases include the Clinical Trial Data Commons, Cancer Data Service, and

Integrated Canine Data Commons [22,65]. Each data commons has their own portal where researchers can explore different projects with available data; filter for specific cancers, file formats, or metadata characteristics; define their own research cohorts; or download a manifest summarizing all available files for a given cohort.

The CRDC 'Cloud Resources' are cloud-based computing environments where investigators can ingest, process, and analyse big data from the cloud for their research [22]. The three available CRDC Cloud Resource platforms include (1) Broad Institute FireCloud, (2) Institute for Systems Biology Cancer Gateway in the Cloud, and (3) Seven Bridges Cancer Genomics Cloud [22]. Both the Broad Institute FireCloud and Institute for Systems Biology platforms currently run on the Google Cloud Platform, while the Seven Bridges platform runs on Amazon Web Services [22]. Each platform provides interactive computing resources (i.e. Jupyter Notebooks), along with their own suite of tools geared to users with different levels of technical and scientific expertise [22,65]. The CRDC Cloud Resources also provide hundreds of preconfigured analytics workflows and bioinformatics pipelines, along with the option to create new pipelines using the Common Workflow Language or Workflow Description Language [22,65,66].

Cancer investigators leveraging the CRDC for research would typically follow a few common steps. Firstly, users would search all available data commons to identify a collection of datasets most relevant for their research. They would then use one of the Cloud Resources to load those datasets into their cloud environment, import needed software tools and libraries, and then perform their planned analyses. As with any research study, users will need to complete required training; gain appropriate review board approval; and demonstrate awareness and respect for all data sharing, access, and privacy policies and procedures [68].

20.7.2. Precision medicine and the NIH 'All of Us' Program

Biomedical informatics and cloud computing have begun to transform the current practice of medicine from a 'one-size-fits-all' approach to a 'precision medicine' model that personalizes care according to the specific needs of patients [69,70]. The National Institutes of Health 'All of Us' Research Program is one example of a growing trend of using big data, informatics, and cloud technologies to accelerate precision medicine research. This initiative is recruiting a diverse population of at least one million participants throughout the United States, particularly from communities historically underrepresented in biomedical research [69,70]. From these participants, the All of Us program collects medical data from EHRs, genomic data from whole genome sequencing, as well as social, environmental, and behavioural data from mobile health and wearable technologies [69,70]. In addition, informatics approaches are being leveraged to harmonize all data into a single repository accessible through the National Institutes of Health (NIH) All of Us Researcher Workbench, a cloud-based platform available to researchers with appropriate training. Ultimately, this initiative is expected to help researchers discover highly effective and personalized approaches for the diagnosis, treatment, and prevention of a broad range of diseases [69,70]. While still in its early stages, the NIH All of Us program has already demonstrated potential to positively impact research for precision medicine, precision oncology, and for patients with cancer [14,15,33,71,72].

20.8. Conclusions

Biomedical 'big data' is capable of accelerating cancer research through large studies of real-world evidence increasingly available through diverse cancer data ecosystems. Cloud computing can serve as the technical infrastructure for the storage and computation of big data. Further, biomedical informatics provides the foundation for translating raw biomedical data into new research discoveries about cancer. The robust and responsible application of informatics and big data will help transform the potential of precision medicine into a widespread practical reality for patients with cancer.

REFERENCES

1. Brennan PF, Chiang MF, Ohno-Machado L. Biomedical informatics and data science: evolving fields with significant overlap. J Am Med Inf Assoc. 2018;**25**(1):2–3.
2. Kulikowski CA, Shortliffe EH, Currie LM, Elkin PL, Hunter LE, Johnson TR, et al. AMIA Board white paper: definition of biomedical informatics and specification of core competencies for graduate education in the discipline. J Am Med Inf Assoc [Internet]. 2012;**19**(6):931–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22683918
3. Fridsma DB. Health informatics: a required skill for 21st century clinicians. BMJ [Internet]. 2018;**362**:k3043. Available from: http://www.bmj.com/lookup/doi/10.1136/bmj.k3043
4. Banerjee R, George P, Priebe C, Alper E. Medical student awareness of and interest in clinical informatics. J Am Med Inf Assoc. 2015;**22**(e1):e42–7.
5. Detmer DE, Munger BS, Lehmann CU. Clinical informatics board certification: history current status and predicted impact on the clinical informatics workforce. Appl Clin Inf. 2010;**1**(1):11–8.
6. American Board of Preventive Medicine. Clinical informatics [Internet]. 2022 [cited 2022 Aug 8]. Available from: https://www.theabpm.org/become-certified/subspecialties/clinical-informatics/
7. Gadd CS, Williamson JJ, Steen EB, Andriole KP, Delaney C, Gumpper K, et al. Eligibility requirements for advanced health informatics certification. J Am Med Inf Assoc. 2016;**23**(4):851–4.
8. Gadd CS, Steen EB, Caro CM, Greenberg S, Williamson JJ, Fridsma DB. Domains, tasks, and knowledge for health informatics practice: results of a practice analysis. J Am Med Inf Assoc. 2020;**27**(6):845–52.
9. Silverman HD, Steen EB, Carpenito JN, Ondrula CJ, Williamson JJ, Fridsma DB. Domains, tasks, and knowledge for clinical informatics subspecialty practice: results of a practice analysis. J Am Med Inf Assoc [Internet]. 2019;**26**(7):586–93. Available from: https://academic.oup.com/jamia/advance-article/doi/10.1093/jamia/ocz051/5481062
10. Gardner RM, Overhage JM, Steen EB, Munger BS, Holmes JH, Williamson JJ, et al. Core content for the subspecialty of clinical informatics. J Am Med Inf Assoc. 2009;**16**(2):153–7.
11. American Medical Informatics Association. AMIA Health Informatics Certification (AHIC) [Internet]. American Medical Informatics Association: Careers and Certifications. 2022 [cited 2022 Sep 9]. Available from: https://amia.org/careers-certifications/amia-health-informatics-certification-ahic
12. Fridsma DB. Strengthening our profession by defining clinical and health informatics practice. J Am Med Inf Assoc. 2019;**26**(7):585.

13. Meyer MA. Healthcare data scientist qualifications, skills, and job focus: a content analysis of job postings. J Am Med Inf Assoc. 2019;**26**(5):383–91.
14. Ronquillo JG, Lester WT. Pharmacogenomic testing and prescribing patterns for patients with cancer in a large national precision medicine cohort. J Med Genet. 2021;**60**(1):81–3.
15. Na J, Zong N, Wang C, Midthun DE, Luo Y, Yang P, et al. Characterizing phenotypic abnormalities associated with high-risk individuals developing lung cancer using electronic health records from the All of Us researcher workbench. J Am Med Inf Assoc. 2021;**28**(11):2313–24.
16. Beck JT, Rammage M, Jackson GP, Preininger AM, Dankwa-Mullan I, Roebuck MC, et al. Artificial intelligence tool for optimizing eligibility screening for clinical trials in a large community cancer center. JCO Clin Cancer Inf. 2020;(4):50–9.
17. Lerman MH, Holmes B, St Hilaire D, Tran M, Rioth M, Subramanian V, et al. Validation of a mortality composite score in the real-world setting: overcoming source-specific disparities and biases. JCO Clin Cancer Inf. 2021;(5):401–13.
18. Hong JC, Fairchild AT, Tanksley JP, Palta M, Tenenbaum JD. Natural language processing for abstraction of cancer treatment toxicities: accuracy versus human experts. JAMIA Open. 2020;**3**(4):513–17.
19. Wei WQ, Teixeira PL, Mo H, Cronin RM, Warner JL, Denny JC. Combining billing codes, clinical notes, and medications from electronic health records provides superior phenotyping performance. J Am Med Inf Assoc. 2016;**23**(e1):e20–7.
20. Chute CG, Ullman-Cullere M, Wood GM, Lin SM, He M, Pathak J. Some experiences and opportunities for big data in translational research. Genet Med [Internet]. 2013 Sep 5;**15**(10):802–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24008998
21. Ohno-Machado L. NIH's big data to knowledge initiative and the advancement of biomedical informatics. J Am Med Inf Assoc [Internet]. 2014 Feb 8;**21**(2):193. Available from: http://jamia.bmj.com/cgi/doi/10.1136/amiajnl-2014-002666
22. Hinkson IV., Davidsen TM, Klemm JD, Kerlavage AR, Kibbe WA. A comprehensive infrastructure for big data in cancer research: Accelerating cancer research and precision medicine. Front Cell Dev Biol. 2017;**5**(83):1–7.
23. Rozenblatt-Rosen O, Regev A, Oberdoerffer P, Nawy T, Hupalowska A, Rood JE, et al. The human tumor atlas network: charting tumor transitions across space and time at single-cell resolution. Cell. 2020 Apr 16;**181**(2):236–49.
24. Kass-Hout TA, Xu Z, Mohebbi M, Nelsen H, Baker A, Levine J, et al. OpenFDA: an innovative platform providing access to a wealth of FDA's publicly available data. J Am Med Inf Assoc. 2015;**23**(3):596–600.
25. Bartlett VL, Dhruva SS, Shah ND, Ryan P, Ross JS. Feasibility of using real-world data to replicate clinical trial evidence. JAMA Netw Open. 2019;**2**(10):1–11.
26. Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA. 2018;**320**(9):867–8.
27. Miksad RA, Abernethy AP. Harnessing the power of real-world evidence (RWE): a checklist to ensure regulatory-grade data quality. Clin Pharmacol Ther. 2018;**103**(2):202–5.
28. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence—What is it and what can it tell us? New Engl J Med. 2016;**375**(23):2293–7.
29. Miksad RA, Samant MK, Sarkar S, Abernethy AP. Small but mighty: the use of real-world evidence to inform precision medicine. Clin Pharmacol Ther. 2019;**106**(1):87–90.
30. Wang W, Krishnan E. Big data and clinicians: a review on the state of the science. JMIR Med Inf [Internet]. 2014;**2**(1):e1. Available from: https://medinform.jmir.org/2014/1/e1/
31. Ni Y, Wright J, Perentesis J, Lingren T, Deleger L, Kaiser M, et al. Increasing the efficiency of trial-patient matching: automated clinical trial eligibility pre-screening for pediatric oncology patients. BMC Med Inf Decis Mak. 2015;**15**(28):1–10.
32. Cronin RM, Halvorson AE, Springer C, Feng X, Sulieman L, Loperena-Cortes R, et al. Comparison of family health history in surveys vs electronic health record data mapped to the observational medical outcomes partnership data model in the All of Us Research Program. J Am Med Inf Assoc. 2021;**28**(4):695–703.
33. Ramirez AH, Sulieman L, Schlueter DJ, Philippakis A, Roen DM. The All of Us Research Program: data quality, utility, and diversity. Patterns (NY). 2022;**3**(8):100570.
34. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;**3**:160018.
35. Carey VJ, Ramos M, Stubbs BJ, Gopaulakrishnan S, Oh S, Turaga N, et al. Global alliance for genomics and health meets bioconductor: toward reproducible and agile cancer genomics at cloud scale. JCO Clin Cancer Inf. 2020;**4**:472–9.
36. Tenenbaum JD, Sansone S-A, Haendel M. A sea of standards for omics data: sink or swim? J Am Med Inf Assoc [Internet]. 2014;**21**(2):200–3. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3932466&tool=pmcentrez&rendertype=abstract
37. Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. New Engl J Med. 2018;**379**(15):1452–62.
38. Kehl KL, Xu W, Lepisto E, Elmarakeby H, Hassett MJ, Van Allen EM, et al. Natural language processing to ascertain cancer outcomes from medical oncologist notes. JCO Clin Cancer Inf. 2020;(4):680–90.
39. Tu SW, Peleg M, Carini S, Bobak M, Ross J, Rubin D, et al. A practical method for transforming free-text eligibility criteria into computable criteria. J Biomed Inf. 2011;**44**(2):239–50.
40. Yim WW, Yetisgen M, Harris WP, Sharon WK. Natural language processing in oncology review. JAMA Oncol. 2016 Jun 1;**2**(6):797–804.
41. Deckard J, McDonald CJ, Vreeman DJ. Supporting interoperability of genetic data with LOINC. J Am Med Inf Assoc. 2015;**22**(3):621–7.
42. Klann JG, Joss MAH, Embree K, Murphy SN. Data model harmonization for the All Of Us Research Program: transforming i2b2 data into the OMOP common data model. PLoS ONE. 2019;**14**(2):e0212463.
43. Vrbin CM. Parametric or nonparametric statistical tests: considerations when choosing the most appropriate option for your data. Cytopathology. 2022;**33**(6):663–7.
44. Van Buren E, Herring AH. To be parametric or non-parametric, that is the question: parametric and non-parametric statistical tests. BJOG. 2020;**127**(5):549–50.
45. Bensken WP, Pieracci FM, Ho VP. Basic introduction to statistics in medicine, part 1: describing data. Surg Infect. 2021;**22**(6):590–6.
46. Ernecoff NC, Wessell KL, Hanson LC, Lee AM, Shea CM, Dusetzina SB, et al. Electronic health record phenotypes for

identifying patients with late-stage disease: a method for research and clinical application. J Gen Intern Med. 2019;**34**(12):2818–23.
47. Jouhet V, Mougin F, Bréchat B, Thiessard F. Building a model for disease classification integration in oncology, an approach based on the national cancer institute thesaurus. J Biomed Semant. 2017;**8**(1):6.
48. Griffith SD, Miksad RA, Calkins G, You P, Lipitz NG, Bourla AB, et al. Characterizing the feasibility and performance of real-world tumor progression end points and their association with overall survival in a large advanced non–small-cell lung cancer data set. JCO Clin Cancer Inf. 2019;(3):1–13.
49. Wang P, Garza M, Zozus M. Cancer phenotype development: a literature review. Stud Heal Technol Inf. 2019;**257**:468–72.
50. Xu H, Fu Z, Shah A, Chen Y, Peterson NB, Chen Q, et al. Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases. AMIA Annu Symp Proc. 2011;**2011**:1564–72.
51. Stein LD. The case for cloud computing in genome informatics. Genome Biol [Internet]. 2010;**11**(5):207. Available from: http://genomebiology.com/2010/11/5/207
52. Krissaane I, De Niz C, Gutiérrez-Sacristán A, Korodi G, Ede N, Kumar R, et al. Scalability and cost-effectiveness analysis of whole genome-wide association studies on Google Cloud Platform and Amazon Web Services. J Am Med Inf Assoc. 2020;**27**(July):1425–30.
53. Ronquillo JG, Lester WT. Practical aspects of implementing and applying health care cloud computing services and informatics to cancer clinical trial data. JCO Clin Cancer Inf. 2021;**5**:826–32.
54. Ronquillo J. An introduction to cloud computing for cancer research [Internet]. Cancer Data Science Pulse. 2022 [cited 2022 Aug 19]. Available from: https://datascience.cancer.gov/news-events/blog/introduction-cloud-computing-cancer-research
55. Krumm N, Hoffman N. Practical estimation of cloud storage costs for clinical genomic data. Pract Lab Med. 2020;**21**:e00168.
56. Yazar S, Gooden GEC, Mackey DA, Hewitt AW. Benchmarking undedicated cloud computing providers for analysis of genomic datasets. PLoS ONE. 2014 Sep 23;**9**(9):e108490.
57. Miller TA, Avillach P, Mandl KD. Experiences implementing scalable, containerized, cloud-based NLP for extracting biobank participant phenotypes at scale. JAMIA Open. 2020;**3**(2):185–9.
58. Google Cloud. Healthcare Natural Language API pricing [Internet]. 2020 [cited 2022 Sep 23]. Available from: https://cloud.google.com/healthcare/healthcare-natural-language-api-pricing
59. National Cancer Institute. National Cancer Institute Genomic Data Commons data portal [Internet]. 2022 [cited 2022 Sep 5]. Available from: https://portal.gdc.cancer.gov/
60. Zeng Y, Zhang J. A machine learning model for detecting invasive ductal carcinoma with Google Cloud AutoML Vision. Comput Biol Med. 2020;**122**:103861.
61. Sharpless NE, Kerlavage AR. The potential of AI in cancer care and research. Biochim Biophys Acta Rev Cancer [Internet]. 2021;**1876**(1):188573. Available from: https://doi.org/10.1016/j.bbcan.2021.188573
62. Zhu L, Zheng WJ. Informatics, data science, and artificial intelligence. JAMA [Internet]. 2018;**320**(11):1103–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28117445
63. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. New Engl J Med [Internet]. 2016;**375**(13):1216–9. Available from: http://www.nejm.org/doi/10.1056/NEJMp1609300
64. Bai J, Jhaney I, Wells J. Developing a reproducible microbiome data analysis pipeline using the amazon web services cloud for a cancer research group: proof-of-concept study. J Med Internet Res. 2019;**7**(4):e14667.
65. Grossman RL. Progress toward cancer data ecosystems. Cancer J. 2018;**24**(3):126–30.
66. Grossman RL. Data lakes, clouds, and commons: a review of platforms for analyzing and sharing genomic data. Trends Genet [Internet]. 2019;**35**(3):223–34. Available from: https://doi.org/10.1016/j.tig.2018.12.006
67. National Cancer Institute. NCI Cancer Research Data Commons [Internet]. Center for Biomedical Informatics and Information Technology. 2022 [cited 2022 Sep 20]. Available from: https://datascience.cancer.gov/data-commons
68. Zayas-Caban T, Chaney KJ, Rogers CC, Denny JC, White PJ. Meeting the challenge: Health information technology's essential role in achieving precision medicine. J Am Med Inf Assoc. 2021;**28**(6):1345–52.
69. The All of Us Research Program Investigators. The "All of Us" Research Program. New Engl J Med. 2019;**381**(7):668–76.
70. National Institutes of Health. All of Us Research Program. [Internet]. 2022 [cited 2022 Sep 5]. Available from: https://allofus.nih.gov/
71. Ronquillo JG, Lester WT. Precision medicine landscape of genomic testing for patients with cancer in the National Institutes of Health All of Us database using informatics approaches. JCO Clin Cancer Inf [Internet]. 2022;**6**:e2100152. Available from: https://ascopubs.org/doi/full/10.1200/CCI.21.00152
72. Coughlin SS, Chen J, Cortes JE. Health care access and utilization among adult cancer survivors: Results from the National Institutes of Health 'All of Us" Research Program. Cancer Med. 2021;**10**(11):3646–54.

21

Single-cell sequencing analysis focused on cancer immunotherapy

Luciane T. Kagohara and Joseph Tandurella

21.1. Introduction

The development of single-cell sequencing technologies leveraged the previous knowledge of tumour heterogeneity and provided tools to better determine the cell types contributing to the intermixed signals in the tumour microenvironment (TME) [1]. The single-cell resolution analysis of the TME had a major impact on tumour immunology studies, as now it is possible to obtain the transcriptional profile of individual immune types without the sorting step to isolate the multiple cell types and profile each population individually [2]. Single-cell RNA sequencing (scRNA-seq) is the major approach to understand tumour immune responses as it allows cell-type annotations. With that, it is possible to examine the heterogeneous cell composition and also the molecular features of the immune responses against the cancer cells in the absence or presence of immunotherapies. Thus, scRNA-seq became a powerful tool to profile the mechanisms of response and resistance to immune checkpoint inhibitors (ICIs) that revolutionized cancer therapy by providing clinical benefits to previously untreatable cancers, but it is still not efficacious against the majority of cases [3,4].

Transcriptomics analysis with single-cell resolution, due to their genome-wide capability, generates high-dimensional datasets that enable a wide range of analysis to determine cell types, molecular perturbations, and cell-fate inferences [2]. Nevertheless, these analyses require robust computational approaches that account for the complexity of scRNA-seq datasets to deliver biologically true results that could potentially lead to the identification of new targetable genes or pathways to develop new therapeutic interventions or to overcome therapeutic resistance [5].

This chapter briefly describes the computational analysis of scRNA-seq data with a focus on understanding immunotherapies. However, the steps described here are applicable to any scRNA-seq dataset. The overall aim is to highlight benchmarked good practices when handling such precious and complex datasets to properly unveil new biological information to improve tumour immunology comprehension.

21.2. Single-cell RNA-sequencing approaches to examine immunotherapy responses

Single-cell gene expression analysis approaches deliver genome-wide RNA expression for each individual cell captured during experimental data generation [1]. The computational analysis of scRNA-seq allows a wide range of analyses that include cell composition characterization, gene expression and pathway analysis, cell-state inference analysis, and intercellular interaction prediction [6–9]. A variety of experimental approaches have been developed for scRNA-seq. These technologies differ on the method of cell capture, cell barcoding, and sequencing library generation [10–15]. All the approaches ultimately allow cellular composition examination and the downstream molecular investigation of biological perturbations.

An increasing number of studies have used scRNA-seq to understand TME heterogeneity and tumour immune responses [16–29]. To better profile tumour immune responses, an experimental strategy is applied to isolate the tumour infiltrating leukocytes (TILs), perform scRNA-seq only in the immune cells present in the tumours, identify immune effector and suppressive cells, and examine their transcriptional profiles to better understand tumour immune responses and mechanisms of resistance to ICIs [16,21,23,30–34]. The application of scRNA-seq to understand immune responses against cancer has driven the discovery of cell-state transitions conserved across response and resistance to ICIs and provided new insights into the development of new therapeutic interventions [5,35]. The power of scRNA-seq to study tumour immune responses and their modulation by immunotherapeutic agents resides in its ability to provide single-cell gene expression profiles of virtually all immune populations that can be recovered from a tumour sample. With that, it is possible to identify the immune types that are changing as a response to immunotherapies from those that remain unaltered. Then, within the group of cells that are modulated by ICIs, scientists and clinicians can uncover the molecular pathways that are relevant to response or resistance in an attempt to identify new targets to increase treatment efficacy.

The next sections of this chapter focus on the current available computational methods to obtain robust and reproducible results from scRNA-seq. Those methods are not only applicable to examine immune cells infiltrating the tumour but can also be applied to any scRNA-seq dataset.

21.3. Computational analysis of scRNA-seq

To fully explore scRNA-seq data and examine the impact of heterogeneity and transcriptional features on cancer biology, computational analysis is a major step on handling these high-dimensional datasets. The analysis starts with scRNA-seq data pre-processing that involves alignment, sequencing quality check, and initial filtering. Next, the pre-processed data will pass through additional quality check steps and normalization. Then, the data will be corrected for clustering and visualization. Finally, the scRNA-seq data will be read from downstream analysis: differential expression analysis, pathways analysis, cell-fate inferences, and interactions and network analysis (Figure 21.1). For all the steps, there are a multitude of open-source software available as R/Bioconductor or Python packages and pipelines that have been peer reviewed and that usually offer detailed tutorials to facilitate community accessibility. Pipelines in R, using Seurat [36–40] and Monocle [7,8], and in Python, with Scanpy [41], have been extensively applied to analyse scRNA-seq data, and the sections in this chapter are fully implemented on those tools.

21.3.1. Pre-processing: alignment and count generation

The pre-processing of the scRNA-seq raw data involves the alignment of the sequences to a reference genome and the generation of

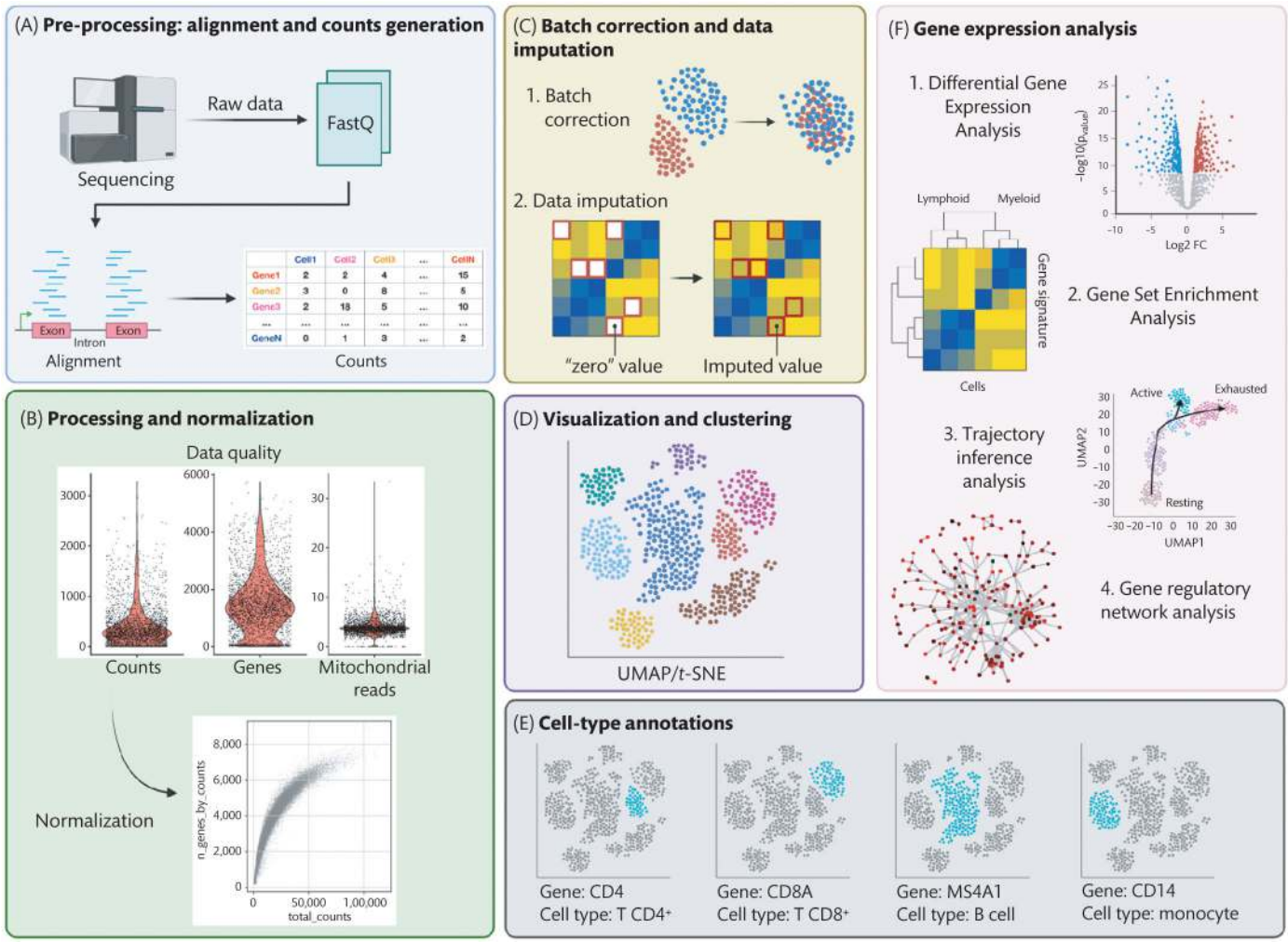


Figure 21.1. Single-cell RNA-sequencing computational analysis workflow. (A) Pre-processing of sequencing raw data (FASTQ files) involves read alignment to a reference genome to generate count matrix of transcript abundances for each cell in the dataset. (B) The quality of the scRNA-seq data is checked for the number of counts, genes, and mitochondrial content per cell; the filtered good quality data is normalized to correct for technical noise. (C) The pre-processed scRNA-seq data is examined for batch effect identification and correction (1) and the sparsity is corrected with imputation (2). (D) The normalized data underdoes dimensionality reduction for visualization and clustering. (E) The clustered data is analysed to identify gene markers for each of the clusters to allow cell-type annotations for downstream analysis. (F) Gene expression analysis is performed to identify differentially expressed genes (1) and pathways enriched (2) under biological conditions, predict cell-state transitions using trajectory inference analysis (3), and identify gene regulatory networks that could be activated as a result of cellular interactions. *Source:* Created with BioRender.com

transcript abundances (counts) for each cell (barcode) sequenced. This pre-processing converts the raw sequencing files (FASTQS) into transcript abundances for each cell captured from the sample and is a critical step in scRNA-seq analysis as the downstream computational pipelines depend on these outputs (Figure 21.1A).

Different open-source software for alignment of single-cell sequencing data are available. The most commonly applied are Kallisto-Bustools [42], Cell Ranger [15], STARsolo [43], and Alevin [44]. In general, these alignment tools

1. read the provided FASTQ files;
2. map the reads to the reference transcriptome chosen, using either classical alignment or pseudo-alignment modalities;
3. adjust for sequencing errors in barcodes that could be introduced during the library preparation;
4. perform unbiased cell identification using restrictive unique molecular identifier mismatch allowances;
5. and assign read counts to the respective genes (features) and cells barcodes.

The output will invariably be a matrix of feature (gene) counts for each barcode or unique cell captured.

Subsequent to alignment and count estimation, the next step is to implement a computational pipeline for data quality check, filtering, normalization, dimensionality reduction, and gene expression analysis. There is a broad range of computational tools for those analyses from more user-friendly automated tools, such as Loupe Browser (10x Genomics), to less automated and more bioinformatics heavy tools that are available as free, open-source software. There is no gold-standard analytical tool for the scRNA-seq analysis. Bioinformaticians in general utilize a software or platform they are more trained on as long as those computational tools incorporate mathematical and statistical methods that make analysis and findings more robust and reproducible. Some examples of software for scRNA-seq analyses are R/Bioconductor packages (Seurat [36–39] and Monocle3 [8,45,46]) and Python packages (Scanpy [41]). These software have been peer reviewed, have detailed publicly available tutorials, and are broadly used by single-cell bioinformatics. Other open-source computational tools are available and have been examined regarding their performance and can be applied for the analysis that is discussed in this chapter [47].

21.3.2. Processing and normalization

Data processing and filtering involves a series of steps to generate a final dataset with transcripts captured from high-quality single cells (Figure 21.1B). Most pipelines have a default set of parameters in place, but for best practices, it is ideal to test different parameters and identify those that work best for the data in hand and hypotheses to be answered.

Verification for sequencing quality is performed by examining the number of counts and features (genes) per cell. The sequencing is usually considered low quality when the number of detected genes per cell is below 200 genes per cell. However, the number of genes expressed might vary depending on the cell type, and the cut-off should be established taking into account the tissue, cell type, and function. Although high number of genes per cell is a sign of good sequencing data, when that number surpasses a few thousand genes, it might suggest that for that specific barcode more than one cell got captured and sequenced. Doublets, as they are called, are a common finding in scRNA-seq datasets and filtered out as it is impossible to separate two or more cells that were labelled with the same barcode. Finally, other parameters to check to filter out cells with no biological relevance to the analysis are those with high mitochondrial gene counts as these reads will frequently be a confounding effect during the downstream gene expression analysis. To determine the ideal maximum percentage of mitochondrial genes present in each cell is very subjective, but it typically ranges from 5% to 20%. Nevertheless, here is another parameter that should take into account the types of cells that are sequenced. Tissues with higher activity, like brain, have higher mitochondrial activity than others, and consequently, the percent of mitochondrial genes will be elevated even among viable cells. This stage of the analysis is probably the most important and should be completed carefully because any misstep could introduce non-biological signal to the data that would interfere with the final interpretation of the results due to the introduction of downstream statistical biases in the following steps of the analysis [48]. The recommendation is that different parameters are tested in each of the pipeline steps to understand the data and adjust and apply the correct cut-offs and thresholds, even if that means going back and forth to the initial step multiple times.

Subsequently, data normalization must be performed to address any bias or noise from technical artefacts potentially introduced during the experimental steps of sequencing library preparation (Figure 21.1B). Unwanted bias and noise can be a result of low sample input or quality, and technical artefacts from amplification during sequencing [48–50]. Thus, to ensure these non-biological variables will not impede proper identification and interpretation of the biological findings, data normalization is crucial to identify and correct the potential unwanted biases that would confound biological interpretation. There are different open-source methods available for scRNA-seq data normalization. Another correction applied by most of these computational tools is gene expression adjustment based on sequence depth in each cell to ensure that gene expression measurement will not be influenced by lower or higher sequencing depth. Some of the computational approaches that have been developed to identify and remove noise from the data include Linnorm (a linear model and normality-based normalizing transformation method for scRNA-seq data) [51], SCnorm (robust normalization of single-cell RNA-seq data) [52], scran [53], and SCTransform [49].

An experimental option to increase robustness of sample processing and normalization is to include engineered spike-in molecules added to each lysate specifically for use in proper normalization as the level of expression should be fairly consistent across cells [54].

21.3.3. Batch correction and data imputation

Experimental variations are commonly detected on next-generation sequencing datasets, and scRNA-seq datasets are not an exception. The variations in the data that introduce unwanted bias are known as batch effects. Batch effects can result from experimental inconsistencies that could happen from sample collection, processing and manipulation, until the final sequencing step. Ideally, protocols should be designed to reduce batch effects and ensuring consistent procedures for sample handling, pipetting, and reagents usage. However, some factors cannot be controlled and thus computational analysis for batch effects identification and correction is a mandatory step for proper single-cell analysis [55,56] (Figure 21.1C).

The data normalization does not remove batch effects, and correction is another essential step in the analysis. Discrepancies in features abundances and variations that are not biologically correlated are fairly common in single-cell data. As mentioned previously, this has been seen with next-generation sequencing technologies, and the single-cell resolution only exacerbates batch effects [57,58]. To correct unwanted batch effects in scRNA-seq data, many algorithms have been developed to account for these discrepancies in the low-dimensional embeddings used during the visualization. Alternatively, manual correction can be performed by preserving biological variation in correlation structures and incorporating batch as a covariate in downstream analytical steps. The presence of batch effects in the data will dramatically impact the downstream gene expression analysis and confound the interpretation of the real biological findings [55,56].

In addition to batch effects, another future on scRNA-seq datasets that can impact the analysis and data interpretation is the high frequency of zeros. The zeros can be interpreted in two ways: true biology (lack of gene expression) or technical dropouts (transcript that is not captured during library preparation). Technical zeros can be introduced when transcripts are rare or when sequencing depth is not adequate [59]. To distinguish biological and technical zeros and adjust the data for the last step called data imputation is performed (**Figure 21.1**C). Overall, data imputation is an additional method used to remove technical noise, more specifically the increased sparsity (high frequency of zeros) present in single-cell data [60]. During data imputation, the sparsity of the data is examined to estimate the real gene expression for all genes, including those with very low abundances. Good computational methods for data imputation should not interfere with the downstream gene expression analysis. There are various imputation methods to correct data sparsity, which can be summarized into three main approaches: (1) those that apply probabilistic models, (2) those that consider the diffusion of gene expression using 'neighbouring' cell profiles, (3) and those that use latent space/deep-learning models of cell representation. An example of a highly utilized imputation tools are the SAVER and Magic due to their robustness [60,61], created specifically for use in single-cell and bulk RNA sequencing. Although data imputation is important for adequate downstream analysis, it is still controversial as it can introduce false positives in the data due to data heterogeneity associated with cell number differences across biological conditions. Another issue from data imputation considers expression differences that when small can be lost during imputation. Nevertheless, comparisons of scRNA-seq data analysis with and without imputation demonstrate that, even if not perfect, data imputation methods improve the robustness of the results [60,61].

21.3.4. Visualization and clustering

The normalized scRNA-seq data is represented by high-dimensional matrices of transcripts abundances for each cell captured. The high dimensionality imposes challenges to the application of standard visualization methods as they mostly rely on 2D or 3D graphs. Thus, dimensionality reduction is performed to allow visualization and downstream analysis (**Figure 21.1**D). Most biological processes (cell types, state transitions, etc.) do not occur at the single feature (gene) level but rather encompass measurements of the correlations across multiple features, allowing the application of dimensionality reduction for scRNA-seq [62–64].

Currently, the most frequently used computational methods for data dimension reduction are the *t*-distributed stochastic neighbour embedding (*t*-SNE) [65] and the uniform manifold approximation and embedding (UMAP) [66]. The resulting plots from both computational methods are a representation of each individual cell in the data with the preserved distances among cells within the axes corresponding to the distances separating them molecularly—meaning that the distance between the cells in the plot represents their similarities. The more similar the cells' gene expression profiles are, the closer they will be represented on the plot. As previously mentioned, the transition from high dimension to lower dimensions is able to retain the genetic profile while minimizing the number of components to plot. The UMAP and *t*-SNE methods work in similar ways. However, while UMAP balances the global structure of points, *t*-SNE allows adjustments to preserve the distance of nearby points with more distant clusters. Typically, a single embedding is used at a time to keep the consistency across analysis and comparisons. Following dimensional reduction, the cell groups (clusters) can be coloured based on their expression profiles or cell-type annotations [65,66].

In summary, dimensionality reduction is applied for visualization and as a tool to examine cells' gene expression variations and similarities, but additional approaches are required for biological inferences in interpretation of the visualized data.

21.3.4.1. Cell-type annotations

Cell-type annotations from scRNA-seq data (**Figure 21.1**E) are the initial step to determine cell proportions and compare changes across biological conditions (e.g. treated vs. untreated; therapeutic response vs. resistance; and long vs. short disease-free survival) as those are usually performed within specific cell types. Clustering is a suitable approach to annotate cell types in scRNA-seq. Gene markers that are specifically expressed by groups of cells are used to identify cell clusters. The number of cell types and subtypes identified in the data depends on the number of clusters obtained and that is tightly correlated with the number of dimensions (resolution). The resolution will determine the granularity of cell-type delineation, which can be taken into account using ensemble-based clustering [67–69]. Thus, a common practice is to test clustering with multiple numbers of dimensions and examine the gene markers identified for each cluster and with that attempt cell-type annotation. Combined with the gene markers, it is possible to apply gating strategies similar to that applied for flow cytometry. This combined approach allows the identification of major cell types (e.g. B cells, T cells, and myeloid cells) using a gating strategy, while the gene markers from the clustering will provide more granularity and information regarding the molecular heterogeneity of the cells captured. For cell clustering, both *t*-SNE and UMAP are suited to provide clustering for the identification of distinct cell types. Many of the clustering algorithms used in single-cell analysis use social network tools to identify groups of cells with similar expression levels [70,71]. Once identified, these clusters contain unique profiles of expressed genes that will serve as markers to annotate the cell-type identifier of the cluster. Sub-clustering within clusters may be additionally completed to delineate specific cell subtypes. A prime example of this would be the clustering of the tumour infiltrating immune cells in a sample, where initial clustering will separate the immune lineages (lymphocytes, monocytes, and granulocytes), but sub-clustering

within the lymphocytes will allow for the annotation of $CD4^+$ and $CD8^+$ subtypes. This is the standard practice in cell-type annotation, in which clustering and sub-clustering are tools that support the identification of the groups of cells with similar transcriptional profiles and states.

The manual annotation of cell types is a strenuous process that requires knowledge of cell-specific markers. Most frequently, the described cell-type markers are defined based on protein markers adding a new challenge to the annotation process as RNA abundance of some markers is not perfectly correlated with protein abundance with genes that are not translated into proteins that still present basal levels detected in the scRNA-seq data. To facilitate annotation, several methods, like Azimuth [38], have been developed to automatically infer the identity of individual cells or clusters with the use of public domain reference cell databases or atlases, such as the Human Tumor Atlas Network [72]. The automated tools will correlate the cell-type expression profiles in the reference data with the profiles in the target data. This approach is considered to be more robust as the results can easily be reproduced across studies, but requires a comprehensive and laborious back end where these cell types were previously annotated at the desired level of granularity in the reference dataset. The automated approach will not be adequate if the desired cell types are not present in the reference or if the granularity expected is not present. In the case that a cluster of cells cannot be annotated using this method, manual annotation will be required. It should be noted that different tissues have variable expression, so one should ideally use a reference of the same tissue within the model organism. Additionally, ideal annotation using a signature-based method will include a brief manual validation of the annotations.

21.3.5. Differential gene expression analysis and pathway analysis

Following cell-type classification, the analysis to examine the biological relevant perturbations can be performed with differential gene expression analysis (DGEA) followed by gene set enrichment analysis (GSEA) or pathway analysis (**Figure 21.1**F). The DGEA will compare the expression levels for each gene between the biological conditions in question within the cell type of interest. For the comparisons, multiple statistical tests are available with the most robust results obtained with negative binomial tests [47,49,73,74].

An alternative method to perform DGEA on scRNA-seq data is to perform pseudobulking. Pseudobulking refers to the aggregation of captured cells gene counts to create a bulk RNA-seq count matrix. This is particularly of use when there are confounding variables, such as different captured cell numbers across patients. To complete this, a sum aggregation of gene counts by cell types and patients will be completed followed by a typical bulk RNA-sequencing analysis protocol [75].

Standard GSEA or pathway analysis tools can then be applied to determine the molecular pathways that were altered based upon the results of these differential expression analyses [76,77]. While comparing gene-level differences due to treatment provides individual gene changes, exploring changes in biological processes is also possible with scRNA-seq datasets. In the case of treatment, gene-level changes might not be as drastic as expected; however, pathway-level changes may be present. During GSEA, changes in predetermined gene lists across the perturbations in the data are examined. The pathway analysis is performed upon DGEA completion and uses the statistics from the differential analysis to rank genes. Preselected sets of pathways or molecular signatures are required as input for the analysis that uses an enrichment score statistic, calculated from a vector of gene-level statistics [77]. This enrichment score is calculated for each pathway provided, where a multiple hypothesis correction is applied to generate adjusted *p*-values.

While DGEA followed by pathway analysis is extremely useful in identifying changes within a cell type, there are additional molecular changes that may not be limited to single-cell types but rather will result in state transitions across all immune cell types. To address this issue, non-negative matrix factorization approaches have been developed. These methods work to uncover overlapping, low-dimensional patterns of additive variations that occur simultaneously to multiple cell types in the data. These patterns, or sets of genes, can then undergo a pathway identification to better understand which genes are present in these patterns and their biological functions. Each pattern has the potential to represent a single biological process that encompasses multiple signalling pathways [78–80].

21.3.6. Trajectory inference analysis

One of the advantages of scRNA-seq analysis is the possibility of inferring cell-state trajectories based on cells RNA content (**Figure 21.1**F). This is possible because the cells, although captured at the same time, do not all present at the same state. Trajectory inference analysis methods allow for the visualization of cellular changes over time and space. The computational approaches to infer the trajectory of the single cells use the gene expression profiles to order the cells in a continuous trajectory that represents a pseudotime of the evolution of each cell during the development of a biological process [46,81,82]. During the analysis, a pseudotime score, that is representative of cells' position in the biological process trajectory, is assigned to each cell. With that, the pseudotime analysis allows the prediction of continuous changes in the gene expression dynamics to identify cell states across biological processes. In recent years, many trajectory inference methods have been developed; each differing in their underlying algorithms and expected topology profiles. Most of these methods work to order cells along an assumed topology (i.e. cyclic, linear, and bifurcating), while recent methods infer the topology of the trajectory [46,81,82].

The accuracy of the cell-fate predictions depends on the algorithm of choice and also on the features of the dataset. For example, TILs encompasses multiple cell types that are heterogeneous regarding their differentiation processes and state transitions. Thus, to obtain realistic cell trajectories, TILs are sub-clustered and the subtypes of interested are extracted from the dataset for the cell-fate inference analysis [Savas 2018]. Cell-fate or trajectory inference analysis is an important step to understand cell-state changes that are induced under specific conditions, such as treatment with ICIs, and can be useful to predict immune cell evolution during cancer evolution.

21.3.7. Gene regulatory network and cellular interaction analysis

Another pertinent information to understand tumour immune responses is to understand if the cells in the TME are interacting. For example, is tumour reduction a result of immune cell killing elicited

by treatment with ICIs? In scRNA-seq datasets, the possibility of isolating each cell allows cellular interaction analysis to infer which cell types are interacting and the molecular consequences of such relations.

While simultaneously analysing molecular changes, gene regulatory network (GRN) inference is an analysis tool that allows for the understanding of the interactions occurring between genes within and between cells to gain an additional perspective of the underlying molecular regulation (Figure 21.1F). GRN tools were originally created for bulk RNA data, but advances in single-cell resolution have allowed for these tools to look within and between cells rather than only quantifying the correlation between pairs of genes within bulk datasets [83]. Overall, GRN computational methods learn the gene networks in the data, providing a snapshot of the activated pathways in each cell. Then, it identifies pairs of genes (e.g. ligand–receptor pairs) that can modify the activity or expression of one another. With that, it is possible to identify the cell types expressing the ligands of interest and those expressing the matched receptor and examine the activation of the pathways between the cell types to infer the result of the potential cellular interaction [84,85].

Newer approaches have been developed to model the heterogeneity of single-cell data, such as extensions for time-course data or comparing changes across treatment arms [9,86–89]. For example, the utilization of temporal ordering by first applying trajectory inference methods allows for GRN inference methods to predict gene expression regulation through the use of temporal changes when each gene begins expression across time, predicting how genes regulate one another within each cell [90–92]. In addition to exploring temporal changes in gene regulation using GRN tools, many regulatory processes also occur between cells through the use of cell-to-cell contact and paracrine signalling. To estimate these interactions, additional GRN tools have been developed to look at co-expressing genes, further referred to as known ligand–receptor pairs, across cell types.

21.4. Conclusions

The recent and quick development of single-cell sequencing delivered a powerful tool to examine cancer biology while accounting for its complex and heterogeneous TME. Among the single-cell approaches, gene expression profile at the single-cell level is the most used as it allows for cell-type identification combined with the profile of the transcriptome. Thus, scRNA-seq provides valuable information to understand molecular perturbations and cell-state transitions. These capabilities are valuable features to understand how the cells infiltrating the tumour are modulated by immunotherapies and how they interact with the tumour cells. The experimental developments were followed by advancements in computational analysis. Using robust algorithms, it is possible to classify cell types, compare gene expression changes, and infer cells evolution and interactions. Therefore, using meticulous established experimental and computational pipelines, it is possible to obtain an informative dataset that will infer how immune cells in the tumour respond to ICIs and how it drives tumour killing or how the immune cells change from an active to an exhausted state that allows tumour regrowth.

Nevertheless, scRNA-seq requires careful experimental design as we demonstrated in this chapter that numerous confounding effects can be introduced in the data. Moreover, assistance from computational scientists is critical to support the experimental design (sample size and number of replicates) and analysis. Without proper selection and application of the computational algorithms based on the hypotheses to be addressed, the full potential of the data cannot be explored. New translatable discoveries from scRNA-seq data require a multi-disciplinary approach from sample handling, through the computation analysis, to the data interpretation. The analysis of scRNA-seq datasets to leverage our knowledge on immunotherapies is already a reality, but it is mostly a profiling approach to profile the TME in response to ICIs. To translate the findings into new immunotherapeutic interventions, the scRNA-seq findings must be mechanistically validated at the bench using preclinical models and gold-standard methodologies.

Future developments in single-cell approaches for multi-omic profiling of the same cells, in computational algorithms to address the complexity of single-cell datasets, and in the creation of multi-disciplinary teams of scientists are critical to unleash the full potential of single-cell analysis and are already a reality. Altogether, it will provide a deeper knowledge of tumour immune responses and its dynamics that will conclude with new therapeutic agents to trigger more efficient responses and increase the number of cancer patients who benefit from immunotherapies.

REFERENCES

1. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009 May;**6**(5):377–82.
2. Davis-Marcisak EF, Deshpande A, Stein-O'Brien GL, Ho WJ, Laheru D, Jaffee EM, et al. From bench to bedside: single-cell analysis for cancer immunotherapy. Cancer Cell. 2021 Aug 9;**39**(8):1062–80.
3. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science. 2018 Mar 23;**359**(6382):1350–5.
4. Giladi A, Amit I. Single-cell genomics: a stepping stone for future immunology discoveries. Cell. 2018 Jan 11;**172**(1–2):14–21.
5. Gohil SH, Iorgulescu JB, Braun DA, Keskin DB, Livak KJ. Applying high-dimensional single-cell technologies to the analysis of cancer immunotherapy. Nat Rev Clin Oncol. 2021 Apr;**18**(4):244–56.
6. Lim B, Lin Y, Navin N. Advancing cancer research and medicine with single-cell genomics. Cancer Cell. 2020 Apr 13;**37**(4):456–70.
7. Trapnell C. Defining cell types and states with single-cell genomics. Genome Res. 2015 Oct;**25**(10):1491–8.
8. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014 Apr;**32**(4):381–6.
9. Cherry C, Maestas DR, Han J, Andorko JI, Cahan P, Fertig EJ, et al. Computational reconstruction of the signalling networks surrounding implanted biomaterials from single-cell transcriptomics. Nat Biomed Eng. 2021 Oct;**5**(10):1228–38.
10. Ramsköld D, Luo S, Wang Y-C, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA

and individual circulating tumor cells. Nat Biotechnol. 2012 Aug;**30**(8):777–82.
11. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods. 2014 Feb;**11**(2):163–6.
12. Xin Y, Kim J, Ni M, Wei Y, Okamoto H, Lee J, et al. Use of the Fluidigm C1 platform for RNA sequencing of single mouse pancreatic islet cells. Proc Natl Acad Sci USA. 2016 Mar 22;**113**(12):3293–8.
13. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015 May 21;**161**(5):1187–201.
14. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell. 2015 May 21;**161**(5):1202–14.
15. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017 Jan 16;**8**:14049.
16. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. Cell. 2018 Aug 23;**174**(5):1293–1308.e36.
17. Ma L, Wang L, Chang C-W, Heinrich S, Dominguez D, Forgues M, et al. Single-cell atlas of tumor cell evolution in response to therapy in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. J Hepatol. 2021 Dec;**75**(6):1397–1408. doi: 10.1016/j.jhep.2021.06.028. Epub 2021 Jun 30. PMID: 3421672.
18. Bernard V, Semaan A, Huang J, San Lucas FA, Mulu FC, Stephens BM, et al. Single-cell transcriptomics of pancreatic cancer precursors demonstrates epithelial and microenvironmental heterogeneity as an early event in neoplastic progression. Clin Cancer Res. 2019 Apr 1;**25**(7):2194–205.
19. Kieffer Y, Hocine HR, Gentric G, Pelon F, Bernard C, Bourachot B, et al. Single-cell analysis reveals fibroblast clusters linked to immunotherapy resistance in cancer. Cancer Discov. 2020 Sep;**10**(9):1330–51.
20. Davidson S, Efremova M, Riedel A, Mahata B, Pramanik J, Huuhtanen J, et al. Single-cell RNA sequencing reveals a dynamic stromal niche that supports tumor growth. Cell Rep. 2020 May 19;**31**(7):107628.
21. Zheng C, Zheng L, Yoo J-K, Guo H, Zhang Y, Guo X, et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. Cell. 2017 Jun 15;**169**(7):1342–56.e16.
22. Peng J, Sun B-F, Chen C-Y, Zhou J-Y, Chen Y-S, Chen H, et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Res. 2019 Sep;**29**(9):725–38.
23. Guo X, Zhang Y, Zheng L, Zheng C, Song J, Zhang Q, et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. Nat Med. 2018 Jul;**24**(7):978–85.
24. Sun Y, Wu L, Zhong Y, Zhou K, Hou Y, Wang Z, et al. Single-cell landscape of the ecosystem in early-relapse hepatocellular carcinoma. Cell. 2021 Jan 21;**184**(2):404–21.e16.
25. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJL, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. Nat Genet. 2017 May;**49**(5):708–18.
26. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014 Jun 20;**344**(6190):1396–401.
27. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. Cell. 2017 Dec 14;**171**(7):1611–24.e24.
28. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016 Apr 8;**352**(6282):189–96.
29. Zhao J, Zhang S, Liu Y, He X, Qu M, Xu G, et al. Single-cell RNA sequencing reveals the heterogeneity of liver-resident immune cells in human. Cell Discov. 2020 Apr 28;**6**:22.
30. Savas P, Virassamy B, Ye C, Salim A, Mintoff CP, Caramia F, et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. Nat Med. 2018 Jul;**24**(7):986–93.
31. Yost KE, Satpathy AT, Wells DK, Qi Y, Wang C, Kageyama R, et al. Clonal replacement of tumor-specific T cells following PD-1 blockade. Nat Med. 2019 Aug;**25**(8):1251–9.
32. Gubin MM, Esaulova E, Ward JP, Malkova ON, Runci D, Wong P, et al. High-dimensional analysis delineates myeloid and lymphoid compartment remodeling during successful immune-checkpoint cancer therapy. Cell. 2018 Nov 1;**175**(4):1014–30.e19.
33. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su M-J, Melms JC, et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. Cell. 2018 Nov 1;**175**(4):984–97.e24.
34. Sade-Feldman M, Yizhak K, Bjorgaard SL, Ray JP, de Boer CG, Jenkins RW, et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. Cell. 2018 Nov 1;**175**(4):998–1013.e20.
35. Guruprasad P, Lee YG, Kim KH, Ruella M. The current landscape of single-cell transcriptomics for cancer immunotherapy. J Exp Med. 2021 Jan 4;**218**(1):e20201574. doi: 10.1084/jem.20201574.
36. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018 Jun;**36**(5):411–20.
37. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015 May;**33**(5):495–502.
38. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021 Jun 24;**184**(13):3573–87.e29.
39. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun 13;**177**(7):1888–902.e21.
40. Stuart T, Satija R. Integrative single-cell analysis. Nat Rev Genet. 2019 May;**20**(5):257–72.
41. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018 Feb 6;**19**(1):15.
42. Melsted P, Booeshaghi AS, Liu L, Gao F, Lu L, Min KHJ, et al. Modular, efficient and constant-memory single-cell RNA-seq preprocessing. Nat Biotechnol. 2021 Jul;**39**(7):813–8.
43. Kaminow B, Yunusov D, Dobin A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. BioRxiv. 2021 May 5.
44. Srivastava A, Malik L, Smith T, Sudbery I, Patro R. Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. Genome Biol. 2019 Mar 27;**20**(1):65.

45. Qiu X, Hill A, Packer J, Lin D, Ma Y-A, Trapnell C. Single-cell mRNA quantification and differential analysis with Census. Nat Methods. 2017 Mar;**14**(3):309–15.
46. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. Nat Methods. 2017 Oct;**14**(10):979–82.
47. Amezquita RA, Lun ATL, Becht E, Carey VJ, Carpp LN, Geistlinger L, et al. Orchestrating single-cell analysis with Bioconductor. Nat Methods. 2020 Feb;**17**(2):137–45.
48. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019 Jun 19;**15**(6):e8746.
49. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 2019 Dec 23;**20**(1):296.
50. Chen G, Ning B, Shi T. Single-cell RNA-Seq technologies and related computational data analysis. Front Genet. 2019 Apr 5;**10**:317.
51. Yip SH, Wang P, Kocher J-PA, Sham PC, Wang J. Linnorm: improved statistical analysis for single cell RNA-seq expression data. Nucleic Acids Res. 2017 Dec 15;**45**(22):e179.
52. Bacher R, Chu L-F, Leng N, Gasch AP, Thomson JA, Stewart RM, et al. SCnorm: robust normalization of single-cell RNA-seq data. Nat Methods. 2017 Jun;**14**(6):584–6.
53. Lun ATL, Bach K, Marioni JC. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. Genome Biol. 2016 Apr 27;**17**:75.
54. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, et al. Synthetic spike-in standards for RNA-seq experiments. Genome Res. 2011 Sep;**21**(9):1543–51.
55. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet. 2010 Oct;**11**(10):733–9.
56. Goh WWB, Wang W, Wong L. Why batch effects matter in omics data, and how to avoid them. Trends Biotechnol. 2017 Jun;**35**(6):498–507.
57. Tung P-Y, Blischak JD, Hsiao CJ, Knowles DA, Burnett JE, Pritchard JK, et al. Batch effects and the effective design of single-cell gene expression studies. Sci Rep. 2017 Jan 3;**7**:39921.
58. Büttner M, Miao Z, Wolf FA, Teichmann SA, Theis FJ. A test metric for assessing single-cell RNA-seq batch correction. Nat Methods. 2019 Jan;**16**(1):43–9.
59. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. Genome Biol. 2020 Feb 7;**21**(1):31.
60. Hou W, Ji Z, Ji H, Hicks SC. A systematic evaluation of single-cell RNA-sequencing imputation methods. Genome Biol. 2020 Aug 27;**21**(1):218.
61. Linderman GC, Zhao J, Roulis M, Bielecki P, Flavell RA, Nadler B, et al. Zero-preserving imputation of single-cell RNA-seq data. Nat Commun. 2022 Jan 11;**13**(1):192.
62. Cleary B, Cong L, Cheung A, Lander ES, Regev A. Efficient generation of transcriptomic profiles by random composite measurements. Cell. 2017 Nov 30;**171**(6):1424–36.e18.
63. Wagner A, Regev A, Yosef N. Revealing the vectors of cellular identity with single-cell genomics. Nat Biotechnol. 2016 Nov 8;**34**(11):1145–60.
64. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2018 Dec 3;**37**:38–44.
65. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008.
66. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv. 2018 Feb 9;
67. Mohammadi S, Davila-Velderrain J, Kellis M. A multiresolution framework to characterize single-cell state landscapes. Nat Commun. 2020 Oct 26;**11**(1):5399.
68. Xu C, Su Z. Identification of cell types from single-cell transcriptomes using a novel clustering method. Bioinformatics. 2015 Jun 15;**31**(12):1974–80.
69. Way GP, Zietz M, Rubinetti V, Himmelstein DS, Greene CS. Compressing gene expression data using multiple latent space dimensionalities learns complementary biological representations. Genome Biol. 2020 May 11;**21**(1):109.
70. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech. 2008 Oct 9;**2008**(10):P10008.
71. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, et al. FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data. Cytometry A. 2015 Jul;**87**(7):636–45.
72. Rozenblatt-Rosen O, Regev A, Oberdoerffer P, Nawy T, Hupalowska A, Rood JE, et al. The human tumor atlas network: charting tumor transitions across space and time at single-cell resolution. Cell. 2020 Apr 16;**181**(2):236–49.
73. Risso D, Perraudeau F, Gribkova S, Dudoit S, Vert J-P. A general and flexible method for signal extraction from single-cell RNA-seq data. Nat Commun. 2018 Jan 18;**9**(1):284.
74. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. Nat Commun. 2021 Sep 28;**12**(1):5692.
75. GitHub—hbctraining/scRNA-seq [Internet]. [cited 2022 Dec 28]. Available from: https://github.com/hbctraining/scRNA-seq
76. Irizarry RA, Wang C, Zhou Y, Speed TP. Gene set enrichment analysis made simple. Stat Methods Med Res. 2009 Dec;**18**(6):565–75.
77. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005 Oct 25;**102**(43):15545–50.
78. Fertig EJ, Ozawa H, Thakar M, Howard JD, Kagohara LT, Krigsfeld G, et al. CoGAPS matrix factorization algorithm identifies transcriptional changes in AP-2alpha target genes in feedback from therapeutic inhibition of the EGFR network. Oncotarget. 2016 Nov 8;**7**(45):73845–64.
79. Stein-O'Brien GL, Clark BS, Sherman T, Zibetti C, Hu Q, Sealfon R, et al. Decomposing cell identity for transfer learning across cellular measurements, platforms, tissues, and species. Cell Syst. 2019 May 22;**8**(5):395–411.e8.
80. Zhu X, Ching T, Pan X, Weissman SM, Garmire L. Detecting heterogeneity in single-cell RNA-Seq data by non-negative matrix factorization. PeerJ. 2017 Jan 19;**5**:e2888.
81. Setty M, Tadmor MD, Reich-Zeliger S, Angel O, Salame TM, Kathail P, et al. Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat Biotechnol. 2016 Jun;**34**(6):637–45.
82. Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, et al. Single-cell RNA-Seq with waterfall reveals molecular cascades underlying adult neurogenesis. Cell Stem Cell. 2015 Sep 3;**17**(3):360–72.
83. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 2008 Dec 29;**9**:559.
84. Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring regulatory networks from expression data using tree-based methods. PLoS ONE. 2010 Sep 28;**5**(9):e12776..

85. Huynh-Thu VA, Sanguinetti G. Combining tree-based and dynamical systems for the inference of gene regulatory networks. Bioinformatics. 2015 May 15;**31**(10):1614–22.
86. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc. 2020 Apr;**15**(4):1484–506.
87. Kumar MP, Du J, Lagoudas G, Jiao Y, Sawyer A, Drummond DC, et al. Analysis of single-cell RNA-Seq identifies cell-cell communication associated with tumor characteristics. Cell Rep. 2018 Nov 6;**25**(6):1458–68.e4.
88. Chan TE, Stumpf MPH, Babtie AC. Gene regulatory network inference from single-cell data using multivariate information measures. Cell Syst. 2017 Sep 27;**5**(3):251–67.e3.
89. Papili Gao N, Ud-Dean SMM, Gandrillon O, Gunawan R. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. Bioinformatics. 2018 Jan 15;**34**(2):258–66.
90. Matsumoto H, Kiryu H, Furusawa C, Ko MSH, Ko SBH, Gouda N, et al. SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics. 2017 Aug 1;**33**(15):2314–21.
91. Deshpande A, Chu L-F, Stewart R, Gitter A. Network inference with Granger causality ensembles on single-cell transcriptomics. Cell Rep. 2022 Feb 8;**38**(6):110333.
92. Specht AT, Li J. LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering. Bioinformatics. 2017 Mar 1;**33**(5):764–6.

22

Application of artificial intelligence to overcome clinical information overload in cancer

Arnulf Stenzl, Jennifer Ghith, and Bob J.A. Schijvenaars

22.1. Needs and challenges of clinical information overload

There has been unprecedented growth in the volume and availability of published biomedical data over the past decade, exceeding 38 million publications in MEDLINE/PubMed by April 2025. Similarly, a basic PubMed literature search using the word 'oncology' identified >80,000 scientific publications related to clinical trials in oncology from January 2010 to May 2025. Furthermore, approximately 25,000 cancer-related articles were accepted in 10 high-impact oncology journals in 2024 (**Figure 22.1**A). Alongside this growth in biomedical data is the emergence of artificial intelligence (AI) literature. The 2025 AI Index Report highlights that the total number of AI journal publications in computer science more than doubled between 2013 and 2023, from approximately 102,000 to 242,000 (**Figure 22.1**B) [1].

The coronavirus disease 2019 (COVID-19) pandemic changed the dynamics of research and science communication. Publishers responded by expanding offerings, with new spinoffs and a surge in sister journals. For example, Springer Nature launched *Nature Cancer* in January 2020, and the *British Medical Journal* (*BMJ*) started publishing online-only *BMJ Oncology* in October 2022. In parallel, publishers are developing AI-focused journals, such as the *New England Journal of Medicine AI*, to accommodate the rapidly growing medical AI literature and create a stronger evidence base for the clinical applications of AI. During the pandemic, journal editors expedited the dissemination of COVID-19 research through fast-tracked peer-review processes, and online preprint repositories expanded. Over 125,000 articles were released on COVID-19 in 2020, of which more than 30,000 appeared on preprint platforms. Publication using preprint platforms continues to grow post pandemic despite concerns over the lack of traditional peer review and the elevated risk of poor-quality research dissemination [2].

In addition to the surge in scientific data, digital advancements enabled a substantive increase in patient-level data, including electronically recorded patient-reported outcomes (PROs) and other health data from wearables, mobile apps, electronic health records (EHRs), and periodically collected imaging, digital pathology, and multi-omic data [3].

22.1.1. Information overload in clinical practice

Clinicians work in fast-paced, data-rich environments that require digital communication to search for medical information [4]. Clinicians select information sources based on relevance to clinical practice, content credibility, accessibility, and usability [5]. In a 2021 internal survey of 150 US and EU urologists and oncologists treating prostate cancer (PCa), respondents identified time-consuming searching for complex information, and lack of relevant and accessible materials as barriers to successful information seeking. Most respondents also considered the use of trusted, comprehensive, and unbiased sources, and automated concept-based rather than keyword-based searches as important features of an effective question-answering (QA) tool. A summary of the electronic knowledge resources used by healthcare professionals (HCPs) and researchers and their key features is presented in **Table 22.1**.

Clinicians are often unaware of different search strategies (e.g. keywords, Boolean operators, medical subject headings [MeSH], and filters) [5] and make their queries in natural language [6]. Further, traditional QA methods are not scalable to medical information overload and fail to filter out irrelevant information and provide tailored answers to queries, which poses challenges to effective information extraction by HCPs [7]. There are concerns, particularly with the emergence of generative AI, that personalization algorithms can introduce or strengthen confirmation and popularity biases, creating filter bubbles, promoting trending information that underrepresents higher-quality content, and facilitating manipulation by social bots and spread of misinformation or low-credibility content [8].

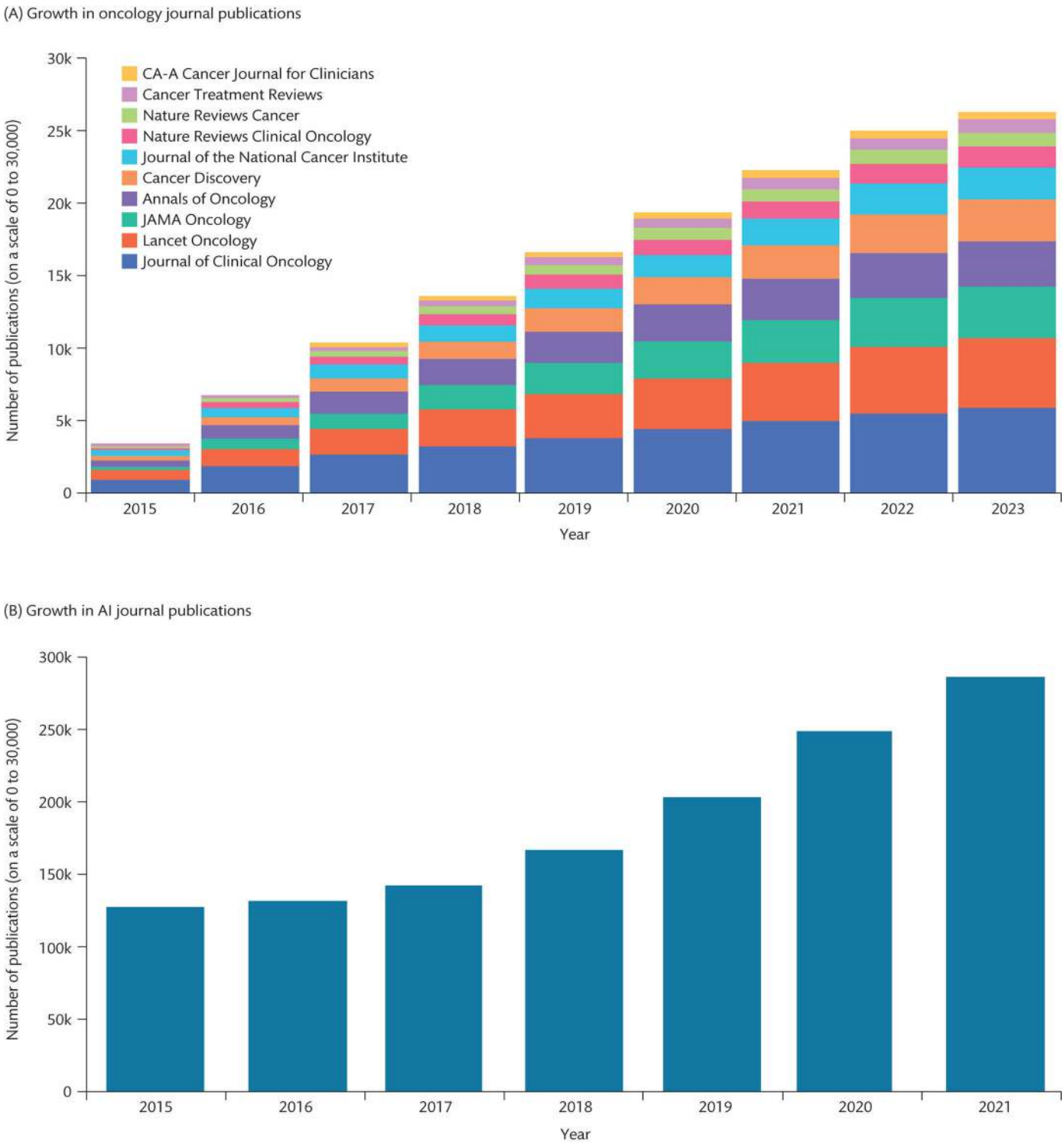


Figure 22.1. The number of (A) PubMed publications in high-impact oncology journals (according to journal impact factor) during 2015–2025, data as of May 2025; and (B) publications in AI journals during 2013–2023. *Source:* Adapted from 'The AI Index 2025 Annual Report' by Maslej N, et al. Stanford (CA): Institute for Human-Centered AI, Stanford University; 2025. CC BY-ND 4.0 [1].

Persistent exposure to information overload leads to unsustainable workload, cognitive burden, inconsistent documentation of care actions, and reduced performance, causing unfavourable impact on clinicians' health and patient outcomes [4, 9]. This also applies to oncology where cancer information overload causes worry and confusion among patients, practice inconsistency, scepticism, uncertainty, distrust of medical evidence, and potential lack of awareness of or adherence to recommendations [10].

Table 22.1. Examples of electronic bibliographic databases, literature search platforms, and medical knowledge resources commonly used by clinicians [5,11,12].

Platform	Organization	Features	Access
Bibliographic databases			
Cochrane Library	Cochrane (published by Wiley)	A collection of six databases that contain high-quality, independent evidence to inform healthcare decision-making, including the Cochrane Database of Systematic Reviews and CENTRAL; CENTRAL includes citations to trials from MEDLINE, Embase, ClinicalTrials.gov, and other sources	Free search; access to the citation records or full text requires a subscription
ClinicalTrials.gov	NIH	The largest searchable registry and results database of clinical trials; contains over 536,900 research studies in 221 countries and all 50 US states	Free access; free registration of clinical trial details
Embase®	Elsevier	A comprehensive bibliographic database of biomedical and pharmacological literature, including journals not indexed in MEDLINE and conference literature; links to the full-text articles, if available	Subscription-based
PubMed®	NLM NCBI	A bibliographic database of life sciences and biomedical information published since 1966; facilitates searching across three NLM literature resources, i.e. MEDLINE, PubMed Central, and Bookshelf; MEDLINE is the primary component of PubMed; records are indexed with MeSH; MEDLINE citations in PubMed contain a link to (free or access-restricted) full texts archived in PubMed Central or at the journal website	Free access
AI-based medical literature search platforms			
Causaly	Causaly	Search and analysis of scientific literature, clinical trials, and side-effect databases; maps relationships within scientific data	Paid access
Dimensions	Digital Science and Research Solutions Inc.	Search and analysis of scientific literature, grants, clinical trials, datasets, patents, and policy documents; graph-based visualization of key elements in the published articles; full-text pdf files of some publications are available to directly view and download; comprehensive coverage, including book chapters and some proceedings papers	Free limited version for personal, non-commercial use; also offers paid access
Dr.Evidence®	Dr.Evidence (Doctor Evidence, LLC)	Search and monitoring of scientific literature, conferences, clinical trials, grants, patents, RSS feeds, and drug labels; identifies and filters studies containing real-world data	Paid access
EVID AI	Genesis Research Group	Gap and landscape analysis of scientific literature; allows comparative effectiveness studies	Paid access
Galactic AI™	Biorelate Ltd	Search and analysis of biomedical literature; drug discovery; causal search capability; detects and categorizes the relationships linked to search terms; lists the relationship type by confidence interval	Paid access
Google Scholar	Google	Search and indexing of scholarly literature; analysis of scientific web content and patents; full-text pdf files of some publications are available to directly view and download	Free access
IRIS.AI	Iris AI AS	Scientific literature processing and research landscape mapping	Paid access
Meta AI	Meta Platforms, Inc.	Scientific literature search and curation of feeds on biomedical topics	Paid access

(*continued*)

Table 22.1. Continued

Platform	Organization	Features	Access
scite_	Scite LLC	Search and analysis; checks if citations have been supported or disputed by others; basic QA functionality to find answers from full-text articles for natural language research queries	Paid access; researchers can access a very limited number of reports and visualizations, and set up author alerts for free; QA function is free to use
SciBite	SciBite from Elsevier	Search and analysis; semantically extracts scientific terminology from unstructured text and converts it into clean, contextualized data	Paid access
Scopus	Elsevier	Search and analysis of scientific literature curated by an advisory board; interface is also available in Chinese and Japanese; full-text pdf files of some publications are available to directly view; links to the repository version of the articles; covers peer-reviewed journal articles with substantial references and citations, conference proceedings, and book chapters	Free access to Scopus; preview features
Semantic Scholar	Ai2, The Allen Institute for Artificial Intelligence	Search and analysis of biomedical literature; highlights and graphically represents the key elements of a paper; provides access to full-text articles; citations cannot be exported; does not search licensed resources	Free access
Web of Science	Clarivate	Search and analysis of scientific literature; contains non-English regional citation indexes; full-text links to the subscribed databases; covers journals, books, proceedings, meeting abstracts, book reviews, and patents; does not include a book citation index	Subscription-based
Medical knowledge and point-of-care learning resources			
epocrates®	epocrates, Inc.	Point-of-care decision app for mobile devices; free version provides drug information, interaction check, pill identifier, clinical practice guidelines, formulary, and dosing calculators; full version (Epocrates+) contains peer-reviewed disease content from the *BMJ*, alternative medicine monographs, ICD 10th Revision and Current Procedural Terminology codes, infectious disease treatment, and laboratory tests	Full version requires a subscription
DynaMed®	EBSCO Industries, Inc.	Point-of-care resource for health information; publishes and monitors NICE guidelines; includes concise summaries and detailed recommendations based on the most current evidence; contains Micromedex® drug content	Subscription-based
UpToDate® Lexidrug™ (formerly Lexicomp®)	Wolters Kluwer	Drug reference tool for healthcare professionals consisting of pharmacological databases, with information on prescription drugs and over-the-counter products, medical calculators, drug interactions, and patient education handouts	Subscription-based
Micromedex®	Merative	Multi-database drug search engine that provides evidence-based information for prescription drugs and over-the-counter products), diseases, toxicology, and alternative medicine; includes patient counselling tools; "Ask Watson" search uses AI to produce answers to simple drug reference questions	Subscription-based
Medscape®	WebMD	Point-of-care drug and disease information for physicians across subspecialities; provides access to MEDLINE and other databases; customizable user profile settings for enhanced learning; provides customized medical news, clinical trial coverage, drug updates, journal articles, and continuous medical education activities	Free to use after initial free registration
QxMD	QxMD Software, Inc.	QxMD platforms (Read and Calculate) include point-of-care medical literature surveillance, clinical calculators, and interactive decision-support tools; provide access to PubMed, with personalized trending literature feeds and subspecialty specific article collections	Free to use after initial free registration; access with no separate registration for Medscape Network users
UpToDate®	Wolters Kluwer	Point-of-care medical resource for physicians; primarily covers internal medicine and its subspecialties; addresses specific clinical issues in the form of topic reviews, although the expert-directed evidence acquisition and synthesis process is not clear	Subscription-based

of Medicine; QA: question answering; RSS: Really Simple Syndication.

22.1.2. Artificial intelligence to address information overload

AI is emerging across oncology in areas such as diagnostics, digital pathology/imaging, multi-omics, drug discovery, and prognostics [13,14]. Although AI applications in biomedical literature mining, knowledge discovery, and writing novel content within academic papers and grants had been a nascent field [3,15], rapidly advancing capabilities in generative AI, such as ChatGPT (Generative Pre-trained Transformer), are making this an increasing area of focus [16–18].

AI can overcome information overload and support decision-making by enabling clinicians to efficiently search and synthesize literature and trial databases, extract meaningful information, identify knowledge gaps, promising research directions, and rapidly changing landscapes [11]. This can be enhanced by AI-assisted analysis of real-world evidence (RWE) through text-mining and processing of real-world data, including clinical records (e.g. EHRs), discussion posts in cancer patient forums, electronic message exchanges between patients and care providers, patient–physician question and answer sites, and narratives in social media feeds [19]. RWE can assist physicians in addressing complex cancer-related queries on demographics, model-based predictions, treatment efficacy and side effects, and PROs [19]. AI tools also factor into improved health literacy in cancer care through text translation and simplification of scientific language into more understandable plain-language summaries, with potentially favourable impacts on patient–provider communication, care experience, and outcomes [20]. Generative AI, in particular, has the potential to be used not only to support clinical practice (e.g. by reducing the administrative burden of note-taking, integration into EHRs, predicting disease risk and outcome, and refining personalized medicine) but also in the evolution of scientific communications and data dissemination (e.g. writing scientific papers, literature searches, and summarization) [21,22].

22.1.3. Takeaways

Increased volume and availability of information as well as improvements in the democratization of medical knowledge are cause for celebration; however, exposure and access to this torrent of information is a double-edged sword.

Widely accessible medical information helps identify precise answers to queries and develop effective strategies for improving patient outcomes. However, processing large amounts of information, as well as potential misinformation, is time-consuming and labour-intensive.

AI technologies can provide accurate and contextualized answers to specific queries of users, driven by their background characteristics and predicted interests. AI developers, however, need to guard against exacerbating bias and incorporating misinformation into their algorithms. When developing AI tools, do so with full constituencies in mind—developers are encouraged to bring in diverse teams to evaluate tools at an early stage and identify potential concerns.

22.2. Information retrieval, QA, and generative AI in the medical domain

Information retrieval systems are algorithms focused on facilitating the processing of large, mostly textual documents to extract information and answer user queries by identifying keywords and matching them with documents. Historically, information retrieval systems relied on a lexical approach using simplified text representation models, which captured the meaning of a text by counting the frequency of each word [23]. In the medical literature, semantic gap and vocabulary mismatch between the query and source document are challenges because of high variability in language and spelling, frequent acronyms and abbreviations, and ambiguity of document contexts. These issues are addressed by semantic searches on structured knowledge sources (MeSH and International Classification of Diseases) that link medical terms to associated meanings or unstructured knowledge, i.e. collections of raw text (MEDLINE and EHRs), to automatically establish semantic relationships between words, phrases, and documents [23]. Linking text documents to knowledge requires semantic annotation that involves identifying medical entities/concepts in the document (named-entity recognition), relationships between entities (relation extraction), and connections between entities/concepts and the whole text document (text classification). Semantic annotation can occur at the document level but can also be assigned to individual terms in a text to allow for searches for two entities (terms) in close proximity (e.g. polyps and prognosis). Relationships can also be leveraged in search, e.g. by searching for general terms (tyrosine kinase inhibitors) and retrieving documents about more specific information (dasatinib, nilotinib, etc.).

Data mined from multiple knowledge sources can be integrated into a graph structure called a knowledge graph. A knowledge graph represents a network of interlinked descriptions of semantically related entities, concepts, and terminologies in a particular domain of knowledge (ontologies). Semantic representation of structured knowledge in knowledge graphs has important applications in semantic search, text analysis, generation of natural language, and complex QA [24].

Modern information retrieval systems include search engines, platforms, QA systems and, recently, large language models (LLMs), and generative AI. Search engines (e.g. PubMed) operate on keyword-based information retrieval and return algorithmically ranked documents for the user to access and review to find answers. Platforms store, organize, and share data in a structured and searchable manner, which enables automated information retrieval and data visualization via web application programming interfaces [25]. QA systems analyse a question posed in natural language, process semantically related documents from existing knowledge sources, extract matching text passages from the documents, and return fitting answers [26]. QA systems use natural language processing (NLP), an AI field that combines linguistics and computer science, to identify natural language rules (syntax, semantics, morphology, and pragmatics) and transform them into a machine-readable format (semantic transformation). Additionally, QA systems employ machine-learning (ML) techniques to automatically mine the literature and extract contextual information related to domain-specific queries [11]. LLMs are pre-trained on large amounts of texts without labelling efforts to learn the rules and associations within a language and then fine-tuned with a small, labelled training set to produce outputs for specific NLP tasks, such as text generation and QA [27]. The advent of LLMs will provide a significant boost in the capabilities of QA systems, although there are still shortcomings to address.

22.2.1. Medical QA systems, LLMs, and generative AI

Medical QA datasets, including biomedical and clinical datasets (Table 22.2), are crucial for training LLMs on biomedical data and benchmarking their capacity to perform NLP tasks, such as QA on smaller datasets. Large datasets, such as PubMed and PubMed Central, contain semi-structured documents with section annotations that cover various medical subfields important for building medical QA models. One limitation is the failure to return accurate answers to clinical queries. In comparison, clinical datasets accommodate semantically annotated unstructured documents (e.g. EHRs and handwritten clinical notes) that cover a wide range of clinical topics and are prominent sources for training deep-learning models. Recently, open-source LLMs, such as Llama 3 and Mistral, have become significant starting points for fine-tuning medical LLMs. This leverages their general language understanding for more efficient adaptation to medical data. Models like BioMistral and the more recent OpenBioLLM-Llama3 [28], which were directly trained or fine-tuned on medical text, demonstrate this trend. While medical data quality still impacts performance, the use of strong open-source foundations could improve results, although it needs further study. These advancements offer promising routes for better medical QA (Table 22.3); however, ongoing evaluation for reliability remains essential.

Table 22.2. Medical datasets for QA systems.

Medical dataset	Description
BioASQ	Semantic biomedical QA dataset derived from the NLP-based BioASQ challenges; includes factoid, list, yes/no, and summary questions, along with expert-provided answers referenced to MEDLINE-indexed documents
emrQA	First large-scale clinical QA dataset; contains large-scale question–answer pairs; automatically generates questions using semantic extractions from the Stanford Question Answering Dataset, medical ontologies from the UML5 database, and expert-level annotations; combines a general-domain language and medical-domain knowledge benchmark for comparing clinical QA models
MedQA	Multiple-choice open-domain QA dataset for real-world medical problem-solving; includes questions collected from the professional medical board exams in the US, Mainland China, and Taiwan with textbook-extracted answers
MEDIQA2021	Extension of MedNLI designed for QA tasks; includes question–answer pairs with paragraph-length context; biomedical language understanding evaluation benchmark for evaluating new clinical QA algorithms; applied to NLP-based extraction of adverse drug reactions data from HCP–patient conversations, with regulatory, pharmaceutical, and clinical implications
MedQuAD	Includes large-scale medical question–answer pairs from 12 NIH websites on diseases, drugs, and other medical entities
PubMedQA	Biomedical QA dataset of questions linked to PubMed abstracts with yes/no/maybe answers, which require reasoning over the context (abstract without the conclusion section)

Abbreviations: HCP: healthcare professional; MedNLI Medical Natural Language Inference: NIH: National Institutes of Health; NLP: natural language processing; QA: question answering; UML: unified modelling language.

Language models (LMs) have evolved over the last 60 years from the early rule-based models to deep-learning models that employ NLP and ML techniques. Subsequently, transformer-based LMs, such as Bidirectional Encoder Representations from Transformer (BERT), with attention/self-attention and unsupervised learning capabilities were developed. This landmark achievement gave rise to emerging/cutting-edge LMs such as OpenAI's GPT models, Google's Gemini, and Meta's LLaMa. These are just a few examples, and due to the fast-evolving nature of the field, there will certainly be more by the time of this book's publication. These models can be fine-tuned with a small, labelled training set for more difficult NLP tasks, such as indexing, ontology matching, semantic reasoning, automatic translation, mathematics, and coding; and natural language generation for summarization or even generation of novel text, sounds, or images, and conversational interaction with users [29–34].

The most recent advances in generative/conversational AI tools include the integration of multimodal LLMs into everyday tools like Microsoft Office and Google Search [16–18]. Multimodal LLMs accept different formats of input (image and text) and generate text outputs. Self-learning AI tools combine GPT with a notable major advance in the field, i.e. the use of reinforcement learning from human feedback, to score the responses to a given prompt by the fine-tuned LM and to optimize the generated text for user intent. This tool uses a dialogue format that allows for answering follow-up questions, while declining questions with inappropriate text [18]. These capabilities may help with hypothesis generation, scientific writing, and RWE analysis. The latest versions of ChatGPT are built into the AI-powered Bing search engine using Microsoft's Prometheus model to provide more accurate answers and offer responses to current events. Microsoft has also integrated the OpenAI technology into Microsoft 365 Office, known as Copilot.

ChatGPT has been listed as a co-author on numerous published research papers, although its authorship is highly controversial [35]. Some editorial boards have stated that inclusion of LLMs in the author list is not appropriate as this does not align with current International Committee of Medical Journal Editors guidance, and recommend against crediting or citing AI-assisted technologies as authors due to the lack of responsibility for the accuracy, integrity, and originality of the submitted material [36,37]. As such, authors should disclose and describe any use of such technologies in the production of their work. ChatGPT-generated research abstracts have also been vague, falsified, and poorly styled to journal specifications despite convincing readability and originality [38]. Furthermore, ChatGPT is not trained on sufficiently specialized scientific literature to generate accurate, evidence-supported, and academically plausible content. Akin to other LLMs, ChatGPT can hallucinate seemingly authoritative but conceptually incorrect, misleading, or nonsensical answers [39]. Cost and computational requirements, copyright and licensing issues, credibility, explainability, safety, responsibility, and unpredictable societal changes are other important considerations [39].

A brief history of the evolution and important characteristics of QA and LMs is presented in Figure 22.2.

Table 22.3. Examples of pre-trained language models and generative AI tools in the medical domain.

Language model	Release/launch date	Description
PubMedBERT	July 2020	First biomedical-domain–specific deep-learning NLP pre-trained model; high accuracy in answering; lacks performance on clinical NLP tasks; low quality in answering longer questions; uses a different vocabulary to BERT
BioELECTRA	June 2021	Biomedical-domain–specific; better performance on benchmark QA tasks; trained on abstracts and articles from PubMed; does not incorporate clinical datasets and has limited performance in a clinical setting
Generative AI		
GENTRL	September 2019	Developed by Insilico Medicine; this generative AI platform has been used to design new drugs for diseases
Med-PaLM 2	December 2022	Google's medical LLM aligned to the medical domain to more accurately and safely answer medical questions; the first AI system to achieve a pass mark (>60%) in USMLE-style questions
Dragon Ambient eXperience (DAX) Express	March 2023	Nuance and Microsoft's clinical documentation tool powered by GPT-4. The tool will enable healthcare workers to automate clinical documentation simply by 'listening' to physician–patient consultations
Glass AI	March 2023	Glass AI 2.0 combines an LLM with a clinical knowledge database, created and maintained by clinicians, to create differential diagnoses and clinical plan outputs
Med-PaLM M	July 2023	A multimodal version of Med-PaLM that generates HCP-interpretable information from imaging files, medical records, and genomics
MedLM	December 2023	Built upon Med-PaLM 2, this set of generative AI models is fine-tuned for healthcare-specific tasks; generally available on Vertex AI to Google Cloud customers in the US; piloted integration into other AI-powered platforms for enhanced ambient medical documentation, drug research and development, and patient access and experience
Starling-LM-7B-beta	March 2024	An open-source LLM fine-tuned with reinforcement learning from AI feedback; good performance (66–70%) on certain medical domains of MMLU benchmark (clinical knowledge, medical genetics, and professional medicine); underperforms the commercial LMs GPT-4 base and Med-PaLM 2
Hermes-2-Pro-Mistral-7B	April 2024	An upgraded version of the open-source Nous Hermes 2 LM with good performance (65–72%) on MMLU clinical knowledge, medical genetics, and professional medicine domains; underperforms GPT-4 base and Med-PaLM 2
Gemini Pro	May 2025	A performance-optimized model from the Google's Gemini family of multimodal models, with broad image and audio-visual processing and reasoning capabilities; strong performance in data-intensive and procedural medical domains, particularly biostatistics, cell biology, and obstetrics and gynaecology
Gemma-7b	May 2025	A more powerful (7b) version of Gemma, a family of lightweight, open-source LMs from Google, based on the same research and technology used to create Gemini models; diverse QA, summarization, and reasoning capabilities; good performance (68–71%) on MMLU clinical knowledge, medical genetics, and professional medicine domains; underperforms GPT-4 base and Med-PaLM 2

Abbreviations: AI: artificial intelligence; BERT: Bidirectional Encoder Representations from Transformers; GENTRL: Generative Tensorial Reinforcement Learning; GPT: Generative Pre-trained Transformer; HCP: healthcare professional; LM: language model; LLM: large language model; MMLU: Measuring Massive Multitask Language Understanding; NLP: natural language processing; QA: question answering; RWD: real-world data; USMLE: US Medical Licensing Examination.

22.2.2. Takeaways

AI has changed medical literature mining, data extraction, and QA, and is a dynamic area that needs bridging to broader audiences to facilitate understanding of strengths, limitations, and responsible use.

Modern QA systems, LLMs, and generative AI tools that use attention mechanisms can deal with increasingly difficult tasks, such as semantic understanding and reasoning.

Next-generation LLMs and generative AI tools are self-learning, multimodal AI interfaces with generative and conversational capabilities that can assist HCPs and researchers with scientific writing, landscape scanning, evidence synthesis, and potentially RWE analysis. The validation of accuracy, clarity, reliability, and completeness of these models is an area of active investigation.

22.3. Applications of AI in literature search and data extraction in oncology

The next section focuses on the utility of AI-based tools for synthesis and automation of cancer literature and clinical trial database searches to address information overload in oncology. AI-driven information processing is increasingly employed in oncology for screening and diagnosis, biomarker phenotyping, genomic characterization of tumours, imaging and digital pathology, optimization of decision-making, individualization of treatment, cancer surveillance, and accelerated drug discovery [13,14]. A review of these applications is beyond the scope of this chapter.

22.3.1. Clinical literature search and synthesis

AI tools for medical literature search and data extraction assist by helping users efficiently conduct systematic searches and obtain focused information from rich datasets, i.e. published literature and clinical trials. In a randomized controlled trial (RCT), scientists with an extensive research background who used the IRIS.AI search engine for the literature search on the research and development needed to implement augmented reality in medical surgical training found more high-quality papers related to the topic compared with their peers' traditional search [40].

Emerging AI platforms also interactively visualize findings to facilitate intuitive understanding of complex information, allowing

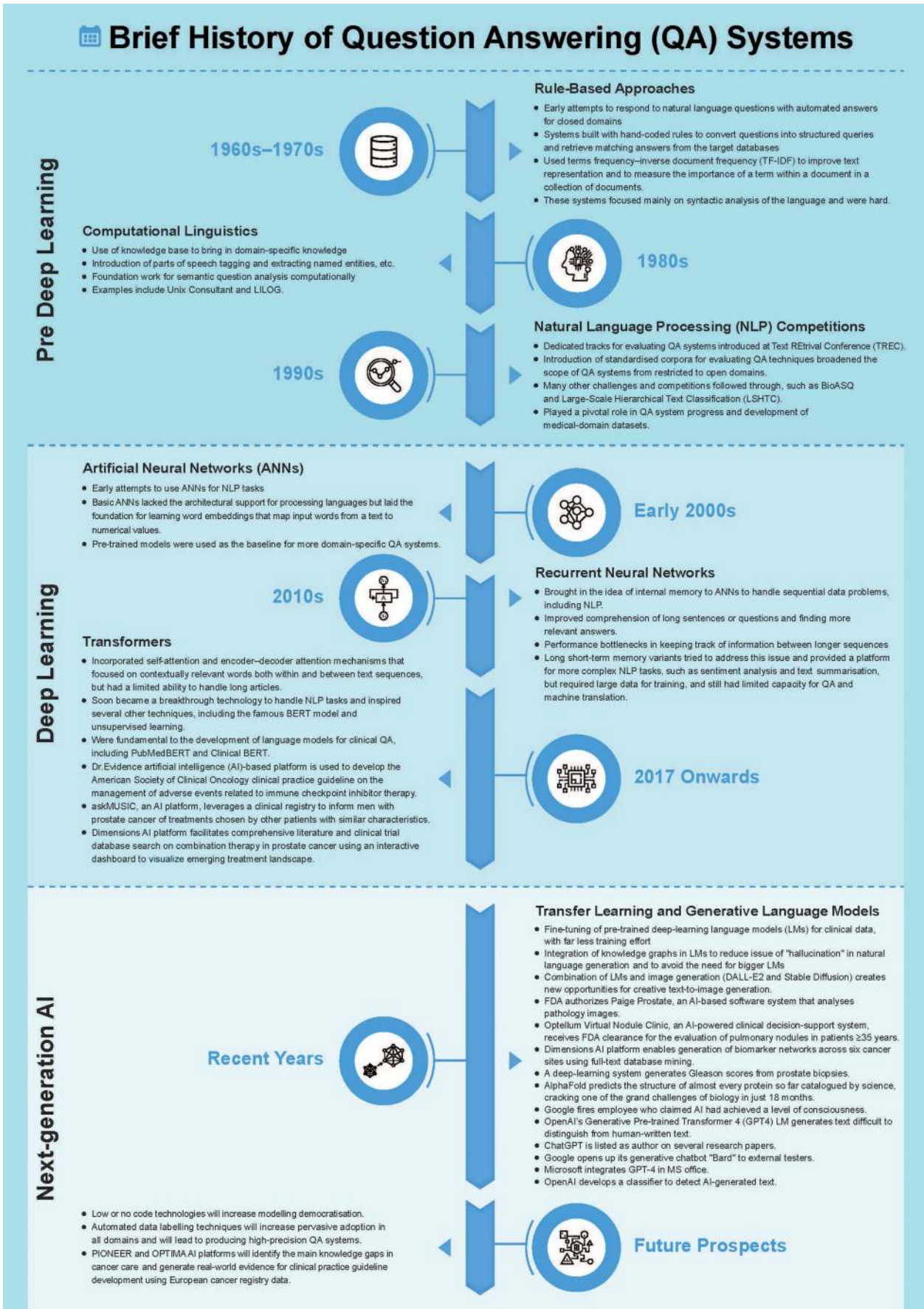


Figure 22.2. Evolution of QA and language models [11,16,30,32,35,41–49]. BERT: Bidirectional Encoder Representations from Transformers; REtrival Conference.

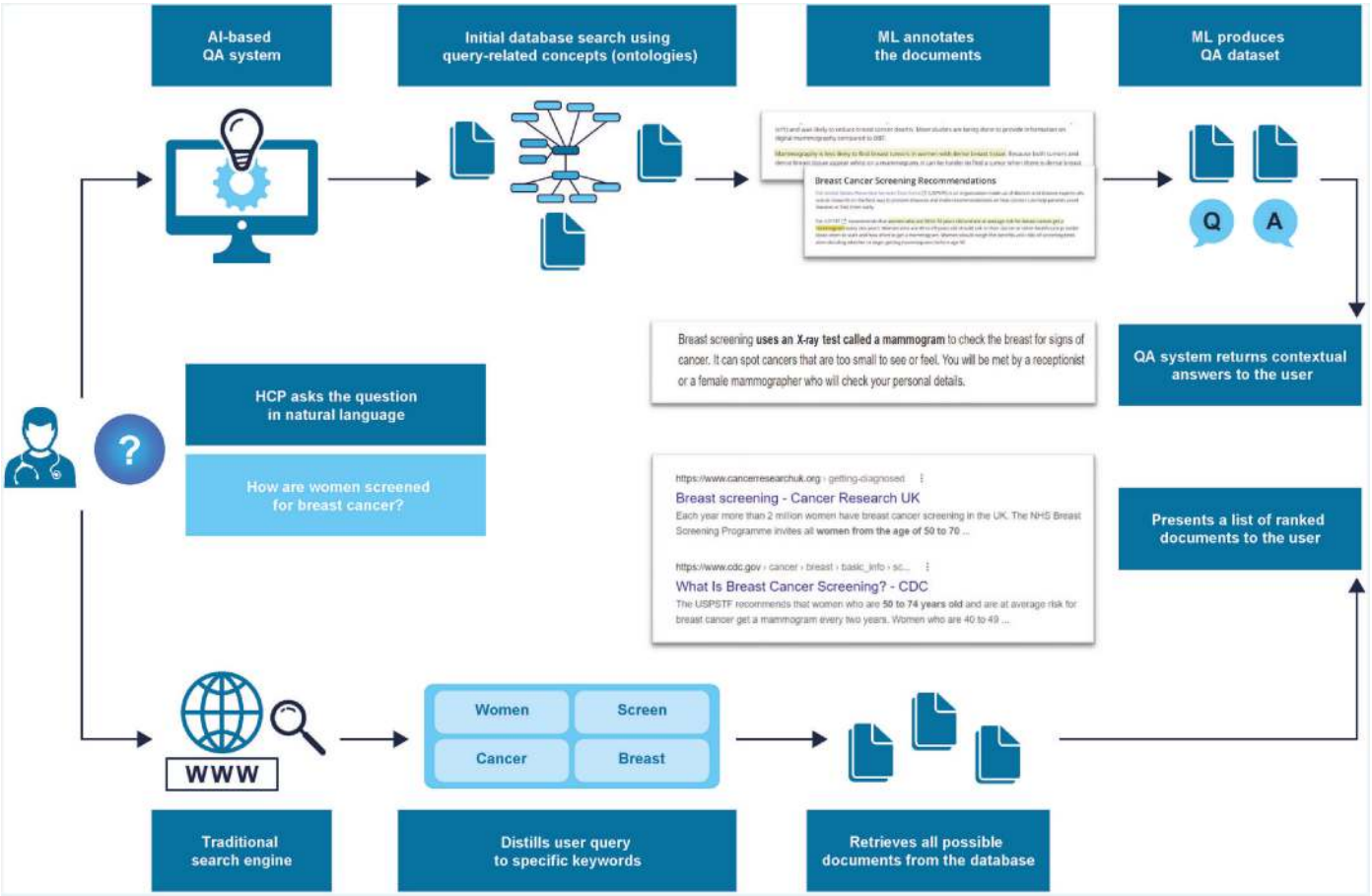


Figure 22.3. A schematic comparison of a typical keyword-based search engine to an AI-powered QA system that uses concept-based data mining and extraction from the biomedical literature and trial databases to answer a clinical query. AI: artificial intelligence; HCP: healthcare professional; ML: machine learning; QA: question answering.

clinicians to delve further into data based on their personal queries, and potentially develop new hypotheses, grant application ideas, or practice guidelines [11,50]. Figure 22.3 compares a typical keyword-based search engine to an AI-powered QA system that uses concept-based literature mining and data extraction to answer a clinical query.

AI-based tools utilize NLP and ML to semantically analyse annotated text documents retrieved from knowledge sources (e.g. titles, abstracts, and pieces of text within an article), sort them into predefined categories of interest (e.g. RCTs), identify snippets of query-related information (e.g. hormone therapies for PCa), filter out irrelevant information, and sometimes produce visual representations of the semantic analysis [11].

AI models have been applied to two key areas of literature search automation. ML-based study design classification systems are currently used in the Cochrane Register of Studies and validated public-access tools have been made available [51]. However, this approach may not be effective when there are multiple topic-specific inclusion criteria with different study designs, and a conventional keyword search is still needed. Literature exploration by concept rather than keyword can be a solution to this issue, but it requires extensive manual annotation. To resolve the need for human intervention, the National Centre for Text Mining has developed a semantic search engine with enhanced concept recognition features (Thalia) that automatically indexes new PubMed articles in the biomedical domain with daily updates and provides a visual interface to interact with the concepts identified [52].

Furthermore, automatic abstract screening systems have demonstrated a high level of accuracy in a systematic search [53]. Select systems (e.g. RobotAnalyst) also allow exploration of search retrieval by automatically grouping abstracts on a similar topic [54]. The key limitation of automated abstract screening is that the optimal stopping point for screening is unclear and their relevant articles may be missed by stopping too early.

Data extraction platforms for systematic reviews are still in the early stages of development and not readily accessible to clinicians. Available prototype systems, such as ExaCT and RobotReviewer, are trained on the annotated full-text articles and automatically extract relevant data on trial characteristics of interest (e.g. study population, descriptions of interventions, and outcomes) [55].

AI capabilities to assess the strength of evidence and the risk of bias in RCTs are growing. For example, RobotReviewer automatically retrieves the text that describes trial conduct relevant to the domains of biases included in the Cochrane Risk of Bias tool and classifies the trial by bias per domain [55]. Although these systems have reasonable accuracy, their performance is below the level achieved by the Risk of Bias tool and requires additional input from the reviewer [56]. Another important limitation is the lack of adequate training data with precise annotations [55].

Smart citation is a function featured in the scite_ tool that differentiates between citation statements (supporting, mentioning, and disputing) for a scientific paper, and reports out on the classification type, publication date, cited paper section, and paper type [57]. Scite_ also launched 'Ask a Question', a QA function that returns reliable answers from full-text scientific articles to research questions asked in plain language [58].

AI tools that incorporate advanced LLMs can be considered for data synthesis and guideline development due to their ability to visualize and automatically summarize large volumes of research evidence [59]. Of note, the Dr.Evidence® AI-based platform was used to systematically review evidence and develop the American Society of Clinical Oncology clinical practice guideline on the management of immune-related adverse events in patients who received immune checkpoint inhibitor therapy [60].

The emergence of AI platforms (**Table 22.1**) with expanded databases, improved semantic reasoning and QA mechanisms, and interactive data visualization capabilities provides an opportunity for clinicians to extract complex, specific, and personalized information, enabling them to achieve the best outcomes for their patients. For example, the Dimensions platform conducted a comprehensive search in clinical trial databases and scientific literature sources on studies that investigated the role of combination therapy and treatment sequencing in PCa [11]. Findings were visualized in an interactive dashboard to represent trial details and the emerging treatment landscape. The output has some limitations, including the absence of semantic analysis to separate the two main topics (drug combination versus treatment sequence) and the need for subject matter expertise to interpret the data [11]. Nevertheless, semantic analysis has been used in the INSIDE-PC project to extract content from scientific publications on clinical outcomes of a specific treatment sequence for advanced PCa. Dimensions has also been used to generate networks of biomarkers across six cancer indications, through biomarker co-occurrence processing [41]. The Michigan Urological Surgery Improvement Collaborative developed a web-based platform (askMUSIC) that utilized data from a prospective registry to inform newly diagnosed males with PCa of the treatments recommended by urologists for other patients with similar characteristics [42].

There is ongoing work on the integration of AI technology to identify and address the main knowledge gaps in cancer care and to generate RWE for the development of clinical practice guidelines. The PIONEER and OPTIMA projects are at the forefront of these efforts. PIONEER is a European consortium of 32 public and private organizations across nine countries that are focused on PCa [43]. The project aims to develop an AI platform that integrates and analyses longitudinal PCa registry data along with clinical and omic data from diverse patient populations across different disease stages. Planned objectives include consensus building on PCa understanding and outcome definitions, identification of critical evidence gaps in PCa management, improved risk stratification, and improved standardized care pathways. Upon implementation, PIONEER will house a central data hub which will support a network of interdisciplinary professionals by addressing PCa-related questions [43]. Similarly, OPTIMA is a European consortium of 36 multidisciplinary expert and public stakeholders that utilizes AI technology to improve care for oncology patients. This project aims to establish a secure, interoperable, and large-scale data platform for prostate, breast, and lung cancer, which includes RWE [44]. These evidence-informed tools can assist clinicians in care decision-making, based on enhanced practice recommendations [44]. OPTIMA will add to the impact of other emerging AI-integrated projects, such as the European Health Data Evidence Network and PIONEER, which currently support the European Health Data Space to implement better exchange and access to different types of health data. The outputs from this project will inform European policy regarding the clinical deployment of AI algorithms in healthcare. Time will tell how and when emerging generative AI capabilities will be integrated into the tools mentioned within this section.

Generative AI has already opened opportunities in other settings. For example, Be My Eyes has developed a tool, Be My AI (formerly Virtual Volunteer™), that integrates OpenAI's GPT models to dynamically assist visually impaired users in their daily tasks, such as description and hazard estimation of an object on the ground [61]. Elsewhere, the Icelandic government has partnered with OpenAI to train GTP models on Icelandic grammar and cultural knowledge through reinforcement learning from human feedback, with the key goal of improving generative AI capabilities in low-resource languages and reducing the existing cultural and language divide in AI [62].

22.3.2. Data extraction from clinical trial databases and EHRs

In the precision medicine era, clinical practice and clinical trials have emerged as valuable sources for AI-powered knowledge extraction and refinement. Clinical notes (e.g. EHRs, pathology and radiology reports, and patient charts) harbour important clinical information, crucial for translational research and patient-centric care [63]. Due to their unstructured or semi-structured text format, clinical notes are not applicable to traditional information retrieval systems due to the time, cost, and errors involved in manual processing. AI tools, including the state-of-the-art transformer-based models, have opportunity in this area and are gaining prominence in oncology practice for data extraction. Promising applications of clinical AI in oncology include text classification, information extraction, text summarization, topic modelling, QA, and, most recently, report generation [64].

These AI methods have been applied to clinical notes from EHRs for different clinical purposes, including cancer screening, diagnosis, staging and risk stratification, tumour description, biomarker association, pre-treatment assessment, and insight into the patient experience [63]. A novel NLP system has used transfer learning capability to automatically extract information from operative and pathology reports of patients with breast cancer. The system extracted outcomes data from tumour characteristics, prognostic factors, and treatment-related variables with high accuracy compared with expert reviews [65]. ML-based text classification and data extraction models trained on narrative text from progress notes and pathology reports have also demonstrated good performance in identifying high-risk breast lesions [66]. NLP has also been applied to extract information on colorectal polyp characteristics from colonoscopy records. Investigators linked colonoscopy records to patient-level data in the Surveillance, Epidemiology, and End Results (SEER) Cancer Registry and their subsequent EHR data to predict the risk of colorectal cancer in patients with conventional adenomas or sessile serrated polyps compared with the polyp-free population [67]. A deep-learning system,

DeepPhe, was developed, which combines ontology and text summarization to extract information pertinent to tumour descriptors in annotated EHRs of patients with breast cancer, and clusters them by their phenotypic profile [68]. NLP-based systems for data extraction from prospective cancer registries and pathology reports have also been applied to risk stratification and staging of solid tumours [69,70]. Additionally, several studies support the utility of ML-based methods in identifying initial treatment options for non-metastatic prostate, oropharyngeal, and oesophageal cancers based on structured and free-text treatment information extraction from cancer registries. At present, a National Cancer Institute—US Department of Energy collaboration project, MOSSAIC (Modeling Outcomes Using Surveillance Data and Scalable Artificial Intelligence for Cancer), applies multitask AI capabilities to automatically extract tumour features from free-text pathology reports using population-level SEER data, in an effort to address gaps in public access to high-quality cancer data and improve cancer surveillance [71].

Although AI shows promise for application in patient-level data, the volume, velocity, veracity, and variety of data present challenges. Most trial and EHR databases underrepresent non-white populations, introducing bias in data analysis; clinical records may contain relevant information for only certain subpopulations, or information may be entered only for select patients; HCPs may enter patient data using inconsistent terms or different formats; and field-entry data could also be incomplete, miscoded, or missing [72]. Furthermore, incorrectly defined selection criteria may lead to inaccuracies in algorithmic outputs. Prospective clinical data often change over time, and this drift in EHR data would deviate the results of AI methods [73]. This highlights the need for continuous quality assessment and refinement of AI algorithms and governance standards, such as the Observational Medical Outcomes Partnership Common Data Model, to harmonize the structure and semantics of disparate real-world data for more efficient and reliable analysis [74]. Apart from technical hurdles, legal and practical barriers (data privacy regulations, storage, and sharing policies) present additional challenges to the availability of RWE, including EHRs [72].

22.3.3. AI in clinical decision-making

Progress in AI-assisted clinical decision-making has been made through automated translation of cancer-related queries into actionable recommendations that can be adjusted by practitioners' feedback to improve the model's output [75]. Traditional clinical decision-support systems comprised software that matched patient characteristics to a clinical knowledge base and presented patient-specific recommendations for point-of-care decisions. However, these were stand-alone systems that integrated poorly into the clinician workflow and generated imprecise or disruptive alerts, and resulted in clinician fatigue and reduced quality of care [76]. AI models can augment efficiency and accuracy of decision-support systems by mining and distilling knowledge from big data sources in oncology. The emerging agentic AI and AI agents have metacognitive-like capabilities to self-regulate and adapt to changing data inputs for refined and relevant executive actions, and offer promising potential in clinical decision-making and care delivery support [50, 77]. However, data-driven bias and limitations in clinical reasoning, adaptability, and contextual comprehension are important challenges for the integration of these systems in the real-world setting [50].

In a healthcare framework where point-of-care decision-making is informed by routine clinical practice and/or clinical trial data, AI-integrated decision-support systems can beneficially contribute to precision medicine, and improve patient classification, tumour control, treatment optimization, and quality of life [78]. For example, IBM's Watson for Oncology (WFO), trained on US cancer treatment guidelines and clinical practice experiences, provided treatment recommendations for multiple malignancies concordant with decisions by multidisciplinary teams [79]. However, WFO needs improvements as the recommendations may not be applicable to other countries with different population characteristics and national guidelines, or may not account for coexisting conditions or reversible drug-related complications [79]. Notably, the Optellum Virtual Nodule Clinic, an AI-powered computer-aiding system for risk stratification of indeterminate pulmonary nodules using nodule descriptors on radiology reports, demonstrated better performance than conventional risk models in a prospective validation cohort from Vanderbilt University Medical Centre. It received FDA clearance for the evaluation of incidentally detected solid and semi-solid pulmonary nodules in patients aged ≥35 years [80].

Despite these encouraging results, a critical challenge is the limited explainability of outputs from complex deep-learning systems (black-box models) [73]. Inadequate data availability for training and heterogeneity of clinical practice data (e.g. EHRs) is another impediment to data mining and integration from multiple sources [75]. Nevertheless, the use of ontologies as a well-defined collective reference for all data sources has provided opportunity for knowledge distillation from semantically interoperable data, which enables standardized disease-specific approaches to AI-integrated decision-support systems in clinical practice [81]. Furthermore, the application of knowledge graphs and reranking to AI-driven recommendation models may improve reasoning and multitask learning potential in cancer research [82].

22.3.4. Takeaways

AI technologies are emerging as ways to enable clinicians to efficiently and systematically search published literature and clinical trial databases to obtain focused and reliable information for complex queries.

Emerging AI platforms offer the opportunity for interactive data visualization, evidence synthesis, insight generation, guideline development, and even scientific writing.

AI-based data extraction from clinical notes is gaining prominence in oncology practice, with promising applications in cancer screening, biomarker association discovery, tumour staging and risk stratification, treatment landscape scanning, and exploration of the patient experience.

AI-powered tools can assist urologists and oncologists in point-of-care decision-making by providing evidence-informed recommendations based on big data sources in oncology.

Integration of AI systems in clinical practice is, however, hindered by important factors, such as data limitations, automation bias, limited explainability, and inadequate AI literacy.

22.4. Enablers and barriers to the application of AI-driven systems in healthcare

22.4.1. Educational needs of the medical workforce

With the exponential growth in medical knowledge, physicians are

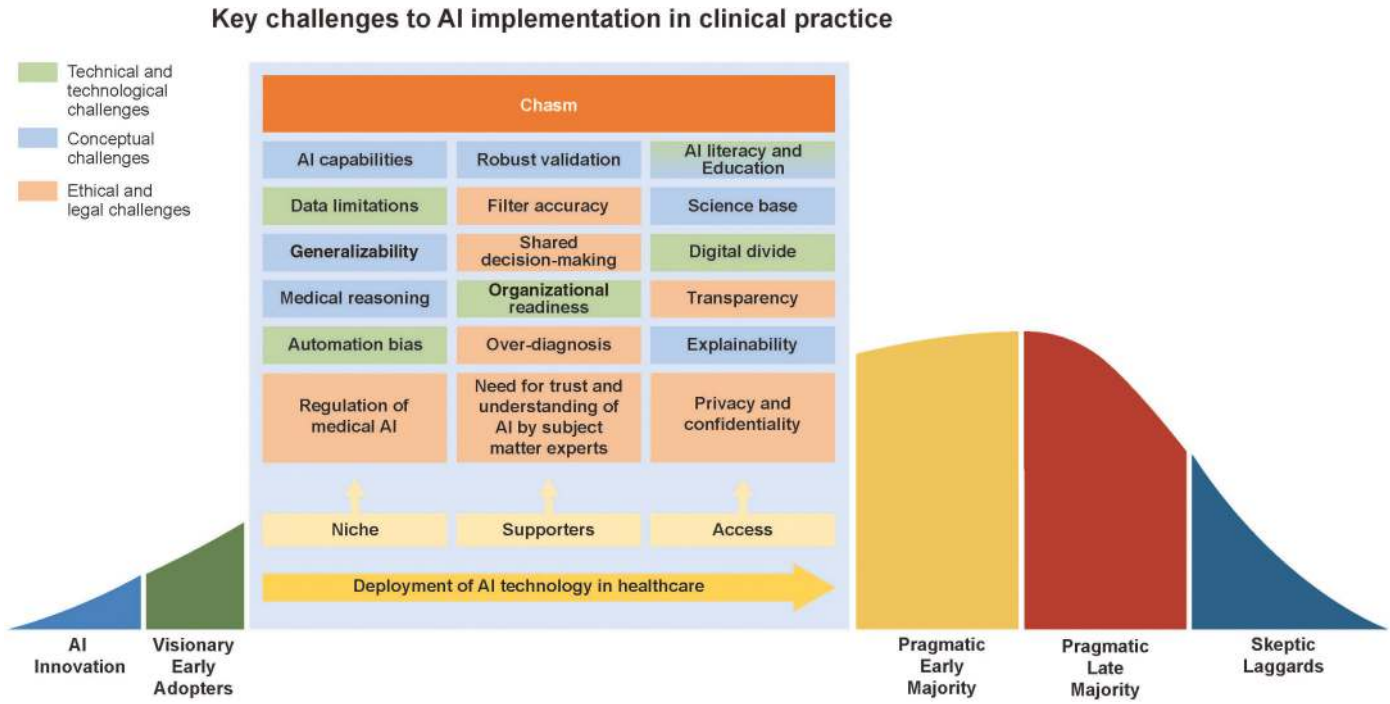


Figure 22.4. A graphical summary of the barriers to the wider implication of AI in the clinical setting based on the technology adoption life cycle model from Rogers' Diffusion of Innovations [83]. The 'chasm' is depicted as a set of technical/technological (green), conceptual (blue), and ethical/legal (peach) challenges that limit access to the use of AI technologies within healthcare niches by HCPs [8,73,84–89]. *Source:* Adapted from 'Crossing the chasm: a "tube-map" for agent-based social simulation of policy scenarios in spatially-distributed systems' by Polhill, JG et al. GeoInformatica. 2019; 23:169–199 [90]. AI: artificial intelligence.

organize, retain, and utilize this information in their practices. However, there is a gap in clinicians' knowledge of AI and its various applications, which puts present and future physicians at a disadvantage. The implementation of AI-based technologies in the clinic requires informed clinician–machine interaction to enhance evidence-based clinical decision-making. Thus, medical practitioners need to develop a good understanding of fundamental AI functions and their practical implications in healthcare, as well as the shortfalls, ethical considerations, and regulatory oversight [91]. The transition to an AI-assisted clinical practice can also be challenging because of resistance to change. This is reinforced by difficulties in changing curricula, and scepticism in the trustworthiness of AI-powered predictions and recommendations [84]. In the absence of evidence-based interventions, a multi-tiered approach to the inclusion of AI across medical education through changes in the national medical education system, institutional programmes, medical school curricula, and individual student and practitioner engagement would be a step forward in addressing these challenges [92].

22.4.2. Barriers to AI implementation in healthcare

At present, there is a chasm preventing the deployment of AI technology in clinical practice (Figure 22.4) [8,14,84–89]. There are conceptual, technical, and humanistic challenges to patient–practitioner trust in medical AI [85]. AI systems currently lack medical reasoning capabilities when using data from a specific patient population to make decisions about patients with different characteristics. Incorrect conceptualization of the medical problem by AI tools adds to this confusion [85]. Use of EHRs for selecting control cohorts, incomplete and non-representative training datasets, and data drifts in AI models and clinical decision-support systems may impair model performance, and produce biased and non-generalizable outputs, with potentially unfavourable patient outcomes [73]. Over-reliance of clinicians on AI algorithms may intensify biases and result in over-testing, overtreatment or undertreatment, or have other implications for clinical decision-making [85]. Lack of legal liability for medical AI and fear of over-reliance among the community, patients, and caregivers further impact the trustworthiness of AI in healthcare settings [84]. We foresee, however, that future legal liability may occur for clinicians who will not use robustly validated AI systems. Lack of transparency in the design and functioning of AI technologies is contrary to patient-centred values of patient autonomy and informed decision-making, privacy, confidentiality, and fairness [14,85,87]. It is critical to address these key challenges using a transdisciplinary approach. For instance, groups of AI experts, HCPs, patient advocates, and medical ethicists can collaborate to develop AI algorithms [85]. On this note, the FDA has released discussion papers on the use of AI and ML technologies in medical product development and drug manufacturing to advance stakeholder engagement and collaboration in this field [93].

There is also increasing demand by clinicians and researchers for transparency to address the black-box nature and incomprehensibility of AI data and models. Multiple technical approaches have been used to get insight into deep-learning models, such as DeepLIFT and Shapley additive explanations [94]. However, most of these explainability techniques only describe how AI models function without providing insights into the validity of their decision-making process [86]. In the absence of suitable explainability methods, rigorous performance evaluation of AI models in RCTs

is crucial for establishing a scientific basis for medical AI [85,86]. This evidence base is growing and several trials have evaluated the effectiveness of AI-based interventions [95]. In conjunction, RWE, such as EHRs, can be leveraged to measure the impact of AI in clinical practice [88]. Additionally, European Parliament is working towards approval of the AI Act, the world's first transparency and risk-management legislation for AI systems [96].

Specific challenges also remain on the application of AI methods in literature search and clinical trial assessment. Many AI models are not trained on full-text articles and lack the capacity to read and extract information from full-text documents or rank the quality of source information. The inclusivity and diversity gap in the literature or clinical trial data for training AI tools is an additional source of bias, which limits data quality and accuracy [73]. Moreover, publisher licences for full-text data mining are mostly restricted to research. Also, there are concerns over the originality of AI-driven manuscripts and non-human authorship, capability of AI systems in making complex decisions, risk of bias in AI-training data, and generation of misinformation; with potential for abuse, lack of appropriate governance, the possibility of AI sentience, and the consequences of deviation from human intelligence for human society [97,98].

22.5. Conclusions and future outlook

AI is an evolving area in oncology that holds promise for the synthesis and automation of literature, trial databases, and EHR searches. AI exhibits broader potential for clinical QA, insight generation, guideline development, interactive data visualization, and decision-making support. However, a wider application of AI tools in this setting requires multi-tiered approaches to tackle existing technical, conceptual, and ethical challenges. There is ongoing rapid progress in AI research, including conversational AI, smart search, and generative QA systems. With these advancements, future AI technologies will be able to extract more complex clinical concepts and specific text from a large body of documents, perform advanced systematic literature reviews, generate novel hypotheses, develop auto-updating oncology guidelines, act as intelligent teaching systems for HCPs, assist in medical writing and peer review, produce ready-to-use plain-language summaries from scientific publications, and help medical practitioners find accurate answers to their complex clinical queries more efficiently. As the use of AI technologies continues to grow, their capabilities will take different directions, and it will be difficult to forecast all future uses, challenges, and misuses of these systems, and their ultimate impact on society.

Acknowledgements

Medical writing and editorial support were provided by Roham Sadeghimakki, MD, PhD; Julie B. Stimmel, PhD, CMPP™, and Rosie Henderson, MSc, all of Onyx (a division of Prime, London, UK) and funded by Pfizer Inc. according to Good Publication Practice guidelines (https://www.tandfonline.com/doi/full/10.1080/03007995.2024.2320298). The sponsor was involved in the study design, collection, analysis, and interpretation of information provided in the manuscript. However, ultimate responsibility for opinions, conclusions, and data interpretation lies with the authors.

Disclosures/conflicts of interests

Arnulf Stenzl reports no current advisory or consultancy roles. Jennifer Ghith is a former employee and shareholder of Pfizer Inc., and a current employee and shareholder of GlaxoSmithKline. Bob J.A. Schijvenaars is an employee of Digital Science & Research Ltd.

Funding

Sponsored by Pfizer Inc.

Author contributions

All authors were involved in the conception and design of the work; data acquisition, analysis, and interpretation; drafting and critically revising the book chapter for important intellectual content; and final approval of the version to be published.

Glossary

Term	Description
Artificial intelligence (AI)	Refers to systems or machines that mimic human intelligence to perform tasks and can iteratively improve themselves based on the information that they collect
Conversational AI	A set of AI technologies, such as chatbots and virtual assistants, that utilize natural language processing and machine learning to interact with human users in a natural way
Deep learning	A subfield of machine learning. Machines are trained on the principles of the human brain, where neurons are triggered by an activity they have previously learned and a path of decision-making is chosen. For example, when humans see a burning surface, our brain immediately relays information not to touch it as it would cause severe burns. The neurons trained on predicting such an event are triggered and a path is chosen from there
Generative AI	A subfield of AI that focuses on generating new data rather than analysing and categorizing existing data
Information extraction	Involves the identification of key snippets of information from structured texts

Term	Description
Information retrieval	Refers to the process, methods, and procedures of searching, locating, and retrieving data or information from a database or set of documents. An information retrieval system is a set of algorithms that facilitate the relevance of displayed documents to searched queries
Language model	A machine-learning model designed to represent a language domain. It can be used for a number of semantically oriented tasks, e.g. extraction of entities or relations, classification of text, or translations. Generative language models are able to generate text, which enables question and answer sessions with users
Machine learning (ML)	A type of AI where machines are trained to learn from historical data and make future predictions or decisions afterwards without being explicitly programmed to perform the task
Natural language processing (NLP)	An area of AI that focuses on problems involving the interpretation and understanding of free text by a non-human system. NLP allows machines to understand human language as it is spoken and written
Reinforcement learning from human feedback	A technique in ML for training large language models (LLMs) that involve incorporating a small amount of feedback from the human evaluator into the learning process to optimize the output of a language model
Semantics	Refers to the meaning that is conveyed by a text
Semantic transformation	A method that enables data sources to be converted into target generic data models by providing abstract mapping rules, which eliminate the effort expended in resolving specific database specifications
Supervised learning and self-supervised learning	Training an ML model using labelled data. This can be data labelled manually but also automatically (e.g. in a teacher–student setting where both are ML models). In the latter case, this is called self-supervised learning
Text clustering	Automated categorization of unlabelled texts into natural clusters (groups) of interest, where texts in the same group are more similar to each other than to those in other groups
Transfer learning	An ML technique that reuses a model trained on one task as a starting point for building a model for a second, related task
Transformer architecture	A transformer takes an input sequence of text (e.g. a question) and produces an output text (e.g. an answer to a question). Transformer architecture is an extension of deep learning that employs a mechanism known as attention. Attention focuses on parts of input sequences that are more relevant to understanding a word, both within its own sequence (self-attention) and with respect to the target output sequence (encoder–decoder attention)
Unsupervised learning	Training an ML model without using labelled data. Self-learning is an example of unsupervised learning. It avoids the bottleneck of the need for annotated data, which is often expensive and time-consuming to produce

REFERENCES

1. Maslej N, Fattorini L, Perrault R, Gil Y, Parli V, Kariuki, N, et al. The AI Index 2025 Annual Report. AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA. 2025.
2. Fraser N, Brierley L, Dey G, Polka JK, Pálfy M, Nanni F, et al. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. PLoS Biol. 2021;**19**(4):e3000959.
3. Gopal G, Suter-Crazzolara C, Toldo L, Eberhardt W. Digital transformation in healthcare-architectures of present and future information technologies. Clin Chem Lab Med. 2019;**57**(3):328–35.
4. Sbaffi L, Walton J, Blenkinsopp J, Walton G. Information overload in emergency medicine physicians: a multisite case study exploring the causes, impact, and solutions in four North England National Health Service Trusts. J Med Internet Res. 2020;**22**(7):e19126.
5. Daei A, Soleymani MR, Ashrafi-Rizi H, Zargham-Boroujeni A, Kelishadi R. Clinical information seeking behavior of physicians: a systematic review. Int J Med Inform. 2020;**139**:104144.
6. Schuers M, Griffon N, Kerdelhue G, Foubert Q, Mercier A, Darmoni SJ. Behavior and attitudes of residents and general practitioners in searching for health information: from intention to practice. Int J Med Inform. 2016;**89**:9–14.
7. Klerings I, Weinhandl AS, Thaler KJ. Information overload in healthcare: too much of a good thing? Z Evid Fortbild Qual Gesundhwes. 2015;**109**(4–5):285–90.
8. Ciampaglia GL, Nematzadeh A, Menczer F, Flammini A. How algorithmic popularity bias hinders or promotes quality. Sci Rep. 2018;**8**:15951.
9. Cohen GR, Friedman CP, Ryan AM, Richardson CR, Adler-Milstein J. Variation in physicians' electronic health record documentation and potential patient harm from that variation. J Gen Intern Med. 2019;**34**(11):2355–67.
10. Jensen JD, Liu M, Carcioppolo N, John KK, Krakow M, Sun Y. Health information seeking and scanning among US adults aged 50–75 years: testing a key postulate of the information overload model. Health Inform J. 2017;**23**(2):96–108.
11. Stenzl A, Sternberg CN, Ghith J, Serfass L, Schijvenaars BJA, Sboner A. Application of artificial intelligence to overcome clinical information overload in urological cancer. BJU Int. 2022;**130**(3):291–300.
12. Aakre CA, Pencille LJ, Sorensen KJ, Shellum JL, Del Fiol G, Maggio LA, et al. Electronic knowledge resources and point-of-care learning: a scoping review. Acad Med. 2018;**93**(11S):S60–7.
13. Shao D, Dai Y, Li N, Cao X, Zhao W, Cheng L, et al. Artificial intelligence in clinical research of cancers. Brief Bioinform. 2022;**23**(1):bbab523.
14. Lotter W, Hassett MJ, Schultz N, Kehl KL, Van Allen EM, Cerami E. Artificial intelligence in oncology: current landscape, challenges, and future directions. Cancer Discov. 2024;**14**(5):711–26.
15. Osmanovic A. We asked GPT-3 to write an academic paper about itself—then we tried to get it published. Scientific American. 2022. Available from: https://www.scientificamerican.com/arti

cle/we-asked-gpt-3-to-write-an-academic-paper-about-itself-then-we-tried-to-get-it-published/.
16. Pease A. Google steps into the chat AI ring with Bard, anthropic investment. Aragon Research. 2023. Available from: https://aragonresearch.com/google-steps-into-the-chat-ai-ring-with-bard/.
17. Warren T. Microsoft is looking at OpenAI's GPT for Word, Outlook, and PowerPoint. The Verge. 2023. Available from: https://www.theverge.com/2023/1/9/23546144/microsoft-openai-word-powerpoint-outlook-gpt-integration-rumor.
18. OpenAI. Introducing ChatGPT. OpenAI. 2022. Available from: https://openai.com/index/chatgpt/#OpenAI.
19. Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. Int J Med Inform. 2019;**125**:37–46.
20. Liu T, Xiao X. A framework of AI-based approaches to improving eHealth literacy and combating infodemic. Front Public Health. 2021;**9**:755808.
21. Loh E. ChatGPT and generative AI chatbots: challenges and opportunities for science, medicine and medical leaders. BMJ Lead. 2023; **8**(1):51–4.
22. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. Healthcare (Basel). 2023;**11**(6):887.
23. Tamine L, Goeuriot L. Semantic information retrieval on medical texts: research challenges, survey, and open issues. ACM Comput Surv. 2021;**54**(7):1–38.
24. Schneider P, Schopf T, Vladika J, Galkin M, Simperl E, Matthes F. A decade of knowledge graphs in natural language processing: a survey. In: Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (volume 1: long papers), online only. Association for Computational Linguistics; 2022: 601–14.
25. Zou D, Ma L, Yu J, Zhang Z. Biological databases for human research. Genomics Proteomics Bioinformatics. 2015;**13**(1):55–63.
26. Utomo FS, Suryana N, Azmi MS. Question answering system: a review on question analysis, document processing, and answer extraction techniques. J Theor Appl Inf Technol. 2017;**95**(14):3158–74.
27. Wiggers K. The emerging types of language models and why they matter. TechCrunch. 2022. Available from: https://techcrunch.com/2022/04/28/the-emerging-types-of-language-models-and-why-they-matter/.
28. Jeong DP, Garg S, Lipton ZC, Oberst M. Medical adaptation of large language and vision-language models: are we making progress? In: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics; 2024:12143–70.
29. Huang H. Second version of Pixar character WALL-E & Salvador Dalí: introduction. 2022. Available from: https://www.researchgate.net/publication/362427293_Second_Version_of_Pixar_character_WALL-E_Salvador_Dali_Introduction.
30. Louis A. A brief history of natural language processing—Part 1. Medium. 2020. Available from: https://medium.com/@antoine.louis/a-brief-history-of-natural-language-processing-part-1-ffbcb937ebce.
31. Souza Dos Reis E, Da Costa CA, Da Silveira DE, Bavaresco RS, Da Rosa Righi R, Victória Barbosa JL, et al. Transformers aftermath: current research and rising trends. Commun ACM. 2021;**64**(4):154–63.
32. Wei J. BERT: why it's been revolutionizing NLP. Medium. 2020. Available from: https://medium.com/data-science/bert-why-its-been-revolutionizing-nlp-5d1bcae76a13.
33. Yang X, Chen A, PourNejatian N, Shin HC, Smith KE, Parisien C, et al. GatorTron: a large clinical language model to unlock patient information from unstructured electronic health records. medRxiv. 2022:2022.02.27.22271257.
34. Collins E, Ghahramani Z. LaMDA: our breakthrough conversation technology. Google. 2021. Available from: https://blog.google/technology/ai/lamda/.
35. Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. Nature. 2023;**613**(7945):620–21.
36. International Committee of Medical Journal Editors. Up-Dated ICMJE Recommendations (May 2023). ICMJE. 2023. Available from: https://www.icmje.org/news-and-editorials/updated_recommendations_may2023.html
37. Thorp HH. ChatGPT is fun, but not an author. Science. 2023;**379**(6630):313.
38. Gao CA, Howard FM, Markov NS, Dyer EC, Ramesh S, Luo Y, Pearson AT. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. NPJ Digit Med. 2023:**6**(1);75.
39. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. Nature. 2023;**614**(7947):214–16.
40. Schoeb D, Suarez-Ibarrola R, Hein S, Dressler FF, Adams F, Schlager D, Miernik A. Use of artificial intelligence for medical literature search: randomized controlled trial using the hackathon format. Interact J Med Res. 2020;**9**(1):e16606.
41. Wager K, Chari D, Ho S, Rees T, Penner O, Schijvenaars BJ. Identifying and validating networks of oncology biomarkers mined from the scientific literature. Cancer Inform. 2022;**21**:11769351221086441.
42. Auffenberg GB, Ghani KR, Ramani S, Usoro E, Denton B, Rogers C, et al. askMUSIC: leveraging a clinical registry to develop a new machine learning model to inform patients of prostate cancer treatments chosen by similar men. Eur Urol. 2019;**75**(6):901–7.
43. Omar MI, Roobol MJ, Ribal MJ, Abbott T, Agapow P-M, Araujo S, et al. Introducing PIONEER: a project to harness big data in prostate cancer research. Nat Rev Urol. 2020;**17**(6):351–62.
44. CORDIS. Optimal treatment for patients with solid tumours in Europe through Artificial intelligence. European Commission. 2022. Available from: https://cordis.europa.eu/project/id/101034347.
45. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;**596**(7873):583–89.
46. Kirchner JH, Ahmad L, Aaronson S, Leike J. New AI classifier for indicating AI-written text. OpenAI. 2023. Available from: https://openai.com/index/new-ai-classifier-for-indicating-ai-written-text/.
47. Nagpal K, Foote D, Tan F, Liu Y, Chen P-HC, Steiner DF, et al. Development and validation of a deep learning algorithm for Gleason grading of prostate cancer from biopsy specimens. JAMA Oncol. 2020;**6**(9):1372–80.
48. Huang KA, Choudhary HK, Kuo PC. Artificial intelligent agent architecture and clinical decision-making in the healthcare sector. Cureus. 2024;**16**(7):e64115.

49. Devnath L, Janzen I, Lam S, Yuan R, MacAulay C. Predicting future lung cancer risk in low-dose CT screening patients with AI tools. In: Proceedings of SPIE Medical Imaging 2025: Computer-Aided Diagnosis. SPIE. 2025:134072L.
50. Howard BE, Phillips J, Miller K, Tandon A, Mav D, Shah MR, et al. SWIFT-Review: a text-mining workbench for systematic review. Syst Rev. 2016;**5**:87.
51. Wallace BC, Noel-Storr A, Marshall IJ, Cohen AM, Smalheiser NR, Thomas J. Identifying reports of randomized controlled trials (RCTs) via a hybrid machine learning and crowdsourcing approach. J Am Med Inform Assoc. 2017;**24**(6):1165–68.
52. Soto AJ, Przybyła P, Ananiadou S. Thalia: semantic search engine for biomedical abstracts. Bioinformatics. 2019;**35**(10):1799–801.
53. Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA. Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In: Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium. Association for Computing Machinery. 2012:819–24.
54. Przybyła P, Brockmeier AJ, Kontonatsios G, Le Pogam M-A, McNaught J, von Elm E, et al. Prioritising references for systematic reviews with RobotAnalyst: a user study. Res Synth Methods. 2018;**9**(3):470–88.
55. Wallace BC, Kuiper J, Sharma A, Zhu MB, Marshall IJ. Extracting PICO sentences from clinical trial reports using supervised distant supervision. J Mach Learn Res. 2016;**17**:132.
56. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. J Am Med Inform Assoc. 2016;**23**(1):193–201.
57. Sterner E. scite: providing additional context for citation statements. Issues Sci Technol Librariansh. 2021;**98**:doi:10.29173/istl74.
58. scite. Ask a question, get answers directly from research article. scite; 2022. Available from: https://scite.ai/blog/2022-11-16_ask_a_question_get_answers_directly.
59. Bao Y, Deng Z, Wang Y, Kim H, Armengol VD, Acevedo F, et al. Using machine learning and natural language processing to review and classify the medical literature on cancer susceptibility genes. JCO Clin Cancer Inform. 2019;**3**:1–9.
60. Brahmer JR, Lacchetti C, Schneider BJ, Atkins MB, Brassil KJ, Caterino JM, et al. Management of immune-related adverse events in patients treated with immune checkpoint inhibitor therapy: American Society of Clinical Oncology Clinical Practice Guideline. J Clin Oncol. 2018;**36**(17):1714–68.
61. Be My Eyes. Introducing Be My AI (formerly Virtual Volunteer) for people who are blind or have low vision, powered by OpenAI's GPT-4. Be My Eyes; 2023. Available from: https://www.bemyeyes.com/news/introducing-be-my-ai-formerly-virtual-volunteer-for-people-who-are-blind-or-have-low-vision-powered-by-openais-gpt-4/.
62. OpenAI. Government of Iceland: How Iceland is using GPT-4 to preserve its language. OpenAI; 2023. Available from: https://openai.com/index/government-of-iceland/.
63. Datta S, Bernstam EV, Roberts K. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. J Biomed Inform. 2019;**100**:103301.
64. López-Úbeda P, Martín-Noguerol T, Aneiros-Fernández J, Luna A. Natural language processing in pathology: current trends and future insights. Am J Pathol. 2022;**192**(11):1486–95.
65. Chen Y, Hao L, Zou VZ, Hollander Z, Ng RT, Isaac KV. Automated medical chart review for breast cancer outcomes research: a novel natural language processing extraction system. BMC Med Res Methodol. 2022;**22**(1):136.
66. Zeng Z, Li X, Espino S, Roy A, Kitsch K, Clare S, et al. Contralateral breast cancer event detection using nature language processing. AMIA Annu Symp Proc. 2018;**2017**:1885–92.
67. Burnett-Hartman A, Newcomb PA, Zeng CX, Zheng Y, Inadomi JM, Fong C, et al. Abstract PR05: using medical informatics to evaluate the risk of colorectal cancer in patients with clinically diagnosed sessile serrated polyps. Cancer Res. 2017;**77**(3_supplement):PR05.
68. Savova GK, Tseytlin E, Finan S, Castine M, Miller T, Medvedeva O, et al. DeepPhe: a natural language processing system for extracting cancer phenotypes from clinical records. Cancer Res. 2017;**77**(21):e115–18.
69. AlAbdulsalam AK, Garvin JH, Redd A, Carter ME, Sweeny C, Meystre SM. Automated extraction and classification of cancer stage mentions from unstructured text fields in a central cancer registry. AMIA Jt Summits Transl Sci Proc. 2018;**2017**:16–25.
70. Mitchell JR, Szepietowski P, Howard R, Reisman P, Jones JD, Lewis P, et al. A question-and-answer system to extract data from free-text oncological pathology reports (CancerBERT Network): development study. J Med Internet Res. 2022;**24**(3):e27210.
71. National Cancer Institute. NCI-Department of Energy Collaboration. National Cancer Institute; 2024. Available from: https://datascience.cancer.gov/collaborations/nci-department-energy-collaboration.
72. Zou KH, Li JZ, Imperato J, Potkar CN, Sethi N, Edwards J, Ray A. Harnessing real-world data for regulatory use and applying innovative applications. J Multidiscip Healthc. 2020;**13**:671–79.
73. Shreve JT, Khanani SA, Haddad TC. Artificial intelligence in oncology: current capabilities, future opportunities, and ethical considerations. Am Soc Clin Oncol Educ Book. 2022;**42**:1–10.
74. Blacketer C. Chapter 4: The Common Data Model. In: The Book of OHDSI: Observational Health Data Sciences and Informatics [Internet]. OHDSI; 2021. Available from: https://ohdsi.github.io/TheBookOfOhdsi/CommonDataModel.html.
75. Adlung L, Cohen Y, Mor U, Elinav E. Machine learning in clinical decision making. Med. 2021;**2**(6):642–65.
76. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020;**3**:17.
77. Sapkota R, Roumeliotis KI, Karkee M. AI agents vs. agentic AI: a conceptual taxonomy, applications and challenges. arXiv. 2025;2505.10468.
78. Walsh S, de Jong EEC, van Timmeren JE, Ibrahim A, Compter I, Peerlings J, et al. Decision support systems in oncology. JCO Clin Cancer Inform. 2019;**3**:1–9.
79. Jie Z, Zhiying Z, Li L. A meta-analysis of Watson for Oncology in clinical application. Sci Rep. 2021;**11**(1):5792.
80. Massion PP, Antic S, Ather S, Arteta C, Brabec J, Chen H, et al. Assessing the accuracy of a deep learning method to risk stratify indeterminate pulmonary nodules. Am J Respir Crit Care Med. 2020;**202**(2):241–49.
81. Benedict SH, Hoffman K, Martel MK, Abernethy AP, Asher AL, Capala J, et al. Overview of the American Society for Radiation Oncology-National Institutes of Health-American Association of Physicists in Medicine Workshop 2015: exploring opportunities for radiation oncology in the era of big data. Int J Radiat Oncol Biol Phys. 2016;**95**(3):873–79.
82. Gogleva A, Polychronopoulos D, Pfeifer M, Poroshin V, Ughetto M, Martin MJ, et al. Knowledge graph-based recommendation

framework identifies drivers of resistance in EGFR mutant non-small cell lung cancer. Nat Commun. 2022;**13**(1):1667.
83. Sahin I. Detailed review of Rogers' diffusion of innovations theory and educational technology-related studies based on Rogers' theory. Turkish J Educational Tech. 2006;**5**(2):14–23.
84. Grunhut J, Wyatt AT, Marques O. Educating future physicians in artificial intelligence (AI): an integrative review and proposed changes. J Med Educ Curric Dev. 2021;**8**:doi: 10.1177/23821205211036836.
85. Quinn TP, Senadeera M, Jacobs S, Coghlan S, Le V. Trust and medical AI: the challenges we face and the expertise needed to overcome them. J Am Med Inform Assoc. 2021;**28**(4):890–94.
86. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;**3**(11):e745–50.
87. Grote T, Berens P. On the ethics of algorithmic decision-making in healthcare. J Med Ethics. 2020;**46**(3):205–11.
88. Marwaha JS, Kvedar JC. Crossing the chasm from model performance to clinical impact: the need to improve implementation and evaluation of AI. NPJ Digit Med. 2022;**5**(1):25.
89. Matheny M, Thadaney Israni S, Ahmed M, Whicher D, Editors. Artificial intelligence in health care: the hope, the hype, the promise, the peril. Washington, DC: National Academy of Medicine; 2019.
90. Polhill JG, Ge J, Hare MP, Matthews KB, Gimona A, Salt D, Yeluripati J. Crossing the chasm: a 'tube-map' for agent-based social simulation of policy scenarios in spatially-distributed systems. Geoinformatica. 2019;**23**:169–99.
91. Lomis K, Jeffries P, Palatta A, Sage M, Sheikh J, Sheperis C, Whelan A. Artificial intelligence for health professions educators. NAM Perspect. 2021;**2021**:10.31478/202109a.
92. Grunhut J, Marques O, Wyatt ATM. Needs, challenges, and applications of artificial intelligence in medical education curriculum. JMIR Med Educ. 2022;**8**(2):e35587.
93. Cavazzoni P. FDA releases two discussion papers to spur conversation about artificial intelligence and machine learning in drug development and manufacturing. US Food & Drug Administration; 2023. Available from: https://www.fda.gov/news-events/fda-voices/fda-releases-two-discussion-papers-spur-conversation-about-artificial-intelligence-and-machine.
94. Linardatos P, Papastefanopoulos V, Kotsiantis S. Explainable AI: a review of machine learning interpretability methods. Entropy (Basel). 2020;**23**(1):18.
95. Zhou Q, Chen Z-H, Cao Y-H, Peng S. Clinical impact and quality of randomized controlled trials involving interventions evaluating artificial intelligence prediction tools: a systematic review. NPJ Digit Med. 2021;**4**(1):154.
96. European Parliament. AI Act: A step closer to the first rules on Artificial Intelligence [press release]. European Parliament; 2023. Available from: https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence
97. Johnson S. A.I. Is mastering language. Should we trust what it says? The New York Times Magazine; 2022. Available from: https://www.nytimes.com/2022/04/15/magazine/ai-language.html.
98. Snoswell AJ, Burgess J. The Galactica AI model was trained on scientific knowledge – but it spat out alarmingly plausible nonsense. The Conversation; 2022. Available from: https://theconversation.com/the-galactica-ai-model-was-trained-on-scientific-knowledge-but-it-spat-out-alarmingly-plausible-nonsense-195445.

23

Application of artificial intelligence in cancer genomics

Xiwei Wu and Supriyo Bhattacharya

23.1. Introduction

Cancer is a complex and heterogenous disease, which can be caused by genetic and epigenetic abnormalities. Identification of these abnormalities in cancer patients not only helps researchers to understand the tumorigenesis process but also guides clinicians on selection of the best treatment. As the genome technologies advanced rapidly in recent years, particularly high-throughput sequencing, single-cell genomics, and spatial transcriptomics (ST), a tremendous amount of genomic data have been generated using specimens from patients in various cancers. The embedded disease signatures within genomic data are often too complex and noise-ridden to be amenable to classical statistical treatment. These large volumes of data, together with clinical records, enabled the development and application of artificial intelligence (AI) methods in cancer research and clinical practice. The applications of AI range from diagnosis and prognosis of cancer, molecular characterization of tumours and their microenvironment, to predicting therapeutic response for patients. In this chapter, we begin by introducing the genomic data types and AI methods. We then focus on summarizing the AI methods for analysis of cancer genomics data as well as its application in cancer biology and clinical practice, mainly published in the past five years. We then discuss the challenges and future perspectives of AI in cancer genomics.

23.2. Genomic data types

Depending on the materials used in the assay, either DNA or RNA, different modalities can be appreciated (**Figure 23.1**). RNA samples can be used to interrogate coding gene expression, and long non-coding RNA expression, microRNA expression, and other small RNA species, including piwi-interacting RNA, small cytoplasmic RNA (scRNA), small nucleus RNA, as well as transfer RNA, can be detected when appropriate library preparation protocols are used. For DNA sequencing, one can obtain single-nucleotide variants (SNVs), small insertions/deletions (INDELs), copy number variants (CNVs), and large structural variants (SVs), which include large insertion, large deletion, translocation, inversion, tandem repeats, etc. These variants can be detected at genome level via whole genome sequencing (WGS) or at targeted regions via whole-exome sequencing (WES) and amplicon sequencing. Chromatin accessibility and three-dimensional interactions, histone modification, and transcription factor binding can also be measured with genomic DNAs enriched for these specific regions of interest. DNA methylation or hydroxymethylation at cytosine residues can be detected at single base resolution with different variants of bisulphite sequencing.

Depending on the goals of the study, samples can be collected from paired tumour and normal tissues of the same patient, or before and after treatment. Liquid biopsy, with cell-free DNA and RNA samples collected from blood and other body fluids, offers a convenient and non-invasive way for early cancer detection and monitoring. Mouse patient-derived xenograft model and primary tissue culture, such as organoid, are also common for easier perturbation.

Cancer is very heterogenous, which frequently includes multiple clones accumulated during evolution, and its microenvironment can also be very diversified, such as tumour infiltrating lymphocytes, myeloid cells, and fibroblasts. Bulk sequencing measures the average expression of variations of all cells in the tumour, providing suboptimal view that is hard to interpret due to mixed cell types and multiple tumour clones. As the single-cell genomics becomes mature and sequencing cost dramatically reduced, it has been increasingly adopted in cancer centres and research communities to study cancer with more granularity. Tumour heterogeneity and tumour microenvironment (TME) can be more appreciated using single-cell technology. However, the tissue context is lost in a typical single-cell analysis, which can be mitigated by ST. Indeed, spatial transcriptomics is so instrumental that it was selected as the technology of the year by *Nature* in 2021 [2].

23.3. Overview of ML methods

ML methods can be classified as unsupervised, supervised, and reinforcement learning. Unsupervised learning refers to the

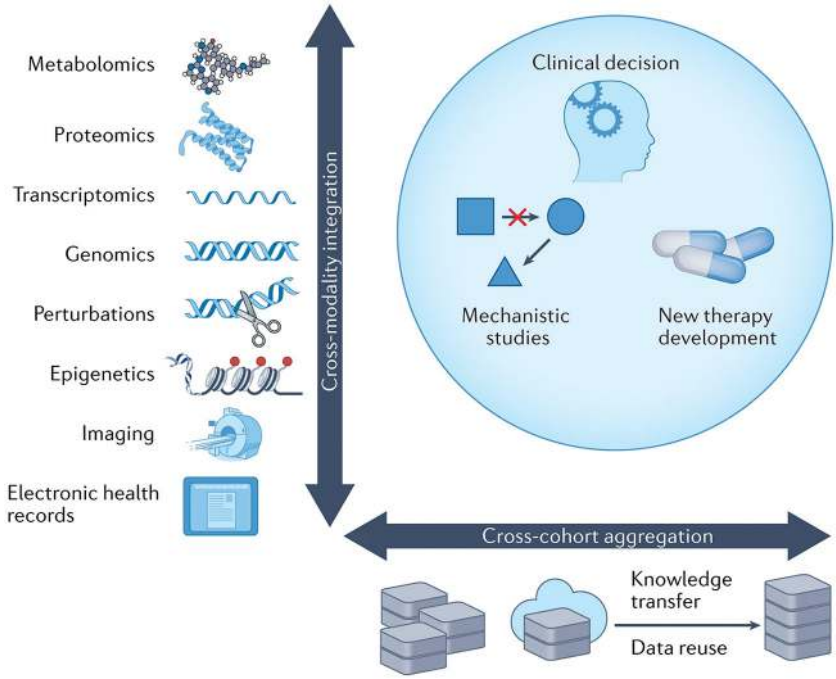


Figure 23.1. Overview of AI applications in cancer genomics. *Source:* Reproduced from [1].

identification of patterns or features by analysing large unlabelled datasets. The goal of unsupervised learning can be, for example, to discover a set of key features, which is then used to reconstruct the input data or to arrange the input data into clusters based on similarity. In contrast with unsupervised learning, supervised learning uses labelled datasets for training purposes that enable the system to classify new data according to learned categories. However, the capacity of supervised learning is limited by the quality and extent of labelled data. In practice, such data may be expensive to annotate, or in some cases, the labels themselves could be uncertain or error prone. Often, the majority of the available training data is unlabelled, with only a small percentage correctly labelled. In such scenarios, therefore, the optimal approach involves combining both supervised and unsupervised learning (i.e., semi-supervised learning) to maximize the performance of the classifier. Another efficient approach called transfer learning can harness an established ML model (source) for a related application (target), where sufficient training data is unavailable or training the model from scratch is time-consuming. Compared to the above approaches, reinforcement learning is less relevant to cancer genomics space.

DL is a special branch of ML that employs artificial neural networks (ANNs). DL agents are comprised of ANNs having many hidden layers (source of the phrase 'deep') that are inspired by the neuronal architecture of the human brain. DL models are capable of identifying key features from large and complex datasets that represent the discriminative properties for tasks such as prediction, classification, or synthetic data generation. This approach, popularly called representation learning [3], is a major strength of DL that separates it from traditional ML methods, which rely on human intervention in feature extraction and are therefore unsuitable for complex and very large datasets or in cases where the training data changes over time. Therefore, DL has found unprecedented success in many fields, such as natural language processing, image analysis, clinical informatics, bioinformatics, and robotics.

23.4. Application of AI in the analysis of cancer genomics data4.1: SNVs and short INDELs

Variant detection is a common task in analysing cancer genome or exome sequencing. Both germline and somatic variants are of interest to understand the mechanisms of tumorigenesis. Besides simple filtering and probabilistic models, AI methods have been popular in variant detection applications. For example, Cerebro uses a specialized random forest classification model that evaluates a large set of decision trees to generate a confidence score for each candidate variant [4]. Several ensemble-based ML approaches have been developed to improve somatic mutation calling accuracy. SMuRF combines predictions from four mutation callers with auxiliary alignment and mutation features using random forest classifier [5]. NeoMutate incorporates seven supervised ML algorithms to exploit the strengths of multiple variant callers, using a non-redundant set of 17 biological and sequence features [6].

DeepVariant is the first attempt to use DL approach for germline SNV calling. The algorithm takes advantage of the robustness of convolutional neural network (CNN) in detecting variants by learning statistical relationships between images of read pileups around putative variant and true genotype calls [7]. HELLO implements meticulously designed deep neural network architectures and customized variant inference functions [8]. It accounts for the underlying nature of sequencing data instead of converting the problem to one of image recognition, with reduced error rate and much smaller model compared to DeepVariant [8].

Deep learning (DL) has also been applied to somatic mutation calling. For example, VarNet is an end-to-end DL approach for the identification of somatic variants from aligned tumour and matched normal DNA reads [9]. Feedforward neural network was evaluated to automate the somatic mutation refinement steps, and it matches or outperforms the time-consuming manual refinement for several cancer types [10].

23.4.1. CNVs

Genome instability is one of the characteristics of many cancers, and the associated CNVs and SVs are as important as other aberrations in cancer diagnosis and prognosis. Although WGS is more affordable now with the rapidly reduced sequencing cost, WES is far more common. Copy number calls based on WES data usually contain false positives due to the noncontiguous nature of the targeted capture. CN-learn, a random-forest-based ensemble approach, was developed to improve the performance of four individual CNV callers [11]. The authors showed that ~58% of all true CNVs recovered by CN-Learn were either singletons or calls that lacked support from at least one caller [11]. After trained using paired WES and WGS 1000 Genomes Project data, DECoNT, a DL model with single hidden layered Bi-LSTM architecture efficiently tripled the duplication call precision and doubled the deletion call precision of the state-of-the-art algorithms [12].

23.4.2. Single-cell RNA-seq (scRNA-seq)

As single-cell genomics became mature in the past few years, more researchers have utilized this technology to better understand tumour heterogeneity and TME. Single-cell data have higher variations, undersampling, and noises compared to bulk data, hence preprocessing and analysis of scRNA-seq data are more challenging. Due to its large data size, ML algorithms, particularly DL, have been widely applied to analyse scRNA-seq data.

Dropout is very common in single-cell data, and it can significantly affect downstream bioinformatics analysis as it decreases the sensitivity and introduces biases into the data. Currently, several machine learning (ML) imputation algorithms have been proposed. Besides LASSO regression model used by ScImpute [13] and SAVER [14], autoencoder (AE) becomes the natural choice because of its data regeneration and denoising properties. For example, AutoImpute uses an overcomplete AE trained on non-zero entries for the imputation of dropout locations [15]. scScope uses an iterative AE that cycles output into input while applying batch effect correction [16]. DCA builds an AE to model the gene distribution using a zero inflated negative binomial prior [17]. scGNN uses three AEs in a cycle—a graph AE, a plain AE, and a cluster-specific collection of AEs [18], and the convergence of a cell graph is used in the regularization of a final AE that performs the imputation. Other deep NNs have also been explored. scIGANs uses a generative adversarial network (GAN) to model the generation of scRNA-seq data using the generated expression data for the imputation [19]. DeepImpute is a deep-neural-network-based imputation algorithm that uses dropout layers and loss functions to learn patterns in the data, allowing for accurate imputation [20].

Transforming raw count data to a lower-dimensional representation of each cell using dimensional reduction (DR) technique is a useful step to remove technical noise in single-cell data and allows easier visualization. DR belongs to unsupervised ML. While the linear principal component analysis (PCA) on log-transformed count data was initially popular for DR, *t*-distributed stochastic neighbour embedding [21] and uniform manifold approximation and projection (UMAP) [22], which are non-linear, graph-based methods, became widely adopted because they can better represent the high-dimensional and non-linear single-cell data. AE and variational autoencoders (VAEs), very popular classes of DL models, have also been used in DR of single-cell RNA-seq data. These include SAUCIE [23], scVI [24], DCA [17], and scVAE [25]. It is interesting to note that when AE contains single layer with two nodes and uses linear activation function, its reduced dimensions are equivalent to PCA [26].

Annotation of cells or cell clusters is one of the most critical tasks in analysing single-cell RNA-seq data. Classical supervised ML algorithms, such as support vector machine (SVM) [27], K-nearest neighbour [28], neural network [29], and random forest (RF) [30,31], are commonly used. Recently, DL-based algorithms began to gain popularity due to their superior accuracy. ItClust uses stacked AE and transfer learning to learn a target network for clustering and cell-type classification [32]. scSemiGAN is a semi-supervised cell-type annotation and dimensionality reduction framework based on GAN [33]. scMRA uses a graphic convolutional network based on a knowledge graph constructed from multiple reference datasets [34]. scNym is another DL method that combines semi-supervised learning and adversarial neural network [35]. sigGCN incorporates gene interactions using a graph convolutional network (GCN) combined with a neural network for cell-type classification task [36]. scDeepSort provides a pre-trained DL model with a weighted graph neural network (GNN) of cells and genes based on large human and mouse single-cell data [37].

23.4.3. Spatial transcriptomics

Currently, the two most common types of ST approaches are spatial barcoding-based techniques, such as Slide-seq and VISIUM from 10X Genomics, and multiplex imaging-based techniques, such as MERFISH (Multiplexed error-robust Fluorescent *in situ* Hybridization). While barcoding-based techniques can resolve the expressions of tens of thousands of genes, their resolution is limited to several cells per spot. On the other hand, multiplex imaging-based methods can achieve cellular to sub-cellular resolution but only for a limited number (up to 1,000) of genes.

These ST maps can provide valuable insight into tumour heterogeneity, immune cell composition of the microenvironment, as well as cellular interactions. However, proper utilization of ST-derived data is associated with several challenges. In barcoding-based methods, the expression information of a single spot can come from multiple cell types, contributing to uncertainties in assigning cell identities to spatial locations. Also, the amount of RNA captured in each spot is limited, thus affecting the accuracy of low expression genes. In multiplex imaging-based methods, expression information is available only for a limited number of genes. These deficiencies can be addressed by supplementing ST data with single and bulk RNA-seq profiles of the same tissue section, as well as anatomical information obtained from histology staining.

Due to their capacity to extract high level features of gene expression covariation among multiple cells and spatial locations, DL-based methods such as DEEPs [38] and Bulk2space [39] have been successfully applied to impute missing expression data in ST

maps, using reference single-cell RNA-seq profiles. In DEEPsc, the gene expression profiles of ST spots are analysed to derive principal components (PCs), and the scRNA-seq data is projected onto this PC space. These low-dimensional feature vectors are then used to train a neural network model to predict the probability of each cell in the scRNA-seq to have originated from a given spot in the ST map. In contrast, Bulk2space uses generative DL (VAE) to synthesize cellular expression data from a bulk RNA-seq dataset using scRNA-seq of the same tissue as reference. It then uses traditional ML (random forest) to assign cells to each spot in the ST map.

Besides mapping scRNA-seq to ST profiles, one important application of DL is to predict the cell-type composition of each ST spot. Methods such as DestVI [40] and DSTG [41] fall in this category. In DestVI, both the cell-type annotated scRNA-seq and the paired ST profiles are converted into low-dimensional latent variables using encoder neural networks. Then, a decoder network trained on the scRNA-seq data is used to deconvolute each ST spot into the constituent cell types and their gene expressions. In DSTG, the reference scRNA-seq is used to generate a pseudo-ST map by randomly selecting a few cells and combining their gene expressions to create an ST spot. The pseudo-ST map is then compared with the real ST map to generate a similarity graph between the spots in both maps and used as input to a GCN model. By learning the topological structure of the similarity map, the GCN predicts the cell-type composition of each spot in the real ST.

While ST captures the spatial variations of multiple gene expressions, imaging methods such as histology staining can provide highly detailed anatomical features in the tissue. Combining ST with anatomical images could therefore improve the identification of functional spatial features in the tissue, leading to better interpretability and clinical prediction. DL-based methods, such as Tangram [42], spaCell [43], and spGCN [44], have been successfully applied in this area. Tangram is a method that integrates scRNA-seq and anatomical imaging with ST. While the mapping of scRNA-seq to ST maps is achieved through traditional approaches, the alignment of anatomical images to ST is performed using a Siamese network model, which is a specialized CNN to compare pairs of pixel maps to analyse their similarity. spaCell uses DL to predict disease states from histology images and corresponding ST maps. It uses an autoencoder model to convert the ST and histology profiles into low-dimensional latent variables and uses them to train a deep neural network to predict disease states. spaGCN integrates between ST and histological maps using GCN, leading to better annotation of spatial domains, such as malignant cells and immune niches in a tumour sample.

Besides the above examples, further application of ML in spatial genomics is in learning cell–cell communication using ST maps, such as SpaOTsc [45]. Here, the ST and paired scRNA-seq data is employed in an optimal transport model to derive cell–cell distance. Next, the cell distance matrix is used in conjunction with random forest models to construct space-constrained cell–cell communication networks that reflect the dynamics of cellular communication within the tissue.

In future, with the continuous improvement of ST technologies, we expect to see further expansion of ML methods, especially in the field of functional interpretation of spatial data. However, with the increase of spatial resolution and sequencing depth per spot, ML methods need to adapt to the increased volume of data so that a balance could be maintained between precision and computational time.

23.5. Application in clinical oncology

Potential applications of AI-developed signatures in clinical oncology span the entire cancer (**Figure 23.1**). Specifically, gene signatures can be used for screening and early detection. Diagnostic tests can help determine the primary tissue of origin and classify the disease subtype, which is critical for clinicians to make therapeutic decisions. Prognostic tests can be used to assess patient risk and predict survival. During therapy, signatures can be used to predict response or toxicity, which can guide the treatment to target cancer more effectively or prevent severe side effects. Minimum residual disease can be detected by genomic tests to identify early signs of recurrence. Many traditional ML and DL approaches have been exploited in these areas. We cover some examples below, with focus on DL models.

23.5.1. Cancer diagnosis and tissue of origin

Cancer diagnosis includes distinguishing between cancer and normal tissues, molecular subtypes, as well as cancer staging and grading. Multiple genomic data types have been used to classify cancers derived from different cell types.

A DL model trained on the most frequent cancer-specific point mutations obtained from WES profiles in The Cancer Genome Atlas (TCGA) can distinguish between healthy and tumour tissue with high accuracy but did not perform well in a multi-class classification task to distinguish 12 cancer types [46]. Similarly, DeepCues, a DL model that utilizes CNNs for cancer classification, only achieved 77% accuracy [47]. One of the reasons for this rather poor performance is that most genetic aberrations are not specific to a single cancer type and hence cannot distinguish carcinomas of different origins.

In contrast, RNA-seq provides a robust, high-throughput transcriptomic platform that represents tumour tissue type and states, and therefore it is considered suitable for this classification task. There are numerous publications on cancer classifications using RNA-seq data in the past 10 years [48], and we focus on a few recent attempts to further improve the classification performance. In one such attempt, a stacking ensemble DL model based on one-dimensional CNN was able to perform a multi-class classification on the five common cancers among women with an F1 score of over 0.99 [49]. A recent study also shows the importance of RNA-seq data preprocessing, including data smoothing, feature selection, and over-sampling, at building classifiers across 22 tumour types in the TCGA database [3]. With optimized data preprocessing steps and random forest model, PanClassif achieves 100% accuracy on both binary and multi-class classification [3].

Patterns of DNA methylation and miRNA expression are cell type specific. This property uniquely positions these data types for cancer classification, especially when combined with liquid biopsy. As new miRNA and DNA methylation interrogation technologies become feasible, it will make minimally invasive approach to cancer classification a viable strategy. For example, a random forest model was

built with four serum miRNAs to screen for 13 cancer types and achieved high accuracy (AUC = 0.98) [50]. In a recent large-scale serum miRNomics analysis, an ensemble classifier, called the hierarchical ensemble algorithm with deep learning model, which combines seven different learners, was able to predict cancer tissue of origin for early-stage diseases [51]. Capper et al. demonstrated that a random forest classifier trained exclusively on tumour DNA methylation profiles can significantly improve the prediction accuracies for the hard to diagnose subclasses of the central nervous system cancers (AUC, 0.99) [52]. In a separate study, four traditional ML algorithms were evaluated to detect the primary tumour in head and neck squamous cell cancers (HNSCs) that present as metastases with an unknown primary (HNSC-CUPs) using DNA methylation data, with an accuracy of 83–89% [53]. DL was also used to classify nine different cancer types based on DNA methylation data from TCGA, with AUC range 0.85–0.89 [54].

23.5.2. Cancer prognosis

Over the past 20 years, genomic features have been repeatedly leveraged to identify clinically useful signatures to predict prognosis and survival. An excellent recent review summarized the ML and DL models involved in providing the prognosis of cancer patients [55]. Due to small sample size in most studies, DL algorithm is rarely used, while classical ML algorithms, such as SVM and RF, are more popular. Many studies used various traditional ML approaches to select features before applying a risk scoring system to stratify patients into high or low risk. As one example, we have established a 4-miRNA signature to predict ccRCC metastasis using logistic regression [56]. SVM was used to generate a risk score using 32-gene that is prognostic for five-year overall survival of colorectal cancers and validate the risk score using three independent datasets [57]. Recently, as TCGA and other large genomic database became available, DL algorithms have been explored and began to show the advantages over traditional ML methods. For example, DL-based model was applied on hepatocellular carcinomas to differentiate survival subpopulations of patients and then validated in six independent cohorts [58]. CNN-Cox model, which combines a special CNN framework with prognosis-related feature selection cascaded Wx, showed higher C-index values and better survival prediction performance across seven cancer type datasets in TCGA [59]. Another DL approach combining a CNN with stationary wavelet transform (SWT-CNN) was developed to predict clinical outcomes, and it overperformed the classical ML algorithms [60]. DeepProg is a novel ensemble framework of deep-learning and machine-learning approaches that robustly predict patient survival subtypes using multi-omic data [61]. It identifies two optimal survival subtypes in most cancers and yields significantly better risk stratification than other traditional ML methods [61].

Despite large number of prognostic signatures that have been developed, most of them lack large independent validation in clinical settings. As a result, only a handful of molecular signatures have been commercialized and clinically validated, such as Oncotype DX [62], Prolaris [63], and colon ColoPrint [64].

23.5.3. Treatment response

One important application of AI is predicting drug response based on molecular profiles. A recent paper provides a systematic review of the literature on monotherapy drug response prediction including more than 70 ML methods in 13 subclasses [65]. Some of the more recent DL models are reviewed here. scDEAL transfers the model trained on bulk RNA-seq data from cell line to predict drug responses in six scRNA-seq datasets, achieving an average AUC of 0.89 [5]. MultiDCP, a DL model, can predict cellular context-dependent gene expressions and cell viability on a specific dosage [66]. The novelties of MultiDCP include a knowledge-driven gene expression profile transformer that enables context-specific phenotypic response predictions of novel cells or tissues, integration of multiple diverse labelled and unlabelled omics data, the joint training of the multiple prediction tasks, and a teacher–student training procedure that allows us to utilize unreliable data effectively [66]. One caveat for drug response prediction analysis is that most of the models are based on data from cell lines or xenograph models, which do not reflect the *in vivo* conditions of human subjects. Therefore, these prediction models are not directly applicable to patients. There were some limited explorations using genomic data of clinical trials in cancers for response prediction purposes. For example, an attempt to use clinical and genotype features of patients with metastatic colorectal cancer to predict the dose-limiting toxicity event after the first cycle of FOLFIRI plus bevacizumab treatment was conducted using classical ML methods, while RF showed the best accuracy of 84% [67]. Transfer learning was evaluated to predict treatment responses using several real clinical datasets pertaining to patients with breast cancer ($N = 24$), triple-negative breast cancer ($N = 169$), and multiple myeloma ($N = 24$) [68]. Future design of large-scale trials with therapeutic intervention along with measurements at genomic scale should warrant more translational studies.

23.5.4. Genomics-guided clinical trials

According to the Precision Medicine Initiative, precision medicine is 'an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person'. This approach will allow doctors and researchers to make the therapeutic decisions more accurately based on individual patient's genomic data. However, whether this precision medicine approach could bring true benefit for cancer patient had been largely under debate. To fill these gaps, NCI launched a number of precision medicine trials to evaluate the benefit of tailored therapy by matching patients in prospective multi-arm clinical trials.

The Trial Assigning Individualized Options for Treatment (TAILORx) was designed to determine whether chemotherapy is beneficial for women with a mid-range recurrence score (RS) of 11–25, using OncoDx test that assesses the expression of 21 genes associated with breast cancer recurrence. Early results showed that adjuvant endocrine therapy and chemoendocrine therapy had similar efficacy in women with hormone-receptor positive, HER2-negative, and axillary node negative breast cancer who had a mid-range RS [69], which was confirmed later with longer follow-up time. NCI-MATCH was launched in 2015 to investigate whether assigning patients to targeted therapies based on tumour genomics would improve outcomes for patients with no standard drug options. Although 38% of the patients whose tumours were analysed had 'actionable' molecular alterations, only 18% had access to a

relevant therapy through one of the thirty single-arm subprotocols. The overall response rates only ranged from 2% to 38%. However, the limited number of therapies tested in NCI-MATCH yielded a good response rate. One example is that 29 patients with various solid tumours harbouring the $BRAF^{V600}$ mutation and receiving dabrafenib plus trametinib yielded 38% response rate [70]. Since launch in 2017, NIH-COG paediatric MATCH trial has recruited more than 1,000 patients. Analysis of these patients showed that tumour sequencing was effective in finding actionable tumour mutations in paediatric and young adult cancer patients. While 31% of tumours have actionable mutations, 28% of patients were assigned to a phase 2 clinical trial treatment arm of the study [71].

Other similar trials, including Adjuvant Lung Cancer Enrichment Marker Identification and Sequencing Trials (ALCHEMIST) for people with EGFR and ALK mutation and Lung-MAP trial for people with advanced *non-small-cell lung cancer* that has continued to grow after treatment, are still at relatively early stage and it will be interesting to see how these precision medicine trials turn out at the completion.

23.6. Conclusions

Although genomic data is considered as high dimensional, its size is much smaller than other data types, e.g. imaging data and electronic medical records. Therefore, overfitting can become an issue with many ML algorithms, including DL algorithms. Robust validation, such as cross-validation and independent cohort validation, which is lacking in many studies, is needed to ensure that the developed ML models are reproducible. Variations and batch effects across different platforms and centres also contribute to the unreproducibility of ML models. Data harmonization, particular preprocessing steps that remove batch effects, is critical towards mitigating these issues.

It should be noted that a single data type such as transcriptomics or epigenomics seldom captures the disease-relevant signatures of a tumour. Multimodal measurements combining genetics, transcriptomics, epigenomics, proteomics, metabolomics, imaging, as well as phenotypes is needed to improve the performance of AI/ML models. The ethnicity bias in TCGA also limits the application of ML models to certain populations. As the cost of genomic analysis drops quickly, more genomic data will be generated for more AI/ML application development. Interrogation of tumours at single-cell level and even sub-cellular levels with advanced spatial transcriptomics and multiplex imaging will provide more granularity of tumour and TME cells. As more patients in multiple ethnicity groups participate in AI-guided clinical trials, the ethnicity bias will also be alleviated. Advances in software and hardware such as large-scale cloud computing (powered by highspeed networks that enable large volume of data transfer) will enable more applications of DL methods in clinical genomic data.

With all the advances mentioned above, translation of AI models to clinical practice still requires long process of assay development, clinical validation, FDA approval, government regulation, and insurance company acceptance. These typically take several years if not longer. With the experience of prior genomic AI models currently used in the clinical practice and maturation of both AI/ML algorithms and genome technology, it is hopeful that clinical translation of AI models will be accelerated in the foreseeable future.

REFERENCES

1. Jiang P, Sinha S, Aldape K, Hannenhalli S, Sahinalp C, Ruppin E. Big data in basic and translational cancer research. Nat Rev Cancer. 2022;**22**(11):625–39.
2. Marx V. Method of the Year: spatially resolved transcriptomics. Nat Methods. 2021;**18**(1):9–14.
3. Mahin KF, Robiuddin M, Islam M, Ashraf S, Yeasmin F, Shatabda S. PanClassif: Improving pan cancer classification of single cell RNA-seq gene expression data using machine learning. Genomics. 2022;**114**(2):110264.
4. Wood DE, White JR, Georgiadis A, Van Emburgh B, Parpart-Li S, Mitchell J, et al. A machine learning approach for somatic mutation discovery. Sci Transl Med. 2018;**10**(457):eaar7939:1–16.
5. Chen J, Wang X, Ma A, Wang QE, Liu B, Li L, et al. Deep transfer learning of cancer drug responses by integrating bulk and single-cell RNA-seq data. Nat Commun. 2022;**13**(1):6494.
6. Anzar I, Sverchkova A, Stratford R, Clancy T. NeoMutate: an ensemble machine learning framework for the prediction of somatic mutations in cancer. BMC Med Genomics. 2019;**12**(1):63.
7. Poplin R, Chang P-C, Alexander D, Schwartz S, Colthurst T, Ku A, et al. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol. 2018;**36**(10):983–7.
8. Ramachandran A, Lumetta SS, Klee EW, Chen D. HELLO: improved neural network architectures and methodologies for small variant calling. BMC Bioinform. 2021;**22**(1):404.
9. Krishnamachari K, Lu D, Swift-Scott A, Yeraliyev A, Lee K, Huang W, et al. Accurate somatic variant detection using weakly supervised deep learning. Nat Commun. 2022;**13**(1):4248.
10. Ainscough BJ, Barnell EK, Ronning P, Campbell KM, Wagner AH, Fehniger TA, et al. A deep learning approach to automate refinement of somatic variant calling from cancer sequencing data. Nat Genet. 2018;**50**(12):1735–43.
11. Pounraja VK, Jayakar G, Jensen M, Kelkar N, Girirajan S. A machine-learning approach for accurate detection of copy number variants from exome sequencing. Genome Res. 2019;**29**(7):1134–43.
12. Özden F, Alkan C, Çiçek AE. Polishing copy number variant calls on exome sequencing data via deep learning. Genome Res. 2022;**32**(6):1170–82.
13. Li WV, Li JJ. An accurate and robust imputation method scImpute for single-cell RNA-seq data. Nat Commun. 2018;**9**(1):997.
14. Huang M, Wang J, Torre E, Dueck H, Shaffer S, Bonasio R, et al. SAVER: gene expression recovery for single-cell RNA sequencing. Nat Methods. 2018;**15**(7):539–42.
15. Talwar D, Mongia A, Sengupta D, Majumdar A. AutoImpute: autoencoder based imputation of single-cell RNA-seq data. Sci Rep. 2018;**8**(1):16329.
16. Deng Y, Bao F, Dai Q, Wu LF, Altschuler SJ. Scalable analysis of cell-type composition from single-cell transcriptomics using deep recurrent learning. Nat Methods. 2019;**16**(4):311–4.
17. Eraslan G, Simon LM, Mircea M, Mueller NS, Theis FJ. Single-cell RNA-seq denoising using a deep count autoencoder. Nat Commun. 2019;**10**(1):1–14.
18. Gu H, Cheng H, Ma A, Li Y, Wang J, Xu D, et al. scGNN 2.0: a graph neural network tool for imputation and clustering of single-cell RNA-Seq data. Bioinformatics. 2022;**38**(23):5322–5.
19. Xu Y, Zhang Z, You L, Liu J, Fan Z, Zhou X. scIGANs: single-cell RNA-seq imputation using generative adversarial networks. Nucleic Acids Res. 2020;**48**(15):e85:1–16.

20. Arisdakessian C, Poirion O, Yunits B, Zhu X, Garmire LX. DeepImpute: an accurate, fast, and scalable deep neural network method to impute single-cell RNA-seq data. Genome Biol. 2019;**20**(1):211.
21. Van Der Maaten L, Hinton G. Visualizing high-dimensional data using t-sne. J Mach Learn Res. 2008;**9**(26):5.
22. Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, Ng LG, et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. 2019;**37**(1):38–44.
23. Amodio M, van Dijk D, Srinivasan K, Chen WS, Mohsen H, Moon KR, et al. Exploring single-cell data with deep multitasking neural networks. Nat Methods. 2019;**16**(11):1139–45.
24. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018;**15**(12):1053–8.
25. Grønbech CH, Vording MF, Timshel PN, Sønderby CK, Pers TH, Winther O. scVAE: variational auto-encoders for single-cell gene expression data. Bioinformatics. 2020;**36**(16):4415–22.
26. Plaut E. From principal subspaces to principal components with linear autoencoders. arXiv preprint arXiv. 2018:180410253.
27. Alquicira-Hernández J, Sathe A, Ji HP, Nguyen Q, Powell JE. scPred: cell type prediction at single-cell resolution. bioRxiv. 2018:369538.
28. Kiselev VY, Yiu A, Hemberg M. scmap: projection of single-cell RNA-seq data across data sets. Nat Methods. 2018;**15**(5):359–62.
29. Li Z, Feng H. A neural network-based method for exhaustive cell label assignment using single cell RNA-seq data. Sci Rep. 2022;**12**(1):910.
30. Johnson TS, Wang T, Huang Z, Yu CY, Wu Y, Han Y, et al. LAmbDA: label ambiguous domain adaptation dataset integration reduces batch effects and improves subtype detection. Bioinformatics. 2019;**35**(22):4696–706.
31. Lieberman Y, Rokach L, Shay T. CaSTLe–classification of single cells by transfer learning: harnessing the power of publicly available single cell RNA sequencing experiments to annotate new experiments. PLoS ONE. 2018;**13**(10):e0205499.
32. Hu J, Li X, Hu G, Lyu Y, Susztak K, Li M. Iterative transfer learning with neural network for clustering and cell type classification in single-cell RNA-seq analysis. Nat Mach Intell. 2020;**2**(10):607–18.
33. Xu Z, Luo J, Xiong Z. scSemiGAN: a single-cell semi-supervised annotation and dimensionality reduction framework based on generative adversarial network. Bioinformatics. 2022;**38**(22):5042–8.
34. Yuan M, Chen L, Deng M. scMRA: a robust deep learning method to annotate scRNA-seq data with multiple reference datasets. Bioinformatics. 2022;**38**(3):738–45.
35. Kimmel JC, Kelley DR. Semisupervised adversarial neural networks for single-cell classification. Genome Res. 2021;**31**(10):1781–93.
36. Wang T, Bai J, Nabavi S. Single-cell classification using graph convolutional networks. BMC Bioinform. 2021;**22**(1):364.
37. Shao X, Yang H, Zhuang X, Liao J, Yang P, Cheng J, et al. scDeepSort: a pre-trained cell-type annotation method for single-cell transcriptomics using deep learning with a weighted graph neural network. Nucleic Acids Res. 2021;**49**(21):e122-e.
38. Maseda F, Cang Z, Nie Q. DEEPsc: a deep learning-based map connecting single-cell transcriptomics and spatial imaging data. Front Genet. 2021;**12**:636743.
39. Liao J, Qian J, Fang Y, Chen Z, Zhuang X, Zhang N, et al. De novo analysis of bulk RNA-seq data at spatially resolved single-cell resolution. Nat Commun. 2022;**13**(1):6498.
40. Lopez R, Li B, Keren-Shaul H, Boyeau P, Kedmi M, Pilzer D, et al. DestVI identifies continuums of cell types in spatial transcriptomics data. Nat Biotechnol. 2022;**40**(9):1360–9.
41. Song Q, Su J. DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence. Brief Bioinform. 2021;**22**(5):1–13.
42. Biancalani T, Scalia G, Buffoni L, Avasthi R, Lu Z, Sanger A, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. Nat Methods. 2021;**18**(11):1352–62.
43. Tan X, Su A, Tran M, Nguyen Q. SpaCell: integrating tissue morphology and spatial gene expression to predict disease cells. Bioinformatics. 2020;**36**(7):2293–4.
44. Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods. 2021;**18**(11):1342–51.
45. Cang Z, Nie Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. Nat Commun. 2020;**11**(1):2084.
46. Sun Y, Zhu S, Ma K, Liu W, Yue Y, Hu G, et al. Identification of 12 cancer types through genome deep learning. Sci Rep. 2019;**9**(1):17256.
47. Zeng Z, Mao C, Vo A, Li X, Nugent JO, Khan SA, et al. Deep learning for cancer type classification and driver gene identification. BMC Bioinform. 2021;**22**(Suppl 4):491.
48. Cieślik M, Chinnaiyan AM. Cancer transcriptome profiling at the juncture of clinical translation. Nat Rev Genet. 2018;**19**(2):93–109.
49. Mohammed M, Mwambi H, Mboya IB, Elbashir MK, Omolo B. A stacking ensemble deep learning approach to cancer type classification based on TCGA data. Sci Rep. 2021;**11**(1):15626.
50. Chen JW, Dhahbi J. Identification of four serum miRNAs as potential markers to screen for thirteen cancer types. PLoS ONE. 2022;**17**(6):e0269554.
51. Matsuzaki J, Kato K, Oono K, Tsuchiya N, Sudo K, Shimomura A, et al. . JNCI Cancer Spectr. 2022;**7**(1):1–13.
52. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. Nature. 2018;**555**(7697):469–74.
53. Leitheiser M, Capper D, Seegerer P, Lehmann A, Schüller U, Müller KR, et al. Machine learning models predict the primary sites of head and neck squamous cell carcinoma metastases based on DNA methylation. J Pathol. 2022;**256**(4):378–87.
54. Eissa NS, Khairuddin U, Yusof R. A hybrid metaheuristic-deep learning technique for the pan-classification of cancer based on DNA methylation. BMC Bioinform. 2022;**23**(1):273.
55. Deepa P, Gunavathi C. A systematic review on machine learning and deep learning techniques in cancer survival prediction. Prog Biophys Mol Biol. 2022;**174**:62–71.
56. Wu X, Weng L, Li X, Guo C, Pal SK, Jin JM, et al. Identification of a 4-microRNA signature for clear cell renal cell carcinoma metastasis and prognosis. PLoS ONE. 2012;**7**(5):e35661.
57. Cheong JH, Wang SC, Park S, Porembka MR, Christie AL, Kim H, et al. Development and validation of a prognostic and predictive 32-gene signature for gastric cancer. Nat Commun. 2022;**13**(1):774.
58. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. Clin Cancer Res. 2018;**24**(6):1248–59.

59. Yin Q, Chen W, Zhang C, Wei Z. A convolutional neural network model for survival prediction based on prognosis-related cascaded Wx feature selection. Lab Invest. 2022;**102**(10):1064–74.
60. Zhao Y, Zhou Y, Liu Y, Hao Y, Li M, Pu X, et al. Uncovering the prognostic gene signatures for the improvement of risk stratification in cancers by using deep learning algorithm coupled with wavelet transform. BMC Bioinform. 2020;**21**(1):195.
61. Poirion OB, Jing Z, Chaudhary K, Huang S, Garmire LX. DeepProg: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data. Genome Med. 2021;**13**(1):112.
62. Cobleigh MA, Tabesh B, Bitterman P, Baker J, Cronin M, Liu ML, et al. Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes. Clin Cancer Res. 2005;**11**(24 Pt 1):8623–31.
63. Shore N, Concepcion R, Saltzstein D, Lucia MS, van Breda A, Welbourn W, et al. Clinical utility of a biopsy-based cell cycle gene expression assay in localized prostate cancer. Curr Med Res Opin. 2014;**30**(4):547–53.
64. Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, et al. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. J Clin Oncol. 2011;**29**(1):17–24.
65. Firoozbakht F, Yousefi B, Schwikowski B. An overview of machine learning methods for monotherapy drug response prediction. Brief Bioinform. 2022;**23**(1):bbab408:1–18.
66. Wu Y, Liu Q, Qiu Y, Xie L. Deep learning prediction of chemical-induced dose-dependent and context-specific multiplex phenotype responses and its application to personalized Alzheimer's disease drug repurposing. PLoS Comput Biol. 2022;**18**(8):e1010367.
67. Bedon L, Cecchin E, Fabbiani E, Dal Bo M, Buonadonna A, Polano M, et al. Machine learning application in a phase I clinical trial allows for the identification of clinical-biomolecular markers significantly associated with toxicity. Clin Pharmacol Ther. 2022;**111**(3):686–96.
68. Turki T, Wang JTL. Clinical intelligence: new machine learning techniques for predicting clinical drug response. Comput Biol Med. 2019;**107**:302–22.
69. Sparano JA, Gray RJ, Makower DF, Pritchard KI, Albain KS, Hayes DF, et al. Adjuvant chemotherapy guided by a 21-gene expression assay in breast cancer. New Engl J Med. 2018;**379**(2):111–21.
70. Salama AKS, Li S, Macrae ER, Park JI, Mitchell EP, Zwiebel JA, et al. Dabrafenib and Trametinib in patients with tumors with BRAF(V600E) mutations: results of the NCI-MATCH trial subprotocol H. J Clin Oncol. 2020;**38**(33):3895–904.
71. Parsons DW, Janeway KA, Patton DR, Winter CL, Coffey B, Williams PM, et al. Actionable tumor alterations and treatment protocol enrollment of pediatric and young adult patients with refractory cancers in the national cancer institute-children's oncology group pediatric MATCH trial. J Clin Oncol. 2022;**40**(20):2224–34.

SECTION 6
Biomechanics

24

A role for mechanical heterogeneity in the tumour microenvironment in driving cancer cell invasion

Madhurima Sarkar, Asadullah, and Shamik Sen

24.1. Introduction

The tumour microenvironment (TME) comprises of the extracellular matrix (ECM), a plethora of ECM anchored growth factors and multiple cell types, including fibroblasts, endothelial cells, and immune cells. It is now increasingly appreciated that in addition to cancer-associated mutations in cancer cells, the spatiotemporal alterations in the TME actively regulate cell–matrix and cell–cell interactions, thereby bringing about the myriad of changes that collectively enable cancer progression. These include alterations in the composition and organization of the ECM, breaching of the basement membrane (BM) and ECM remodelling, reprogramming of fibroblasts and immune cells, and the development of an aberrant blood vessel network that satisfies the oxygen and nutrient demands of tumour cells and enables escape of cancer cells from their primary site.

In this chapter, we first discuss how physicochemical alterations in the ECM associated with cancer give rise to a spectrum of mechanical cues that regulate different aspects of cancer progression. Next, we discuss the challenges to migration and the strategies adopted by cancer cells. In the third section, we highlight the role of matrix metalloproteinases (MMPs) in mediating cancer invasion. In the last section, we discuss computational studies that have contributed to our understanding of cancer invasion.

24.2. ECM alterations in cancer

The ECM is a three-dimensional (3D) network of proteins and polysaccharides which confers structural integrity to tissues and plays an active role in regulating cell behaviour, both in physiological and pathological contexts. The ECM is comprised of the BM and the interstitial matrix. The BM is made up of collagen IV and laminins and is interconnected with the interstitial matrix through multiple proteins and proteoglycans [1]. The interstitial matrix, enriched in collagens (I, III, V, etc.), fibronectin and elastin, forms a porous 3D protein network around cells interconnecting the stroma and the BM.

Remodelling of the interstitial matrix involves several biochemical and biophysical changes leading to ECM stiffening and altered cell signalling, resulting in tumour progression and cancer metastasis [2]. ECM stiffening serves as the basis for detection of tumours in soft tissues using simple palpation, ultrasound, and MRI elastography. Tumour-associated collagen signatures (TACS) correspond to alterations in collagen organization with breast cancer progression [3]. Collagen fibres in soft normal tissues are curly and positioned parallel to the epithelium layer. Interestingly, in non-tumorous breast tissues, high deposition of collagen inhibits tumour formation by the up-regulation of cell–cell adhesion genes (tumour suppressive genes, e.g. cell–cell adhesion) and by down-regulation of mesenchymal genes. In desmoplastic stroma, at the TACS-3 stage, collagen fibres near the tumour boundary are linearized, oriented perpendicularly to the tumour boundary, and actively drive tumour invasion [4].

Stromal alterations are mediated by altered expression of collagen I (Col I), Col III, and ECM-modifying enzymes, such as lysyl oxidases (LOX) and LOX-like proteins. By secreting pro-fibrotic and inflammatory growth factors, such as transforming growth factor β (TGF-α) and transforming growth factor β (TGF-β), fibroblast growth factor 2 (FGF2), platelet-derived growth factor, and epidermal growth factor, tumour cells recruit and activate stromal cells, the major ECM depositors in the TME [5]. The tumour-derived factors also induce the differentiation of stromal cells to cancer-associated fibroblasts (CAFs). CAFs are from multiple origin, specifically tissue-resident or bone-marrow-derived fibroblasts, and they can act as myofibroblasts, remodel the ECM, and support tumour progression [6]. In pancreatic ductal carcinoma, a direct interaction

of CAFs with adjacent tumour cells drive TGF-β signalling and collagen deposition. CAFs that are at a greater distance are also activated by tumour cells but are unresponsive to TGF-β; instead, they deposit hyaluronic acid (HA) and by expressing IL-6 establish a tumorigenic, pro-inflammatory environment [7].

HA can function as tumour suppressor or tumour promoter, depending on its molecular weight [8]. High molecular weight HA and CD44 signalling induce tumour suppression via cell cycle arrest. In several tumours, such as prostate, colorectal, and breast cancer, low molecular weight HA (LMM-HA) is associated with poor prognosis [9]. Dysregulation of HA synthase and hyaluronidase leads to accumulation of LMM-HA, which interacts with cell surface receptors and regulates pro-tumorigenic signalling cascades, such as glycolysis, and also promotes migration [10]. Signalling of LMW-HA through CD44 increases the resistance to cellular stress to promote tumour development [11,12].

24.3. Mechanical heterogeneity of the TME

As the tumour stroma undergoes gradual changes from a TACS-1 state to a TACS-3 state [3], the compositional and organizational heterogeneity of the TME provides several mechanical cues to cancer and stromal cells. Of these cues, matrix stiffness is the most widely studied [13]. Using matrix-coated hydrogels as synthetic ECMs, the elastic modulus of the substrate has been shown to influence cell spreading, proliferation, cell migration, and malignancy. In a landmark study, Weaver and co-workers showed that acinar structures formed by normal mammary epithelial cells on soft gels are disrupted with increase in stiffness [14]. By extending this work in a rat model, her group showed that breast tumorigenesis was associated with matrix stiffening and increased integrin signalling [4]. Using interpenetrating networks of reconstituted BM matrix and alginate that allows independent control of stiffness, composition, and architecture, Mooney and co-workers demonstrated that in addition to ECM mechanics ECM composition is also an important determinant [15].

Since collagen is one of the most abundant ECM proteins whose expression increases during cancer progression, 3D collagen gels that better mimic the TME have been employed to better understand biophysical properties of the TME. Collagen gels are formed from triple-helical strands of $\alpha_1 = \alpha_2$ subunits that self-assemble to form collagen fibres. Rheological studies of 3D collagen gels have demonstrated several non-linear responses exhibited by these gels, including strain stiffening, viscoelasticity, and plasticity. These properties are determined by the combination of fibre density, orientation, and extent of crosslinking. Strain stiffening corresponds to increased resistance to deformation by collagen fibres and is attributed to strain-induced alignment of individual fibres in the direction of applied strain [16]. Similar to soft tissues, such as brain, adipose, breast, and muscle tissues, collagen gels are viscoelastic, i.e. they exhibit a combination of elastic solid-like and viscous fluid-like behaviour. Stress relaxation and creep recovery experiments have been widely used for characterizing viscoelastic properties of collagen gels. In stress relaxation experiments, the resisting stress in response to a constant strain is measured over time. In creep recovery experiments, the strain is measured in response to an initial step stress and monitored after removal of the stress [17]. Plasticity of collagen gels can be measured using creep recovery experiments wherein the degree of plasticity corresponds to the ratio of the residual strain at long times to that of the maximum strain [18].

In addition to ECM alterations, uncontrolled cell proliferation in tumours leads to increased cell density that further leads to build-up of confinement-induced stresses. The mechanical environment witnessed by cells changes considerably from initial stages of tumour formation wherein tumour cells are able to divide, migrate, and proliferate within the lumen in an unconstrained manner. However, at later stages, when luminal space decreases, dividing cells are increasingly subjected to compressive stresses till the BM is breached. Further, the growth of tumour cells induces compression of the surrounding stroma and has been shown to lead to collapse of the blood and lymphatic vessels [19]. This increased interstitial fluid pressure in turn leads to increased fluid flow away from the tumour core to the surrounding healthy tissues.

24.4. Mechanisms of cancer cell invasion

The spatiotemporal alterations of the ECM associated with cancer progression can both impede or promote cancer invasion. Increase in ECM density leads to reduction in average pore size of the matrix, thereby hindering cancer invasion. Depending on the extent of epithelial-to-mesenchymal transition (EMT), cancer cells migrate as single cells, as loosely connected cell streams or as cell clusters with intact cell–cell adhesions [20]. In the mesenchymal mode of migration, cancer cells exhibit fibroblast-like elongated morphologies stabilized by integrin-based focal adhesions and rely on protease-mediated matrix degradation for generating migration paths [21]. Seminal work by Friedl and co-workers showed that upon inhibition of proteolysis, cells transition from a protease-dependent to a protease-independent mode of migration characterized by rapid changes in cell shape and the absence of adhesions [22]. In a landmark study, Piel and co-workers showed that confinement and low adhesion can switch slow-moving mesenchymal cells to a fast-moving amoeboidal mode of migration suggesting that these modes of migration are dictated by the balance between cell protrusions, adhesions, and actomyosin contractility [23]. Recently, in an elegant study on the role of matrix plasticity in mediating cancer invasion, Chaudhuri and co-workers mimicked the plasticity of breast tumour tissue using nanoporous hydrogels wherein matrix plasticity can be tuned independent of stiffness. Using this system, they established a protease-independent mode of migration wherein cell-generated forces mechanically opened up migration tracks that remained open if the matrix was plastic [24].

Collective invasion has been reported in epithelial cancers, such as breast and colorectal cancer. Clusters of circulating tumour cells have also been found to be present in circulation in several cancers and have been associated with higher colonization efficiency [25]. The presence of such clusters suggests that instead of undergoing complete EMT, cells undergoing partial or hybrid EMT might be fitter in countering the multitude of different stresses associated with cancer metastasis [26]. Apart from EMT, another determinant of collective invasion may be ECM organization. The crosslinked and highly linearized collagen fibrils oriented perpendicularly to the tumour periphery serve as contact guidance tracks for cells to

disseminate [3]. This has been corroborated by intravital imaging studies using breast cancer models wherein motile cells were localized close to the collagen bundles and migrated in directions parallel to the collagen fibres [27]. Similar to the topography-guided migration by collagen fibres, cancer cells also exhibit fast collective migration along muscle fibres, nerves, and perivascular spaces. In contrast, adipose tissue drives single cell migration. High collagen density has also been shown to induce a switch from single cell to collective migration in breast cancer and melanoma cells when embedded in dense collagen matrices [28].

Cancer invasion through 3D matrices requires constant deformation of the cell body and the nucleus—the largest (≈5–15 μm) and stiffest organelle of the cell. Nuclear translocation through pores is the rate limiting factor for cancer cell invasion through 3D matrices [29]. In the absence of matrix proteolysis, the migration rate decreases with decreasing pore size with the nucleus capable of being deformed up to 10% of its undeformed size [29]. For mediating nuclear deformation during 3D migration, forces are transmitted to the nucleus through the linker of nucleoskeleton and cytoskeleton (LINC) complex [30]. Nuclear translocation through confined spaces is mediated by a combination of pulling from the cell front and pushing from the cell rear. While pulling from the cell front is primarily mediated by non-muscle myosin IIA (NMMIIA)-containing actomyosin bundles by binding to nesprin-3 of the LINC complex [31,32], pushing from the cell rear is mediated by perinuclear localized non-muscle myosin IIB (NMMIIB) that squeezes the nucleus [33]. In an elegant study, Robinson and co-workers showed that among the numerous actin binding proteins, NMMIIA and NMMIIB are two mechanoresponsive proteins that exhibit enriched localization at sites of physical stress [34]. Intriguingly, the actin crosslinking protein α-actinin-4, which is also mechanoresponsive and exhibits nucleocytoplasmic shuttling, has been shown to regulate NMMIIB expression transcriptionally and associates with NMMIIA at the cell periphery, thereby modulating focal adhesion turnover [35]. Compartmentalization of the nucleus during confined migration gives rise to osmotic pressure differences due to water permeation; polarized distribution of ion channels/aquaporins on the plasma membrane at the front and back of the cell can drive nuclear translocation [36].

Mechanical properties of the cell nucleus are collectively determined by properties of the nuclear lamina and the chromatin. The nuclear lamina corresponds to the meshwork of proteins that line the inner nuclear membrane and is composed of type A lamins (A, C, and C2) and type B lamins (B1 and B2), which form independent non-overlapping networks and serve distinct nuclear functions [37]. While B-type lamins are present in all cells, Lamin A expression has been correlated with cell-fate determination. Sub-type-specific post-translational modifications including farnesylation, phosphorylation, SUMOylation, and glycosylation further alter assembly and dynamics of lamin sub-types that collectively determine the bulk properties of the nucleus [38,39]. Seminal work by Discher and co-workers showed that the Lamin A:Lamin B ratio scales with tissue stiffness with stiffer tissues possessing higher Lamin A levels [40]. Micropipette aspiration of nuclei suggests a stronger impact of Lamin A/C on nuclear stiffness, cells with lower Lamin A/C being more deformable and more migratory in confined spaces. Consistent with this, the loss of Lamin A/C as well as LINC complex proteins reported in multiple cancers enables cancer cells with increased invasiveness. In MDA-MB-231 and HT-1080 cells that exhibit mesenchymal-to-amoeboidal transition upon inhibition of matrix proteolysis, nuclear softening mediated by increased phosphorylation of Lamin A/C necessary for sustaining non-proteolytic migration highlights tuning of nuclear properties depending on the mode of invasion [41]. Moreover, nuclear stiffness is also a determinant of migration-induced DNA damage in cancer cells, with greater extent of DNA damage in cells with higher Lamin A levels [42]; interestingly, nuclear softening has been shown to limit the extent of DNA damage [43]. In addition to alterations in lamin levels and/or phosphorylation, nuclear stiffness is also dictated by the euchromatin-to-heterochromatin levels. Increase in euchromatin levels induced by treatment with the deacetylase inhibitor trichostatin A (TSA) induces nuclear softening [44]. Interestingly, mechanical stretch has also been shown to confer mechanoprotective properties to the cell by nuclear softening mediated by loss of H3K9me3-marked heterochromatin [45].

24.5. MMPs in cancer invasion

After breaching the BM, cancer cells must manoeuvre through the surrounding collagen-rich stroma that provides substantial steric hindrance. This is achieved by matrix degradation mediated by a wide array of matrix degrading enzymes that generate paths amenable for cell migration; in addition, the ECM fragments also support cell adhesion by engaging cell surface integrins. MMPs represent one of the most prominent family of proteinases associated with cancer progression [46]. All MMPs contain an N-terminal signal peptide, a pro-domain that interacts with the active site, and a catalytic domain with a zinc ion at the active site. In addition, some MMPs have a C-terminal hemopexin domain that is linked to the catalytic domain through a flexible hinge. Based on their substrate specificity, 24 different types of MMPs discovered in humans are broadly classified as collagenases, gelatinases, stromelysins, matrilysins, or membrane-anchored MMPs. Secretory MMPs are synthesized in the latent form and become functional upon proteolytic activation by serine proteases, membrane-anchored MMPs, or by other activated MMPs. In addition to degrading abundant ECM proteins, such as collagen, fibronectin, and laminin, MMPs are also capable of cleaving cell surface receptors involved in growth factor signalling. Newer intracellular roles of MMPs are also being discovered [47].

MMPs can broadly be divided into secreted MMPs and membrane-anchored MMPs. Of the four membrane anchored MMPs, MT1-MMP (or MMP14) is one of the most widely studied MMPs and plays a key role at different stages of cancer metastasis. Though cells express a wide range of MMP molecules, the importance of MT1-MMP in cancer invasion can be appreciated from studies wherein MT1-MMP silencing failed to induce the transition from ductal carcinoma *in situ* to invasive phenotype in a tumour xenograft model [48]. MT1-MMP-mediated ECM degradation in cancer cells is mediated by actin-rich protrusions at the cell periphery known as invadopodia. In cells cultured on 2D ECM-coated substrates, invadopodia assemble on the ventral surface and can be detected by the formation of punctate degradation spots that colocalize with F-actin and actin regulatory proteins, such as cortactin, cofilin, and N-Wasp. In a landmark study, Weaver and co-workers demonstrated the importance of matrix stiffness in regulating invadopodia formation

and dynamics, with both the number and size of invadopodia being larger on stiffer substrates [49]. Another recent study showed that during confined migration through pores smaller than 7 μm^2, nuclear deformation drives invadopodia-mediated matrix degradation by triggering recycling of MT1-MMP from endosomes to the invasive front [50].

In comparison to membrane-anchored MMPs, soluble MMPs secreted out diffuse through the extra-cellular space and mediated matrix degradation. Of the soluble MMPs, MMP-2 and MMP-9 are gelatinases that are involved in further processing of collagen-rich matrices after initial processing by other MMPs, including MT1-MMP and MMP-1. Interestingly, sampling of conditioned media from non-invasive MCF-7 cells, metastatic MDA-MB-231 breast cancer cells, and HT-1080 fibrosarcoma cells cultured on collagen-coated soft and stiffpolyacrylamide gels revealed a prominent stiffness-dependent increase in MMP-9 activity in invasive cancer cells [51]. Elimination of stiffness-dependent increase in MMP activity by the actomyosin inhibitor blebbistatin illustrates the importance of stiffness sensing in the regulation of MMP activity. Treatment with the broad spectrum inhibitor GM6001 abolished the prominent mesenchymal phenotype observed on stiff surfaces highlighting the importance of MMP-mediated remodelling in sustaining cancer invasion. Together, MMP-2 and MMP-9 are capable of degrading other ECM proteins, such as type III, IV, V, VII, X and XI collagens, laminin, and fibronectin. In addition, MMP-2 and MMP-9 regulate different aspects of cancer invasion by cleaving a gamut of growth factors, including insulin-like growth factor, vascular endothelial growth factor, FGF, TGF-β, and tumour necrosis factor α. While MMP-2 is activated by MT1-MMP, MMP-9 can in turn be activated by activated MMP-2 highlighting the close coordination between membrane-anchored MT1-MMP and the soluble MMPs. It is likely that degradation at sites of invadopodia is achieved through stepwise activation of soluble MMPs at the vicinity of the invadopodia that then diffuse into the surrounding space, degrade the matrix, and contribute to invadopodia stabilization and growth. Consistent with this line of thought, degradation of invadopodia was near completely abolished by both the soluble MMP-2/MMP-9 inhibitor SB3-CT and the MT1-MMP inhibitor NSC405020 [52].

MMP-mediated matrix degradation can be visualized using microscopy by culturing cells on substrates coated with fluorescently labelled ECM proteins wherein zones lacking fluorescence correspond to degradation spots. Intriguingly, the degradation zones often span several microns across. For achieving such large yet focused degradation, MMPs must preferably be packaged together and secreted as extra-cellular vesicles (EVs). EVs correspond to vesicles of different sizes that traffic bioactive molecules (proteins, glycans, metabolites, DNA, RNA, and miRNA) between cells and include microvesicles, exosomes, and exomeres. In a recent study, ECM stiffness was shown to be a regulator of exosome composition and secretion in breast cancer cells, with an increase in stiffness leading to increased secretion, mediated by stiffness-dependent activation of the YAP/TAZ pathway [53]. Both MMP-2 and MMP-9 were found to be present in these exosomes and exhibited stiffness-dependent packaging in exosomes. MMPs were activated by thrombospondin-1 (THBS1) that exhibited exclusive exosomal localization; cell invasiveness was markedly reduced in the presence of exosomes harvested from THBS1 knockdown cells. The presence of both integrins and fibronectin in exosomes raises the possibility that integrin–fibronectin binding across exosomes may serve as a mechanism to concentrate exosomes locally, thereby mediating focused MMP-mediated degradation.

Apart from their matrix remodelling function, MMPs regulate cancer invasion in part by regulating integrin dynamics. In an elegant study, Bissell and co-workers showed that even in sparse collagen matrices that do not require matrix remodelling, the transmembrane/cytosolic domains of MT1-MMP are essential in driving mammary branching morphogenesis by stabilizing β_1 integrins at the cell surface [54]. Similarly, in invasive MDA-MB-231 breast cancer cells and HT-1080 fibrosarcoma cells, loss of focal adhesions and stiffness-dependent spreading abolished by the broad spectrum MMP inhibitor GM6001 are attributed to reduction in levels of β_1 integrins. Increased recycling of membrane-bound integrins confirmed that using an AFM-based adhesion assay suggests that MMP proteolytic activity stabilizes integrin at the cell membrane, thereby preventing its internalization and degradation [51]. This stabilization might be driven by MMP-mediated matrix remodelling exposing cryptic domains for integrin engagement as observed with laminin-5 and collagen IV cleavage. In addition to this, MMPs might provide direct binding sites for integrins, thereby restricting their dynamics. Indeed, MMP-2 has been shown to regulate invasiveness of endocardial cells by binding $\alpha_v\beta_3$ integrins via its hemopexin domain [55]. Similarly, in chronic lymphocytic leukaemia (B-CLL), hemopexin domain of MMP-9 has been shown to engage $\alpha_4\beta_1$ integrins [56]. Apart from the hemopexin domains, the fibronectin type II inserts present in the catalytic domains of MMP-2 and MMP-9 may also participate in anchoring integrins onto the cell membrane.

24.6. Cancer invasion: insights from computational modelling

Cancer invasion is a multiscale phenomenon collectively determined by physicochemical attributes of the ECM, the spatiotemporal kinetics of MMP-mediated ECM remodelling, alterations in cell–cell adhesions brought about by EMT, crosstalk between distinct cell types within the TME, and cell migration that involves nuclear deformation. These factors are in turn dependent on heterogeneity in protein expression and localization across different cell types. Given the differences in timescales and length scales associated with these processes, it may not be possible to study all these processes within the same experimental framework or experimentally measure one or more of these processes. Computational studies pave the way to study the individual and collective contributions of disparate elements of the complete TME as it provides better control of the parameters. In this section, we present some of the computational studies that have contributed to our understanding of cancer invasion.

1. *Invadopodia growth, dynamics, and function*: The dependence of ECM fibre density and crosslinking on invadopodia dynamics was probed in an elegant study by Alissa Weaver and co-workers using a combination of experiments and simulations [57]. The authors simulated the dynamics of invadopodia penetration into a fibrillar matrix using a rule-based cellular automata model that accounted for preferential invadopodia growth as well as retraction and lateral sliding depending on the extent of crosslinking. Their model predicted reduced

invadopodia penetration as well as shorter invadopodia timescales with increase in matrix crosslinking. In a separate work, by coupling actin dynamics, EGFR signalling, MMP synthesis and delivery, and ECM degradation using a set of non-linear partial differential equations, Suzuki and co-workers were able to successfully recapitulate the formation of micron-long protrusions with maturation timescales of ≈ 1 h [58]. In another study combining experiments and simulations, two distinct FRAP timescales (≈ 30 s and ≈ 260 s) of MT1-MMP recovery at invadopodia were attributed to vesicular transport and MT1-MMP turnover with rapid turnover essential for matrix degradation. MT1-MMP is known to activate soluble MMP2 that in turn activates MMP9 [59]. To parse the relative contributions of membrane-bound MT1-MMP and soluble MMPs in mediating matrix degradation at invadopodia, a discrete cellular automata model was integrated with reaction–diffusion dynamics [60]. Assuming an ECM density-dependent MMP secretion profile, this model was able to recapitulate the experimental observation of increased matrix degradation at higher ECM densities and established the role of soluble MMPs in mediating matrix degradation after activation by MT1-MMP. Further, by modelling multiple invadopodia, it was shown that optimal inter-invadopodia spacing led to the formation of degradation zones or pores amenable for cell invasion. By accounting for the cross-talk between the branched actin network within invadopodia and the fibrillar ECM network, as well as cytoskeletal connections to the nucleus, Asada and co-workers showed that invadopodia growth is collectively determined by actin bundling and MMP-mediated matrix degradation, both of which are dependent on ECM stiffness [61]. The study also identified myosin turnover time as a critical parameter regulating nuclear movement towards the leading edge.

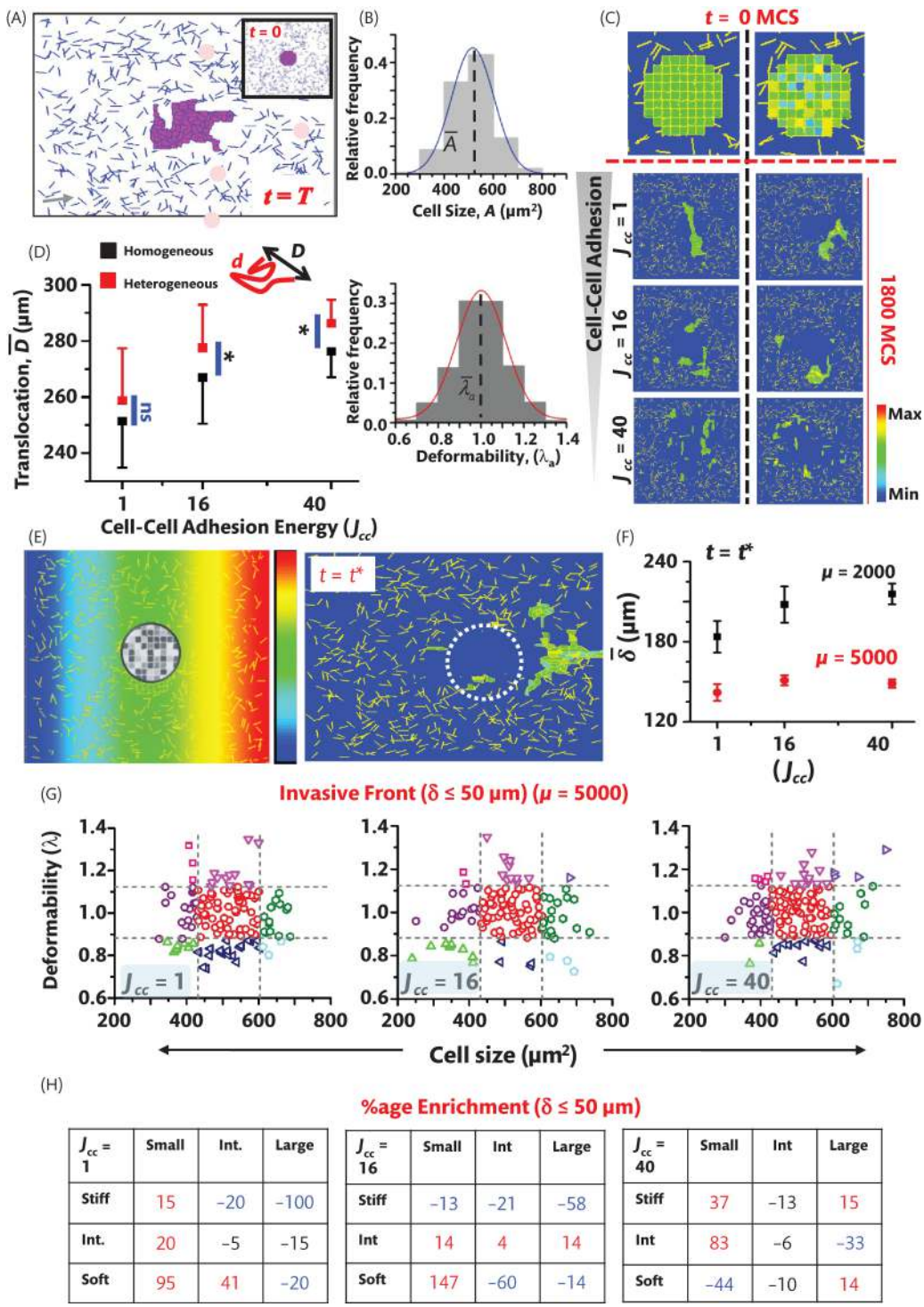
2. *Cell–matrix interactions in cancer invasion*: While computational modelling invacdopodia dynamics provides insight into early stages of cancer invasion, the timescales of cell migration are much longer and involve dramatic alterations in cell shape with biophysical properties of cells being an important determinant of invasion efficiency. In this context, the Cellular Potts Modelling (CPM) framework has emerged as a simple yet attractive framework for studying complex multicellular behaviour integrating phenomena across different timescales and length scales [62]. In this framework, cells, sub-cellular organelles, and ECM can be modelled as domains with preferred sizes, shapes, and biophysical properties. In the CPM framework, spatiotemporal evolution of the simulation lattice is based on minimization of a generalized energy given by the following equation:

$$E_0 = \sum_{\forall i,j} J_{\tau(\sigma(i)),\tau(\sigma(i))} + \sum_{\forall \sigma} \lambda_a \left(a(\sigma) - a_0\right)^2 + \sum_{\forall \sigma} \lambda_p \left(p(\sigma) - p_0\right)^2 + \omega(\sigma) - \mu \times \left[\upsilon_{(\text{target})} - \upsilon_{(\text{source})}\right] \quad (24.1)$$

In the above expression, $\sigma(i)$ and $\tau(\sigma)$ represent the ID of pixel i and the cell type, respectively. The first term in the energy expression (i.e. J_{τ_1,τ_2}) represents the boundary energy per unit length between cells of type τ_1 and τ_2, and is indicative of the adhesion energy between two cells. By assigning different interfacial energies to cell–cell and cell–matrix interactions, it is possible to study the process of EMT. The second term represents the energy associated with change in area of a given cell from its preferred area (a_0), with the area constraint (λ_a) representing the bulk stiffness or inverse compressibility of the cell. Thus, this energy term penalizes for large variations in cell size. Similarly, the third term accounts for variation in cell perimeter from its preferred perimeter (p_0), with the perimeter constraint (λ_p) reflective of line tension. Thus, this energy term penalizes for large variations in cell perimeter. The fourth energy term ($\omega(\sigma)$) is associated with active motility of a cell and is typically given by the expression $\omega(\sigma) = -\mu_0 \hat{P}$, where μ_0 represents the strength of motility and $\hat{P}$ represents the extent of cell polarization. The last term in the energy expression models chemotaxis of cells, with μ representing the effective chemical potential, and υ(target) and υ(source) representing the concentrations of chemoattractant at target and source pixels, respectively.

The ease of integrating the CPM framework with physical processes such as reaction diffusion kinetics makes it particularly attractive to study cancer invasion. By assigning different interfacial energies to cell–cell and cell–matrix interactions to mimic different extents of EMT, the MMP secretion rate and ECM density/organization were shown to collectively regulate the patterns of single-cell and collective cell invasion [63]. Interestingly, collective invasion was found to require less ECM degradation compared to single-cell invasion. While higher MMP secretion rates were required to sustain invasion through dense matrices, matrix proteolysis was not required for invasion through aligned matrices, wherein cells squeeze in between the ECM fibres. By incorporating MMP interactions with tissue inhibitors of metalloproteinases and matrix secretion by cancer cells and two distinct layers of ECM networks, Bhat and co-workers showed that when both cell–cell and cell–matrix adhesions are intact, cells invaded collectively [64].

While insightful, the importance of cell mechanics was not addressed in these studies. However, the importance of cell properties, particularly stiffness, has been implicated in enhancing confined migration through small pores. In a recent study, we probed the importance of biophysical heterogeneity by first mapping cell size and cell deformability distributions of MCF-7 and MDA-MB-231 breast cancer cells. To assess the functional relevance of such heterogeneity in cancer invasion, we developed a CPM-based formalism wherein we tracked invasiveness of a cell cluster (cells shown in pink) surrounded by a fibrillar matrix (blue lines indicate fibres) (**Figure 24.1**A, inset) [65]. Cell scattering was mediated by MMP molecules secreted by cells that diffuse into the surrounding matrix and degrade matrix fibres, thereby creating migration paths. To assess the importance of phenotypic heterogeneity, simulations were performed both for a homogeneous cell cluster with cells having identical size and deformability and for a heterogeneous cell cluster with size and deformability distributions approximated as Gaussians with mean sizes and deformabilities set to that of the homogeneous cluster (**Figure 24.1**B). These simulations were performed in the absence of any chemokine gradient, i.e. the last term in Equation (24.1) was not accounted for. Different extents of EMT were simulated by choosing different values of cell–cell adhesion energies (J_{cc}) with J_{cc} = 1 simulating collective migration and J_{cc} = 40 simulating scattered single-cell migration (**Figure 24.1**C). Interestingly, tracking of end-to-end distance (D) travelled by cells revealed increased



J_{cc} = 1	Small	Int.	Large
Stiff	15	-20	-100
Int.	20	-5	-15
Soft	95	41	-20

J_{cc} = 16	Small	Int	Large
Stiff	-13	-21	-58
Int	14	4	14
Soft	147	-60	-14

J_{cc} = 40	Small	Int	Large
Stiff	37	-13	15
Int	83	-6	-33
Soft	-44	-10	14

Figure 24.1. Effect of phenotypic heterogeneity on cancer invasiveness. (A) Schematic of cancer invasion model. Invasion was simulated by studying scattering of a cell cluster (pink pixels) positioned at the centre of a 2D space mimicking the extra-cellular matrix (ECM) 11 mm^2 in size (inset). The ECM consists of randomly positioned ECM fibres (blue lines). ECM degradation is mediated by cell-secreted MMP molecules that diffuse and degrade ECM fibres. (B) Simulations were performed for a homogeneous cell cluster, i.e. all cells were of the same size and deformability, and a heterogeneous cell cluster. For the heterogeneous cluster, cell size (A) and cell deformability (or area constraint) were approximated as normal distributions with means corresponding to the cell size and deformability of the homogeneous cell cluster. (C) Time-dependent invasion of a homogeneous and a heterogeneous cell cluster for three different values of cell–cell adhesion energy (J_{cc} = 1, 16, 40). (D) Quantification of average cell invasion of homogeneous and heterogeneous cell clusters measured by end-to-end distance distances (D) travelled by cells. (E) Directed migration of cancer cells was simulated by doing simulations in the presence of a stable chemokine gradient (highest concentration at the rightmost edge). Simulations were stopped at time $t = t^*$, i.e. when the first cell reaches the right edge in a given simulation. (F) Quantification of δ for two different values of chemotactic strength μ. δi corresponds to the distance of ith cell from right edge at $t = t^*$. (G) Plot of cell size versus deformability for cells at the invasive front (i.e. $\delta \leq 50$ m at $t = t^*$). Grey dotted lines correspond to $(\bar{A} \pm \sigma_A)/(\lambda \pm \sigma\lambda)$ and allow the population to be segregated into none subpopulations based on size and deformability. (H) Percentage enrichment of individual subpopulation at the invasive front relative to the whole population. *Source:* Adapted from Asadullah et al. [65].

population-level invasion for the heterogeneous cluster for all values of J_{cc} (Figure 24.1D).

To probe how cells of varying size and deformabilities get positioned during invasion, simulations were performed in the presence of a stable chemokine gradient for different values of chemotactic strength (μ) (Figure 24.1E). In the presence of this gradient, cells migrated towards the right side, i.e. side with higher chemokine concentration. Simulations were stopped at time $t = t^*$ at which the first cell reached the right boundary of the lattice. At this time point, the distribution of cell size and deformability were determined for cells that reached within 50 μm from the right edge, i.e. $\delta \leq 50\,\mu$m, where δ_i corresponds to distance of ith cell from right edge at $t = t^*$ (Figure 24.1F and G). By performing enrichment analysis, i.e. comparison of proportion of cells of a given subpopulation at the invasive front in comparison to that in the entire population, three distinct cell subpopulations of increased invasiveness were identified (Figure 24.1H). These include small cells of varying deformabilities, small and intermediate sized soft cells, and intermediately stiff cells of varying cell sizes. By doing further experiments, we showed that small and soft cells at the invasive front correspond to cancer stem cells. In addition to establishing phenotypic heterogeneity as an enabler of cancer invasion, these results illustrate how by combining systems-level computational modelling with experiments we can obtain novel insights into the mechanisms of cancer invasion.

3. *Nuclear mechanics in cancer invasion*: Several experimental studies have established the importance of nuclear properties in regulating cancer invasion. In an earlier work, we had developed a CPM formalism wherein both cell and nucleus properties were taken into account [66]. For this case, the expression for the generalized energy (E_{total}) was modified from the earlier energy term (E_0) as follows:

$$E_{\text{total}} = E_0 + \sum_{\substack{\forall c_1, c_2 \\ \in \text{components}}} k_{c_1, c_2}(d_{c_1, c_2} - r_{c_1, c_2})^2 \qquad (24.2)$$

In these simulations, to make sure the nucleus remains within the cytoplasm, high positive interface energies were assigned for nucleus–matrix and nucleus–fluid interfaces, while high negative values were assigned to cytoplasm–cytoplasm, cytoplasm–nucleus, and nucleus–nucleus interfaces. In the second term, k_{c_1,c_2}, d_{c_1,c_2}, and r_{c_1,c_2} represent the strength of connectivity between the compartments c_1 and c_2, the desired distance between the compartments, and the Euclidean distance, respectively. This term was added to eliminate the possibility of cell fragmentation by constraining the inter-compartment distance between two neighbouring compartments. By simultaneously accounting for cell and nuclear stiffness, we showed that both cell and nuclear softening are necessary for migrating through sub-nuclear-sized channels with confinement history another determinant of invasion efficiency.

While insightful, these models are unable to capture several complex behaviour of nuclei including strain stiffening and plasticity. In this context, continuum mechanics and molecular dynamics models that account for distinct mechanical properties of nuclear lamins and chromatin better recapitulate the response of nuclei to physiologically relevant forces. We addressed this question in an earlier work wherein we probed the importance of nuclear and tissue properties in dictating the dynamics of confined cell migration by accounting for properties of the cell membrane, cytoplasm (E_c), and nucleus (E_n) [43]. The cell membrane, cell cytoplasm, nuclear membrane, and tissue (E_T) were modelled as viscoelastic Kelvin–Voigt materials; for these materials, the stress in the system depends on the strain (ε) and strain rate ($\dot{\varepsilon}$) and is given by the equation $\sigma = +K_\varepsilon + \eta\dot{\varepsilon}$, where K corresponds to the solid stiffness and η represents the fluid viscosity. The nucleus was assumed to be an elastoplastic ($\sigma = a + b\varepsilon^n$) solid to model the collective behaviour of the nuclear membrane and the chromatin fibres. Nuclear translocation was mediated by protrusive forces at the front edge of the cell (Figure 24.2A). For $E_c = 1$ Pa, $E_n = 1$ kPa, and $0.5 \leq E_T \leq 2$ kPa, forces required for pore entry were mostly dependent on the ratio of the undeformed nucleus size (D_0) to pore size (ϕ), with entry into small pores (i.e. $D_0/\phi = 1.67$) requiring higher forces (Figure 24.2B). A parameter study over a wider range of nuclear to tissue stiffness ($0.1 \leq E_n/E_T \leq 10$) revealed sharp drop in nuclear circularity (i.e. D/L) for moderately stiff nuclei ($E_n = 1$ kPa) and $E_n/E_T < 2$ (Figure 24.2C). This can be attributed to nuclei undergoing plastic deformation as apparent from the build-up of plastic strain for specific parameter combination of nuclear stiffness, tissue stiffness, and extent of confinement (Figure 24.2D and E). In addition, plastically deformed nuclei exhibited kink formation at the front edge indicative of nuclear rupture. Collectively, these simulations establish a physical basis for nuclear rupture during confined migration. The absence of kink formation in softer nuclei suggests that cells may dynamically soften their nuclei by Lamin A/C phosphorylation as a strategy to minimize nuclear rupture.

Similar to nuclear softening achieved by modulation of Lamin A/C levels, nuclear softening via TSA-induced chromatin decondensation also leads to faster confined migration and is associated with lesser force required for pore entry [67]. A molecular dynamics simulation set-up comprising of an inner crosslinked polymer (i.e. chromatin) surrounded by a polymer shell (i.e. lamina) was able to capture strain-stiffening behaviour of isolated nuclei as well as buckling of the nuclear lamina in the absence of chromatin [68]. While initial models of nuclear blebbing identified Lamin A:Lamin B levels and their spatial separation as important determinants of blebbing, a recent computational model suggests that blebbing occurs at sites of chromatin tethering to the nuclear lamina at lamin-associated domains [69].

24.7. Conclusion

Cancer invasion is a highly complex process dictated by multiple aspects of the ECM, the extent of cell–cell adhesion, and genetic/phenotypic heterogeneity. Successful recapitulation of the complex TME is expected to be incorporated in lab-on-chip systems amenable for high-throughput assays. While engineered systems that control one or more aspect of ECM properties have contributed significantly to our understanding of cancer invasion, these reductionist approaches fail to capture the complex interactions between distinct cell types within the TME that is now increasingly understood to play key roles in driving cancer progression. With single-cell RNAseq technologies being more accessible and rapid integration of AI/ML-based approaches for spatiotemporal mapping of genetic heterogeneity and epigenetic reprogramming, incorporation of this information within CPM modelling formalisms can make powerful predictions pertaining to the nature of cancer invasion and identify

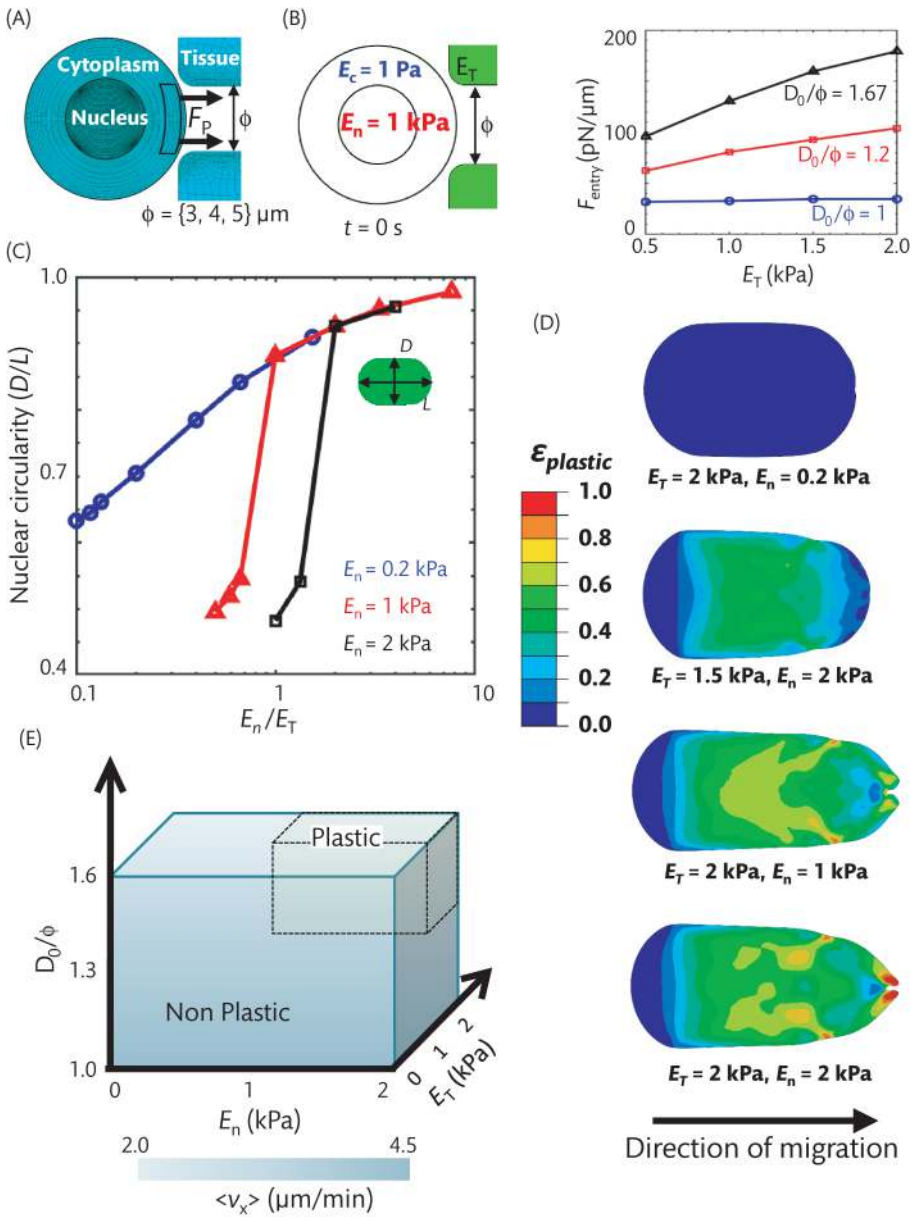


Figure 24.2. Nuclear mechanics during confined migration. (A) Schematic of a cell squeezing through a pore in a given tissue. (B) Force (F_{entry}) required for a cell to enter a pore of given size and its dependence on tissue stiffness ($0.5 \le E_T \le 2$ kPa) for different extents of confinement (D_0/ϕ). E_c and E_n were kept constant at 1 Pa and 1 kPa, respectively. (C) Dependence of nuclear circularity on $E_n = E_T$ for different values of E_n. (D) Spatial map of plastic strain accumulated in the nucleus just after pore entry for different values of E_n and E_T. (E) Predicted phase diagram depicting the zones of non-plastic and plastic nuclear deformation during pore entry for different values of E_n, $E_n = E_T$, and D_0/ϕ. *Source:* Adapted from Mukherjee et al. [43].

signalling nodes that may be explored for therapeutic targeting of cytoskeletal, matrix remodelling, and immune evasion pathways.

Competing interests

No competing interest is declared.

Author contributions

M.S., A., and S.S. conceived the outline of the book chapter. M.S. and A. wrote the initial draft. S.S. edited and finalized the manuscript.

Acknowledgements

The authors thank the anonymous reviewers for their valuable suggestions.

REFERENCES

1. Pozzi A, Yurchenco PD, Iozzo RV. The nature and biology of basement membranes. Matrix Biol. 2017;**57–58**:1–11.
2. Egeblad M, Rasch MG, Weaver VM. Dynamic interplay between the collagen scaffold and tumor evolution. Curr Opin Cell Biol. 2010;**22**(5):697–706.

3. Conklin MW, Eickhoff JC, Riching KM, Pehlke CA, Eliceiri KW, Provenzano PP, et al. Aligned collagen is a prognostic signature for survival in human breast carcinoma. Am J Pathol. 2011;**178**:1221–32.
4. Levental KR, Yu H, Kass L, Lakins JN, Egeblad M, Erler JT, et al. Matrix crosslinking forces tumor progression by enhancing integrin signaling. Cell. 2009;**139**(5):891–906.
5. Heneberg, P. Paracrine tumor signaling induces transdifferentiation of surrounding fibroblasts. Crit. Rev. Oncol. Hematol. 2016;**97**:303–11.
6. Ishii G, Ochiai A, Neri S, Phenotypic and functional heterogeneity of cancer-associated fibroblast within the tumor microenvironment. Adv Drug Deliv Rev. 2016;**99**(Pt B):186–96.
7. Elyada E, Bolisetty M, Laise P, Flynn WF, Courtois ET, Burkhart RA, et al. Cross-species single-cell analysis of pancreatic ductal adenocarcinoma reveals antigen-presenting cancer-associated fibroblasts. Cancer Discov. 2019;**9**(8):1102–23.
8. Bohaumilitzky L, Huber A-K, Stork, EM, Wengert S, Woel F, Boehm H. A trickster in disguise: Hyaluronan's ambivalent roles in the matrix. Front Oncol. 2017;**7**:242.
9. Schmaus A, Bauer J, Sleeman JP. Sugars in the microenvironment: the sticky problem of HA turnover in tumors. Cancer Met Rev. 2014;**33**(4):1059–79.
10. Chanmee T, Ontong P, Itano N. Hyaluronan: A modulator of the tumor microenvironment. Cancer Lett. 2016;**375**(1):20–30.
11. Itano N, Sawai T, Atsumi F, Miyaishi O, Taniguchi S, Kannagi R, et al. Selective expression and functional characteristics of three mammalian hyaluronan synthases in oncogenic malignant transformation. J Biol Chem. 2004;**279**(18):18679–87.
12. Sullivan WJ, Mullen PJ, Schmid EW, Flores A, Momcilovic M, Sharpley M.S, et al. Extracellular matrix remodeling regulates glucose metabolism through TXNIP destabilization. Cell. 2018;**175**(1):117–32.
13. Discher DE, Janmey PA, Wang Y-L. Tissue Cells Feel and Respond to the Stiffness of Their Substrate. Science 2005, 310 (5751): 1139–1143.
14. Paszek MJ, Zahir N, Johnson KR, Lakins JN, Rozenberg GI, Gefen A, et al. Tensional homeostasis and the malignant phenotype. Cancer Cell. 2005;**8**(3):241–54.
15. Chaudhuri O, Koshy ST, Branco da Cunha C, Shin JW, Verbeke CS, Allison KH, et al. Extracellular matrix stiffness and composition jointly regulate the induction of malignant phenotypes in mammary epithelium. Nat Mater. 2014;**13**(10):970–8.
16. Vader D, Kabla A, Weitz D, Mahadevan L. Strain-induced alignment in collagen gels. PLoS ONE. 2009;**4**(6):e5902.
17. Nam S, Hu KH, Butte MJ, Chaudhuri O. Strain-enhanced stress relaxation impacts nonlinear elasticity in collagen gels. Proc Natl Acad Sci U S A. 2016;**113**(20):5492–7.
18. Nam S, Lee J, Brownfield DG, Chaudhuri O. Viscoplasticity enables mechanical remodeling of matrix by cells. Biophys J. 2016;**111**:2296–308.
19. Stylianopoulos T, Martin JD, Snuderl M, Mpekris F, Jain SR, Jain RK. Coevolution of solid stress and interstitial fluid pressure in tumors during progression: implications for vascular collapse. Cancer Res. 2013;**73**(13):3833–41.
20. Dongre A, Weinberg RA. New insights into the mechanisms of epithelial–mesenchymal transition and implications for cancer. Nat Rev Mol Cell Biol. 2019;**20**(2):69–84.
21. Friedl P, Wolf K. Proteolytic interstitial cell migration: a five-step process. Cancer Met Rev. 2009;**28**:129–135.
22. Wolf K, Mazo I, Leung H, Engelke K, von Andrian UH, Deryugina EI, et al. Compensation mechanism in tumor cell migration: mesenchymal–amoeboid transition after blocking of pericellular proteolysis. J Cell Biol. 2003;**160**(2):267–77.
23. Liu YJ, Le Berre M, Lautenschlaeger F, Maiuri P, Callan-Jones A, Heuzé M, et al. Confinement and low adhesion induce fast amoeboid migration of slow mesenchymal cells. Cell. 2015;**160**(4):659–72.
24. Wisdom KM, Adebowale K, Chang J, Lee JY, Nam S, Desai R, et al. Matrix mechanical plasticity regulates cancer cell migration through confining microenvironments. Nat Commun. 2018;**9**(1):4144.
25. Satelli A, Mitra A, Brownlee Z, Xia X, Bellister S, Overman MJ, et al. Epithelial-mesenchymal transitioned circulating tumor cells capture for detecting tumor progression. Clin Cancer Res. 2015;**21**:899–906.
26. Liao T-T, Yang, M-H. Hybrid epithelial/mesenchymal state in cancer metastasis: clinical significance and regulatory mechanisms. Cells. 2020;**9**:623.
27. Giampieri S, Manning C, Hooper S, Jones L, Hill C.S, Sahai E. Localized and reversible TGFbeta signalling switches breast cancer cells from cohesive to single cell motility. Nat Cell Biol. 2009;**11**:1287–96.
28. Haeger A, Krause M, Wolf K, Friedl P. Cell jamming: collective invasion of mesenchymal tumor cells imposed by tissue confinement. Biochim Biophys Acta. 2014;**1840**:2386–95.
29. Wolf K, te Lindert M, Krause M, Alexander S, te Riet J, Willis LA, et al. Physical limits of cell migration: control by ECM space and nuclear deformation and tuning by proteolysis and traction force. J Cell Biol. 2013;**201**:1069–84.
30. Crisp M, Liu Q, Roux K, Rattner JB, Shanahan C, Burke, B, et al. Coupling of the nucleus and cytoplasm: role of the LINC complex. J Cell Biol. 2006;**172**:41–53.
31. Wu J, Kent IA, Shekhar N, Chancellor TJ, Mendonca A, Dickinson RB, et al. Actomyosin pulls to advance the nucleus in a migrating tissue cell. Biophys J. 2014;**106**:7–15.
32. Petrie RJ, Koo H, Yamada KM. Generation of compartmentalized pressure by a nuclear piston governs cell motility in a 3D matrix. Science. 2014;**345**:1062–65.
33. Thomas DG, Yenepalli A, Denais CM, Rape A, Beach JR, Wang Y-L, et al. Non- muscle myosin IIB is critical for nuclear translocation during 3D invasion. J Cell Biol. 2015;**210**:583–94.
34. Schiffhauer ES, Luo T, Mohan K, Srivastava V, Qian X, Griffis ER, et al. Mechanoaccumulative elements of the mammalian actin cytoskeleton. Curr Biol. 2016;**26**(11):1473–79.
35. Barai A, Mukherjee A, Das A, Saxena N, Sen S. α-actinin-4 drives invasiveness by regulating myosin IIB expression and myosin IIA localization. J Cell Sci. 2021; **134**(23):jcs.258581.
36. Stroka KM, Jiang H, Chen S-H, Tong Z, Wirtz D, Sun SX et al. Water permeation drives tumor cell migration in confined microenvironments. Cell. 2014;**157**(3):611–23.
37. de Leeuw R, Gruenbaum Y, Medalia O. Nuclear lamins: thin filaments with major functions Trends Cell Biol. 2018;**28**:34–45.
38. Machowska M, Piekarowicz K, Rzepecki R. Regulation of lamin properties and functions: does phosphorylation do it all? Open Biol. 2015;**5**:150094.
39. Zheng M, Jin G, Zhou Z. Post-Translational modification of lamins: mechanisms and functions. Front Cell Dev Biol. 2022;**10**:864191.
40. Swift J, Ivanovska IL, Buxboim A, Harada T, Dingal PC, Pinter J, et al. Nuclear lamin—a scales with tissue stiffness and enhances matrix-directed differentiation. Science. 2013;**341**(6149):1240104.

41. Das A, Barai A, Monteiro M, Kumar S, Sen S. Nuclear softening is essential for protease-independent migration. Matrix Biol. 2019;**82**:4–19.
42. Harada T, Swift J, Irianto J, Shin JW, Spinler KR, Athirasala A, et al. Nuclear lamin stiffness is a barrier to 3D migration, but softness can limit survival. J Cell Biol. 2014;**204**(5):669–82.
43. Mukherjee A, Barai A, Singh RK, Yan W, Sen S. Nuclear plasticity increases susceptibility to damage during confined migration. PLoS Comput Biol. 2020;**16**(10):e1008300.
44. Stephens AD, Liu PZ, Banigan EJ, Almassalha LM, Backman V, Adam SA, et al. Chromatin histone modifications and rigidity affect nuclear morphology independent of lamins. Mol Biol Cell. 2018;**29**(2):220–33.
45. Nava MM, Miroshnikova YA, Biggs LC, Whitefield DB, Metge F, Boucas J, et al. Heterochromatin-driven nuclear softening protects the genome against mechanical stress-induced damage. Cell. 2020;**181**(4):800–17.e22.
46. Kessenbrock K, Plaks V, Werb Z. Matrix metalloproteinases: regulators of the tumor microenvironment. Cell. 2010;**141**(1):52–67.
47. Bassiouni W, Ali MAM, Schulz R. Multifunctional intracellular matrix metalloproteinases: implications in disease. FEBS J. 2021;**288**(24):7162–82.
48. Lodillinsky C, Infante E, Guichard A, Chaligne R, Fuhrmann L, Cyrta J, et al. p63/MT1-MMP axis is required for in situ to invasive transition in basal-like breast cancer. Oncogene. 2016;**35**(3):344–57.
49. Alexander NR, Branch KM, Parekh A, Clark ES, Iwueke IC, Guelcher SA et al. Extracellular matrix rigidity promotes invadopodia activity. Curr Biol. 2008;**18**(17):1295–99.
50. Infante E, Castagnino A, Ferrari R, Monteiro P, Agüera-Gonzalez S, Paul-Gilloteaux P, et al. LINC complex-Lis1 interplay controls MT1-MMP matrix digest-on-demand response for confined tumor cell migration. Nat Commun. 2018 June **22**;**9**(1):2443.
51. Das A, Monteiro M, Barai A, Kumar S, Sen S. MMP proteolytic activity regulates cancer invasiveness by modulating integrins. Sci Rep. 2017;**7**(1):14219.
52. Kumar S, Das A, Barai A, Sen S. MMP secretion rate and inter-invadopodia spacing collectively govern cancer invasiveness. Biophys J. 2018;**114**(3):650–62.
53. Patwardhan S, Mahadik P, Shetty O, Sen S. ECM stiffness-tuned exosomes drive breast cancer motility through thrombospondin-1. Biomaterials. 2021;**279**:121185.
54. Mori H, Lo AT, Inman JL, Alcaraz J, Ghajar CM, Mott JD, et al. Transmembrane/cytoplasmic, rather than catalytic, domains of Mmp14 signal to MAPK activation and mammary branching morphogenesis via binding to integrin β1. Development. 2013;**140**(2):343–52.
55. Rupp PA, Visconti RP, Czirok A, Cheresh DA, Little CD. Matrix metalloproteinase 2-integrin alpha(v)beta3 binding is required for mesenchymal cell invasive activity but not epithelial locomotion: a computational time-lapse study. Mol Biol Cell. 2008;**19**(12): 5529–40.
56. Ugarte-Berzal E, Bailon E, Amigo-Jiménez I, Vituri CL, Hernández del Cerro M, Terol MJ, et al. A 17-residue sequence from the Matrix Metalloproteinase-9 (MMP-9) hemopexin domain binds 41 integrin and inhibits MMP-9-induced functions in chronic lymphocytic leukemia B cells. J Biol Chem. 2012;**287**(33):27601–13.
57. Enderling H, Alexander NR, Clark ES, Branch KM, Estrada L, Crooke C, et al. Dependence of invadopodia function on collagen fiber spacing and cross-linking: computational modeling and experimental evidence. Biophys J. 2008;**95**:2203–18.
58. Saitou T, Rouzimaimaiti M, Koshikawa N, Seiki M, Ichikawa K, Suzuki T. Mathematical modeling of invadopodia formation. J Theor Biol. 2012;**298**:138–46.
59. Hoshino H, Koshikawa N, Suzuki T, Quaranta V, Weaver AM, Seiki M. et al. Establishment and validation of computational model for MT1-MMP dependent ECM degradation and intervention strategies. PLoS Comput Biol. 2012;**8**(4):e1002479.
60. Kumar S, Das A, Barai A, Sen S. MMP secretion rate and inter-invadopodia spacing collectively govern cancer invasiveness. Biophys J. 2018;**114**(3):650–62.
61. Kim M-C, Li R, Abeyaratne R, Kamm RD, Asada H. A computational modeling of invadopodia protrusion into an extracellular matrix fiber network. Sci Rep. 2022;**12**(1):1231.
62. Hirashima T, Rens EG, Merks RMH. Cellular Potts modeling of complex multicellular behaviors in tissue morphogenesis. Dev Growth Differ. 2017;**59**(5):329–39.
63. Kumar S, Kapoor A, Desai S, Inamdar MM, Sen S. MMP secretion rate and inter-invadopodia spacing collectively govern cancer invasiveness. Sci Rep. 2016;**6**:19905.
64. Pally D, Pramanik D, Bhat R. An interplay between reaction-diffusion and cell-matrix adhesion regulates multiscale invasion in early breast carcinomatosis. Front Physiol. 2019;**10**:790.
65. Asadullah Kumar S, Saxena N, Sarkar M, Barai A, Sen S. Combined heterogeneity in cell size and deformability promotes cancer invasiveness. J Cell Sci. 2021;**134**(7):jcs250225.
66. Kumar S, Das A, Sen S. Multicompartment cell-based modeling of confined migration: regulation by cell intrinsic and extrinsic factors. Mol Biol Cell. 2018;**29**(13):15991610.
67. Heo S-J, Song KH, Thakur S, Miller LM, Cao X, Peredo AP, et al. Nuclear softening expedites interstitial cell migration in fibrous networks and dense connective tissues. Sci Adv. 2020;**6**:eaax5083.
68. Stephens AD, Banigan EJ, Adam SA, Goldman RD, Marko JF. Chromatin and lamin A determine two different mechanical response regimes of the cell nucleus. Mol Biol Cell. 2017;**28**:1984–96.
69. Lionetti MC, Bonfanti S, Fumagalli MR, Budrikis Z, Font-Clos F, Costantini G, et al. Chromatin and cytoskeletal tethering determine nuclear morphology in progerin-expressing cells. Biophys J. 2020;**118**:2319–32.

25

Adaptation of cancer cells to altered stiffness of the extra-cellular matrix

Christina R. Dollahon, Ting-Ching Wang, Srinikhil S. Vemuri, Suchitaa Sawhney, and Tanmay P. Lele

25.1. Introduction

Cancer is a complex disease that involves changes not only to the biochemical properties of cells and tissues but also to the mechanical properties of tissues. For example, palpation of tissue for variations in mechanical stiffness has been used in clinical diagnosis of different types of cancer for a long time [1–5]. Quantitative measurements of mechanical stiffness with techniques such as atomic force microscopy (AFM) have revealed a general 'stiffening' of tissue in cancer (Table 25.1). While cancer tissues tend to be stiffer than their normal counterparts, cancer cells themselves tend to be softer [21]. Instead, mechanical stiffening of tissue is caused in part by increased deposition of extra-cellular matrix (ECM) proteins in the tumour microenvironment and subsequent crosslinking of ECM fibrils and networks [22–27]. Importantly, cells in tissues sense and respond to the stiff tissue microenvironment in a way that enhances their growth rate and their migratory potential [28]; both these phenotypes contribute to disease progression [22,29,30]. As such, an increasing number of studies are beginning to focus on the mechanisms for stiffening of tumour tissue and the resulting cellular responses.

That cells can differentiate between mechanically soft and stiff extra-cellular microenvironments was demonstrated through studies starting in the late 1990s from the groups of Wang and Dembo [28,31–33]. One of the first papers on this topic reported imaging of motile fibroblasts on polyacrylamide hydrogels that were conjugated covalently with the matrix protein type I collagen [34]. Fibroblasts crawled faster on soft polyacrylamide substrates than on stiff polyacrylamide substrates. Epithelial cells seeded on the gels did not migrate much, but the examination of the dynamics of epithelial cell focal adhesion assembly on these gels revealed that vinculin-labelled focal adhesions underwent rapid turnover on soft but not stiff gels. These studies demonstrated that cell phenotype was sensitive to the *mechanical* stiffness of the underlying substrate.

These initial discoveries motivated many studies that have continued to this day (Figure 25.1), which seek to address the underlying causes for cellular sensitivity to the mechanical stiffness of substrates, to identify *in vivo* contexts in which cellular sensitivity to matrix stiffness could be potentially important [33,35], and to understand the role of tissue stiffness in human diseases [36]. It is now known that many cell types, including fibroblasts [34], endothelial cells [37], neurons [38], epithelial cells [34], cardiomyocytes [39], mesenchymal stem cells [40], and, important for this chapter, cancer cells [41–43], are profoundly affected by substrate stiffness. These studies span a range of cellular responses in *in vitro* model systems that have been developed over the years to present controlled mechanical substrate stiffness to cells, *in vivo* studies in animal models, and emerging approaches to quantify changes in tissue mechanical stiffness in humans [44].

This chapter examines the emerging role of ECM mechanical stiffness in the context of cancer. We describe some of the model systems that can be used to study cancer cell responses to ECM stiffness. We present an overview of the impact of changing mechanical stiffness on cancer cell phenotypes, such as migration and proliferation, and the molecular mechanisms that mediate sensing of ECM stiffness.

25.2. Cellular response to model substrates of controlled stiffness

25.2.1. Cellular response to 2D substrates

Polyacrylamide gels that were used in early studies by the Wang group have remained the workhorse for the field of ECM stiffness sensing. Polyacrylamide gels can easily be synthesized through the polymerization of acrylamide with bis-acrylamide. High ratios of bis to acrylamide result in stiff hydrogels, while low ratios result in soft hydrogels. Hydrogel stiffness is commonly quantified in terms of the Young's modulus, which is a ratio of the extensional stress applied to a hydrogel block divided by the strain at steady state. The resulting hydrogels must be covalently conjugated with matrix proteins of choice, such as fibronectin or collagen, in order to allow cells to adhere and spread on these substrates [45]. Despite some

Table 25.1. Comparison of tissue stiffness in cancer.

Species of origin	Diagnosis	Parameter[a]	Method	Reference
Human	Control colon tissue (collagen-rich regions)	EM = 0.8 ± 0.4 kPa	AFM	[6]
	Colon carcinoma (collagen-rich regions)	EM = 2.40 ± 1.83 kPa		
Human	Normal colon tissue	SM = 1.52 kPa	Strain rheometry	[7]
	Cancerous colon tissue	SM = 9.60 kPa		
Human	Benign breast lesion	YM = 28 kPa	Shear wave elastography (SWE)	[8]
	Ductal breast carcinoma *in situ*	YM = 76 kPa		
	Lobular breast carcinoma *in situ*	YM = 82 kPa		
	Invasive breast cancer	YM = 140 kPa		
Human	Healthy fibroglandular breast tissue	SS = 7.5 ± 3.6 kPa	Magnetic resonance elastography (MRE)	[9]
	Breast tumour	SS = 33 kPa		
Human	Adjacent healthy bladder tissue	YM ≈ 3 kPa	AFM	[10]
	Newly diagnosed bladder cancer tissue	YM ≈ 8 kPa		
	Recurrent bladder cancer tissue	YM ≈ 13 kPa		
Mouse xenograft	Non-mesenchymal ovarian tumours	YM ≤ 60 kPa	SWE	[11]
	Ovarian cancer	YM = 120–140 kPa		
Human	Benign prostate tissue	YM = 74.9 ± 47.3 kPa	Shear wave imaging (SWI)	[12]
	prostate intraepithelial neoplasia/atypia	YM = 83.3 ± 38.6 kPa		
	Prostate cancer	YM = 133.7 ± 57.6 kPa		
Human	Normal prostate tissue	CYM = 15.9 ± 5.9 kPa	Stress relaxation tests	[13]
	Prostate cancer	CYM = 40.4 ± 15.7 kPa		
Human	Non-tumour brain gliosis	YM = 10–180 Pa	AFM	[14]
	Lower grade glioma	YM = 50–1,400 Pa		
	Glioblastoma	YM = 70–13,500 Pa		
Human	Normal brain tissue	YM = 7.3 ± 2.1 kPa	SWE	[15]
	Meningioma	YM = 33.1 ± 5.9 kPa	SWE	[15]
	Low-grade glioma	YM = 23.7 ± 4.9 kPa		
	High-grade glioma	YM = 11.4 ± 3.6 kPa		
	Brain metastases	YM = 16.7 ± 2.5 kPa		
Human	Normal liver tissue	SS = 2.3 ± 0.3 kPa	MRE	[16]
	Normal liver tissue	*S* = 2.1 kPa		[17]
	Fibrotic liver tissue	SS = 5.9 ± 2.5 kPa		[16]
	Benign liver tumour	SS = 2.7 ± 0.4 kPa		
	Malignant liver tumour	SS = 10.1 ± 3.6 kPa		
	Well/moderately differentiated hepatocellular carcinoma (HCC)	*S* = 6.5 ± 1.2 kPa		[18]
	Poorly differentiated HCC	*S* = 4.9 ± 1.2 kPa		
Human	Normal pancreatic tissue	SS = 2.47 ± 0.11 kPa	MRE	[19]
	Pancreatic cancer	SS = 6.06 ± 0.49 kPa		
Human	Lymph node without metastasis	*S* = 1.23 ± 0.50 g/cm	Tactile sensor	[20]
	Lymph node with metastasis	*S* = 3.35 ± 1.57 g/cm		

[a]EM: elastic modulus; SM: storage modulus; YM: Young's modulus; SS: shear stiffness; CYM: complex Young's modulus; *S*: stiffness.

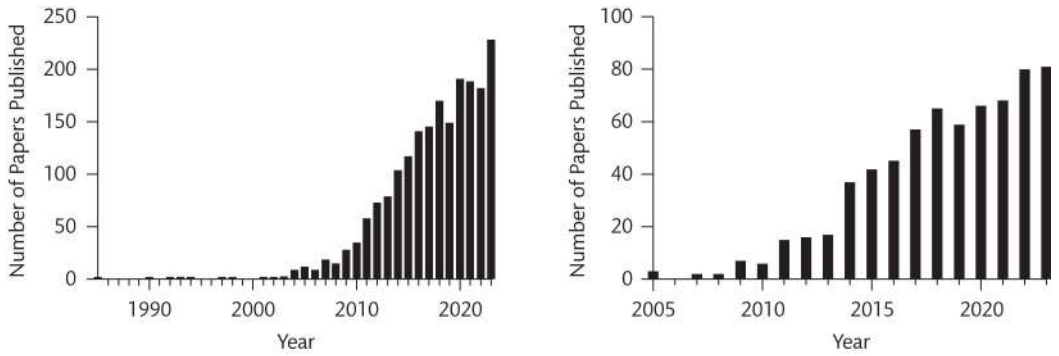


Figure 25.1. Left: Results of a search at https://pubmed.ncbi.nlm.nih.gov/ with keywords ' "ECM stiffness" OR "matrix stiffness" ' OR "matrix rigidity" '. Right: Results of a search at https://pubmed.ncbi.nlm.nih.gov/ with keywords ' "ECM stiffness" OR "matrix stiffness" OR "matrix rigidity" AND "cancer" '.

controversy about whether polyacrylamide gel stiffness can be controlled independently from ligand tethering and porosity [46], we and others have shown that cells do indeed sense stiffness in this system [47,48].

When cultured on soft and stiff polyacrylamide hydrogels, the canonical response of most normal cell types (with the exception of neutrophils [49] and Schwann cells [50]) is that they spread less on the soft gels and more on stiff gels [40,42,49,51–56]. Higher spreading on stiff substrates compared to soft substrates was also observed in different cancer cell lines, including MDA-MB-231 cells (breast cancer), HN and BHY cells (oral squamous cell carcinoma), A549 cells (lung cancer), BxPC-3 and PANC-1 cells (pancreatic cancer), HT-1080 cells (fibrosarcoma), and multiple glioblastoma cell lines [41,57,58] (see **Figure 25.2** for an example with MDA-MB-231 human breast cancer cells). There are exceptions to this rule. For example, the prostate cancer cell line PC-3 spread equally well on both soft (150 Pa) and stiff (4,800 Pa) substrates, while the pancreatic cancer cell line mPanc-96 did not spread well on either soft or stiff substrates [41]. Likewise, H-ras-transformed NIH 3T3 fibroblasts spread the same on substrates of Young's moduli ranging from 14 to 33 kPa, unlike normal NIH 3T3 fibroblasts [32]. The H-ras-transformed cells did begin to spread less once the substrate stiffness was lowered to 10 kPa.

The motor-clutch model proposed by Mitchison and Kirschner [59], Chan and Odde [60], and later refined by Roca-Cusachs and co-workers [61] explains cell spreading in general, and more specifically explains why cells tend to spread less on soft ECM and more on stiff ECM. In the motor-clutch model, myosin motors generate force by binding to F-actin filaments newly assembled at the leading edge of the cell. Matrix–protein-bound transmembrane integrin receptors act as the molecular clutches that transmit this force to the compliant substrate's ECM proteins, which deform in response to the force. On stiff substrates, talin unfolds before the integrin–ECM bonds dissociate. This exposes its cryptic binding sites for further

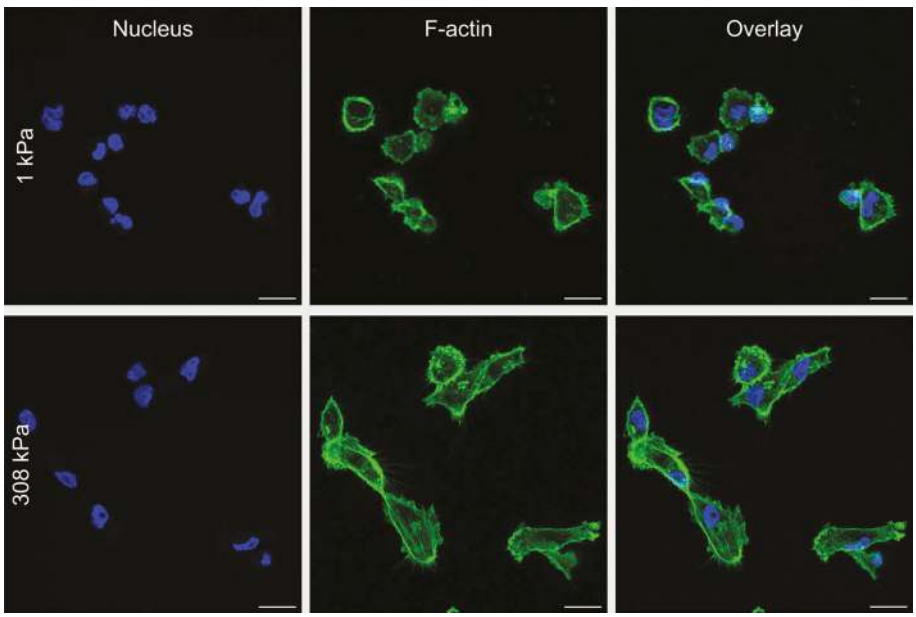


Figure 25.2. MDA-MB-231 human breast cancer cells seeded on polyacrylamide gels coated with collagen I of 1 kPa (5% acrylamide and 0.1% bis-acrylamide; top panel) and 308 kPa (15% acrylamide and 1.2% bis-acrylamide; bottom panel). Cells were fixed after three days of culture and stained with Hoechst 33342 and Phalloidin. Scale bar is 20 μm.

F-actin and vinculin binding and reinforcement, crosslinking the cytoplasmic integrin tails. This crosslinking reinforces adhesion and therefore increases cell spreading. On soft substrates, the integrin–ECM protein bond dissociates before talin can unfold and expose its binding sites. Thus, the motor-clutch model predicts less traction stress on softer ECM compared to stiffer ECM, which correlates with decreased spreading.

Traction force microscopy was developed in the 1990s to quantify the traction stress cells exert on the underlying substrate [31,62–64]. The technique involves culture of cells on soft polyacrylamide gels of a known Young's modulus containing sub-micron-sized fluorescent beads. Computational analysis of the motion of gel-embedded beads when cells migrate or spread on the gel allows the calculation of spatially distributed, cellular traction stress on the substrate. Alternative methods have been developed and refined over the years that directly measure the force of deformation by culturing cells on a bed of vertical pillars made of compliant poly-di-methoxy-silane (PDMS) or of flexible nanowires [65–68]. The top and bottom of the micropillars can be imaged with fluorescence microscopy (or scanning electron microscopy in the case of nanowires) and deflection of the pillar quantified. Knowledge of the Young's modulus and the geometry of the pillar allow calculation of the force applied by cells on top of the pillars. Another technique images deformations in fibronectin-printed patterns on a substrate of known area and modulus to calculate traction forces [69]. Such techniques have revealed that cancer cells exerted lower traction on soft substrates, and higher traction on stiff substrates, consistent with the motor-clutch model [70,71]. H-ras-transformed NIH 3T3 fibroblasts were an exception to this rule as they exerted roughly the same amount of traction force on soft (14 kPa) and stiff (33 kPa) substrates [32]. Furthermore, when presented with a gradient of ECM stiffness, cancer cells migrated towards the ECM stiffness that allows them to generate maximal traction, which is also predicted by the computational motor-clutch model [72].

The way traction force varies between healthy, less metastatic, and more metastatic cancer cells is still unclear. Conflicting results have been reported in studies that compare traction forces exerted by healthy and cancerous cells. One study found that MDA-MB-231 breast cancer cells, PC3 prostate cancer cells, and A549 lung cancer cells exerted more traction force than healthy cells independent of substrate stiffness [71]. Other studies have found that MCF7 breast cancer cells, transformed cells, and H-ras-transformed NIH 3T3 fibroblasts exerted less traction force on the substrates [73,74]. There are also conflicting reports regarding differences in traction force exerted by cancerous cells of different metastatic abilities. When murine breast cancer cells of varying metastatic potential were compared, the more metastatic cell lines exerted less traction force than the less metastatic cell lines [75]. Contrary to this finding, MDA-MB-231 cells transferred more strain energy on micropatterned deformable substrates than the less metastatic MCF-7 cell line [69], in alignment with other studies [71,73].

Correlated with the inability of cells to spread on soft substrates, cells proliferate less on soft gels compared to stiff gels [41,76–79]. Proliferation is insensitive to substrate stiffness in cell lines in which spreading is insensitive to substrate stiffness [41]. Again, there are exceptions to the rule. For example, Schwann cells proliferated the most on an intermediate stiffness (7.45 kPa), which is where they also spread best [50]. Furthermore, the proliferation of cancerous and transformed cells was less affected by substrate stiffness than the corresponding normal cells [73].

Other two-dimensional (2D) model systems to study cell responses include PDMS, 2D collagen gels, hyaluronic acid, and azobenzene gels. The stiffness of PDMS can be controlled by varying the crosslinking ratio of vinyl-terminated base to methyl hydrogen siloxane and using different curing methods, achieving a range of stiffnesses from 0.1 kPa to 10 MPa [46,80]. Collagen gel stiffness can be controlled by synthesizing gels that have low concentration of collagen (e.g. 1 and 3 mg/ml corresponding to 124 ± 8 and 502 ± 48 Pa stiffness, respectively) [81]; however, this method also changes the density of collagen molecules that are presented to cells. Another approach is to change substrate stiffness by varying the height of PDMS microposts because shorter posts are stiffer to horizontal traction forces generated by cells [65,82]. In these systems also, similar responses have been observed as with polyacrylamide gels—cells in general spread more on stiff substrates compared with soft substrates and proliferated more on stiff substrates compared with soft substrates [82–84].

25.2.2. Cellular response to 3D substrates

While 2D substrates have proven invaluable in fundamental studies of cellular sensing and response to stiffness, cells *in vivo* rarely see flat 2D surfaces. Instead, cells inhabit complex microenvironments that present three-dimensional (3D) cues to cells. Recognizing this limitation of 2D substrates, numerous studies over the past two decades have developed 3D gels of controlled stiffness (reviewed in [85]).

Early studies used 3D collagen gel models in which the concentration of collagen was altered systematically to stiffen the gel. Culture of human mammary epithelial cells in 3D basement membrane/collagen gels of low stiffness (167–170 Pa) resulted in hollow acinar structures, which are typical of glandular structures *in vivo*. However, culture of these cells in stiffer gels (170–1,200 Pa) led to a loss of the lumen in the acinar structures, and at the highest stiffness values, disrupted the cells' basal polarity [86]. One drawback of using collagen gel models is that increasing collagen concentration to increase gel stiffness can have unintended effects on the extent of integrin binding to the gel [86]. To avoid this problem, basement membrane–polyacrylamide gels were used with a stiffness range from 150 to over 5,000 Pa, with similar results [86]. These results are compatible with measurements of mice mammary tumours and healthy mouse mammary glands revealing higher elastic moduli of the cancerous tissue and surrounding stroma [86].

Stiffening of 3D gels can switch cells between differentiation or proliferation. For example, breast epithelial cells cultured in 'floating' gels, which are unattached to the underlying substrate and permit deformation of the gel (effectively acting as soft ECM), assembled tubules, reflecting a differentiated phenotype [87]. When cultured in adherent gels, the cells did not assemble tubules but proliferated. Thus, stiffening of the ECM may promote a loss of tissue structure and uncontrolled proliferation that can contribute to tumorigenesis.

Alternative approaches to tune stiffness of 3D culture systems include 3D hyaluronic acid gels crosslinked with genipin [88], gelatin methacryloyl (GelMA) gels stiffened by nanoparticles [89,90], poly(ethylene glycol) (PEG)-based gels, and alginate gels. 3D gels can also be made of synthetic compounds [91,92]. Experiments with cancer cells seeded in 3D gels have revealed that cancer cells often respond differently to stiffness than on 2D gels. A study culturing

MDA-MB-231 breast cancer cells in 3D GelMA hydrogels observed more cell spreading in softer gels (1–3 kPa) compared to stiffer gels (10–15 kPa) [58]. Other studies showed that cancer cells proliferated more in soft 3D gels, contrary to results in 2D gel studies [93,94]. For example, when U87 glioblastoma cells were cultured in 3D PEG-based hydrogels of different stiffnesses, cell proliferation and spreading was greater in the softer gels [92]. In addition to increased proliferation, cancer cells invaded softer gels at a faster rate [95]. However, Huh7-transformed cells were found to proliferate more in stiffer gels, suggesting that cancer cell response to 3D gel stiffness may be dependent on cell type [96].

While most studies of cancer cell interactions with substrates of tunable stiffness have focused on static stiffness values, methods have emerged to dynamically tune stiffness. These studies are motivated by the fact that the stiffness inside a developing tumour *in vivo* undergoes a gradual change over time. In one such approach, mammary epithelial cells were seeded on partially crosslinked 2D methacrylated glycosaminoglycan hyaluronic acid (MeHA) hydrogels of tunable stiffness, coated with collagen [97]. After seeding, the cells were covered in Matrigel, which polymerizes around cells, allowing their development into 3D acinar structures. Next, exposure to UV light and free radical donors was used to stiffen the gels dynamically. This caused the acinar structure to lose its integrity with dissociation of spheroids into individual cells. When a 3D alginate–Matrigel model system was used to dynamically tune stiffness, breast epithelial cells became more invasive and proliferated more in a stiffened substrate [98]. Another system used azobenzene hydrogels stiffened dynamically with UV light to investigate the response of MCF7 breast cancer cells to stiffening in 2D culture. The cells were cultured on gels stiffened at different timepoints after seeding: cell aggregates grew more on gels stiffened 12 h after seeding and reduced on gels stiffened 36 h after seeding [99].

Studies that tune the stiffness of 3D hydrogels can be complicated by additional effects of gel stiffening in three dimensions, including a smaller pore size, changes to the matrix accessibility of ECM proteins, and the degradability of the matrix [55,88,100,101]. Also, emerging studies show that viscoelasticity is an additional variable present in many tissues *in vivo* which impacts cell spreading and proliferation [102]. For example, MDA-MB-231 and MCF7 breast cancer cell spheroids, as well as HT-1080 fibrosarcoma spheroids, grew larger in fast-relaxing (less viscous) 3D gels [103]. The impact of viscoelasticity can be seen in 2D culture as well as 3D: U20S osteosarcoma cells spread significantly differently on viscoelastic substrates than purely elastic substrates of the same Young's modulus [104]. Huh7 liver cancer cells spread more on viscoelastic substrates than purely elastic ones, in contrast to normal hepatocytes [105].

25.3. A systems biology perspective of cellular adaptation to ECM stiffness

As mentioned above, the most robust and canonical response of cells to stiffness in 2D model systems is that they spread less on soft ECM and more on stiff ECM. The lack of spreading, caused by an inability to exert sufficient traction on the ECM, can be explained by the motor-clutch model. Particularly relevant to cancer is the fact that cancer cells tend to proliferate less when less spread in otherwise identical 2D systems. Cell behaviour also varies on soft and stiff substrates in terms of differentiation. For example, stem cells differentiate into neuronal lineages on soft ECM, muscle cells on intermediate stiffness, and bone cells on the stiffest substrates [40]. All this occurs in a constant soluble environment of growth factors, cytokines, and nutrients.

Because differences in cell spreading on soft and stiff ECM are a robust and reproducible response in normal and cancer cells, it is likely that cell shape is a major mediator of the effect of ECM stiffness on cell behaviour, including differentiation and proliferation. For example, it has been shown in other contexts that controlling the degree of cell spreading can switch cells between different fates. Ingber and co-workers demonstrated this principle by culturing endothelial cell shapes on square fibronectin-coated patterns that forced them to spread less in area or allowed them to spread more [78]. Endothelial cells that were spread less underwent apoptosis, while cells that spread more proliferated more. Again, all this occurred in an otherwise identical soluble environment of growth factors and cytokines. Likewise, Chen and co-workers found that human mesenchymal stem cells differentiated to adipocytes when forced to spread less on micropatterned islands, while they differentiated to osteoblasts when allowed to spread more [106].

How might controlling cell shape regulate cell fate? The motor-clutch model holds that cell shape is established by F-actin polymerization pushing on the cell membrane when the F-actin filaments engage with the integrin receptor clutches. Integrin engagement, recruitment of adhesion proteins to integrins, and eventual stable assembly of focal adhesions can trigger signalling pathways in cells. This is because focal adhesions are themselves extremely complex assemblies of more than 50 different proteins, many of which are enzymes that can modulate signalling pathways [107]. For example, breast epithelial cells in floating collagen gels initially contract the floating (soft) gels, but the contraction results in smaller adhesions with decreased phosphorylation of focal adhesion kinase, and decreased Rho and ROCK activity, resulting in an eventual decrease in contractility and tubulogenesis [87]. Conversely, cells cultured in adherent (hence stiff) collagen gels had an increased phosphorylation of focal adhesion kinase and associated high Rho–ROCK-mediated contractility, which promoted proliferation [87]. As another example, when mouse embryonic fibroblasts were cultured on stiff matrices, focal adhesion kinase was activated, which in turn activated Rac, and Cyclin D1, triggering the cell cycle and associated proliferation [108].

Cell shape can also modulate cell fate independently of adhesion assembly. Ingber and co-workers micropatterned spread or unspread/round morphologies of cells while keeping the adhesive area constant [78]. Spread cells still proliferated while round cells did not, which led them to propose that cell shape alone can control cell proliferation. Because controlling cell shape also modulates the polymerization of F-actin, and its bundling into actomyosin stress fibres, it is possible that these changes may mediate the effect of cell shape on cell proliferation [109,110]. Another consequence of cell shape control is the control of nuclear shape because nuclear shape generally conforms to cell shape [111,112]. Changes in nuclear shape can also alter chromatin conformation that can in turn alter gene expression [113–116].

Cell spreading can induce nuclear translocation and activation of the transcriptional co-activator yes-associated protein (YAP) independently of the Hippo pathway [117]. This phenomenon,

first reported by Dupont et al. [117], has been extensively studied. Different mechanisms have been proposed to explain YAP sensitivity to mechanical cues. In one proposed mechanism, talin-depleted mouse embryonic fibroblasts spread on stiff ECM have compressed nuclei with stretched nuclear pores, which contributes to increased nuclear import of YAP [118]. Tension in the nuclear envelope has been suggested to cause YAP nuclear localization in mesenchymal stem cells in 3D methacrylated hyaluronic acid (MeHA) hydrogels [119].

In an alternative mechanism, cell spreading on stiff ECM promotes YAP nuclear localization by increasing the availability of unphosphorylated, cytoplasmic YAP for import by the nuclear transport receptor importin-7 (Imp7) [120]. Cortical actomyosin structures assemble in spread cells on stiff ECM and inhibit the Hippo kinases MST1/2 that normally phosphorylate YAP [120], or sequester Amot, a protein that phosphorylates YAP [121]. Shifting the balance of cortical actin to cytoplasmic actin results in phosphorylation of YAP, preventing its nuclear entry [121]. YAP mechanosensing is also observed in diverse cancer cell lines with YAP being present in cells cultured on stiff, but not soft, hydrogels [58]. YAP mechanosensitivity is abrogated in these cancer cell lines upon depletion of Lamin A/C, through mechanisms that are not fully understood. A loss of nuclear YAP caused by a decrease in Lamin A/C has also been shown to occur in embryonic development [121]. The loss of Lamin A/C may promote YAP phosphorylation by modulating the balance of cortical versus cytoplasmic f-actin filaments [121]. Since Lamin A/C is down-regulated in some cancers, it may modulate cell fate by perturbing YAP mechanosensing in these contexts [122].

As the foregoing discussion shows, the mechanisms by which cell shape control alters cell fate are diverse and complex. The expression of many hundreds of genes can become altered when cells are cultured on soft ECM compared to stiff ECM [123]. These alterations are reflective of the modulation of many diverse signalling pathways. And yet, alterations in so many variables result in ECM stiffness-mediated switching between relatively few cell fates: proliferation, differentiation, or apoptosis. This is consistent with the concept that cell fates may be attractors in cell regulatory networks [124,125]. That is, cell fates correspond to minima of potential energy in an N-dimensional potential energy landscape, where N refers to the number of independent variables that describe regulatory networks (gene expression, enzyme activity, protein states, etc.). In this conceptual picture, cell shape has been proposed to be a physical parameter that selects between different potential energy minima. Such a 'systems-level' understanding of the relation between cell shape and cell fate may be useful because cell shape control can switch between cell fates, and cell fates correspond to large differences in protein levels and signalling pathways [124]. Cell shape as a control parameter appears to override inputs from soluble factors in the environment as well. As such, this property of cells influences the integration of the biochemical signalling machinery that ultimately determines fate.

25.4. Cancer heterogeneity and response to ECM stiffness

Cancer is fundamentally a genetic disease [126]. Proliferating cancer cells in a growing tumour have remarkable heterogeneity in genetic makeup [127]. Yet, studies examining the role of ECM stiffness in cancer have not accounted for genetic tumour heterogeneity, representing a crucial knowledge gap because individual genetic variants might respond to ECM stiffness differently from the responses reported from population-level studies.

Selection of individual tumour cell clones has been observed in tumours [128] both due to selection pressure from the changing microenvironment and pressure imposed by drug treatment [129]. It is therefore reasonable to expect that changes in the ECM stiffness in a growing tumour can result in selection of genetic clones from the growing, heterogeneous population of tumour cells. Likewise, the selection of clones on soft ECM may occur *in vivo* when cells migrate from a stiff tumour to tissues that are softer than the tissue of origin, such as cells that migrate from the stiffer breast to the soft brain or from stiff bone to softer lung. Selection in the new microenvironment could lead to functional consequences, such as comparable cell proliferation even in the new tissue with distinct mechanical properties from the tissue of origin. Selected metastatic cells may also exhibit differential responses to pharmacological targeting. However, such possibilities are *invisible* in population-level studies of cancer cell responses to ECM stiffness, which focus on average properties and do not account for clone-to-clone genetic variations.

The possibility of evolution of cell populations due to selection by changes in ECM stiffness is supported by our recent study that demonstrated evolution of genetically variable fibroblast and myoblast populations [130]. We found that not all cells exhibit canonical phenotypic plasticity on soft vs. stiff ECM. Instead, some mutant clones within populations of fibroblasts exhibit phenotypic plasticity behaviours opposite those of the overall cell population, in which clones spread and proliferate well on soft ECM but poorly on stiff ECM (Figure 25.3). Our studies established two new concepts. (1) Mutation reverses the phenotypic plasticity of cells in response to ECM stiffness. (2) Soft ECM consistently selects deviant clones that 'do the wrong thing', resulting in cell populations with plasticity and

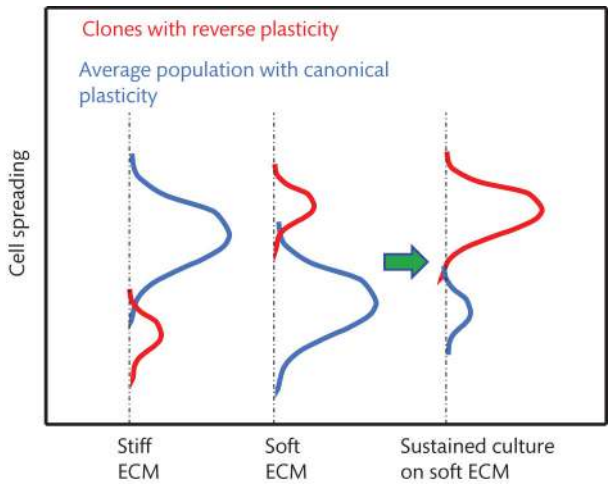


Figure 25.3. Reverse plasticity in the cellular response to ECM stiffness. Mutant clones depicted in red spread less on stiff ECM and more on soft ECM, which is opposite from the canonical plasticity of the average population (blue). The red clones proliferate more on soft ECM than blue clones due to increased spreading and eventually outcompete blue clones, giving rise to a population dominated by deviant clones that drive tumour behaviours.

cellular behaviours opposite those found in the canonical pattern. We observed similar selection in two cancer cell lines: MDA-MB-231 cancer cells and HT-1080 fibrosarcoma cells in another recent study [131] although we did not observe reverse plasticity in these cells. Instead, selected cells spread equally well on both soft and stiff ECM. Overall, these results indicate that future studies of cancer cell behaviour on soft and stiff ECM should take into account the genetic heterogeneity of starting cell populations, and that laboratory evolution is a useful technique for such studies. It is important to understand the relationship between genotype and cancer cellular plasticity in response to ECM stiffness, if we are to fully understand the impact of mechanical changes in the progression of cancer, and in the development of new therapies [132,133].

Acknowledgements

This work was supported by NIH U01 CA225566 (T.P.L.), a CPRIT-established investigator award (grant number RR200043; T.P.L.), and NSF grant numbers 2412520 and 2226157.

REFERENCES

1. Mahoney L, Csima A. Efficiency of palpation in clinical detection of breast cancer. Can Med Assoc J. 1982;**127**(8):729–30.
2. Palacio-Torralba J, Reuben RL, Chen Y. A novel palpation-based method for tumor nodule quantification in soft tissue—computational framework and experimental validation. Med Biol Eng Comput. 2020;**58**(6):1369–81.
3. Shaw J. Diagnosis and treatment of testicular cancer. Am Fam Physician. 2008;**77**(4):469–74.
4. Baird DC, Meyers GJ, Hu JS. Testicular cancer: diagnosis and treatment. Am Fam Physician. 2018;**97**(4):261–8.
5. Borkenhagen JF, Eastwood D, Kilari D, See WA, Van Wickle JD, Lawton CA, et al. Digital rectal examination remains a key prognostic tool for prostate cancer: a national cancer database review. J Natl Compr Cancer Netw. 2019;**17**(7):829–37.
6. Brauchle E, Kasper J, Daum R, Schierbaum N, Falch C, Kirschniak A, et al. Biomechanical and biomolecular characterization of extracellular matrix structures in human colon carcinomas. Matrix Biol. 2018;**68–69**:180–93.
7. Deptuła P, Łysik D, Pogoda K, Cieśluk M, Namiot A, Mystkowska J, et al. Tissue rheology as a possible complementary procedure to advance histological diagnosis of colon cancer. ACS Biomater Sci Eng. 2020;**6**(10):5620–31.
8. Evans A, Whelehan P, Thomson K, McLean D, Brauer K, Purdie C, et al. Quantitative shear wave ultrasound elastography: initial experience in solid breast masses. Breast Cancer Res. 2010;**12**(6):R104.
9. McKnight AL, Kugel JL, Rossman PJ, Manduca A, Hartmann LC, Ehman RL. MR elastography of breast cancer: preliminary results. Am J Roentgenol. 2002;**178**(6):1411–7.
10. Ghasemi H, Mousavibahar SH, Hashemnia M, Karimi J, Khodadadi I, Mirzaei F, et al. Tissue stiffness contributes to YAP activation in bladder cancer patients undergoing transurethral resection. Ann N Y Acad Sci. 2020;**1473**(1):48–61.
11. Mieulet V, Garnier C, Kieffer Y, Guilbert T, Nemati F, Marangoni E, et al. Stiffness increases with myofibroblast content and collagen density in mesenchymal high grade serous ovarian cancer. Sci Rep. 2021;**11**(1).
12. Ahmad S, Cao R, Varghese T, Bidaut L, Nabi G. Transrectal quantitative shear wave elastography in the detection and characterisation of prostate cancer. Surg Endosc. 2013;**27**(9):3280–7.
13. Zhang M, Nigwekar P, Castaneda B, Hoyt K, Joseph JV, di Sant'Agnese A, et al. Quantitative characterization of viscoelastic properties of human prostate correlated with histology. Ultrasound Med Biol. 2008;**34**(7):1033–42.
14. Miroshnikova YA, Mouw JK, Barnes JM, Pickup MW, Lakins JN, Kim Y, et al. Tissue mechanics promote IDH1-dependent HIF1α–tenascin C feedback to regulate glioblastoma aggression. Nat Cell Biol. 2016;**18**(12):1336–45.
15. Chauvet D, Imbault M, Capelle L, Demene C, Mossad M, Karachi C, et al. In vivo measurement of brain tumor elasticity using intraoperative shear wave elastography. Ultraschall Med–Eur J Ultrasound. 2015;**37**(06):584–90.
16. Venkatesh SK, Yin M, Glockner JF, Takahashi N, Araoz PA, Talwalkar JA, et al. MR elastography of liver tumors: preliminary results. Am J Roentgenol. 2008;**190**(6):1534–40.
17. Trout AT, Anupindi SA, Gee MS, Khanna G, Xanthakos SA, Serai SD, et al. Normal liver stiffness measured with MR elastography in children. Radiology. 2020;**297**(3):663–9.
18. Thompson SM, Wang J, Chandan VS, Glaser KJ, Roberts LR, Ehman RL, et al. MR elastography of hepatocellular carcinoma: correlation of tumor stiffness with histopathology features-Preliminary findings. Magn Reson Imaging. 2017;**37**:41–5.
19. Itoh Y, Takehara Y, Kawase T, Terashima K, Ohkawa Y, Hirose Y, et al. Feasibility of magnetic resonance elastography for the pancreas at 3T. J Magn Reson Imaging. 2016;**43**(2):384–90.
20. Miyaji K, Furuse A, Nakajima J, Kohno T, Ohtsuka T, Yagyu K, et al. The stiffness of lymph nodes containing lung carcinoma metastases: a new diagnostic parameter measured by a tactile sensor. Cancer. 1997;**80**(10):1920–5.
21. Deville SS, Cordes N. The extracellular, cellular, and nuclear stiffness, a trinity in the cancer resistome—a review. Front Oncol. 2019;**9**:1376.
22. Levental KR, Yu H, Kass L, Lakins JN, Egeblad M, Erler JT, et al. Matrix crosslinking forces tumor progression by enhancing integrin signaling. Cell. 2009;**139**(5):891–906.
23. Acerbi I, Cassereau L, Dean I, Shi Q, Au A, Park C, et al. Human breast cancer invasion and aggression correlates with ECM stiffening and immune cell infiltration. Integr Biol. 2015;**7**(10):1120–34.
24. Luthold C, Hallal T, Labbé DP, Bordeleau F. The extracellular matrix stiffening: a trigger of prostate cancer progression and castration resistance? Cancers. 2022;**14**(12):2887.
25. Lopez JI, Kang I, You W-K, McDonald DM, Weaver VM. *In situ* force mapping of mammary gland transformation. Integr Biol. 2011;**3**(9):910–21.
26. Lampi MC, Reinhart-King CA. Targeting extracellular matrix stiffness to attenuate disease: From molecular mechanisms to clinical trials. Sci Transl Med. 2018;**10**(422):eaao0475.
27. Wang JP, Hielscher A. Fibronectin: how its aberrant expression in tumors may improve therapeutic targeting. J Cancer. 2017;**8**(4):674–82.
28. Lo C-M, Wang H-B, Dembo M, Wang Y-L. Cell movement is guided by the rigidity of the substrate. Biophys J. 2000;**79**(1):144–52.
29. Provenzano PP, Inman DR, Eliceiri KW, Knittel JG, Yan L, Rueden CT, et al. Collagen density promotes mammary tumor initiation and progression. BMC Med. 2008;**6**(1):11.

30. Yu H, Mouw JK, Weaver VM. Forcing form and function: biomechanical regulation of tumor evolution. Trends Cell Biol. 2011;**21**(1):47–56.
31. Wang Y-L, Pelham RJ, Jr. [39] Preparation of a flexible, porous polyacrylamide substrate for mechanical studies of cultured cells. Methods Enzymol. 1998;**298**: 489–96. Elsevier.
32. Wang H-B, Dembo M, Wang Y-L. Substrate flexibility regulates growth and apoptosis of normal but not transformed cells. Am J Physiol–Cell Physiol. 2000;**279**(5):C1345–C50.
33. Discher DE, Janmey P, Wang Y-l. Tissue cells feel and respond to the stiffness of their substrate. Science. 2005;**310**(5751):1139–43.
34. Pelham RJ, Jr, Wang Y. Cell locomotion and focal adhesions are regulated by substrate flexibility. Proc Natl Acad Sci U S A. 1997;**94**(25):13661–5.
35. Chin L, Xia Y, Discher DE, Janmey PA. Mechanotransduction in cancer. Curr Opin Chem Eng. 2016;**11**:77–84.
36. Kraning-Rush CM, Reinhart-King CA. Controlling matrix stiffness and topography for the study of tumor cell migration. Cell Adh Migr. 2012;**6**(3):274–9.
37. Sieminski AL, Hebbel RP, Gooch KJ. The relative magnitudes of endothelial force generation and matrix stiffness modulate capillary morphogenesis in vitro. Exp Cell Res. 2004;**297**(2):574–84.
38. Flanagan LA, Ju YE, Marg B, Osterfield M, Janmey PA. Neurite branching on deformable substrates. Neuroreport. 2002;**13**(18):2411–5.
39. Bajaj P, Tang X, Saif TA, Bashir R. Stiffness of the substrate influences the phenotype of embryonic chicken cardiac myocytes. J Biomed Mater Res Part A. 2010;**95A**(4):1261–9.
40. Engler AJ, Sen S, Sweeney HL, Discher DE. Matrix elasticity directs stem cell lineage specification. Cell. 2006;**126**(4):677–89.
41. Tilghman RW, Cowan CR, Mih JD, Koryakina Y, Gioeli D, Slack-Davis JK, et al. Matrix rigidity regulates cancer cell growth and cellular phenotype. PLoS ONE. 2010;**5**(9):e12905.
42. Wang M, Chai N, Sha B, Guo M, Zhuang J, Xu F, et al. The effect of substrate stiffness on cancer cell volume homeostasis. J Cellular Physiol. 2018;**233**(2):1414–23.
43. Miyazawa A, Ito S, Asano S, Tanaka I, Sato M, Kondo M, et al. Regulation of PD-L1 expression by matrix stiffness in lung cancer cells. Biochem Biophys Res Commun. 2018;**495**(3):2344–9.
44. Larin KV, Sampson DD. Optical coherence elastography—OCT at work in tissue biomechanics [Invited]. Biomed Opt Express. 2017;**8**(2):1172–202.
45. Georges PC, Janmey PA. Cell type-specific response to growth on soft materials. J Appl Physiol. 2005;**98**(4):1547–53.
46. Trappmann B, Gautrot JE, Connelly JT, Strange DG, Li Y, Oyen ML, et al. Extracellular-matrix tethering regulates stem-cell fate. Nat Mater. 2012;**11**(7):642–9.
47. Lovett DB, Shekhar N, Nickerson JA, Roux KJ, Lele TP. Modulation of nuclear shape by substrate rigidity. Cell Mol Bioeng. 2013;**6**(2):230–8.
48. Wen JH, Vincent LG, Fuhrmann A, Choi YS, Hribar KC, Taylor-Weiner H, et al. Interplay of matrix stiffness and protein tethering in stem cell differentiation. Nat Mater. 2014;**13**(10):979–87.
49. Yeung T, Georges PC, Flanagan LA, Marg B, Ortiz M, Funaki M, et al. Effects of substrate stiffness on cell morphology, cytoskeletal structure, and adhesion. Cell Motil Cytoskeleton. 2005;**60**(1):24–34.
50. Gu Y, Ji Y, Zhao Y, Liu Y, Ding F, Gu X, et al. The influence of substrate stiffness on the behavior and functions of Schwann cells in culture. Biomaterials. 2012;**33**(28):6672–81.
51. Peyton SR, Putnam AJ. Extracellular matrix rigidity governs smooth muscle cell motility in a biphasic fashion. J Cell Physiol. 2005;**204**(1):198–209.
52. Califano JP, Reinhart-King CA. Substrate stiffness and cell area predict cellular traction stresses in single cells and cells in contact. Cell Mol Bioeng. 2010;**3**(1):68–75.
53. Guo W-h, Frey MT, Burnham NA, Wang Y-l. Substrate rigidity regulates the formation and maintenance of tissues. Biophys J. 2006;**90**(6):2213–20.
54. Engler AJ, Richert L, Wong JY, Picart C, Discher DE. Surface probe measurements of the elasticity of sectioned tissue, thin gels and polyelectrolyte multilayer films: correlations between substrate stiffness and cell adhesion. Surf Sci. 2004;**570**(1–2):142–54.
55. Pathak A, Kumar S. Independent regulation of tumor cell migration by matrix stiffness and confinement. Proc Natl Acad Sci. 2012;**109**(26):10334–9.
56. Moshayedi P, da F Costa L, Christ A, Lacour SP, Fawcett J, Guck J, et al. Mechanosensitivity of astrocytes on optimized polyacrylamide gels analyzed by quantitative morphometry. J Phys: Condens Matter. 2010;**22**(19):194114.
57. Ulrich TA, De Juan Pardo EM, Kumar S. The mechanical rigidity of the extracellular matrix regulates the structure, motility, and proliferation of glioma cells. Cancer Res. 2009;**69**(10):4167–74.
58. Wang TC, Abolghasemzade S, McKee BP, Singh I, Pendyala K, Mohajeri M, et al. Matrix stiffness drives drop like nuclear deformation and lamin A/C tension-dependent YAP nuclear localization. Nat Commun. Forthcoming 2024.
59. Mitchison T, Kirschner M. Cytoskeletal dynamics and nerve growth. Neuron. 1988;**1**(9):761–72.
60. Chan CE, Odde DJ. Traction dynamics of filopodia on compliant substrates. Science. 2008;**322**(5908):1687–91.
61. Elosegui-Artola A, Oria R, Chen Y, Kosmalska A, Perez-Gonzalez C, Castro N, et al. Mechanical regulation of a molecular clutch defines force transmission and transduction in response to matrix rigidity. Nat Cell Biol. 2016;**18**(5):540–8.
62. Harris AK, Wild P, Stopak D. Silicone rubber substrata: A new wrinkle in the study of cell locomotion. Science. 1980;**208**(4440):177–9.
63. Lee J, Leonard M, Oliver T, Ishihara A, Jacobson K. Traction forces generated by locomoting keratocytes. J Cell Biol. 1994;**127**(6):1957–64.
64. Munevar S, Wang Y, Dembo M. Traction force microscopy of migrating normal and H-ras transformed 3T3 fibroblasts. Biophys J. 2001;**80**(4):1744–57.
65. Fu J, Wang Y-K, Yang MT, Desai RA, Yu X, Liu Z, et al. Mechanical regulation of cell function with geometrically modulated elastomeric substrates. Nat Methods. 2010;**7**(9):733–6.
66. Suyatin DB, Hällströ M W, Samuelson L, Montelius L, Prinz CN, Kanje M. Gallium phosphide nanowire arrays and their possible application in cellular force investigations. J Vac Sci Technol B: Microelectron Nanometer Struct. 2009;**27**(6):3092.
67. Tan JL, Tien J, Pirone DM, Gray DS, Bhadriraju K, Chen CS. Cells lying on a bed of microneedles: an approach to isolate mechanical force. Proc Natl Acad Sci U S A. 2003;**100**(4):1484–9.
68. Saez A, Buguin A, Silberzan P, Ladoux B. Is the mechanical activity of epithelial cells controlled by deformations or forces? Biophys J. 2005;**89**(6):L52–4.
69. Ghagre A, Amini A, Srivastava LK, Tirgar P, Khavari A, Koushki N, et al. Pattern-based contractility screening, a reference-free alternative to traction force microscopy methodology. ACS Appl Mater Interf. 2021;**13**(17):19726–35.

70. Li Z, Persson H, Adolfsson K, Abariute L, Borgström MT, Hessman D, et al. Cellular traction forces: a useful parameter in cancer research. Nanoscale. 2017;**9**(48):19039–44.
71. Kraning-Rush CM, Califano JP, Reinhart-King CA. Cellular Traction Stresses Increase with Increasing Metastatic Potential. PLoS ONE. 2012;**7**(2):e32572.
72. Isomursu A, Park K-Y, Hou J, Cheng B, Mathieu M, Shamsan GA, et al. Directed cell migration towards softer environments. Nat Mater. 2022;**21**(9):1081–90.
73. Lin H-H, Lin H-K, Lin IH, Chiou Y-W, Chen H-W, Liu C-Y, et al. Mechanical phenotype of cancer cells: cell softening and loss of stiffness sensing. Oncotarget. 2015;**6**(25):20946–58.
74. Munevar S, Wang Y-L, Dembo M. Traction force microscopy of migrating normal and H-ras transformed 3T3 fibroblasts. Biophys J. 2001;**80**(4):1744–57.
75. Indra I, Undyala V, Kandow C, Thirumurthi U, Dembo M, Beningo KA. An in vitro correlation of mechanical forces and metastatic capacity. Phys Biol. 2011;**8**(1):015015.
76. Mih JD, Sharif AS, Liu F, Marinkovic A, Symer MM, Tschumperlin DJ. A multiwell platform for studying stiffness-dependent cell biology. PLoS ONE. 2011;**6**(5):e19929.
77. Purkayastha P, Pendyala K, Saxena AS, Hakimjavadi H, Chamala S, Dixit P, et al. Reverse plasticity underlies rapid evolution by clonal selection within populations of fibroblasts propagated on a novel soft substrate. Mol Biol Evol. 2021;**38**(8):3279–93.
78. Chen CS, Mrksich M, Huang S, Whitesides GM, Ingber DE. Geometric control of cell life and death. Science. 1997;**276**(5317):1425–8.
79. Sun M, Chi G, Li P, Lv S, Xu J, Xu Z, et al. Effects of matrix stiffness on the morphology, adhesion, proliferation and osteogenic differentiation of mesenchymal stem cells. Int J Med Sci. 2018;**15**(3):257–68.
80. Seghir R, Arscott S. Extended PDMS stiffness range for flexible systems. Sens Actuat A: Phys. 2015;**230**:33–9.
81. Byfield FJ, Reen RK, Shentu T-P, Levitan I, Gooch KJ. Endothelial actin and cell stiffness is modulated by substrate stiffness in 2D and 3D. J Biomech. 2009;**42**(8):1114–9.
82. Aw Yong KM, Sun Y, Merajver SD, Fu J. Mechanotransduction-induced reversible phenotypic switching in prostate cancer cells. Biophys J. 2017;**112**(6):1236–45.
83. Narkhede AA, Crenshaw JH, Manning RM, Rao SS. The influence of matrix stiffness on the behavior of brain metastatic breast cancer cells in a biomimetic hyaluronic acid hydrogel platform. J Biomed Mater Res Part A. 2018;**106**(7):1832–41.
84. Mubarok W, Elvitigala KCML, Nakahata M, Kojima M, Sakai S. Modulation of cell-cycle progression by hydrogen peroxide-mediated cross-linking and degradation of cell-adhesive hydrogels. Cells. 2022;**11**(5):881.
85. Huang G, Wang L, Wang S, Han Y, Wu J, Zhang Q, et al. Engineering three-dimensional cell mechanical microenvironment with hydrogels. Biofabrication. 2012;**4**(4):042001.
86. Paszek MJ, Zahir N, Johnson KR, Lakins JN, Rozenberg GI, Gefen A, et al. Tensional homeostasis and the malignant phenotype. Cancer Cell. 2005;**8**(3):241–54.
87. Wozniak MA, Desai R, Solski PA, Der CJ, Keely PJ. ROCK-generated contractility regulates breast epithelial cell differentiation in response to the physical properties of a three-dimensional collagen matrix. J Cell Biol. 2003;**163**(3):583–95.
88. Bonnesoeur S, Morin-Grognet S, Thoumire O, Le Cerf D, Boyer O, Vannier JP, et al. Hyaluronan-based hydrogels as versatile tumor-like models: tunable ECM and stiffness with genipin-crosslinking. J Biomed Mater Res A. 2020;**108**(5):1256–68.
89. Jaiswal MK, Xavier JR, Carrow JK, Desai P, Alge D, Gaharwar AK. Mechanically stiff nanocomposite hydrogels at ultralow nanoparticle content. ACS Nano. 2016;**10**(1):246–56.
90. Kersey AL, Cheng DY, Deo KA, Dubell CR, Wang TC, Jaiswal MK, et al. Stiffness assisted cell-matrix remodeling trigger 3D mechanotransduction regulatory programs. Biomaterials. 2024;**306**:122473.
91. Kraehenbuehl TP, Zammaretti P, Van der Vlies AJ, Schoenmakers RG, Lutolf MP, Jaconi ME, et al. Three-dimensional extracellular matrix-directed cardioprogenitor differentiation: systematic modulation of a synthetic cell-responsive PEG-hydrogel. Biomaterials. 2008;**29**(18):2757–66.
92. Wang C, Tong X, Yang F. Bioengineered 3D brain tumor model to elucidate the effects of matrix stiffness on glioblastoma cell behavior using PEG-based hydrogels. Mol Pharm. 2014;**11**(7):2115–25.
93. Fang JY, Tan S-J, Wu Y-C, Yang Z, Hoang BX, Han B. From competency to dormancy: a 3D model to study cancer cells and drug responsiveness. J Transl Med. 2016;**14**(1).
94. Zhang M, Xu C, Wang H-Z, Peng Y-N, Li H-O, Zhou Y-J, et al. Soft fibrin matrix downregulates DAB2IP to promote Nanog-dependent growth of colon tumor-repopulating cells. Cell Death Dis. 2019;**10**(3).
95. Vasudevan J, Lim CT, Fernandez JG. Cell Migration and breast cancer metastasis in biomimetic extracellular matrices with independently tunable stiffness. Adv Funct Mater. 2020;**30**(49):2005383.
96. Bomo J, Ezan F, Tiaho F, Bellamri M, Langouët S, Theret N, et al. Increasing 3D matrix rigidity strengthens proliferation and spheroid development of human liver cells in a constant growth factor environment. J. Cell Biochem. 2016;**117**(3):708–20.
97. Ondeck MG, Kumar A, Placone JK, Plunkett CM, Matte BF, Wong KC, et al. Dynamically stiffened matrix promotes malignant transformation of mammary epithelial cells via collective mechanical signaling. Proc Natl Acad Sci. 2019;**116**(9):3502–7.
98. Stowers RS, Allen SC, Sanchez K, Davis CL, Ebelt ND, Van Den Berg C, et al. Extracellular Matrix Stiffening Induces a Malignant Phenotypic Transition in Breast Epithelial Cells. Cell Mol Bioeng. 2017;**10**(1):114–23.
99. Homma K, Chang AC, Yamamoto S, Tamate R, Ueki T, Nakanishi J. Design of azobenzene-bearing hydrogel with photoswitchable mechanics driven by photo-induced phase transition for in vitro disease modeling. Acta Biomater. 2021;**132**:103–13.
100. Pathak A, Kumar S. Biophysical regulation of tumor cell invasion: moving beyond matrix stiffness. Integr Biol. 2011;**3**(4):267–78.
101. González-Díaz E, Varghese S. Hydrogels as extracellular matrix analogs. Gels. 2016;**2**(3):20.
102. Chaudhuri O, Cooper-White J, Janmey PA, Mooney DJ, Shenoy VB. Effects of extracellular matrix viscoelasticity on cellular behaviour. Nature. 2020;**584**(7822):535–46.
103. Nam S, Gupta VK, Lee H-p, Lee JY, Wisdom KM, Varma S, et al. Cell cycle progression in confining microenvironments is regulated by a growth-responsive TRPV4-PI3K/Akt-p27^{Kip1} signaling axis. Sci Adv. 2019;**5**(8):eaaw6171.
104. Chaudhuri O, Gu L, Darnell M, Klumpers D, Bencherif SA, Weaver JC, et al. Substrate stress relaxation regulates cell spreading. Nat Commun. 2015;**6**(1):6365.

105. Mandal K, Gong Z, Rylander A, Shenoy VB, Janmey PA. Opposite responses of normal hepatocytes and hepatocellular carcinoma cells to substrate viscoelasticity. Biomater Sci. 2020;**8**(5):1316–28.
106. McBeath R, Pirone DM, Nelson CM, Bhadriraju K, Chen CS. Cell shape, cytoskeletal tension, and RhoA regulate stem cell lineage commitment. Dev Cell. 2004;**6**(4):483–95.
107. Geiger B, Bershadsky A, Pankov R, Yamada KM. Transmembrane crosstalk between the extracellular matrix and the cytoskeleton. Nat Rev Mol Cell Biol. 2001;**2**(11):793–805.
108. Klein EA, Yin L, Kothapalli D, Castagnino P, Byfield FJ, Xu T, et al. Cell-cycle control by physiological matrix elasticity and in vivo tissue stiffening. Curr Biol. 2009;**19**(18):1511–8.
109. Théry M, Racine V, Piel M, Pépin A, Dimitrov A, Chen Y, et al. Anisotropy of cell adhesive microenvironment governs cell internal organization and orientation of polarity. Proc Natl Acad Sci. 2006;**103**(52):19771–6.
110. Brock A, Chang E, Ho C-C, LeDuc P, Jiang X, Whitesides GM, et al. Geometric determinants of directional cell motility revealed using microcontact printing. Langmuir. 2003;**19**(5):1611–7.
111. Lele TP, Dickinson RB, Gundersen GG. Mechanical principles of nuclear shaping and positioning. J Cell Biol. 2018;**217**(10):3330–42.
112. Dickinson RB, Katiyar A, Dubell CR, Lele TP. Viscous shaping of the compliant cell nucleus. APL Bioeng. 2022;**6**(1):010901.
113. Liu W, Padhi A, Zhang X, Narendran J, Anastasio MA, Nain AS, et al. Dynamic heterochromatin states in anisotropic nuclei of cells on aligned nanofibers. ACS Nano. 2022;**16**(7):10754–67.
114. Vergani L, Grattarola M, Nicolini C. Modifications of chromatin structure and gene expression following induced alterations of cellular shape. Int J Biochem Cell Biol. 2004;**36**(8):1447–61.
115. Dahl KN, Ribeiro AJS, Lammerding J. Nuclear Shape, Mechanics, and Mechanotransduction. Circ Res. 2008;**102**(11):1307–18.
116. Nava MM, Miroshnikova YA, Biggs LC, Whitefield DB, Metge F, Boucas J, et al. Heterochromatin-driven nuclear softening protects the genome against mechanical stress-induced damage. Cell. 2020;**181**(4):800–17.e22.
117. Dupont S, Morsut L, Aragona M, Enzo E, Giulitti S, Cordenonsi M, et al. Role of YAP/TAZ in mechanotransduction. Nature. 2011;**474**(7350):179–83.
118. Elosegui-Artola A, Andreu I, Beedle AEM, Lezamiz A, Uroz M, Kosmalska AJ, et al. Force triggers YAP nuclear entry by regulating transport across nuclear pores. Cell. 2017;**171**(6):1397–410 e14.
119. Cosgrove BD, Loebel C, Driscoll TP, Tsinman TK, Dai EN, Heo SJ, et al. Nuclear envelope wrinkling predicts mesenchymal progenitor cell mechano-response in 2D and 3D microenvironments. Biomaterials. 2021;**270**:120662.
120. Garcia-Garcia M, Sanchez-Perales S, Jarabo P, Calvo E, Huyton T, Fu L, et al. Mechanical control of nuclear import by Importin-7 is regulated by its dominant cargo YAP. Nat Commun. 2022;**13**(1):1174.
121. Skory RM, Moverley AA, Ardestani G, Alvarez Y, Domingo-Muelas A, Pomp O, et al. The nuclear lamina couples mechanical forces to cell fate in the preimplantation embryo via actin organization. Nat Commun. 2023;**14**(1):3101.
122. Singh I, Lele TP. Nuclear morphological abnormalities in cancer: a search for unifying mechanisms. Results Probl Cell Differ. 2022;**70**:443–67.
123. Medina SH, Bush B, Cam M, Sevcik E, Delrio FW, Nandy K, et al. Identification of a mechanogenetic link between substrate stiffness and chemotherapeutic response in breast cancer. Biomaterials. 2019;**202**:1–11.
124. Huang S, Ingber DE. Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. Exp Cell Res. 2000;**261**(1):91–103.
125. Huang S, Ernberg I, Kauffman S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. Semin Cell Dev Biol. 2009;**20**(7):869–76.
126. Obuch JC, Ahnen DJ. Colorectal cancer: genetics is changing everything. Gastroenterol Clin North Am. 2016;**45**(3):459–76.
127. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. Nature. 2013;**501**(7467):338–45.
128. Colom B, Herms A, Hall MWJ, Dentro SC, King C, Sood RK, et al. Mutant clones in normal epithelium outcompete and eliminate emerging tumours. Nature. 2021;**598**(7881):510–4.
129. Gutierrez C, Al'Khafaji AM, Brenner E, Johnson KE, Gohil SH, Lin Z, et al. Multifunctional barcoding with ClonMapper enables high-resolution study of clonal dynamics during tumor evolution and treatment. Nat Cancer. 2021;**2**(7):758–72.
130. Purkayastha P, Pendyala K, Saxena AS, Hakimjavadi H, Chamala S, Dixit P, et al. Reverse plasticity underlies rapid evolution by clonal selection within populations of fibroblasts propagated on a novel soft substrate. Mol Biol Evol. 2021;**38**(8):3279–93.
131. Wang TC, Sawhney S, Morgan D, Bennett RL, Rashmi R, Estecio MR, et al. Genetic variation drives cancer cell adaptation to ECM stiffness. Proc Natl Acad Sci U S A. 2024;**121**(39):e2403062121.
132. Atanasova KR, Chakraborty S, Ratnayake R, Khare KD, Luesch H, Lele TP. An epigenetic small molecule screen to target abnormal nuclear morphology in human cells. Mol Biol Cell. 2022;**33**(6).
133. Shen Y, Wang X, Lu J, Salfenmoser M, Wirsik NM, Schleussner N, et al. Reduction of liver metastasis stiffness improves response to bevacizumab in metastatic colorectal cancer. Cancer Cell. 2020;**37**(6):800–17.e7.

26

Decoding mechano-oncology principles through microfluidic devices and biomaterial platforms

Alka Kumari, Abhishek Goswami, and Ajay Tijore

26.1. Introduction

Cancer is one of the leading causes of death across the world [1]. In particular, cancer metastasis (tumour spread from primary tumour sites to other body parts) is the major reason behind cancer-associated mortality. Historically, oncogenes and biochemical cues have been considered the drivers of cancer due to their significant role in tumour development [2,3]. However, recent studies have shown that the mechanical properties of tumour cells and tumour microenvironment (TME) substantially contribute to tumour development and progression [4–6]. Interestingly, it is now possible to measure the mechanical properties of tumour cells, including stiffness and contractility, *in vitro* and *in vivo* due to the availability of techniques, such as traction force microscopy, atomic force microscopy, optical tweezers, micropillar platforms, elastic modulus Pen, and real-time elastography [7–9]. These techniques thus enable the diagnosis of tumours with aberrant mechanical stiffness and the spatial distribution of tumour markers using palpation and imaging [7,10].

Traditionally, conventional tumour models have been used to study tumour development and progression. For instance, *in vitro* tumour models, such as trans-well assays, wound healing assays, and chemotaxis assays, were used [11]. However, these tumour models cannot recapitulate the complexities of TME, including different types of mechanical forces generated within TME (stretch, compression, and fluid shear stress) and extra-cellular matrix (ECM) stiffness (Figure 26. 1A) [12,13]. On the other hand, animal-based tumour models provide an excellent platform to mimic *in vivo* complexities of TME. But these models have their disadvantages, e.g. the requirement of immunodeficient animals, long tumour latency, difficulty in monitoring individual steps of metastasis, expensive, and time-consuming [14]. Recent advancements in the field of biomaterials, tissue engineering, and microfabrication, as well as an increase in the basic understanding of cancer biology, have contributed to the development of complex *in vitro* tumour models that recapitulate TME by incorporating different cell types, vasculature, and ECM composition along with spatiotemporal control on biochemical growth factor addition [15]. This has led to the development of a wide range of models ranging from tumour spheroids, tumour organoids, tumour vasculature models and complex tumour-on-chip models mimicking *in vivo* TME, and normal physiological conditions. These 2D and 3D *in vitro* tumour models have enabled the real-time monitoring of critical steps involved in cancer cell progression, including primary tumour formation, invasion, intravasation, extravasation, and secondary tumour formation at the distant organ (**Figure 26.1**B) [16]. Tumours are known for their heterogeneity, and this feature makes it difficult to find common treatments against different tumour types. Thus, precision medicine has increasingly become popular in which patient-specific *in vitro* tumour models have been developed to assess clinical treatment efficacy and management [17].

In this chapter, we broadly categorize contemporary *in vitro* tumour models based on their applications to study the major steps of tumour progression. Also, we summarize the mechano-oncology principles of these tumour models, which will serve as the guideline for selecting tumour models to study the involvement of various mechanobiology principles in tumour progression and develop an effective cancer treatment.

26.2. Decoding tumour cell progression and survival using microfluidic platforms

Metastasis of tumour cells involves five major steps: invasion, intravasation, circulating tumour cell formation, extravasation, and colonization at the distant organ [18]. It has been well established that the ECM and TME play a critical role in the tumour progression [16]. Here, we describe *in vitro* microfluidic and biomaterials platforms that mimic *in vivo* TME architecture and different types of mechanical forces experienced by tumour cells and are used to study tumour metastasis steps and tumour cell survival during metastasis (**Table 26.1**).

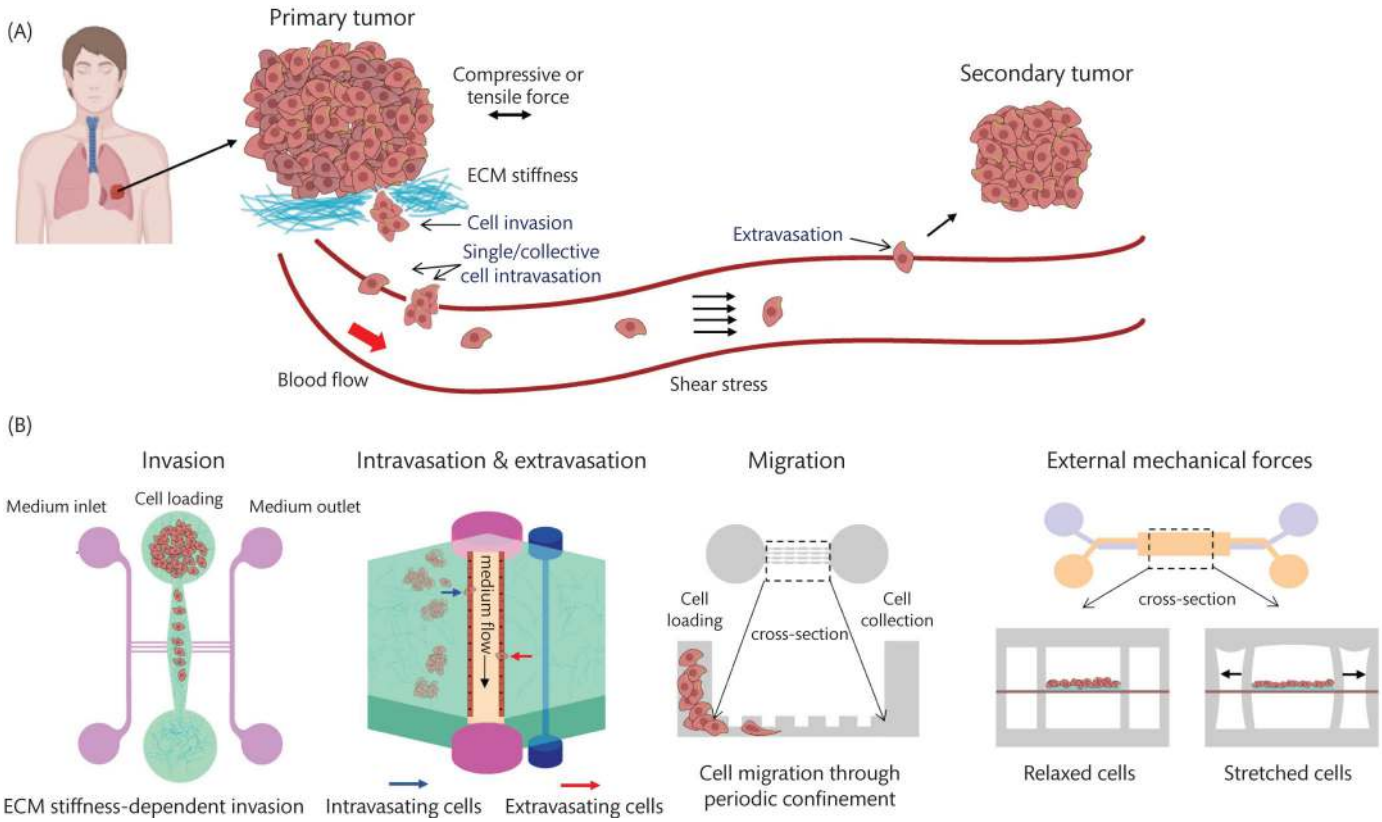


Figure 26.1. Schematic showing (A) steps involved in tumour cell metastasis and different types of mechanical forces generated by the TME contributing to the metastasis and (B) microfluidic/biomaterial platforms used to systematically study the steps involved in metastasis by mimicking TME.

Table 26.1. The list of microfluidic devices/biomaterials to study tumour progression.

Cancer type	Cell line	Microfluidic/biomaterial platforms and their application	Study outcome(s)	Reference
Breast cancer	MDA-MB-231	A PDMS-based microchannel platform to study the effect of fluid shear stress on cell intravasation.	Mechanosensitive receptor TRPM7 is activated by fluid shear stress and reverses the direction of cell migration.	[19]
Breast cancer	MDA-MB-231, MCF7	The concentric three-layered hydrogel-based microfluidic device to explore cell invasion, intravasation, and angiogenesis.	Several signalling cytokines involved in invasion, intravasation, and angiogenesis were identified.	[20]
Breast cancer	MDA-MB-231	PDMS-based microchannel platforms with varying channel widths to study cell invasion.	Narrow channels promoted mesenchymal to ameboid transition.	[21]
Breast cancer	MDA-MB-231	Collagen–alginate hydrogel-based microchannel platform to study confined cancer cell migration.	Confinement and stiffness jointly promoted mesenchymal to ameboid transition	[22]
Breast cancer	MX-1, MCF7	PDMS-based 3D microfluidic device used to study cancer cell migration and invasion.	Real-time observation of cancer cell migration was carried out to monitor metastasis in the presence of drugs.	[23]
Breast cancer	MDA-MB-231	PDMS-based microchannel platform to illustrate the physiological significance of the Osmotic Engine Model in confined cell migration.	Polarization of NHE1 and SWELL1 at the leading and trailing edge respectively controlled cell volume and confined migration.	[24]
Oral cancer	HN12, HN13, HN30, CAL27	Chitosan-based microfluidic arrays used for the quantitative analysis of metastatic biomarker, desmoglein 3 in head and neck cancer.	Development of a microfluidic-based single-cell analysis technique for cancer biomarker detection.	[25]
Colorectal cancer	SW620	Single-channel microfluidic-based chip to measure the pharmacokinetic profile of drugs.	Development of a tumour-on-chip model to provide an alternative to the animal model.	[26]
Colorectal cancer	Cancer stem cell (CSC)	Alginate hydrogel with droplet microfluidic device to capture and propagate cancer stem cells.	Isolation of cancer stem cells to develop personalized CSC-targeted therapy.	[27]

Table 26.1. Continued

Cancer type	Cell line	Microfluidic/biomaterial platforms and their application	Study outcome(s)	Reference
Colon cancer	HT29- MTX-E12	3D tumour spheroid model combined with a PDMS-based microfluidic chip to study the uptake of nanoparticles.	Development of a novel platform to test the efficacy of nanoparticle-based drug delivery.	[28]
Glioblastoma	U87MG, HUVECs	Silver nanowire and collagen I-based electro-responsive hydrogel for photothermal therapy application.	Development of hydrogel system for nanomaterial delivery and photothermal therapy application.	[29]
Glioblastoma	G55	PDMS-based microchannel platform to test the response of chemotherapeutic drugs during the invasion–metastasis process.	Confined cells showed drug resistance and cancer stem cell-like properties.	[30]
Prostate cancer	LNCaP	PDMS-based cluster wells to detect circulating tumour cells (CTCs).	Isolation of CTC cluster from unprocessed whole blood.	[31]
Melanoma	Murine B16, Human MP-1	PDMS-based microfluidic system to isolate soft and stiff populations of cancer cells from heterogeneous cell population.	Soft cancer cells were isolated and found that softness is an intrinsic characteristic of cancer stem cells.	[32]
Lung cancer	Patient-derived organoid	PDMS-based one-stop microfluidic devices to mimic *in vivo* conditions for drug testing.	Developed 3D organoids to check the efficacy of the chemotherapeutic drug.	[33]
Lung cancer	SCLC	PDMS-based multi-flow microfluidic system to detect CTCs.	Successful isolation of patient-derived CTCs.	[34]
Lung cancer-liver metastasis	A549 and HFL-1	PDMS-based 3D multiorgan microfluidic device to induce hypoxia condition *in vitro*.	The device provides an alternative to animal models and used as drug screening platform under hypoxic condition.	[35]
Pancreatic cancer	Blood from cancer patient	Combination of magnetic field and microfluidic device to isolate cancer generated exosomes with magnetic nanoparticle-tagged exosome antibodies.	Successful isolation of exosomes from cancer patient's blood.	[36]

26.2.1. Tumour cell invasion

Cancer cells secret several enzymes, such as matrix metalloproteinases and lysyl oxidases, which remodel the ECM, create 3D microchannel tracks, and increase matrix stiffness that promotes the invasion [37]. Traditionally, 2D chemotaxis assays, trans-well assays, and 3D invasion wound healing assays have been used to study the invasion. However, these methods do not mimic complex 3D TME that accommodates the effect of different types of mechanical forces, fluid shear stress, and the vasculature [13,38].

In recent years, micro-engineered physiological systems, including 3D microfluidic devices, have emerged rapidly to study mechanical aspects of tumour invasion along with associated TME parameters having both spatial and temporal control [39,40]. Cell morphological phenotype is a simple but promising biomarker for tumour diagnosis. The cell size and shape deformability can be a label-free marker that can give information about the mechanisms of invasion. Still, the primary constraint in developing the device using cell phenotypic characters is the lack of standardization and throughput. Recently, researchers have developed a single-cell microfluidic system based on the quantitative deformability cytometry principle and machine-learning algorithm by which the analysis of tumour cell invasion was performed [41]. Next, the degree of cellular deformability or plasticity varies between less invasive to highly invasive tumour cells. However, the measurement of the plastic deformation ability of tumour cells is limited due to the lack of quantitative methods. To address this problem, Yan Z. et al. developed a microfluidic device that can precisely impose cyclic deformation on the cells, resulting in the significant accumulation of more deformable and highly invasive late-stage lung cancer cells. At the same time, less plasticity was observed in the early-stage, less invasive cell lines [42]. Transformation in tumour cells disrupts cells' mechanical and electrochemical properties. Recently, a microfluidic chamber-based 'dielectrophoretic stretch assay' has been developed to investigate the correlation between dielectric properties and the mechanical response of the cells. A significantly different stretch pattern was observed in cancer and normal cells. This platform provides a promising tool to detect micro-invasion in the tissue sample without the need to label them [43].

Highly invasive tumour cells have been seen to secrete large amounts of protons during glycolysis or acidosis pathways to their surrounding ECM which further activate the NF-kB pathway in neighbouring stromal cells. However, the correlation between metabolic pathways and metastasis has not been well understood. Recently, researchers using the 3D microfluidic system showed that acidosis activated IL6 and MSC, further promoting osteosarcoma invasiveness [44]. Neutrophils have both tumour-promoting and tumour-limiting properties. 3D tumour-immune microenvironment-on-chip device was recently developed to better understand the intercellular dynamic and different modalities between neutrophils and tumour cells (**Figure 26.2**A). In brief, both the chemotaxis and formation of neutrophil extra-cellular traps (NETs) induced the response of neutrophils towards the tumour spheroid. It was further reported that the location-dependent mechanism of NETosis was responsible for the induction of collective invasion of ovarian tumour cells (**Figure 26.2**A) [45]. Cancer-associated fibroblast (CAF) functions

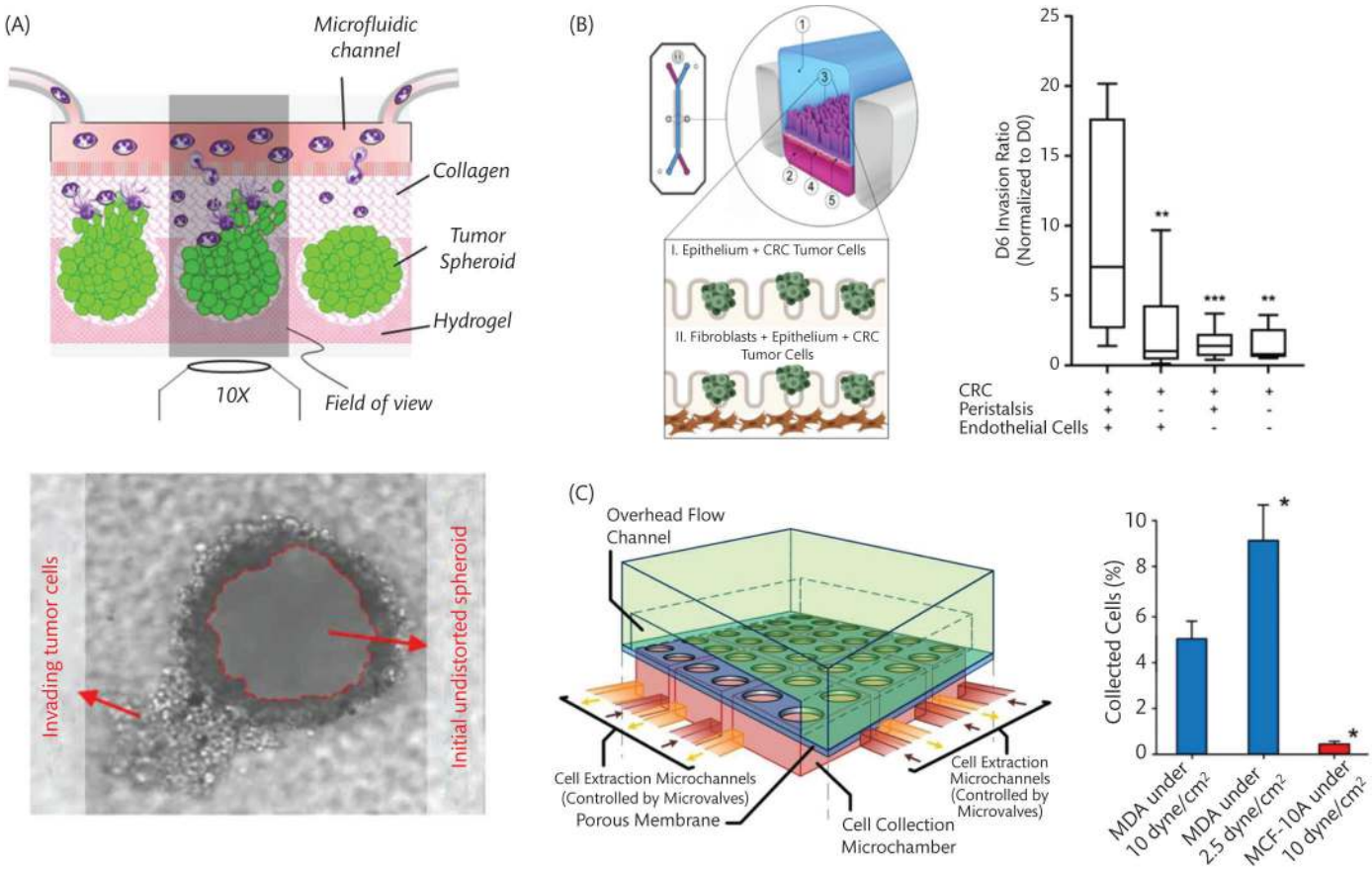


Figure 26.2. (A) Image showing *in vitro* TIME chip used to check neutrophil invasion in tumour spheroid embedded within the hydrogel and the collagen matrix. Bright-field image shows the distorted invasion domain of collective cell migration induced by NETosis. Red arrows show the original undistorted periphery of the spheroid before the invasion. (B) Schematic shows a CRC-on-chip platform for studying intravasation of CRC tumour cells. The bar diagram shows the rate of trans-endothelial migration of the tumour cells with and without peristalsis and quantified by determining the invasion ratio (number of tumour cells in endothelial channel per number of tumour cells in epithelial channel). (C) Schematic of microfluidic platform for investigating trans-endothelial migration of tumour cells, having an endothelial monolayer on a porous membrane and a cell collection chamber underlying the porous membrane. The bar diagram shows percentage of breast cancer cells (MDA-MB-231) and normal breast epithelial cells (MCF-10A) collected after subjecting cells to varying shear stress in the overhead flow channel. *Sources:* Part (A): Adapted with permission from [45], IOP Publishing, Ltd. Part (B): Adapted with permission from [59], Elsevier. Part (C): Adapted with permission from [61], AIP Publishing.

as a tumour ECM remodelling machine and facilitates tumour cell invasion. Cancer-cell-derived exosomes have been reported to promote endothelial cell differentiation into CAF. To explore this differentiation mechanism, a 3D microfluidic model was developed for real-time monitoring of CAF development in the presence of tumour-cell-derived exosomes at the tumour invasion site [46]. Several ECM proteins present in TME can alter tumour cell migration. However, developing the heterogeneous ECM proteins composition *in vitro* is challenging. A research group recently developed a microfluidic device in which ECM heterogeneity can be regulated and reported that fibronectin-rich ECM induced highly invasive breast tumour cell migration and observed the presence of cell micro-track within the ECM [47]. E-cadherin is a cytoskeleton protein that helps in cell–cell adhesion and plays a significant role in tumour invasion. The expression level of E-cadherin determines the invasion mode of tumour cells. A microfluidic approach (micro-fibrous polycaprolactone mesh in the chip with stable chemotactic gradient) was used to study the invasion mode and found that highly invasive tumour cells showed single-cell migration. In contrast, less invasive tumour cells showed collective cell migration by maintaining their E-cadherin expression level [48]. Bacteria, too, are a component of TME and exert a significant effect on tumour progression. To study the impact of extra-tumoural bacteria, researchers developed a microfluidic device that provides access to evaluate the effect of extra-tumoural bacteria on cancer progression. It was observed that biofilms formed by bacteria increased tumour progression, viability, and cancer stem cell population [49].

26.2.2. Tumour cell intravasation and extravasation

Due to the enhanced invasive capability, the metastatic tumour cells can infiltrate through the walls of the blood vessels or the lymphatic vessels. This process is known as intravasation. It is a crucial step in the metastatic cascade as it enables the tumour cells to enter the circulatory system and disseminate to (often distant) secondary tissue sites in the body. The vessel diameter has been identified as one of the critical parameters affecting the tumour cell intravasation [50]. Once the cancer cells enter circulation, they are termed circulating tumour cells (CTCs) [51,52]. The CTCs in the blood vessels are subjected to mechanical stresses and haemodynamic shear forces. The tumour cells in the bloodstream are also susceptible to anti-tumour cell responses of the immune system [53]. It has been found that zones of reduced fluid flow, which results in

and their survival in the bloodstream [51]. The surviving tumour cells eventually exit the circulatory system by migrating across the endothelium to reach the underlying secondary tissue site. This process is known as extravasation, resulting in tumour growth in a secondary tissue site in the body [53]. Studies have reported that physical and mechanical cues, such as diameter and architecture of the vessel/capillary and shear stress experienced by the CTCs due to blood flow, are important determinants of the location of extravasation of the tumour cells in the body [54,55].

In vitro models, ranging from comparatively simple Boyden chambers to complex 3D microfluidic devices, have been deployed to study metastasis mechanisms, including tumour cell intravasation and extravasation [56,57]. However, a major drawback of simplified models such as the Boyden chamber is that it is difficult to analyse complex tumour-cell–endothelial-cell crosstalk in such systems. Several 3D microfluidic platforms have been used to investigate the role of physical and biomechanical factors in influencing the intravasation and extravasation of tumour cells [16]. Wu et al. fabricated a 3D vessel-on-a-chip microfluidic device with reconstituted blood vessels that could recapitulate the crosstalk between CTCs and blood vessels *in vitro* [58]. The morphology of blood vessels in this microfluidic model was fine-tuned to remodel blood vessels of various geometries, such as aneurysm, stenosis, and bifurcation. In another study, researchers fabricated a microfluidic device with a set of parallel collagen-coated microchannels of varying widths to migrate cells between two orthogonal media channels in which shear stress was produced by regulating the fluid flow [19]. In particular, it was found that the activation of the mechanosensitive channel, TRPM7, under fluid shear stresses allowed Ca^{+2} entry inside the cell, which resulted in the reversal of the direction of normal cell migration. In contrast, fibrosarcoma cells, due to low expression of TRPM7, displayed 10-fold lower sensitivity to shear stress during intravasation. In another study, Strelez et al. investigated the role of physical forces in the intravasation of colorectal cancer (CRC) using a CRC-on-chip platform [59]. The CRC-on-chip platform comprised two channels, an epithelial channel and an endothelial channel, with an intervening porous membrane (**Figure 26.2**B). The authors found that the tumour cell intravasation across channels was markedly enhanced in conditions mimicking peristalsis.

Follain et al. developed a microfluidic platform to decipher how blood flow influences the phenomena of CTC arrest, adhesion to the endothelium, and extravasation to the underlying tissue at the site of metastasis [55]. Decreased flow profiles with optimal CTC adhesion were found to facilitate the stable CTC attachment to the endothelium. Also, it was observed that the fluid shear stresses were important for remodelling the endothelium, which stimulates the extravasation of CTCs. Jeon et al. established a microfluidic platform having a monolayer of endothelial cells for extravasation studies, using which they noted cancer cell extravasation within 24 h of introducing cancer cells to the system and a significant rise in the permeability of endothelial monolayer in the presence of cancer cells [60]. In a different study, Jeon et al. configured a microfluidic platform to study the extravasation of metastatic breast cancer cells in a bone-mimetic microenvironment under fluid shear stresses [56]. A significant reduction in tumour cell extravasation and vascular permeability was observed in the presence of fluid shear stresses. However, an increase in the migration distance of the extravasating tumour cells in the surrounding ECM was noted. In another study, Cui et al. established a microfluidic platform for extravasation studies (**Figure 26.2**C) [61]. Using this set-up, the authors demonstrated that trans-endothelial migration of tumour cells was facilitated by lower shear stress (2.5 dyne/cm^2) and significantly impeded at higher shear stress (10 dyne/cm^2). Mei et al. recently designed a microfluidic device to investigate the osteocyte-mediated mechano-regulation of bone metastasis to the breast tissue [62]. The microfluidic set-up contained two adjacent microchannels, one having tumour cells in a lumen lined with endothelial cells and the other cultured with osteocyte-like MLO-Y4 cells, with an intervening hydrogel matrix. The oscillatory fluid-flow-mediated mechanical stimulation (1 Pa, 1 Hz) of osteocytes resulted in a significant decline in the percentage of extravasating cancer cells and the distance migrated by the extravasating cells.

26.2.3. Tumour cell survival

In recent years, a growing body of literature has suggested that external mechanical forces can influence tumour growth and survival. In particular, the latest studies found that physiologically relevant mechanical forces promote tumour cell apoptosis or inhibit tumour cell growth. For example, a microfluidic platform was used to develop fluid shear stress generated during exercise conditions [63]. These high-shear stresses caused the significant killing of metastatic and drug-resistant tumour cells. This study highlighted the benefits of exercise in generating high-shear stress that can prevent metastasis. In another study, the cone and plate viscometer-generated fluid shear force induced TRAIL death receptor-mediated tumour cell apoptosis [64]. Takao et al. found that dynamic compressive forces generated using a mechanical stress-loading device caused tumour cell necrosis [65]. Basu et al. showed that cytotoxic T-cell killing of tumour cells was potentiated using stiff hydrogel compared to a soft hydrogel platform [66].

In support of mechanical force-induced tumour cell killing (mechanoptosis), we recently observed selective tumour cell mechanoptosis in many tumour types after cyclic stretch application using a cell stretching device [67]. In addition, low-frequency ultrasound-mediated mechanical forces were used to promote selective mechanoptosis in tumour cells, tumour organoids, and CAM models without damaging normal cells [68–72]. Mechanistic molecular studies demonstrated the mechanism of mechanoptosis that depends on mechanosensitive Piezo1 channel activation that allows calcium entry inside cells upon mechanical activation and initiates mitochondria-mediated apoptosis. In general, these studies indicate that mechanical forces can be harnessed to develop a mechanical force-based tumour treatment [73, 74].

26.3. Conclusions and future directions

With the advent of microfabrication and an increase in the fundamental understanding of tumour biology, the development of microfluidic-based *in vitro* tumour models recapitulating TME has become possible. The microfluidic platforms significantly contributed to studying the effect of mechanical forces on tumour progression and deciphering the mechanobiological principles involved in it. While microfluidic platforms will continue to evolve to be better indicators of *in vivo* biophysical forces, we envision that combining the state-of-the-art microfluidic platforms with existing tumour models will develop more realistic *in vitro* tumour models to further probe the effect of mechanical forces at molecular levels. Due to tumour heterogeneity, nowadays, precision medicine has become popular with a focus on developing patient-specific *in vitro* tumour models for clinical assessment and treatment. Similarly, in the future, microfluidic platforms could help to develop patient-specific mechanical-force-

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;**71**(3):209–49.
2. Bister K. Discovery of oncogenes: The advent of molecular cancer research. Proc Natl Acad Sci. 2015;**112**(50):15259–60.
3. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. Science. 2013;**339**(6127):1546–58.
4. Chaudhuri PK, Low BC, Lim CT. Mechanobiology of tumour growth. Chem Rev. 2018;**118**(14):6499–515.
5. Montagner M, Dupont S. Mechanical forces as determinants of disseminated metastatic cell fate. Cells. 2020;**9**(1):250.
6. Riehl BD, Kim E, Bouzid T, Lim JY. The role of microenvironmental cues and mechanical loading milieus in breast cancer cell progression and metastasis. Front Bioeng Biotechnol. 2021;**8**.
7. Hu X, Liu Y, Qian L. Diagnostic potential of real-time elastography (RTE) and shear wave elastography (SWE) to differentiate benign and malignant thyroid nodules: A systematic review and meta-analysis. Medicine (Baltimore). 2017;**96**(43):e8282.
8. Iskratsch T, Wolfenson H, Sheetz MP. Appreciating force and shape — the rise of mechanotransduction in cell biology. Nat Rev Mol Cell Biol. 2014;**15**(12):825–33.
9. Li Z, Tofangchi A, Stavins RA, Emon B, McKinney RD, Grippo PJ, et al. A portable pen-sized instrumentation to measure stiffness of soft tissues in vivo. Sci Rep. 2021;**11**(1):378.
10. Tiernan JP, Perry SL, Verghese ET, West NP, Yeluri S, Jayne DG, et al. Carcinoembryonic antigen is the preferred biomarker for in vivo colorectal cancer targeting. Br J Cancer. 2013;**108**(3):662–7.
11. Katt ME, Placone AL, Wong AD, Xu ZS, Searson PC. In vitro tumour models: advantages, disadvantages, variables, and selecting the right platform. Front Bioeng Biotechnol. 2016;4:12.
12. Polacheck WJ, Li R, Uzel SG, Kamm RD. Microfluidic platforms for mechanobiology. Lab Chip. 2013;**13**(12):2252–67.
13. Wan L, Neumann CA, LeDuc PR. Tumour-on-a-chip for integrating a 3D tumour microenvironment: chemical and mechanical factors. Lab Chip. 2020;**20**(5):873–88.
14. Golovko D, Kedrin D, Yilmaz Ö, Roper J. Colorectal cancer models for novel drug discovery. Exp Opin Drug Discov. 2015;**10**:1–13.
15. Beshay PE, Cortes-Medina MG, Menyhert MM, Song JW. The biophysics of cancer: emerging insights from micro- and nanoscale tools. Adv NanoBiomed Res. 2022;**2**(1):2100056.
16. Mehta P, Rahman Z, ten Dijke P, Boukany PE. Microfluidics meets 3D cancer cell migration. Trends Cancer. 2022;**8**(8):683–97.
17. Pauli C, Hopkins BD, Prandi D, Shaw R, Fedrizzi T, Sboner A, et al. Personalized in vitro and in vivo cancer models to guide precision medicine. Cancer Discov. 2017;7(5):462–77.
18. Hapach LA, Mosier JA, Wang W, Reinhart-King CA. Engineered models to parse apart the metastatic cascade. NPJ Precis Oncol. 2019;**3**(1):20.
19. Yankaskas CL, Bera K, Stoletov K, Serra SA, Carrillo-Garcia J, Tuntithavornwat S, et al. The fluid shear stress sensor TRPM7 regulates tumour cell intravasation. Sci Adv. 2021;**7**(28):eabh3457.
20. Nagaraju S, Truong D, Mouneimne G, Nikkhah M. Microfluidic tumour–vascular model to study breast cancer cell invasion and intravasation. Adv Healthc Mater. 2018;**7**(9):1701257.
21. Holle AW, Govindan Kutty Devi N, Clar K, Fan A, Saif T, Kemkemer R, et al. Cancer cells invade confined microchannels via a self-directed mesenchymal-to-amoeboid transition. Nano Lett. 2019;**19**(4):2280–90.
22. Wang M, Cheng B, Yang Y, Liu H, Huang G, Han L, et al. Microchannel stiffness and confinement jointly induce the mesenchymal-amoeboid transition of cancer cell migration. Nano Lett. 2019;**19**(9):5949–58.
23. Toh Y-C, Raja A, Yu H, Van Noort D. A 3D microfluidic model to recapitulate cancer cell migration and invasion. Bioengineering. 2018;**5**(2):29.
24. Zhang Y, Li Y, Thompson KN, Stoletov K, Yuan Q, Bera K, et al. Polarized NHE1 and SWELL1 regulate migration direction, efficiency and metastasis. Nat Commun. 2022;**13**(1):6128.
25. Sharafeldin M, Chen T, Ozkaya GU, Choudhary D, Molinolo AA, Gutkind JS, et al. Detecting cancer metastasis and accompanying protein biomarkers at single cell levels using a 3D-printed microfluidic immunoarray. Biosens Bioelectron. 2021;**171**:112681.
26. Petreus T, Cadogan E, Hughes G, Smith A, Pilla Reddy V, Lau A, et al. Tumour-on-chip microfluidic platform for assessment of drug pharmacokinetics and treatment response. Commun Biol. 2021;**4**(1):1001.
27. Lin D, Chen X, Liu Y, Lin Z, Luo Y, Fu M, et al. Microgel single-cell culture arrays on a microfluidic chip for selective expansion and recovery of colorectal cancer stem cells. Anal Chem. 2021;**93**(37):12628–38.
28. Elberskirch L, Knoll T, Königsmark R, Renner J, Wilhelm N, von Briesen H, et al. Microfluidic 3D intestine tumour spheroid model for efficient in vitro investigation of nanoparticular formulations. J Drug Deliv Sci Technol. 2021;**63**:102496.
29. Ha JH, Shin HH, Choi HW, Lim JH, Mo SJ, Ahrberg CD, et al. Electro-responsive hydrogel-based microfluidic actuator platform for photothermal therapy. Lab Chip. 2020;**20**(18):3354–64.
30. Shen Q, Hill T, Cai X, Bui L, Barakat R, Hills E, et al. Physical confinement during cancer cell migration triggers therapeutic resistance and cancer stem cell-like behavior. Cancer Lett. 2021;**506**:142–51.
31. Boya M, Ozkaya-Ahmadov T, Swain BE, Chu C-H, Asmare N, Civelekoglu O, et al. High throughput, label-free isolation of circulating tumour cell clusters in meshed microwells. Nat Commun. 2022;**13**(1):3385.
32. Lv J, Liu Y, Cheng F, Li J, Zhou Y, Zhang T, et al. Cell softness regulates tumourigenicity and stemness of cancer cells. EMBO J. 2021;**40**(2):e106123.
33. Jung DJ, Shin TH, Kim M, Sung CO, Jang SJ, Jeong GS. A one-stop microfluidic-based lung cancer organoid culture platform for testing drug sensitivity. Lab Chip. 2019;**19**(17):2854–65.
34. Zhou J, Kulasinghe A, Bogseth A, O'Byrne K, Punyadeera C, Papautsky I. Isolation of circulating tumour cells in non-small-cell-lung-cancer patients using a multi-flow microfluidic channel. Microsyst Nanoeng. 2019;**5**(1):8.
35. Zheng L, Wang B, Sun Y, Dai B, Fu Y, Zhang Y, et al. An oxygen-concentration-controllable multiorgan microfluidic platform for studying hypoxia-induced lung cancer-liver metastasis and screening drugs. ACS Sens. 2021;**6**(3):823–32.

36. Sancho-Albero M, Sebastián V, Sesé J, Pazo-Cid R, Mendoza G, Arruebo M, et al. Isolation of exosomes from whole blood by a new microfluidic device: proof of concept application in the diagnosis and monitoring of pancreatic cancer. J Nanobiotechnol. 2020;**18**(1):150.
37. Eslami Amirabadi H, SahebAli S, Frimat JP, Luttge R, den Toonder JMJ. A novel method to understand tumour cell invasion: integrating extracellular matrix mimicking layers in microfluidic chips by 'selective curing'. Biomed Microdevices. 2017;**19**(4):92.
38. Roussos ET, Condeelis JS, Patsialou A. Chemotaxis in cancer. Nat Rev Cancer. 2011;**11**(8):573–87.
39. Ahn J, Sei YJ, Jeon NL, Kim Y. Tumour microenvironment on a chip: the progress and future perspective. Bioengineering. 2017;**4**(3):64.
40. Huang YL, Segall JE, Wu M. Microfluidic modeling of the biophysical microenvironment in tumour cell invasion. Lab Chip. 2017;**17**(19):3221–33.
41. Nyberg KD, Bruce SL, Nguyen AV, Chan CK, Gill NK, Kim T-H, et al. Predicting cancer cell invasion by single-cell physical phenotyping. Integr Biol. 2018;**10**(4):218–31.
42. Yan Z, Xia X, Cho WC, Au DW, Shao X, Fang C, et al. Rapid plastic deformation of cancer cells correlates with high metastatic potential. Adv Healthc Mater. 2022;**11**(8):2101657.
43. Shalileh S, Khayamian MA, Ghaderinia M, Abadijoo H, Hassanzadeh-Moghadam H, Dalman A, et al. Label-free mechanoelectrical investigation of single cancer cells by dielectrophoretic-induced stretch assay. Sens Actuat B: Chem. 2021;**346**:130409.
44. Avnet S, Lemma S, Cortini M, Di Pompo G, Perut F, Lipreri MV, et al. The release of inflammatory mediators from acid-stimulated mesenchymal stromal cells favours tumour invasiveness and metastasis in osteosarcoma. Cancers. 2021;**13**(22):5855.
45. Surendran V, Rutledge D, Colmon R, Chandrasekaran A. A novel tumour-immune microenvironment (TIME)-on-chip mimics three dimensional neutrophil-tumour dynamics and neutrophil extracellular traps (NETs)-mediated collective tumour invasion. Biofabrication. 2021;**13**(3):035029.
46. Kim K, Sohn YJ, Lee R, Yoo HJ, Kang JY, Choi N, et al. Cancer-associated fibroblasts differentiated by exosomes isolated from cancer cells promote cancer cell invasion. Int J Mol Sci. 2020;**21**(21):8153.
47. Lugo-Cintrón KM, Gong MM, Ayuso JM, Tomko LA, Beebe DJ, Virumbrales-Muñoz M, et al. Breast fibroblasts and ECM components modulate breast cancer cell migration through the secretion of MMPs in a 3D microfluidic co-culture model. Cancers. 2020;**12**(5):1173.
48. Eslami Amirabadi H, Tuerlings M, Hollestelle A, SahebAli S, Luttge R, van Donkelaar CC, et al. Characterizing the invasion of different breast cancer cell lines with distinct E-cadherin status in 3D using a microfluidic system. Biomed Microdevices. 2019;**21**(4):101.
49. Deng Y, Liu SY, Chua SL, Khoo BL. The effects of biofilms on tumour progression in a 3D cancer-biofilm microfluidic model. Biosens Bioelectron. 2021;**180**:113113.
50. Tsuji T, Sasaki Y, Tanaka M, Hanabata N, Hada R, Munakata A. Microvessel Morphology and Vascular Endothelial Growth Factor Expression in Human Colonic Carcinoma With or Without Metastasis. Lab Invest. 2002;**82**(5):555–62.
51. Chiang SPH, Cabrera RM, Segall JE. Tumour cell intravasation. Am J Physiol–Cell Physiol. 2016;**311**(1):C1–C14.
52. Zavyalova MV, Denisov EV, Tashireva LA, Savelieva OE, Kaigorodova EV, Krakhmal NV, et al. Intravasation as a key step in cancer metastasis. Biochemistry (Moscow). 2019;**84**(7):762–72.
53. Strilic B, Offermanns S. Intravascular survival and extravasation of tumour cells. Cancer Cell. 2017;**32**(3):282–93.
54. Azevedo AS, Follain G, Patthabhiraman S, Harlepp S, Goetz JG. Metastasis of circulating tumour cells: favorable soil or suitable biomechanics, or both? Cell Adh Migr. 2015;**9**(5):345–56.
55. Follain G, Osmani N, Azevedo AS, Allio G, Mercier L, Karreman MA, et al. Hemodynamic forces tune the arrest, adhesion, and extravasation of circulating tumour cells. Dev Cell. 2018;**45**(1):33–52.e12.
56. Jeon JS, Bersini S, Gilardi M, Dubini G, Charest JL, Moretti M, et al. Human 3D vascularized organotypic microfluidic assays to study breast cancer cell extravasation. Proc Natl Acad Sci. 2015;**112**(1):214–9.
57. Zhang Q, Liu T, Qin J. A microfluidic-based device for study of transendothelial invasion of tumour aggregates in realtime. Lab Chip. 2012;**12**(16):2837–42.
58. Wu Y, Zhou Y, Paul R, Qin X, Islam K, Liu Y. Adaptable microfluidic vessel-on-a-chip platform for investigating tumour metastatic transport in bloodstream. Anal Chem. 2022;**94**(35):12159–66.
59. Strelez C, Chilakala S, Ghaffarian K, Lau R, Spiller E, Ung N, et al. Human colorectal cancer-on-chip model to study the microenvironmental influence on early metastatic spread. iScience. 2021;**24**(5):102509.
60. Jeon JS, Zervantonakis IK, Chung S, Kamm RD, Charest JL. In vitro model of tumour cell extravasation. PLoS ONE. 2013;**8**(2):e56910.
61. Cui X, Guo W, Sun Y, Sun B, Hu S, Sun D, et al. A microfluidic device for isolation and characterization of transendothelial migrating cancer cells. Biomicrofluidics. 2017;**11**(1):014105.
62. Mei X, Middleton K, Shim D, Wan Q, Xu L, Ma Y-HV, et al. Microfluidic platform for studying osteocyte mechanoregulation of breast cancer bone metastasis. Integr Biol. 2019;**11**(4):119–29.
63. Regmi S, Fu A, Luo KQ. High shear stresses under exercise condition destroy circulating tumour cells in a microfluidic system. Sci Rep. 2017;**7**:39975.
64. Mitchell MJ, King MR. Fluid shear stress sensitizes cancer cells to receptor-mediated apoptosis via trimeric death receptors. New J Phys. 2013;**15**:015008.
65. Takao S, Taya M, Chiew C. Mechanical stress-induced cell death in breast cancer cells. Biol Open. 2019;**8**(8):bio043133.
66. Basu R, Whitlock BM, Husson J, Le Floc'h A, Jin W, Oyler-Yaniv A, et al. Cytotoxic T cells use mechanical force to potentiate target cell killing. Cell. 2016;**165**(1):100–10.
67. Tijore A, Yao M, Wang Y-H, Hariharan A, Nematbakhsh Y, Lee Doss B, et al. Selective killing of transformed cells by mechanical stretch. Biomaterials. 2021;**275**:120866.
68. Kumari A, Kumar A, Xu G, Roy K, Pratap R, Holle A, et al. Confinement suppresses mechanical force-mediated cancer cell apoptosis. Small. 2025:e2504261.
69. Singh A, Tijore A, Margadant F, Simpson C, Chitkara D, Low BC, et al. Enhanced tumor cell killing by ultrasound after microtubule depolymerization. Bioengineering & Translational Medicine. 2021;6(3):e10233.

70. Tijore A, Margadant F, Dwivedi N, Morgan L, Yao M, Hariharan A, et al. Ultrasound-mediated mechanical forces activate selective tumor cell apoptosis. Bioengineering & Translational Medicine. 2025;10(2):e10737.
71. Tijore AS, Margadant F, Yao M, Sheetz MP. Systems and methods for cancer treatment. US Patent App. 2022;17607819.
72. Veena SM, Chen D, Kumar A, Pratap R, Young JL, Tijore A. Nanoscale ligand spacing regulates mechanical force-induced cancer cell killing. Nano Letters. 2025;25(6):2418–25.
73. Kumari A, Veena SM, Luha R, Tijore A. Mechanobiological strategies to augment cancer treatment. ACS Omega. 2023;8(45):42072–85.
74. Tijore A, Yang B, Sheetz M. Cancer cells can be killed mechanically or with combinations of cytoskeletal inhibitors. Front Pharmacol. 2022;13.

27

Understanding contribution of fibroblasts in inception of cancer metastasis from an evolutionary perspective

Yasir Suhail, Wenqiang Du, Günter Wagner, and Kshitiz

27.1. Introduction

In recent years, our understanding of cancer as a disease of proliferating tumour cells acquiring mutations leading to metastasis has undergone significant revision [1–3]. It is now widely appreciated that the progression of many cancers towards metastasis is an outcome of both cancer autonomous changes [4,5,6] and the cancer's interaction with its microenvironment [3,7,8]. Cancer cells exist in a continuum of states within the metastatic cascade; the breach of basal lamina and the stromal trespass constituting the earliest steps towards distal metastasis [1]. Fibroblasts are the most abundant cells in the stroma and respond to the injury to basal lamina by activating the wound healing response, repairing the disrupted extracellular matrix, and facilitating wound resolution [9–13]. However, as cancer is a perpetual wound that never heals, the physiological wound healing response is taken over by a more inflammatory stromal reaction, benefiting metastatic progression [14]. Although evidence is building upon the role of stromal microenvironment in regulating and predicting early dissemination and onset of metastasis [15,16], little is known about the mechanisms by which fibroblasts either assist or resist metastatic initiation.

Stromal response to the collective invasion of epithelial-like cells is not only limited to cancer metastasis but also occurs in other physiological contexts, notably placental invasion during pregnancy. Analogies between cancer and placentation have been drawn for nearly a century, largely centred around the invading cells, either trophoblasts or cancer cells, respectively [17–19]. Indeed, there are many parallels drawn between the transformed epithelial invasion into the surrounding stroma, and the trophoblasts invasion into the maternal stroma [20,21]. Evolution has resulted in a remarkable diversity of placentation phenotypes across mammals, presenting a wide phenotypic landscape to understand invasive processes in a physiological context. In the past few years, we have identified parallels between the contribution of stromal fibroblasts to both placental invasion and early stages of cancer metastasis. This has led to the development of an evolutionary framework to elucidate selected stromal mechanisms contributing to cancer metastasis. Here, we explain this framework and our developing understanding of how the selected fibroblast-specific genetic and molecular factors may also play a role in the metastatic transition of human cancers.

27.2. Evolution of stromal control of placental invasion in mammals

Pregnancy and reproduction, due to their essentiality in transferring life across generations, are obviously under very high selective pressure in evolution. In mammals, the foetus develops within a confined maternal cavity, the uterus, and interfaces with the maternal tissue through placenta, an organ of extra-embryonic origin. While the developing foetus is dependent on maternal life to continue its development and nutrition, greater foetal growth at some cost of maternal health can be to its advantage. The maternal–foetal interface (MFI), due to, firstly, the genetic distinctiveness of the foetus and the maternal tissues and, secondly, both their competing and cooperative interests, is also the site of continuous evolution and maternal–foetal conflict [22,23]. This is most famously represented in genetic imprinting, wherein the maternal and paternal lineage of epigenetics would favour growth-limiting and growth-enhancing phenotypes, respectively [22,24,25]. Theoretically, cell proliferation of the developing foetal tissue, invasiveness of the placental trophoblasts, and any associated energetic or nutritional demands should all be under competing selection of matrilineal and patrilineal expressed genes. Interestingly, a large fraction of genes found to be differentially imprinted between the two lineages are expressed in placenta, suggesting the centrality of implantation and placental invasion in this conflict [26]. While the sites of genetic imprinting are concentrated in the imprinting control regions that are CpG-rich [27], the imprinted genes are phenotypically enriched for fetoplacental

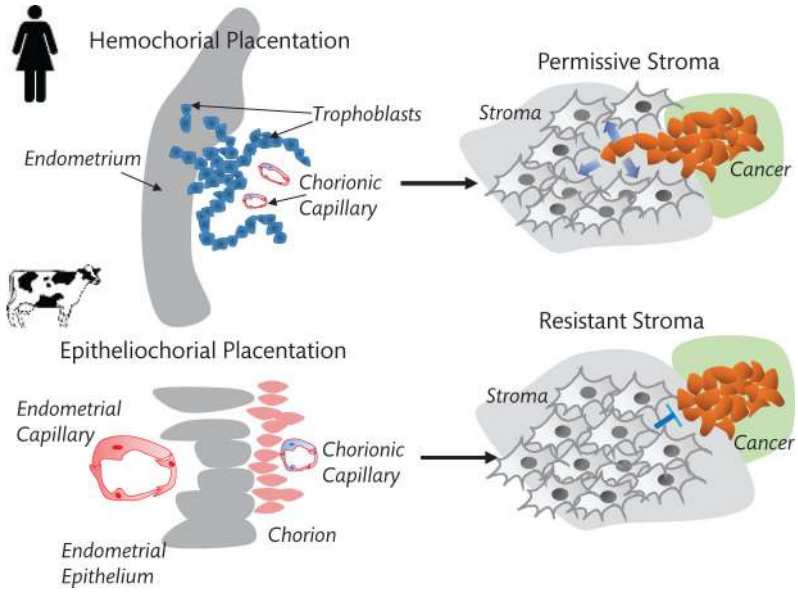


Figure 27.1. Evolved Levels of Invasibility (ELI). Recent evolution of non-invasive epitheliochorial placentation in cows is an outcome of increased resistance in endometrial stromal fibroblasts. Increased resistance is conferred in secondary tissues, resulting in decreased cancer malignancy.

growth [28]. One of the earliest described imprinted genes was IGF2 in mice and is associated with placental growth [29,30].

However, the maternal–foetal cooperation and competition, which we term as 'coopetition' in the game-theory parlance [31], plays out not only in fetoplacental growth but also in the placental invasion as a phenotype. The distinct genetic make-up of the tissues constituting the MFI, and the resulting coopetition, has resulted in a large evolutionary solution space presenting an immense diversity at every biological scale: genetic, signalling, and tissue constitution to anatomical [32]. The MFI has been anatomically classified in three distinct categories with large variations within each category [33,34]. Towards one extreme are *epitheliochorial* species, wherein the chorion is juxtaposed to the maternal endometrial epithelium. In these species (e.g. pig, cow, dugong, etc.), the nutrient transfer at the MFI primarily occurs via diffusion. To note, placentation in cattle is described as a distinct sub-category, called *synepitheliochorial*, wherein the epithelium on the maternal endometrium is digested by the placenta, but there is no invasion at the MFI. Towards the other extreme is *hemochorial* placentation, characterized by deep invasion of placental trophoblasts into the maternal stroma. This type of placentation occurs in rodents and primates, including the humans, and is most similar to the early stages in cancer metastasis. Among primates, placental invasion is particularly invasive, led by specialized extravillous trophoblasts (EVTs), which not only invade into the stromal compartment but also digest the smooth muscle surrounding the high-resistance spiral arteries in the endometrium and take over control of the blood supply to the foetus [35]. Again, there are parallels here with cancer metastasis where cancer cells can undergo endothelial transformation called vascular mimicry. Endotheliochorial placentation that is present in carnivores stands phenotypically in between the two, characterized by moderate stromal invasion, but no direct placental access to maternal blood.

While parallels exist in tumour growth and fetoplacental growth related to genetic imprinting, comparative differences in placental invasion also show a remarkable correlation with the extent of cancer malignancy. D'Souza and Wagner showed in 2014 that multiple tumour types compared across species showed a much higher frequency of malignancy in mammals with invasive placentation using a dataset of 8,500 tumours in a cohort of 200,000 animals [36,37]. It was posited that the increased invasion among hemochorial species (including in humans) is attributed to more invasive trophoblasts [19,38]. However, recent genomic phylogenetic construction has shown that contrary to the prevailing anthropocentric view that invasive hemochorial placentation is older, it is the non-invasive placentation that has evolved more recently [39]. Indeed, the invasive hemochorial placentation is present in the stem lineage of the placental mammals. Non-invasive placentation has evolved multiple times independently, suggesting that there may be a selective advantage to reduced placental invasion [32]. Wagner, Levchenko, and our group therefore proposed to explain these findings in an evolutionary framework, termed 'Evolved Levels of Invasibility (ELI)', which posits that it is the endometrial stromal fibroblasts that have evolved to resist trophoblast invasion in certain mammals [40] (Figure 27.1). Secondary acquirement of these resistive characteristics in stromal fibroblasts of other tissues has resulted in reduced chances of cancer metastasis.

27.3. Quantitative tools to measure 'stromal invasibility' as a phenomenon

As the role of stromal components regulating invasion is being more appreciated, new methods have been developed to quantitatively measure the phenomenon. These include invasion of tumour explant or spheroid stromal environment, which may be either composed of matrix or cells from the connective tissue, in 2D (Figure 27.2A) or in 3D (Figure 27.2B). Boyden chambers with pre-seeded layers of fibroblasts and cancer cells, or microfluidic chambers to spatially define stroma–cancer interface, have also been increasingly used (Figure 27.2C and D). We have ourselves

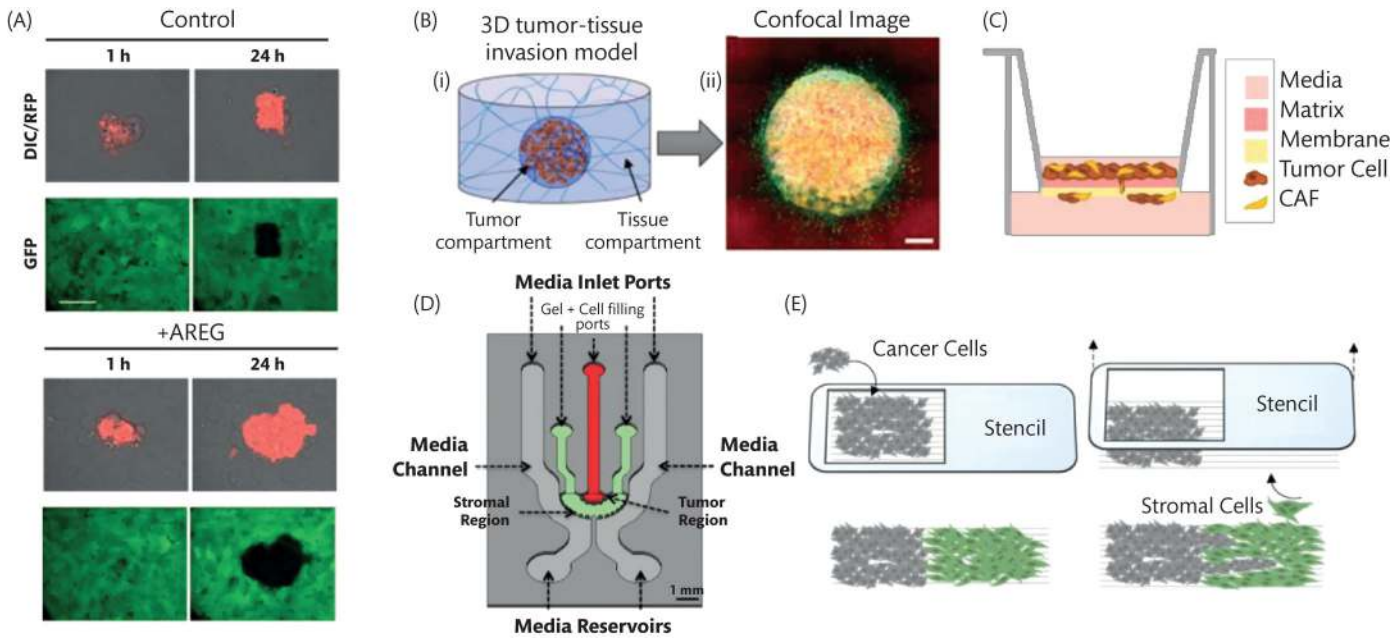


Figure 27.2. Methods to measure stromal response to invasion. (A) Representative differential interference contrast (DIC) and pseudocoloured confocal fluorescence images of ability of vehicle- or AREG-treated CL31 cell clusters expressing red fluorescent protein (RFP) to clear a mesothelial monolayer expressing green fluorescent protein (GFP) at the indicated time points. Scale bar: 50 μm [4]. (B) Overview of 3D tumour–tissue invasion model and fabrication system. (i) Schematic of 3D tumour–tissue invasion model and (ii) a representative image of tumour cells (Panc-1) invading into the surrounding matrix. Image represents 16 fields of view, each of which is a maximum projection of a 400 μm z-stack (20 μm step; 21 slices) after 5 days of culture; green: actin (phalloidin); blue: nuclei (Hoechst 33342); and red: fibrillar collagen (confocal reflectance). Scale bar: 400 μm [10]. (C) Trans-well invasion of tumour cells mixed with cancer-associated fibroblasts (CAFs) [12]. (D) Spatial organization of ECM and cells in a microfluidic chip. The tumour region is represented by the red colour and the stroma is represented by the green. The depth of the channels is 200 μm [20]. (E) Schematic showing set-up to measure stromal invasibility, Accelerated Nanopatterned Stromal Invasion Assay (ANSIA); fluorescently labelled cancer cells are patterned with stromal fibroblasts to create a collective interface between the two orthogonal to the underlying anisotropic nanopatterned fibre direction; cancer invasion is measured for 24 h using live cell fluorescent microscopy. Here, invading cancer cells are the constant, while the invaded fibroblasts are perturbed with CRISPR/Cas9-mediated gene silencing [21].

used stencil-based cell patterning [41] to create interface of cancer cells and fibroblasts, allowing live cell microscopy to observe dynamics of invasive processes. However, the process of stromal invasion is very slow, taking days to establish measurable differences across well-established models (e.g. benign and metastatic cancer cells). Introducing aligned nanostructured ridges on the surface, orthogonal to the initial direction of the interface aligns the actomyosin assemblage of cells. This topographical feature accelerates the process of invasion, specifically by aligning the migratory movement of cells in a single direction—orthogonal to the interface (Figure 27.2E). We were able to establish measurable and large phenotypic differences across benign WM35 and malignant 1205Lu melanoma cells [40]. This platform is called *Accelerated Nanopatterned Stromal Invasion Assay* (*ANSIA*), consisting of an anisotropic nanofabricated substratum created using polyurethane, or poly-ethylene glycol and its derivatives, or any other UV-curable biomaterial using capillary force lithography. Either of the two cell types could be genetically perturbed prior to patterning allowing stromal-specific mechanisms to be investigated. Furthermore, image analysis of the pattern of collective invasion can potentially provide plausible mechanisms driving stromal invasion [42]. As noted earlier, stromal invasion of cancer is a complex phenotype and can be a composite outcome of many sub-processes, each likely driven by different signalling mechanisms. These include increased migration (e.g. through chemotactic mechanism towards stromal fibroblasts), changes in matrix production, degradation or remodelling of extracellular matrix, mechanical pull or push force generation by the stromal fibroblasts, and coupled fibroblast–cancer displacement driven by covalent cell–cell adhesions. Even image analysis without any perturbation can potentially provide opportunities to constrain the possible mechanisms driving stromal invasion in a given experiment, allowing more focused hypothesis generation that can then be experimentally tested [42].

27.4. Identifying genes correlating with stromal invasibility

We have already shown experimentally that it is the identity of the endometrial fibroblasts from different species that determine the extent of invasion of EVTs, more so than the trophoblasts. It is pertinent to point out that we thus label this phenomenon *invasibility* (i.e. the propensity or opposition of the stromal fibroblasts to allow invasion by another cell population), in contrast to the more well-known concept of *invasiveness* (the ability of the invading cell population to invade the stroma). As we have demonstrated that stromal invasibility itself is a selected phenotype, and has therefore evolved among mammals to prevent excessive placental invasion, the obvious question is whether it is possible to identify genes associated with invasibility.

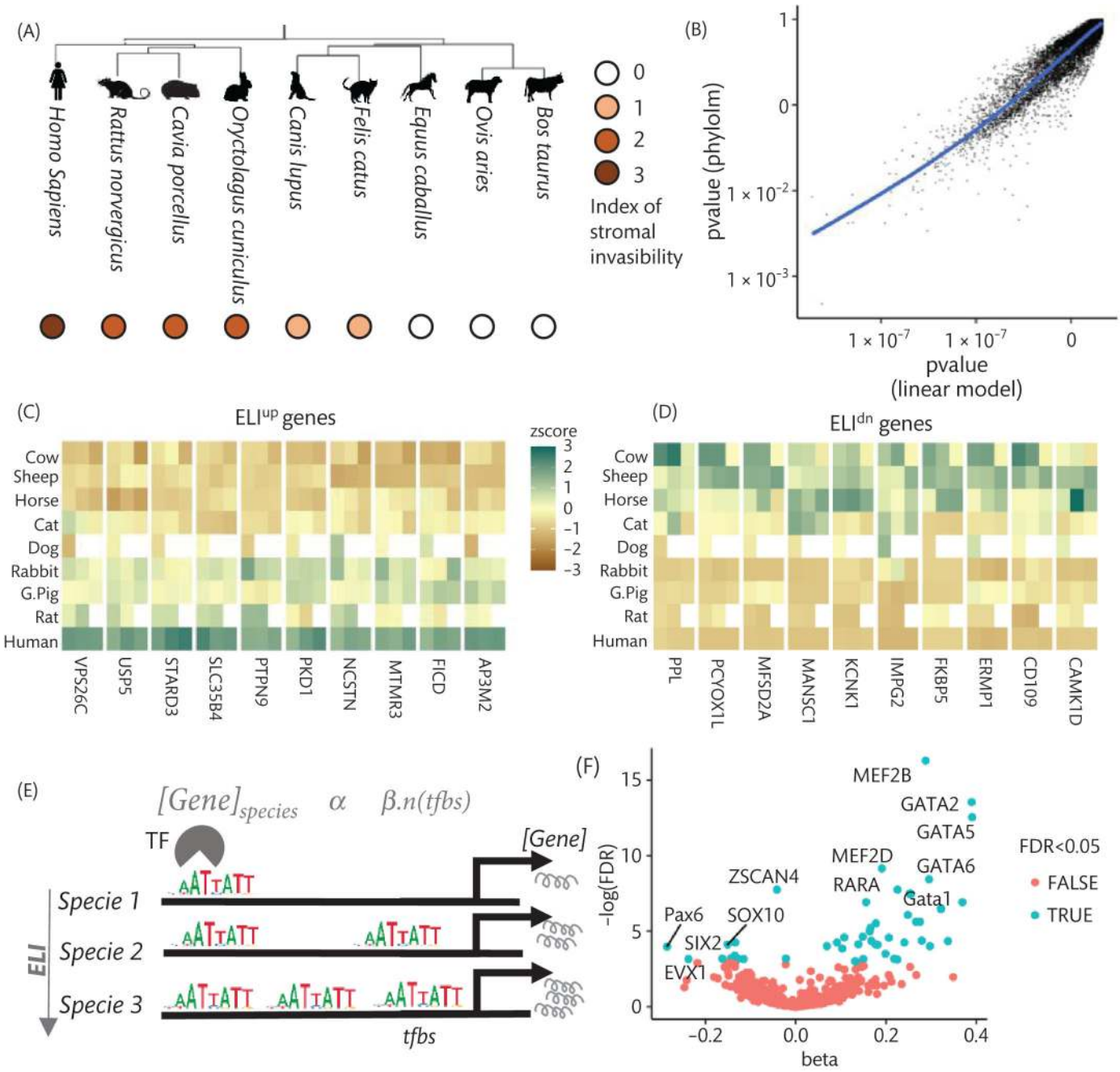


Figure 27.3. Identifying the transcriptional regulators explaining variation in stromal invasibility across mammals. (A) Hierarchical phylogenetic representation of transcripts of endometrial fibroblasts from various mammalian species with varying degrees of placentation invasion (ranked 0–3). (B) Correlation of individual gene expression with invasibility evaluated using the linear model described; the *x*-axis shows *p*-value for linear model, while the *y*-axis shows it for a phylogenetic linear model. (C, D) Gene expression of top ELI^{up} and ELI^{dn} genes in the analysed species. (E) Schematic showing the model explaining change in downstream gene expression in mammalian fibroblasts based copy number of transcription factor binding sites in the cis-regulatory region. (F) Scatter plot showing transcription factors whose binding sites affect expression of downstream gene expression for ELI-related genes. The *x*-axis shows the effect size, beta, while the *y*-axis shows statistical significance (significant TFs are coloured green). *Source:* Adapted from Suhail et al. [11].

Transcriptomic profiling of the stromal fibroblasts from the endometrium of a number of eutherian (i.e. placental) mammals has afforded an opportunity to explore the genetic and transcriptional basis of the modulation of stromal invasibility [43]. This approach limits the identification of genes that are selected within mammalia to prevent trophoblast invasion. It is likely that there are many more genes, and potentially other mechanisms, which may explain stromal contribution to cancer malignancy. But the strong selective pressure during pregnancy, and which has resulted in innovation of this particular phenotype (a large increase in stromal resistance), gives confidence that the genes we identify through this approach are more likely to be connected to the phenotype of stromal invasibility (or resistance).

Wagner group has collected transcriptomic data from skin and endometrial fibroblasts from various mammals with differing extent of placental invasion [44]. We used these transcriptomic data comprising of fibroblasts from mammals with invasive (rabbit, guinea pig, rat, and humans) and less invasive (cat, horse, sheep, dog, and cow) placentation (**Figure 27.3**A). Based on well-understood anatomical presentation of placentation (r_s for species s), we identified stromal genes likely involved in modulating the degree of invasibility by fitting the linear model $r_s = \gamma_g \cdot e_{g,s} + \epsilon_{g,s}$, relating the expression $e_{g,s}$ of gene g to the placental invasibility index r_s, of species s. In this model, the fitted parameter γ_g denotes the causality of gene g to the stromal invasibility. In addition to the magnitude and sign (positive for pro-invasable genes, and negative for anti-invasable genes), the p-value of the fit gives the statistical confidence in the correlation of gene g to stromal invasibility (**Figure 27.3**B).

The most likely causal genes (based on the p-value of the correlation and the sign of effect size γ_g) were thus labelled as ELI^{up} (if putatively pro-invasable) and ELI^{dn} (if putatively anti-invasable). This approach has provided us with a putative ELI score for each gene. Figure 27.3C and D shows examples of genes regressively changing in expression along with the placental invasion index. Without any validation, one would hypothesize that ELI^{up} genes are likely to promote stromal invasibility, while ELI^{dn} genes are likely to promote stromal resistance in endometrial fibroblasts. ELI framework also posits that these gene to phenotype relationships are likely to be maintained in fibroblasts from other tissues, including in the skin. These genes are good candidates to test for their potential to affect changes in stromal invasibility, both to cancer or trophoblast invasion.

As noted earlier, stromal invasibility is a composite phenomenon, and it is not our case that evolution of placental invasion is the only framework to observe and investigate the phenomenon. From a human disease perspective, it is also possible that these evolved mechanisms may not be the primary mechanisms involved, although evidence is building that these genes are causal in regulating stromal invasibility in human cancers too (see next sections). However, ELI sets this question in a sharp relief, vastly reducing the phenotypic landscape to study the particular phenomenon of stromal contribution to epithelial invasion, and if not to explain it, then to identify strategies to modulate human stromal environment increasing its resistance to invasion by a metastasizing cancer.

27.5. Identifying transcriptional correlates and their genetic regulators explaining stromal invasibility

Having identified genes that are correlated to stromal invasibility phenotype across mammals, it is a logical question to ask if there are transcriptional regulators explaining the phenomenon. Species-specific information should be primarily encoded in the genomic sequence (ignoring epigenetic regulation). Gene regulatory network databases do not exist for non-standard eukaryotic species to infer transcriptional regulation with gene expression changes and then linking them to the invasibility phenotype. We therefore have used transcription factor (TF) binding sites (TFBS) in the cis-regulatory regions of the genes as a proxy to the strength of a given TF-mediated transcriptional regulation. By connecting number of TFBS in the cis-regulatory region of each gene, in each observed species with the gene expression values, we could identify key TFs contributing to changes in ELI genes.

Specifically concentrating on the binding of the TFs in the promoter regions of genes, we hypothesized that gain or loss of TFBS proximal to the gene of interest encodes differences in transcriptional levels across species. While these genomic changes should affect all cell types, the tissue-specific transcriptional levels would be set using the gene expression of transcriptional factors in this model.

In order to mitigate the baseline variation in expression among genes, we first shrink the gene expression $e_{g,s}$ of gene g in species s in transcripts per million using a square-root transform to obtain scaled expression $h_{g,s}$:

$$h_{g,s} = \sqrt{e_{g,s}} - \frac{1}{N_{species}} \sum_{s'} \sqrt{e_{g,s'}}$$

The scaled gene expression matrix (for each gene X species) is then fitted using a linear model on the number of transcription factor binding sites $n_{t,g,s}$ for transcription factor t in the promoter region of gene g in the genome of species s,

$$h_{g,s} = \beta_t n_{t,g,s} + \epsilon_{g,s}$$

Here, β_t represents the effect on (scaled) gene expression $h_{g,s}$ from each additional occurrence of a binding site of transcription factor t. The sign of the coefficient β_t represents the direction of the effect, i.e. positive for enhancers and negative for repressors.

A total of 572 known eukaryotic TFBS motifs were downloaded from JASPAR, and genomic sequences of the eutherian mammals were downloaded from Ensembl. The genomic region 5kb upstream of the translation start site to 1 kb downstream was considered as the promoter region for our purpose. The occurrences of the TFBS sequences were counted in these promoters using the MEME suite to fit the model. A model is fitted independently for each motif t, allowing identification of 187 TFs (FDR < 0.05) significantly explaining inter-species differences in whole transcriptome in the stroma. A preliminary analysis showed that there are many TFs that exhibit an antagonistic species-wide effect on stromal genes related to invasibility (**Figure 27.3**F). These TFs present attractive targets for stromal screening to establish their functional effect on regulating stromal integrity.

We surmise that many of these TFs may be central players in regulating fibroblast activation or fibroblast-mediated transformation of cancer into a metastatic disease. We identified GATA2 and TFDP1 as two major TFs showing an enhancer effect on pro-invasible (ELI^{up}) genes. Effect of both these TFs in the target gene expression was demonstrated, and their role in regulating stromal resistance was also demonstrated using ANSIA.

27.6. Signature of evolved stromal resistance in human cancers

While we have identified gene expression changes that occur in parallel to the evolved phenotypic changes in endometrial resistance, we have advanced the ELI framework to essentially explain the reported differences in cancer malignancies across mammals. A key question

is whether the differences in endometrial stromal invasibility has any bearing in human cancers.

The answer is not obvious and is also quite difficult to answer without fibroblast-specific data in large numbers. Firstly, picking on the obvious part of argument, it is rational to argue that endometrial changes in stromal resistance to placental invasion, even if it is paralleled in skin fibroblasts against melanoma, may be quite unimportant in humans. Humans are not cows, and therefore, it is not obvious to expect that stromal fibroblasts in humans may behave similar to mammals with epitheliochorial placentation. The variation in human stromal transcriptome may occur at axes different from the ELI axis we have identified through natural evolution in stromal resistance. Secondly, to test these hypotheses requires fibroblast-specific data in human cancers, correlated with a phenotype related to early stages of metastasis. Single-cell sequencing data is only now beginning to be collected in large numbers to allow sufficient large statistical power to test these hypotheses directly.

We therefore resorted to the second best approach, using what data is currently available, to test if the evolutionary changes in stromal resistance have any purchase in human biology of cancer. The largest central database of gene expression in tumour samples is The Cancer Gene Atlas (TCGA). We downloaded the tumour gene expression data from TCGA for all cancer types and compared the expression to tissue-matched normal samples from the Genotype Tissue Expression database. ELI^{up} genes, i.e. whose greater expression is associated with hemochorial placentation, were prominently up-regulated in tumour samples for a number of cancer types, including PAAD, ESCA, HNSC, LIHC, etc. (**Figure 27.4**A). Conversely, the putatively anti-invasable ELI^{dn} genes were prominently down-regulated in SKCM (melanoma) tumours compared to normal skin (Figure 27.4B). This is indicative of a statistical confirmation that many genes correlated with placental invasion in the endometrial fibroblasts are also associated with cancer invasion (Figure 27.4C).

As humans have hemochorial placentation, we investigated whether the ELI^{dn} genes (for which humans would have low expression) are associated with increased resistance. Does higher expression of these genes confer epitheliochorial-like resistance to invasion? We looked deeper in the melanoma dataset and investigated the inter-patient variation in gene expression of ELI^{dn} genes. Collating the gene expression with patient clinical data (specifically, the years survived after diagnosis), we found that patient survival significantly correlated with increased expression of ELI^{dn} genes. Indeed, loss of ELI^{dn} genes was associated with decreased survival in patients of melanoma (Figure 27.4D). Effectively, it means that the more cow like a human melanoma fibroblast are (presumably ELI genes are mostly expressed in fibroblasts that are mixed with cancer cells), more likely it is for the patient to survive. These correlative analyses have only become possible now with the availability of new transcriptomic data for a large number of patients. It is also remarkable that evolutionarily selected changes are reflected in human diseases.

Although a large amount of patient data is not existent in single-cell sequencing (wherein any signal could be correlated to a broader phenotype like survival), moderate-sized data with cancers in different stages of metastasis ia beginning to be available. We chose pancreatic cancer, as it is the most rich in stromal fibroblasts, and is also incidentally, as one of the deadliest cancers. Interestingly, even within stromal component, pancreatic ductal cell carcinoma, the most deadliest of pancreatic cancers, there is remarkable heterogeneity [9,45]. As single-cell RNA sequencing allows direct measurement of gene expression within fibroblasts, we assessed the changes associated in fibroblasts and stellate cells at each stage of cancer [9]. A remarkable finding emerged. In pre-metastatic pancreatic ductal adenocarcinoma cancer (PDAC) samples, fibroblasts showed remarkable homogeneity, expressing markers associated with fibroblast activation (Figure 27.4E and F). This is akin to the classic wound healing response, wherein fibroblasts undergo rapid proliferation in response to an injury, increase their contractile machinery, and produce matricellular proteins. However, when we focused on the transition of cancer from pre-metastatic to an early metastatic stage, where only one or two proximal lymph nodes are infiltrated, we found a sudden burst of heterogeneity in fibroblasts (Figure 27.4F). While the activated fibroblasts nearly disappeared, many new clusters of fibroblasts emerged, both associated with inflammation and others. However, one of the largest subpopulation was the one associated with higher enrichment of ELI^{up} genes (Figure 27.4G). Effectively, the onset of metastasis and stromal trespass was associated with a dramatic increase in fibroblast subtypes with relative pro-invasive phenotype.

27.7. Conclusions

Our understanding of cancer metastasis has been focused on tumour cells and their transformation into highly mesenchymal, invading, extravasating cell types. Although there is an increased appreciation for the role of non-cancer cells, and certainly cancer fibroblasts in contributing to cancer metastasis, there is no thematic or systems understanding of why and how fibroblasts may contribute to cancer metastasis. It appears that continually new markers are discovered for cancer-associated fibroblasts, with new stromal genes identified which could confer in fibroblasts the capability to enhance metastasis. We consider that for such systems-level phenotype, a single marker or a single gene may never be found (unlike in cancer, where driver mutations may be possible to identify). Instead, it may be worthwhile considering if fibroblasts are contributing (or resisting) to metastasis by changes in their phenotype, which may have a basis in multiple genes and pathways.

There are many potential ways to approach this problem beyond evolution. Early events in development where embryonic mesenchymal cells are crucial agents in determining many invasive processes (e.g. gastrulation) may provide us with occasions to study, which are likely repeated in cancers. Similarly, wound-healing response, both scarless and regenerative, can provide other themes to explore. We have thematically explored a remarkable change in collective invasion that has occurred in mammals, to extrapolate mechanisms for human diseases. As these changes in mammalian placentation were also correlated with rates of cancer malignancy in these mammals, hinting that the ELI framework we have advanced may have a thing or two to say about cancer metastasis.

Why is it important to study the evolutionary history of the stromal control of invasion in pregnancy? Tracing the evolutionary trajectories of selected phenotypes of interest can provide a powerful and focused approach to investigate the genetic basis of the same phenotypes in humans. The genetic basis of evolved phenotypes may be similar, or dissimilar in either cases, but these generated hypotheses could be tested against well-designed experiments measuring

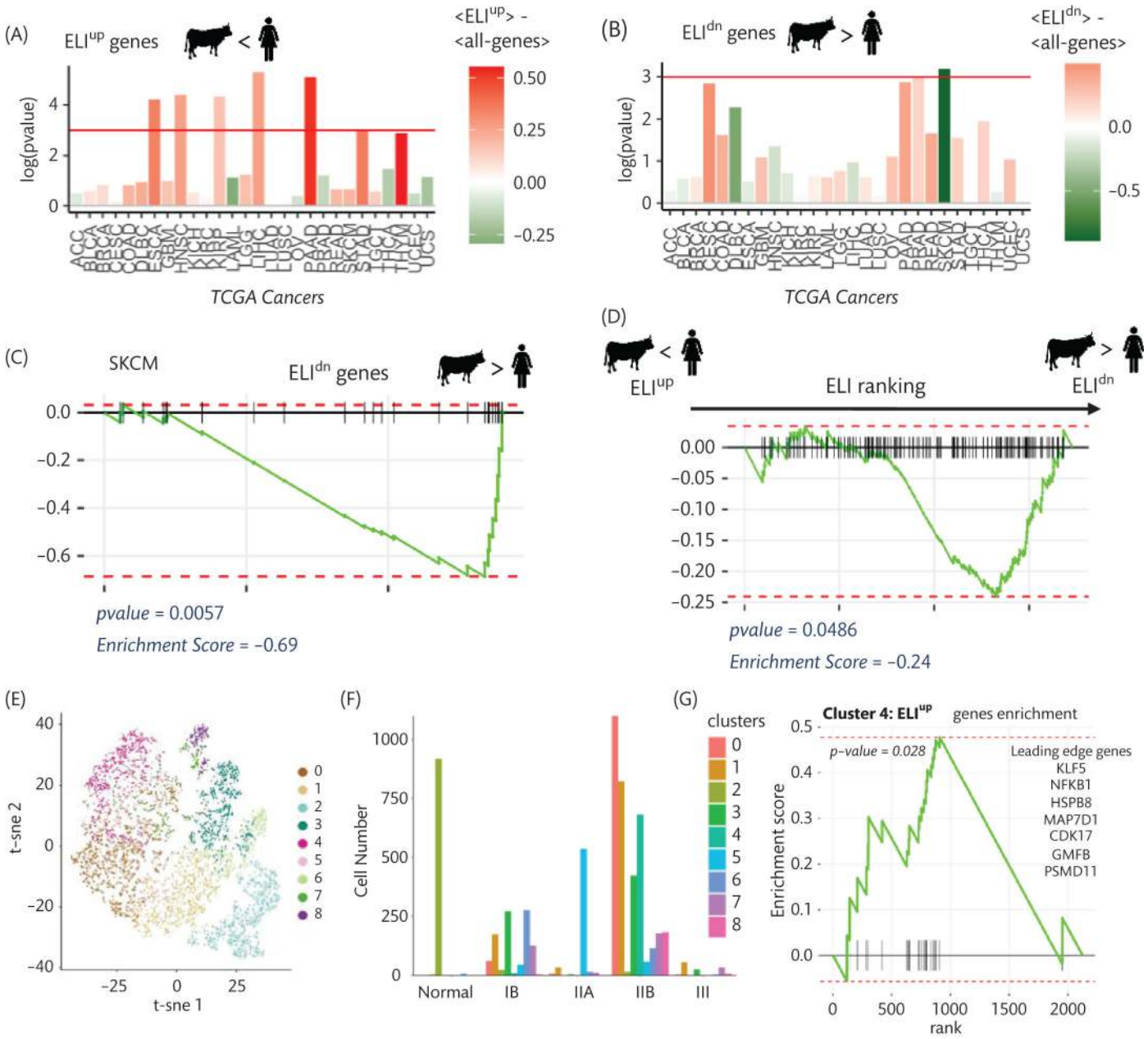


Figure 27.4. Genes evolved with increased stromal resistance to invasion in placentation show a pro-survival and anti-metastatic signal in human cancers. Differential expression of ELI^{up} (A) and ELI^{dn} (B) genes in human cancer samples vs. normal tissue control; *t*-tests were conducted for difference between ELI genes and all other genes for each cancer in the The Cancer Genome Atlas (TCGA); bar colours refer to difference in mean expression of ELI vs. other genes (red: increase; green: decrease). (C) Gene set enrichment analysis (GSEA) of ELI^{dn} genes in skin carcinoma samples vs. control. (D) GSEA analysis of top 500 ranking pro-survival genes in skin carcinoma ordered for their ELI score. (E) t-sne analysis of fibroblast transcriptomics from pancreatic patient data [9] pooled for all cancer metastatic stages, with (F) relative cell numbers in each identified fibroblast clusters for all cancer metastatic stages. (G) GSEA enrichment analysis of ELI^{up} genes within cluster 4 (which is increased in stage IIB) with the leading edge genes highlighted. *Source:* Parts (A)–(D): Modified from Suhail et al. [13]. Parts (E)–(G): Modified from Liu et al. [21].

these phenotypes. However, if there are commonalities in the mechanisms between evolved phenotypes and the pathological presentation of these phenotypes, it could provide a focused approach to look for signal amid much noise. If genes that have changed in correlation with increased stromal resistance to invasion among mammals are implicated in their contribution to keeping metastasis in check, it could provide many new orthogonal opportunities to target cancer metastasis.

REFERENCES

1. Suhail Y, Cain MP, Vanaja K, Kurywchak PA, Levchenko A, Kalluri R, et al. Systems biology of cancer metastasis. Cell Syst. 2019;**9**:109–27.
2. Harper KL, Sosa MS, Entenberg D, Hosseini H, Cheung JF, Nobre R, et al. Mechanism of early dissemination and metastasis in Her2(+) mammary cancer. Nature. 2016;**540**: 588–92.

3. Te Boekhorst V, Friedl P. Plasticity of cancer cell invasion-mechanisms and implications for therapy. Adv Cancer Res. 2016;**132**:209–64.
4. Naffar-Abu Amara S, Kuiken HJ, Selfors LM, Butler T, Leung ML, Leung CT, et al. Transient commensal clonal interactions can drive tumor metastasis. Nat Commun. 2020;**11**:5799.
5. Padmanaban V, Krol I, Suhail Y, Szczerba BM, Aceto N, Bader JS, et al. E-cadherin is required for metastasis in multiple models of breast cancer. Nature. 2019;**573**:439–44.
6. Wellenstein MD, Coffelt SB, Duits DEM, van Miltenburg MH, Slagter M, de Rink I, et al. Loss of p53 triggers WNT-dependent systemic inflammation to drive breast cancer metastasis. Nature. 2019;**572**:538–42.
7. Lambert AW, Pattabiraman DR, Weinberg RA. Emerging biological principles of metastasis. Cell. 2017;**168**:670–91.
8. Linde N, Casanova-Acebes M, Sosa MS, Mortha A, Rahman A, Farias E, et al. Macrophages orchestrate breast cancer early dissemination and metastasis. Nat Commun. 2018;**9**:21.
9. Peng J, Sun BF, Chen CY, Zhou JY, Chen YS, Chen H, et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. Cell Res. 2019;**29**:725–38.
10. Puls TJ, Tan X, Husain M, Whittington CF, Fishel ML, Voytik-Harbin SL. Development of a novel 3D tumor-tissue invasion model for high-throughput, high-content phenotypic drug screening. Sci Rep. 2018;**8**:13039.
11. Suhail Y, Maziarz JD, Dighe A, Wagner GP, Kshitiz. Cis-regulatory differences explaining evolved levels of endometrial invasibility in eutherian mammals. bioRxiv. 2020.
12. Poon S, Ailles LE. Modeling the role of cancer-associated fibroblasts in tumor cell invasion. Cancers (Basel). 2022;**14**:962.
13. Suhail Y, Afzal J, Kshitiz. Evolved resistance to placental invasion secondarily confers increased survival in melanoma patients. J Clin Med. 2021;**10**.
14. LeBleu VS, Kalluri R. A peek into cancer-associated fibroblasts: origins, functions and translational impact. Dis Model Mech. 2018;**11**:dmm029447.
15. Entenberg D, Pastoriza JM, Oktay MH, Voiculescu S, Wang Y, Sosa MS, et al. Time-lapsed, large-volume, high-resolution intravital imaging for tissue-wide analysis of single cell dynamics. Methods. 2017;**128**:65–77.
16. Neelakantan D, Zhou H, Oliphant MUJ, Zhang X, Simon LM, Henke DM, et al. EMT cells increase breast cancer metastasis via paracrine GLI activation in neighbouring tumour cells. Nat Commun. 2017;**8**:15773.
17. Wagner GP, Kshitiz, Dighe A, Levchenko A. The coevolution of placentation and cancer. Annu Rev Anim Biosci. 2022;**10**:259–79.
18. Ross CA. The trophoblast model of cancer. Nutr Cancer. 2015;**67**:61–7.
19. Murray MJ, Lessey BA. Embryo implantation and tumor metastasis: common pathways of invasion and angiogenesis. Semin Reprod Endocrinol. 1999;**17**:275–90.
20. Truong D, Puleo J, Llave A, Mouneimne G, Kamm RD, Nikkhah M. Breast cancer cell invasion into a three dimensional tumor-stroma microenvironment. Sci Rep. 2016;**6**: 34094.
21. Liu S, Suhail Y, Novin A, Perpetua L, Kshitiz. Metastatic transition of pancreatic ductal cell adenocarcinoma is accompanied by the emergence of pro-invasive cancer-associated fibroblasts. Cancers (Basel). (2022);**14**:2197.
22. Haig, D. Cooperation and conflict in human pregnancy. Curr Biol. 2019;**29**:R455–58.
23. Haig, D. Maternal-fetal conflict, genomic imprinting and mammalian vulnerabilities to cancer. Philos Trans R Soc Lond B Biol Sci. 2015;**370**:20140178.
24. Haig, D. Colloquium papers: Transfers and transitions: parent-offspring conflict, genomic imprinting, and the evolution of human life history. Proc Natl Acad Sci U S A. 2010;**107**(Suppl 1):1731–35.
25. Trivers R, Burt A. In Genomic imprinting. An interdisciplinary approach results and problems in cell differentiation (ed Ohlsson R). Springer; 1999, p. 1–21.
26. Tunster SJ, Jensen AB, John RM. Imprinted genes in mouse placental development and the regulation of fetal energy stores. Reproduction. 2013;**145**:R117–37.
27. Buiting K, Saitoh S, Gross S, Dittrich B, Schwartz S, Nicholls RD, et al. Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. Nat Genet. 1995;**9**:395–400.
28. Bressan FF, De Bem TH, Perecin F, Lopes FL, Ambrosio CE, Meirelles FV, et al. Unearthing the roles of imprinted genes in the placenta. Placenta. 2009;**30**:823–34.
29. Sibley CP, Coan PM, Ferguson-Smith AC, Dean W, Hughes J, Smith P, et al. Placental-specific insulin-like growth factor 2 (Igf2) regulates the diffusional exchange characteristics of the mouse placenta. Proc Natl Acad Sci U S A. 2004;**101**:8204–208.
30. Constancia M, Hemberger M, Hughes J, Dean W, Ferguson-Smith A, Fundele R, et al. Placental-specific IGF-II is a major modulator of placental and fetal growth. Nature. 2002;**417**:945–48.
31. Carfi D, Okura M. Coopetition and game theory. J Appl Econ Sci. 2014;**9**:458–69.
32. Carter AM, Enders AC, Kunzle H, Oduor-Okelo D, Vogel P. Placentation in species of phylogenetic importance: the Afrotheria. Anim Reprod Sci. 2004;**82-83**:35–48.
33. Carter AM, Pijnenborg R. Evolution of invasive placentation with special reference to non-human primates. Best Pract Res Clin Obstet Gynaecol. 2011;**25**:249–57.
34. Silva JF, Serakides R. Intrauterine trophoblast migration: a comparative view of humans and rodents. Cell Adh Migr. 2016;**10**:88–110.
35. Carter AM, Enders AC, Pijnenborg R. The role of invasive trophoblast in implantation and placentation of primates. Philos Trans R Soc Lond B Biol Sci. 2015;**370**:20140070.
36. D'Souza AW, Wagner GP. Malignant cancer and invasive placentation: a case for positive pleiotropy between endometrial and malignancy phenotypes. Evol Med Public Health. **2014**;2014:136–45.
37. Priester WA, Mantel N. Occurrence of tumors in domestic animals. Data from 12 United States and Canadian colleges of veterinary medicine. J Natl Cancer Inst. 1971;**47**;1333–44.
38. Costanzo V, Bardelli A, Siena S, Abrignani S. Exploring the links between cancer and placenta development. Open Biol. 2018;**8**:180081.
39. Wildman DE, Chen C, Erez O, Grossman LI, Goodman M, Romero R. Evolution of the mammalian placenta revealed by phylogenetic analysis. Proc Natl Acad Sci U S A. 2006;**103**:3203–208.
40. Kshitiz, Afzal J, Maziarz JD, Hamidzadeh A, Liang C, Erkenbrack EM, Kim HN, et al. Evolution of placental invasion and cancer metastasis are causally linked. Nat Ecol Evol. 2019;**3**:1743–53.
41. Kshitiz, Afzal J, Suhail Y, Ahn EH, Goyal R, Hubbi ME, Hussaini Q, et al. Control of the interface between heterotypic

cell populations reveals the mechanism of intercellular transfer of signaling proteins. Integr Biol. (Camb) 2015;**7**:364–72.
42. Novin A, Suhail Y, Ajeti V, Goyal R, Wali K, Seck A, et al. Diversity in cancer invasion phenotypes indicates specific stroma regulated programs. Hum Cell. 2021;**34**:111–21.
43. Suhail Y, Maziarz JD, Novin A, Dighe A, Afzal J, Wagner G, et al. Tracing the cis-regulatory changes underlying the endometrial control of placental invasion. Proc Natl Acad Sci U S A. 2022;**119**:e2111256119.
44. Ba Q, Hei Y, Dighe A, Li W, Maziarz J, Pak I, et al. Proteotype coevolution and quantitative diversity across 11 mammalian species. Sci Adv 2022;**8**:eabn0756.
45. Hosein AN, Brekken RA, Maitra A. Pancreatic cancer stroma: an update on therapeutic targeting strategies. Nat Rev Gastroenterol Hepatol. 2020;**17**:487–505.

28

Cell competition in tumorigenesis and epithelial defence against cancer

Amrapali Datta and Medhavi Vishwakarma

28.1. Introduction

Cells reside within our body as structured communities abiding by the rules that allow them to carry out relevant physiological functions. Mutations allowing cells to break free from these regulations often lead to uncontrolled proliferation of cells, giving rise to the pathology of cancer. This cell-intrinsic ability of unrestricted growth, e.g. by evading cell cycle control or growth restrictions imposed by cellular damage, is considered to be an important hallmark of cancer progression. Most of our knowledge on cancer is built on research from advanced stages of cancer, and thus the events governing the onset of tumorigenesis remain obscure. In this respect, an obvious but less well-recognized aspect is that tumours grow into space that is already occupied by other cells. Are tumour cells just relying on slow tissue turnover to take advantage of space being freed or are they able to clear their way in an active manner? Increasing evidence suggests that an incipient tumour's fate is largely determined by the complex cellular interactions [1]. Such interactions are often competitive in nature, allowing host cells to kill misfit tumour cells, thereby providing a defence mechanism against cancer [2] [Figure 28.1]. Conversely, when tumour is fitter than the host, similar interactions could lead to host-cell death, allowing for tumour growth and expansion [1,3] Therefore, while cell competition in healthy tissue functions as a surveillance mechanism for inspecting and eliminating the aberrant members, the flip side seems to be exploited by cells with oncogenic mutations that expand at the expense of their neighbours, marking the initiation of tumours. Hence, understanding how cells compete will not only expand our understanding of how tumours manifest into the host but will also allow exploration into new routes of cancer therapeutics involving tumour–host interactions.

28.2. History of cell competition

In the multi-cellular community, cells with different properties often compete with each other for survival and space. This process is named cell competition and was originally discovered in *Drosophila melanogaster*. Through cell competition, cells that are relatively less fit or those that harbour certain detrimental mutations are eliminated and hence called 'losers', when in the vicinity of fitter cells, termed 'winners'. This results in the progressive elimination of the

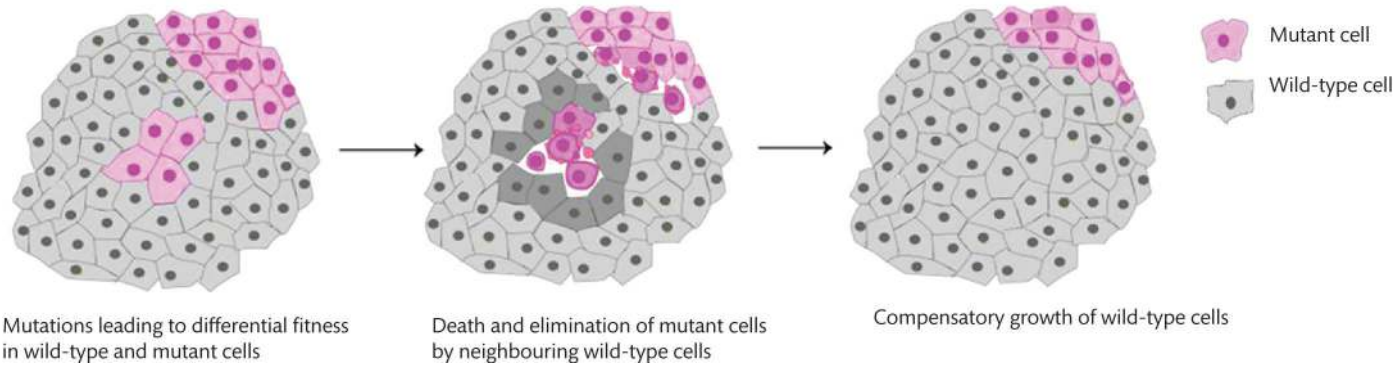


Figure 28.1. Cell competition. Cell competition is a form of cell–cell interaction that acts as a quality control mechanism by selectively eliminating less fit or 'loser cells', upon their interaction with fitter or 'winner cells'. Elimination of loser cells by mechanisms, such as apoptosis or extrusion, is subsequently followed by compensatory growth of winner cells to replace them and take over the tissue.

population of loser cells and their replacement through active tissue colonization by winner cells [Figure 28.1]. A key defining feature of cell competition is that the elimination of loser cells is context dependent, as these cells are viable on their own. The first reports describing cell competition studied cells carrying heterozygous mutations in the *Minute* gene (mutations in ribosomal protein synthesis) in Drosophila [4,5]. In a series of landmark articles, it was shown that while homozygous mutant flies did not survive due to lack of functional ribosomes, heterozygous mutants, despite their slow growth rate, developed and reached near normal body sizes. Remarkably, these heterozygous mutants were eliminated when surrounded by wild-type neighbours demonstrating context-dependent elimination of mutants [4,5]. Interestingly, a similar type of competition was subsequently also observed in mouse embryonic [6] as well as adult cells [7] indicating that it is conserved across species and exists across the lifespan of the individual. Following Minute, several other mutations affecting pathways of basic survival functions have also been shown to trigger competition, e.g. cell growth (mutations affecting Hippo pathway) [8–10], cell polarity (mutations in disc large, scribble, or lethal giant larvae) [10–14], cell anabolism (Myc, RasV12, and v-Src mutations) [15–17], endocytosis (Rab5 mutations) [18], and cell cycle (p53 mutations) [19,20], all lead to the elimination of less fit when fitter neighbours were in proximity. Thus, cell competition acts as a quality control mechanism for maintaining tissue health by elimination of unfit cells from the tissue.

28.3. Cancer cells as super-competitors

Initial studies had attributed the role of cell competition to identifying and correcting developmental defects [4,21,22]; however, the active participation of cells to compete for resources and space in a tissue is also very relevant to tumour progression. Discoveries in the past few decades identifying the involvement of several proto-oncogenes in cell competition [8,9,23,24] have fuelled the hypothesis that cell competition acts as a key driver of tumorigenesis by allowing cancer cells to expand at the expense of the host and suggested that competitive interactions foster expansion of premalignant cells much before morphological alterations are clinically detectable, and hence play a major role in the onset of tumorigenesis [23,25–29].

28.3.1. Oncogenes providing super-competition status to cells

The first proto-oncogene discovered in the context of cell competition—Myc, encoded by the diminutive (dm) gene, is a functionally conserved transcriptional regulator that promotes cell growth and metabolism [30– 32] and is de-regulated in many human cancers [33]. Mutants expressing low copies of Myc show phenotypic similarity with Minute mutants and are therefore eliminated from the tissue by cell competition [30]. Interestingly, slight elevation of Myc levels in mosaic tissues had a striking effect of causing death in nearby wild-type cells, thereby turning Myc overexpressing cells into 'super-competitors' [23,24]. Cell competition driven by Myc activity is also shown to exist in mammalian systems—in mouse embryonic stem cells [6,34] and adult tissues [35]—demonstrating Myc competition is conserved across species. The super-power of myc overexpressing cells stems from their enhanced translational capacity and increased growth potential [30,36] and is beneficial for the tumour to grow and invade into a tissue. Intuitively, then, more pathways involved in cell growth control, such as Hippo pathway, Wingless/Wnt signalling, and JAK/STAT pathway, were explored for their role in conferring cells with a super-competitor status. Hippo pathway is a growth regulatory mechanism that restricts cellular growth by preventing nuclear accumulation of transcriptional coactivator Yorki (Yki/Yap for mammals) through a phosphorylation cascade [37,38]. When Hippo pathway is de-regulated, Yki translocates to the nucleus and turns cells into super-competitors by enhancing growth and proliferation [8,39,40] [Figure 28.2B]. Interestingly, Yki overexpression alone is sufficient for turning cells into super-competitors [40], and myc, being a transcriptional target of Yki, is expressed downstream and facilitates overgrowth of mutants [8,40]. Both myc overexpression and yki nuclear accumulation are frequently observed in human cancers [33,41], suggesting cancer cells are super-competitors. Interestingly, mutations directly affecting myc or Hippo pathway are infrequent and are not known cancer driver mutations, indicating that they mainly act in concert with other oncogenic mutations and help an originally transformed cell to acquire a super-competitor status and establish its territory within the host [33,42]. For example, neoplastic transformation by activating epithelial growth factor receptor (EGFR) mutations, which are known to drive many types of human cancer [43], is mediated by conferring super-competitor status to mutants by increased myc levels [44] [Figure 28.2B]. Similarly, Hippo pathway activity is perturbed through crosstalk with other signalling pathways that frequently harbour oncogenic mutations, e.g. defects in Wnt signalling, transforming growth factor b-bone morphogen protein signalling (TGFb-BMP), Notch signalling, insulin pathway, and mTOR, all facilitate Yki nuclear translocation and are frequently mutated in different human cancers [45–47] indicating the involvement of Hippo pathway downstream of several oncogenic mutations. Other super-competitors genes include those encoding regulators of Wnt/Wingless signalling-Axin and Adenomatous polyposis coli (APC) and genes involved in JAK/STAT pathway. Similar to myc competition, local differences in Wg signalling also trigger competitive interactions wherein cells with high Wg signalling adapt a super-competitor phenotype [48]. These local differences can arise between wild-type cells and clones harbouring mutations, such as Axin or APC, causing hyperactivation of the Wnt pathway [49,50] where neoplastic transformation of these mutants is regulated by hippo pathway [3]. In parallel, JAK/STAT pathway, which is persistently activated in many cancers, is also involved in cell competition—cells with hyper-activated STAT become super-competitors and kill surrounding cells in a manner independent of Myc or Hippo pathway [51] [Figure 28.2B].

Another important gene that is frequently misregulated in many types of human cancers is tumour suppressor p53 [52]. Although, mutations in p53 are mainly observed in the mid/late stages of tumour and are therefore considered to be a product of other oncogenic mutations that precede de-regulation in p53 during tumour formation [52,53]. Recent studies however also show the involvement of p53 in the initial stages of tumorigenesis, more specifically in squamous cell carcinoma, commonly known as skin cancer. Clones lacking p53 in human skin show a competitive advantage

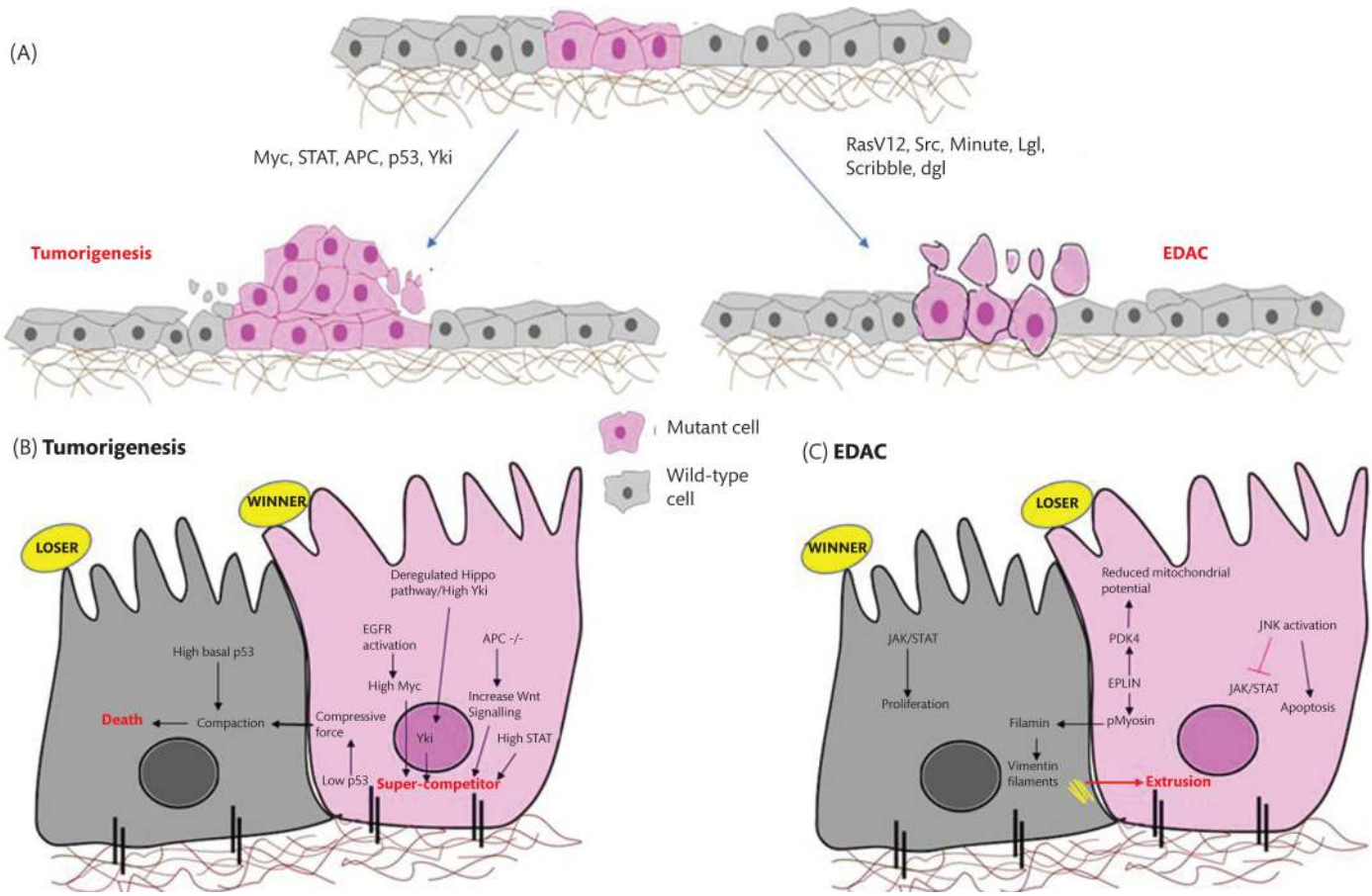


Figure 28.2. Cell competition as a driver of tumorigenesis vs. epithelial defence against cancer (EDAC). (A). Mutations initiating competitive interactions within the cells. Certain mutations lead to extrusion of cells harbouring those mutations via EDAC, while other mutations confer super-competitor status to cells leading to tumorigenesis. Mechanisms that are known to be involved in cell competition leading to tumorigenesis (B, left panel) or cell competition leading to EDAC (C, right panel).

over wild-type cells but only in skin irradiated with UV light [46], suggesting that it is the p53 elevation upon irradiation that makes wild-type clones susceptible to competitive elimination by p53 lacking mutants. Mechanistically, p53 elevation make cells hypersensitive to compaction, leading to apoptotic elimination upon compression [54] [Figure 28.2B]. In consensus with these experiments, another study showed that loss of function p53 mutation confers loser status to Madin–Darby canine kidney (MDCK) epithelial cells [55], unless, preceded with an existing oncogenic mutation (Ras V12), thereby showing that the pre-existing RasV12 transformation affects the cell competition-mediated eradication of mutant p53-expressing cells [55]. In contrast, RasV12-expressing cells were shown to be eliminated from either normal or p53 mutant epithelia, suggesting that the order in which oncogenic mutations occur directs the outcome of cell competition [55]. Such death induction in wild-type cells suggests that super-competitors actively clear space for their growth by killing their neighbours and facilitating tumour formation. Excitingly, this has recently been directly demonstrated in the Drosophila intestine where tumours driven by APC mutations actively kill surrounding cells for their own growth, and by protecting wild-type clones from death, tumour growth could be contained. Therefore, the ability of tumours to metastasize seems to be linked with the initial success in cell competition that confers a Darwinian advantage to the primary tumour and sets off the stage for cancer to progress.

28.4. Epithelial defence against cancer

Carcinogenesis initiates with the transformation of a single cell within the organized epithelia. However, if cancer is driven and dominated by oncogenic mutations [56], then the individual's genetic susceptibility [57] topped with damage from environmental factors (such as smoking, pollution, radiation, obesity, and ageing) increases the number and frequency of these mutations, so much that probability favours getting cancer much more than not getting it. Yet, a majority of people live cancer-free for decades. Why are we not getting more cancer? [58]. The pioneering work by Bissell and co-worker [59] first asked this question when wings of chicken embryo injected with Rous sarcoma virus (RSV) developed cancer free *in vivo* but developed tumour *in vitro*. Thus, by demonstrating the power of tissue microenvironment to override the ability of oncogenes, they suggested that a related defence mechanism against cancer exists in our body. Whether a single transformed cell is

sufficient to drive cancer progression is majorly dependent on what happens at the interface between the transformed cell and the surrounding normal cells during this process.

28.4.1. Initial studies demonstrating EDAC in epithelial tissues

Cell competition experiments with wild-type and RasV12-transformed and with wild-type and Src-transformed MDCK epithelial cell lines have provided insights into the fate of transformed cells when surrounded by wild-type cells [15,16]. These studies have revealed that the transformed cells are apically extruded from the epithelial monolayer when surrounded by normal cells but are viable when they are cultured alone, indicating that the elimination of transformed cells is context dependent and requires the presence of wild-type cells. Similar results were obtained when other types of transformed cells interacted with their wild-type counterparts in an epithelial monolayer. For example, single cells expressing constitutively active form of Cdc42 [60], and constitutively active Yes-associated protein (YAP) [61] within a monolayer of normal epithelial cells, are apically extruded from the epithelium. These studies highlighted the fact that in epithelial tissues, normal cells can sense and actively eliminate the neighbouring transformed cells by a process known as epithelial defence against cancer [62]. This defence mechanism of epithelial tissues against tumorigenesis works on competitive cellular interactions such that the host epithelial cells act as 'winner cells' and oncogenic mutants as 'loser cells' [**Figure 28.2**A] [1,2]. On the flip side, when oncogenic cells become fitter than the host, EDAC is impaired, thus leading to tumour growth and expansion [**Figure 28.2**] [1].

28.4.2. Mechanisms behind EDAC

Recent research have started to provide a mechanistic understanding of the signalling pathways involved in EDAC. For instance, interaction of Src-transformed cells with wild-type cells in a monolayer was shown to activate myosin-II and focal adhesion kinase in the transformed cells, leading to activation of the MAPK pathway [16]. MAPK pathway was also shown to be involved in the apical extrusion of the RasV12-transformed cells when surrounded by *wild-type* cells. The fate of RasV12 cells is majorly influenced by ROCK-dependent activation of myosin-II in RasV12 cells and by drastic cytoskeletal changes in both transformed cells and wild-type cells [15]. Accumulation of cytoskeletal protein filamin in wild-type cells recruits the intermediate filament protein vimentin at the basal side of cell–cell contacts, and these vimentin filaments generate contractile forces to provide wild-type cells with the mechanical strength for squeezing out the transformed cells into apical lumen [63] [**Figure 28.2**C]. Recently, it has been discovered that a matrix stiffness-dependent differential localization of filamin determines the success or failure of EDAC on soft versus stiff matrix and revealed that pathological matrix stiffening, which happens during fibrosis, leads to a failed EDAC, and sets the stage for tumour to grow at the initial stage of oncogenesis [64]. In addition, cell's metabolic machinery is also shown to have an influence on elimination of these transformed cells. For instance, accumulation of EPLIN in Ras-transformed cells leads to increased levels of phospho-dehydrogenase kinase-4, which leads to decreased mitochondrial potential, increased aerobic glycolysis, and cell death [65] [**Figure 28.2**C]. This effect is similar to the conventional Warburg effect that is usually observed in advanced stages of tumour in adaptation to the harsh environment and genetic insults, and is known to promote tumour progression. At the initial stages, however, it seems to act against tumorigenesis by aiding in removal of transformed cells from the epithelium. Such apical extrusion is suggestive of cancer preventive because the direction is opposite from basal invasion that is required for cancer metastasis [2,58]. Diet has also been shown to affect EDAC, and it was shown that a high-fat diet results in substantial attenuation of the frequency of apical elimination of RasV12-transformed cells from intestinal and pancreatic epithelia [66]. A phenomenon similar to EDAC was also observed in three- dimensional (3D) organotypic culture system of mammary breast acini, demonstrating that upon de-regulation of cell–cell adhesions, oncogenic cells that were otherwise contained within the epithelia are translocated into the apical lumen of mammary acini suggesting that active cellular communication between wild-type and oncogenic mutants determines the fate of oncogenic mutants [67]. A recent study has also found that epithelial cells recognize the RasV12-transformed cells via interaction between leukocyte immunoglobulin-like receptor B3 (LILRB3) expressed on non-transformed epithelial cells and major histocompatibility complex class I (MHC class I) expressed on transformed cells. This MHC class I–LILRB3 interaction generates the required mechanical force for extruding the precancerous cells from the epithelia via SHP2-ROCK2 pathway mediated by filamin accumulation in [68]. Mutants defected in cell polarity, such as scribble, lgl, and dlg, also undergo apoptotic death in the presence of wild-type cells, and studies show that their elimination can be mediated by both mechanical signals and by ligand–receptor interactions [8,10,14,54,69]. Nevertheless, common to all these systems is that transformed cells alone survive, over-proliferate, and develop into large masses of tumours, emphasizing that elimination of these mutants is contextual and the presence of surrounding normal cells is necessary for eliminating them. Excitingly, similar defence mechanism has recently been reported in mouse skin—using intra-vital imaging of entire skin tissue. It was shown that the growth of mutated cells (Wnt/B-catenin/Hras mutations) and of structurally deformed cells is shown to be suppressed by surrounding normal skin in a differentiation-dependent manner [70]. Wild-type cells actively eliminate these aberrant cells, revealing unanticipated plasticity of the adult skin epithelium when faced with mutational and non-mutational insults.

28.4.3. Aneuploid cells are extruded out to prevent cancer initiation

Implications of cell competition in EDAC can also be appreciated in the classical Minute competitive environments when cells defective in ribosomal protein genes interact with wild-type cells. Since genes expressing ribosomal proteins are scattered across the chromosome, genomic rearrangements and aneuploidy will lead to ribosomal imbalance that shall induce a loser-like state. Apoptosis of these Minute mutants mainly occur by JNK pathway activation in loser cells mediated by ligand–receptor interactions. Since more than 90% of advanced stage tumours are shown to be aneuploid [71], removal of aneuploid cells by cell competition is suggestive to be tumour preventive [71,72]. This was demonstrated in experiments on mouse haematopoietic stem cells showing that the fitness of induced aneuploid cells is reduced over time and hence tumour formation was prevented [72]. It is now increasingly realized that signals from

the tissue microenvironment strongly dictate the strength and direction of EDAC [73]. Such a crosstalk between extra-cellular matrix and cell–cell interactions is yet to be explored in the context of cell competition and is important for a comprehensive understanding of 'epithelial defence against cancer'. Nevertheless, it can be stated that as long as the microenvironment is tumour suppressing and the architecture of tissue homeostasis is controlled, cell competition appears to be a clean way of providing defence against cancer without the involvement of the immune system.

28.4.3.1. Role of tissue microenvironment in cell competition

How do stimuli, such as smoking, high-fat diet, tissue fibrosis and others, tune the tissue microenvironment to break the tissue homeostasis and set off cancer in the first place? Can the tissue microenvironment shape the strength and direction of cell competition? Can we modulate microenvironment of our tissue to boost EDAC and prevent cancer progression? Excitingly initial studies focusing on this aspect suggest that tissue microenvironment influences the fitness of transformed cells already from the initial genetic hit and play a major role in determining the fate of tumour. For example, study in MDCK cells show that extrusion of RasV12-transformed cells from the tissue is impaired when the substrate is fibrotic [64], indicating that fibrosis may act as a pre-cancer condition by supporting the growth of oncogenic mutants. Additionally, recent mice *in vivo* study demonstrates how inflammation and metabolic changes caused by high-fat diet change the direction of cell competition to facilitate tumour formation [Figure 28.3]. Mice fed with high-fat diet demonstrate impaired defence against cancer in their intestinal and pancreatic epithelia, leading to tumour formation [66]. RasV12-transformed cells, which are normally eliminated due to their low mitochondrial potential in the presence of wild type, cannot be eliminated upon high-fat diet treatment—excess fatty acids are converted to acetyl-CoA, restoring mitochondrial potential in RasV12-transformed cells and protecting them from elimination. EDAC is further attenuated when fatty acid diet that induces chronic inflammation is introduced and is partially rescued upon treatment with anti-inflammatory drugs such as aspirin [66]. Inflammation likely promotes tumour growth by recruiting tumour-promoting cytokines [74]; however, it is not clear how they impact on cell competition to attenuate EDAC [Figure 28.3]. Studies now reveal the involvement of Toll-related receptors (TRR) in initiating death in loser cells when cytokine ligand Spatzle released from neighbouring cells bind to TRR—do inflammatory cytokines released in tumour microenvironment participate in similar interactions to eliminate wild-type cells? While this needs to be tested, recent studies indicate involvement of reactive oxygen species in cell competition and show that activation of oxidative stress pathway turn cells into losers [75]. Since tumour microenvironments are well established to have large amounts of reactive oxygen species [76,77], it is tentative to speculate that by simply activating oxidative stress response in the host tissue, cancer cells actively kill the host cells. Reciprocally, activation of oxidative stress in tumour cells could be an Achille's heel to be exploited for therapy. It is therefore becomes clearly understood that cancer evolution is governed by both occurrence of oncogenic mutations and selection of mutants. While it is undebatable that mutations caused by individual's genetic susceptibility and regulated by environmental factors initiate the phenotypic diversification of tumours, much less attention has been paid to the selection for mutants. In a tumour-promoting microenvironment, it seems that the selection of tumour cells is mediated by cell competition through which tumour cells actively kill their neighbours and establish their territory in the host.

28.4.3.2. Cell competition in cancer therapeutics

Concepts of cell competition demonstrate that both primary and metastasizing tumours actively kill the host tissue cells to expand, indicating that therapies that negatively interfere with the ability of tumour cells to kill the host tissue, or those that revert cell competition in the favour of host tissue, could be sufficient to contain the growth of tumour. This kind of approach is especially relevant for

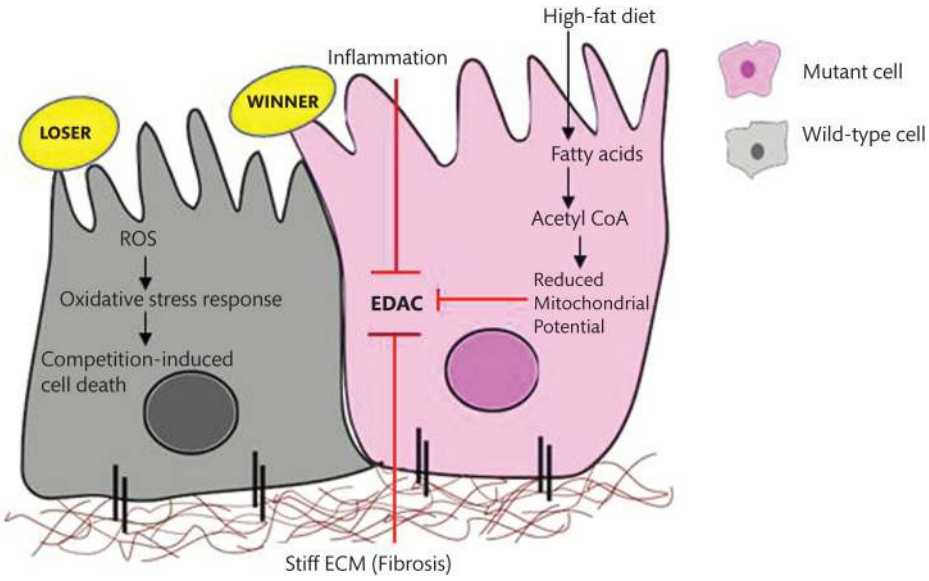


Figure 28.3. Tissue microenvironment modulates cell competition. Tumour-promoting microenvironment, such as oxidative stress, high-fat diet, or tissue fibrosis, leads to impaired EDAC and subsequently allows for tumour growth and expansion.

patients with familial polyposis, whose genetic predisposition gives them a virtually 100% chance of developing cancer. Initial efforts revealed intriguing findings, e.g. Drosophila intestinal tumours, caused by mutations in APC genes, compete with and induce death in surrounding cells [44]. Importantly, tumour growth was dramatically reduced upon preventing cell competition by apoptosis inhibitors. Furthermore, partial inhibition of hippo signalling had a remarkable effect blocking tumour growth [Figure 28.4]. Considering that Hippo signalling is a tumour suppressor, this is a rather counter-intuitive result and is shown to be mainly due to the protective effect of hippo signalling on the host tissue, thus suggesting the rather provocative hypothesis that low-dose hippo signalling inhibition might actually contain tumour growth [44]. On a similar line of thinking, targeting regulators of STAT, e.g. protein inhibitor of activated STAT (PIAS) may also be protective towards host tissue—since persistent activation of STAT is a hallmark of various human cancer [78], giving

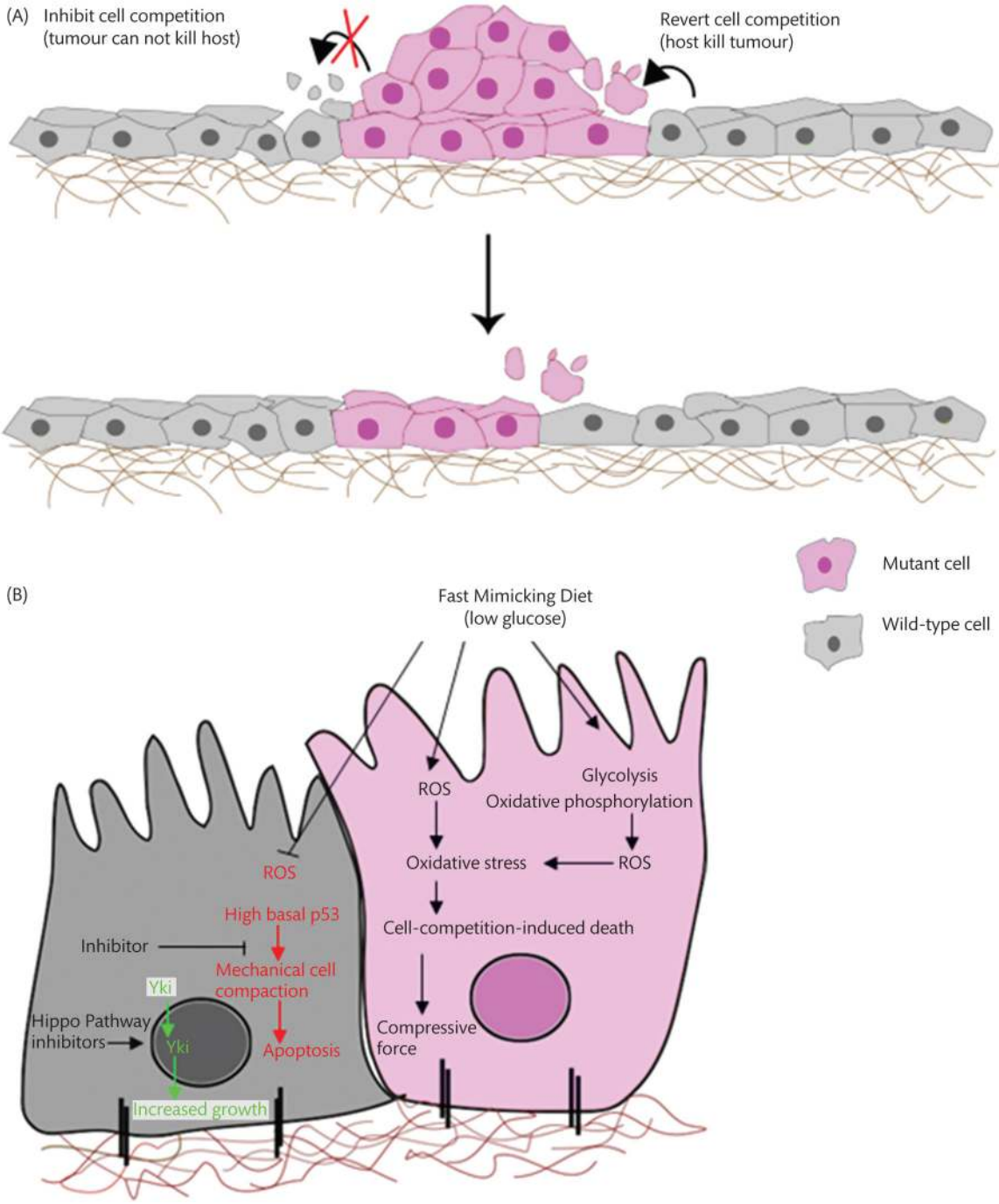


Figure 28.4. Exploiting cell competition as an anticancer therapy. (A) Modulating the strength and direction of cell competition could be harnessed as a therapeutic strategy to contain tumour growth and progression. (B) On the basis of current knowledge of cell competition, different therapeutic strategies could be designed that could contain or delay tumour growth either by augmenting tumour-suppressive cell competition or by curtailing the ability of tumour cells to kill host cells.

tumour cells a status of super-competitor [51], low-dose inhibition of PIAS rendering slight activation of STAT might protect the host cells from being outcompeted by the tumour. Inhibiting apoptosis of p53 driven tumours also seem to restrict tumour growth, suggesting again that protecting wild-type cells from cell competition-mediated death can stop tumours to grow and expand. Indeed, apoptosis inhibition would be a rather crude and pleiotropic treatment, but identifying downstream effectors of cell competition could lead to the identification of novel anticancer drug targets. For example, by attacking the pathway that makes wild-type cells with elevated p53, hypersensitive to compaction in the presence of p53 mutants [54], we might be able to stop p53 mutants from killing the wild-type cells and prevent or revert tumour growth [Figure 28.4]. Additionally, reduced mitochondrial potential and increased oxidative stress response both are shown to potentially turn cells into losers in a competitive environment [75] suggesting an approach that target these basic bioenergetic pathways might induce death in tumour cells. Cancer cells turn their metabolism into well-known 'Warburg effect', i.e. by reducing oxidative phosphorylation and shifting their metabolic pathway to directly ferment glucose into lactate (aerobic glycolysis) [79], they cope with the harsh tissue microenvironment and regulate their high-paced proliferation. Anti-Warburg strategies, such as fast mimicking diets (FMDs), that reduce glucose availability and impair metabolism of cancer cells have recently been suggested in combination with other therapies to enhance treatment efficiency [Figure 28.4]. Interestingly, lack of glucose during FMD shifts metabolism in cancer cells towards oxidative phosphorylation, leading to increased ROS production and high oxidative stress response [80]. Since increased oxidative stress response is linked with loser status [75], excitingly, such a strategy enhances the death of tumour cells by also involving surrounding healthy cells. In addition, neighbouring cells further potentiate this effect by themselves relishing on antioxidant effects due to their protective metabolic shift in the presence of FMD [81,82]. Notably, it is likely that cancer cells can acquire resistance to FMD by circumventing metabolic changes; it therefore becomes more important to use this approach in combination with increases in fitness of wild-type cells, thereby enhancing their competitive ability to fight against the tumour cells. Meanwhile, another important aspect that requires attention is the high toxicity of many existing cancer treatments that cause widespread cell death in host tissue and therefore, from a cell-competition perspective, could have an undesired effect of further promoting tumorigenesis.

28.5. Conclusions

In conclusion, changes in the tumour microenvironment may influence the strength and direction of cell competition (i.e. host kills tumour versus tumour kills host) and hence the likelihood of a cancer to grow. On the basis of recent provocative findings, we suggest that modulating cell competition could provide an orthogonal therapeutic strategy to contain the growth and spread of tumours. With the acknowledgement of many remaining unknowns, we believe that the idea of harnessing cell competition in cancer holds huge promise towards exploration into new routes of therapeutics because the fundamental strength of this concept, 'protecting host tissue cells from being killed by tumour' [Figure 28.4], is undeniable. Philosophically, by employing such an approach, we could potentially revert the Darwinian advantage that the tumour has over the host tissue, and by combining such an approach with an ideal tissue microenvironment, cell competition brings hope not only towards cancer treatment but also towards cancer prevention.

REFERENCES

1. Vishwakarma M, Piddini E. Outcompeting cancer. Nat Rev Cancer. 2020 Mar;**20**(3):187–98.
2. Kajita M, Fujita Y. EDAC: epithelial defence against cancer--cell competition between normal and transformed epithelial cells in mammals. J Biochem (Tokyo). 2015 Jul 1;**158**(1):15–23.
3. Suijkerbuijk SJE, Kolahgar G, Kucinski I, Piddini E. Cell competition drives the growth of intestinal adenomas in drosophila. Curr Biol. 2016 Feb;**26**(4):428–38.
4. Morata G, Ripoll P. Minutes: mutants of drosophila autonomously affecting cell division rate. Dev Biol. 1975 Feb;**42**(2):211–21.
5. Simpson P. Parameters of cell competition in the compartments of the wing disc of Drosophila. Dev Biol. 1979 Mar;**69**(1):182–93.
6. Clavería C, Giovinazzo G, Sierra R, Torres M. Myc-driven endogenous cell competition in the early mammalian embryo. Nature. 2013 Aug;**500**(7460):39–44.
7. Martins VC, Busch K, Juraeva D, Blum C, Ludwig C, Rasche V, et al. Cell competition is a tumour suppressor mechanism in the thymus. Nature. 2014 May 22;**509**(7501):465–70.
8. Neto-Silva RM, de Beco S, Johnston LA. Evidence for a growth-stabilizing regulatory feedback mechanism between Myc and Yorkie, the Drosophila homolog of Yap. Dev Cell. 2010 Oct;**19**(4):507–20.
9. Chen CL, Schroeder MC, Kango-Singh M, Tao C, Halder G. Tumor suppression by cell competition through regulation of the Hippo pathway. Proc Natl Acad Sci. 2012 Jan 10;**109**(2):484–9.
10. Menéndez J, Pérez-Garijo A, Calleja M, Morata G. A tumor-suppressing mechanism in *Drosophila* involving cell competition and the Hippo pathway. Proc Natl Acad Sci. 2010 Aug 17;**107**(33):14651–6.
11. Tamori Y, Bialucha CU, Tian AG, Kajita M, Huang YC, Norman M, et al. Involvement of Lgl and Mahjong/VprBP in cell competition. Baker NE, editor. PLoS Biol. 2010 Jul 13;**8**(7):e1000422.
12. Igaki T, Pagliarini RA, Xu T. Loss of cell polarity drives tumor growth and invasion through JNK activation in Drosophila. Curr Biol. 2006 Jun;**16**(11):1139–46.
13. Brumby AM. Scribble mutants cooperate with oncogenic Ras or Notch to cause neoplastic overgrowth in Drosophila. EMBO J. 2003 Nov 3;**22**(21):5769–79.
14. Norman M, Wisniewska KA, Lawrenson K, Garcia-Miranda P, Tada M, Kajita M, et al. Loss of Scribble causes cell competition in mammalian cells. J Cell Sci. 2012 Jan 1;**125**(1):59–66.
15. Hogan C, Dupré-Crochet S, Norman M, Kajita M, Zimmermann C, Pelling AE, et al. Characterization of the interface between normal and transformed epithelial cells. Nat Cell Biol. 2009 Apr;**11**(4):460–7.
16. Kajita M, Hogan C, Harris AR, Dupre-Crochet S, Itasaki N, Kawakami K, et al. Interaction with surrounding normal epithelial cells influences signalling pathways and behaviour of Src- transformed cells. J Cell Sci. 2010 Jan 15;**123**(2):171–80.
17. Takagi M, Ikegawa M, Shimada T, Ishikawa S, Kajita M, Maruyama T, et al. Accumulation of the myosin-II-spectrin complex plays a positive role in apical extrusion of Src-transformed epithelial cells. Genes Cells. 2018 Nov;**23**(11):974–81.

18. Ballesteros-Arias L, Saavedra V, Morata G. Cell competition may function either as tumour- suppressing or as tumour-stimulating factor in Drosophila. Oncogene. 2014 Aug 28;**33**(35):4377–84.
19. Zhang G, Xie Y, Zhou Y, Xiang C, Chen L, Zhang C, et al. p53 pathway is involved in cell competition during mouse embryogenesis. Proc Natl Acad Sci. 2017 Jan 17;**114**(3):498–503.
20. Bowling S, Di Gregorio A, Sancho M, Pozzi S, Aarts M, Signore M, et al. P53 and mTOR signalling determine fitness selection through cell competition during early mouse embryonic development. Nat Commun. 2018 May 2;**9**(1):1763.
21. Abrams JM. Competition and compensation. Cell. 2002 Aug;**110**(4):403–6.
22. Milán M. Survival of the fittest: cell competition in the *Drosophila* wing. EMBO Rep. 2002 Aug;**3**(8):724–5.
23. Moreno E, Basler K. dMyc transforms cells into super-competitors. Cell. 2004 Apr 2;**117**(1):117–29.
24. de la Cova C, Abril M, Bellosta P, Gallant P, Johnston LA. Drosophila Myc regulates organ size by inducing cell competition. Cell. 2004 Apr;**117**(1):107–16.
25. Johnston LA. Socializing with MYC: Cell competition in development and as a model for premalignant cancer. Cold Spring Harb Perspect Med. 2014 Apr 1;**4**(4):a014274.
26. Baker NE, Li W. Cell competition and its possible relation to cancer. Cancer Res. 2008 Jul 15;**68**(14):5505–7.
27. Moreno E. Is cell competition relevant to cancer? Nat Rev Cancer. 2008 Feb;**8**(2):141–7.
28. Di Gregorio A, Bowling S, Rodriguez TA. Cell competition and its role in the regulation of cell fitness from development to cancer. Dev Cell. 2016 Sep;**38**(6):621–34.
29. Grifoni D, Bellosta P. Drosophila Myc: a master regulator of cellular performance. Biochim Biophys Acta BBA–Gene Regul Mech. 2015 May;**1849**(5):570–81.
30. Johnston LA, Prober DA, Edgar BA, Eisenman RN, Gallant P. Drosophila myc regulates cellular growth during development. Cell. 1999 Sep;**98**(6):779–90.
31. Dang CV. c-Myc target genes involved in cell growth, apoptosis, and metabolism. Mol Cell Biol. 1999 Jan;**19**(1):1–11.
32. de la Cova C, Johnston LA. Myc in model organisms: a view from the flyroom. Semin Cancer Biol. 2006 Aug;**16**(4):303–12.
33. Pelengaris S, Khan M, Evan G. c-MYC: more than just a matter of life and death. Nat Rev Cancer. 2002 Oct;**2**(10):764–76.
34. Sancho M, Di-Gregorio A, George N, Pozzi S, Sánchez JM, Pernaute B, et al. Competitive interactions eliminate unfit embryonic stem cells at the onset of differentiation. Dev Cell. 2013 Jul;**26**(1):19–30.
35. Villa del Campo C, Clavería C, Sierra R, Torres M. Cell competition promotes phenotypically silent cardiomyocyte replacement in the mammalian heart. Cell Rep. 2014 Sep;**8**(6):1741–51.
36. Wu DC, Johnston LA. Control of wing size and proportions by Drosophila Myc. Genetics. 2010 Jan 1;**184**(1):199–211.
37. Tapon N, Harvey KF, Bell DW, Wahrer DCR, Schiripo TA, Haber DA, et al. Salvador promotes both cell cycle exit and apoptosis in Drosophila and is mutated in human cancer cell lines. Cell. 2002 Aug;**110**(4):467–78.
38. Pan D. Hippo signaling in organ size control. Genes Dev. 2007 Apr 15;**21**(8):886–97.
39. Tyler DM, Li W, Zhuo N, Pellock B, Baker NE. Genes affecting cell competition in Drosophila. Genetics. 2007 Feb 1;175(2):643–57.
40. Ziosi M, Baena-López LA, Grifoni D, Froldi F, Pession A, Garoia F, et al. dMyc functions downstream of Yorkie to promote the supercompetitive behavior of hippo pathway mutant cells. Barsh GS, editor. PLoS Genet. 2010 Sep 23;**6**(9):e1001140.
41. Harvey KF, Zhang X, Thomas DM. The Hippo pathway and human cancer. Nat Rev Cancer. 2013 Apr;**13**(4):246–57.
42. Nesbit CE, Tersak JM, Prochownik EV. MYC oncogenes and human neoplastic disease. Oncogene. 1999 May 13;**18**(19):3004–16.
43. Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. *EGFR* mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science. 2004 Jun 4;**304**(5676):1497–500.
44. Eichenlaub T, Cohen SM, Herranz H. Cell competition drives the formation of metastatic tumors in a drosophila model of epithelial tumor formation. Curr Biol. 2016 Feb;**26**(4):419–27.
45. Irvine KD. Integration of intercellular signaling through the Hippo pathway. Semin Cell Dev Biol. 2012 Sep;**23**(7):812–7.
46. White BD, Chien AJ, Dawson DW. Dysregulation of Wnt/β-Catenin signaling in gastrointestinal cancers. Gastroenterology. 2012 Feb;**142**(2):219–32.
47. Bellam N, Pasche B. TGF-β signaling alterations and colon cancer. In: Pasche B, editor. Cancer Genetics [Internet]. Boston, MA: Springer US; 2010 [cited 2023 Feb 24]. p. 85–103. (Cancer Treatment and Research; vol. 155). Available from: https://link.springer.com/10.1007/978-1-4419- 6033-7_5
48. Vincent JP, Kolahgar G, Gagliardi M, Piddini E. Steep differences in wingless signaling trigger Myc-independent competitive cell interactions. Dev Cell. 2011 Aug;**21**(2):366–74.
49. Ikeda S. Axin, a negative regulator of the Wnt signaling pathway, forms a complex with GSK-3beta and beta -catenin and promotes GSK-3beta-dependent phosphorylation of beta -catenin. EMBO J. 1998 Mar 2;**17**(5):1371–84.
50. Akong K, Grevengoed EE, Price MH, McCartney BM, Hayden MA, DeNofrio JC, et al. Drosophila APC2 and APC1 play overlapping roles in wingless signaling in the embryo and imaginal discs. Dev Biol. 2002 Oct;**250**(1):91–100.
51. Rodrigues AB, Zoranovic T, Ayala-Camargo A, Grewal S, Reyes-Robles T, Krasny M, et al. Activated STAT regulates growth and induces competitive interactions independently of Myc, Yorkie, Wingless and ribosome biogenesis. Development. 2012 Nov 1;**139**(21):4051–61.
52. Muller PAJ, Vousden KH. Mutant p53 in cancer: new functions and therapeutic opportunities. Cancer Cell. 2014 Mar;**25**(3):304–17.
53. Walther A, Johnstone E, Swanton C, Midgley R, Tomlinson I, Kerr D. Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer. 2009 Jul;**9**(7):489–99.
54. Wagstaff L, Goschorska M, Kozyrska K, Duclos G, Kucinski I, Chessel A, et al. Mechanical cell competition kills cells via induction of lethal p53 levels. Nat Commun. 2016 Apr 25;**7**(1):11373.
55. Watanabe H, Ishibashi K, Mano H, Kitamoto S, Sato N, Hoshiba K, et al. Mutant p53-expressing cells undergo necroptosis via cell competition with the neighboring normal epithelial cells. Cell Rep. 2018 Jun;**23**(13):3721–9.
56. Nowell PC. The clonal evolution of tumor cell populations: acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression. Science. 1976 Oct;**194**(4260):23–8.
57. Laurie CC, Laurie CA, Rice K, Doheny KF, Zelnick LR, McHugh CP, et al. Detectable clonal mosaicism from birth to old age and its relationship to cancer. Nat Genet. 2012 Jun;**44**(6):642–50.

58. Bissell MJ, Hines WC. Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression. Nat Med. 2011 Mar;**17**(3):320–9.
59. Dolberg DS, Bissell MJ. Inability of Rous sarcoma virus to cause sarcomas in the avian embryo. Nature. 1984 Jun;**309**(5968):552–6.
60. Grieve AG, Rabouille C. Extracellular cleavage of E-cadherin promotes epithelial cell extrusion. J Cell Sci. 2014 Jan 1;jcs.147926.
61. Chiba T, Ishihara E, Miyamura N, Narumi R, Kajita M, Fujita Y, et al. MDCK cells expressing constitutively active Yes-associated protein (YAP) undergo apical extrusion depending on neighboring cell status. Sci Rep. 2016 Jun 21;**6**(1):28383.
62. Tanimura N, Fujita Y. Epithelial defense against cancer (EDAC). Semin Cancer Biol. 2020 Jun;**63**:44–8.
63. Kajita M, Sugimura K, Ohoka A, Burden J, Suganuma H, Ikegawa M, et al. Filamin acts as a key regulator in epithelial defence against transformed cells. Nat Commun. 2014 Jul 31;**5**(1):4428.
64. Pothapragada SP, Gupta P, Mukherjee S, Das T. Matrix mechanics regulates epithelial defence against cancer by tuning dynamic localization of filamin. Nat Commun. 2022 Jan 11;**13**(1):218.
65. Kon S, Ishibashi K, Katoh H, Kitamoto S, Shirai T, Tanaka S, et al. Cell competition with normal epithelial cells promotes apical extrusion of transformed cells through metabolic changes. Nat Cell Biol. 2017 May;**19**(5):530–41.
66. Sasaki A, Nagatake T, Egami R, Gu G, Takigawa I, Ikeda W, et al. Obesity suppresses cell- competition-mediated apical elimination of RasV12-transformed cells from epithelial tissues. Cell Rep. 2018 Apr;**23**(4):974–82.
67. Leung CT, Brugge JS. Outgrowth of single oncogene-expressing cells from suppressive epithelial environments. Nature. 2012 Feb;**482**(7385):410–3.
68. Ayukawa S, Kamoshita N, Nakayama J, Teramoto R, Pishesha N, Ohba K, et al. Epithelial cells remove precancerous cells by cell competition via MHC class I–LILRB3 interaction. Nat Immunol. 2021 Nov;**22**(11):1391–402.
69. Yamamoto M, Ohsawa S, Kunimasa K, Igaki T. The ligand Sas and its receptor PTP10D drive tumour-suppressive cell competition. Nature. 2017 Feb 9;**542**(7640):246–50.
70. Brown S, Pineda CM, Xin T, Boucher J, Suozzi KC, Park S, et al. Correction of aberrant growth preserves tissue homeostasis. Nature. 2017 Aug;**548**(7667):334–7.
71. Rajagopalan H, Lengauer C. Aneuploidy and cancer. Nature. 2004 Nov;**432**(7015):338–41.
72. Pfau SJ, Silberman RE, Knouse KA, Amon A. Aneuploidy impairs hematopoietic stem cell fitness and is selected against in regenerating tissues in vivo. Genes Dev. 2016 Jun 15;**30**(12):1395–408.
73. Ohoka A, Kajita M, Ikenouchi J, Yako Y, Kitamoto S, Kon S, et al. EPLIN is a crucial regulator for extrusion of RasV12-transformed cells. J Cell Sci. 2015 Jan 1;jcs.163113.
74. Mantovani A, Allavena P, Sica A, Balkwill F. Cancer-related inflammation. Nature. 2008 Jul 24;**454**(7203):436–44.
75. Kucinski I, Dinan M, Kolahgar G, Piddini E. Chronic activation of JNK JAK/STAT and oxidative stress signalling causes the loser cell status. Nat Commun. 2017 Jul 26;**8**(1):136.
76. Radisky DC, Levy DD, Littlepage LE, Liu H, Nelson CM, Fata JE, et al. Rac1b and reactive oxygen species mediate MMP-3-induced EMT and genomic instability. Nature. 2005 Jul;**436**(7047):123–7.
77. Sosa V, Moliné T, Somoza R, Paciucci R, Kondoh H, LLeonart ME. Oxidative stress and cancer: an overview. Ageing Res Rev. 2013 Jan;**12**(1):376–90.
78. Groner B, von Manstein V. Jak Stat signaling and cancer: opportunities, benefits and side effects of targeted inhibition. Mol Cell Endocrinol. 2017 Aug;**451**:1–14.
79. Liberti MV, Locasale JW. The Warburg effect: how does it benefit cancer cells? Trends Biochem Sci. 2016 Mar;**41**(3):211–8.
80. Nencioni A, Caffa I, Cortellino S, Longo VD. Fasting and cancer: molecular mechanisms and clinical application. Nat Rev Cancer. 2018 Nov;**18**(11):707–19.
81. van der Horst A, Burgering BMT. Stressing the role of FoxO proteins in lifespan and disease. Nat Rev Mol Cell Biol. 2007 Jun;**8**(6):440–50.
82. Cheng Z, Guo S, Copps K, Dong X, Kollipara R, Rodgers JT, et al. Foxo1 integrates insulin signaling with mitochondrial function in the liver. Nat Med. 2009 Nov;**15**(11):1307–1.

SECTION 7

Translational mathematical oncology

29

Modelling cell population dynamics during chimeric antigen receptor T-cell therapy

Philipp M. Altrock, Guranda Chitadze, Arne Traulsen, and Frederick L. Locke

29.1. The use of mathematical modelling of cancer and the immune system

Mathematical models of biological systems, such as cancer, can integrate mechanistic assumptions with prior knowledge and new data, e.g. building on the theoretical frameworks of dynamical systems theory and applied probability theory. Many such models are in the form of agent-based modelling, where a general analysis is no longer feasible and large-scale computer simulations are necessary [1]. Much like experimental model systems, mathematical models of biological systems spread over multiple scales, including molecules, cells, populations of cells, tissues and organs, organisms, populations of organisms, and population systems. Mathematical models of cell populations are of particular significance to modelling the dynamics of complex tissue phenomena, such as haematopoiesis, immune system dynamics, inflammation or malignant tumours, anti-cancer treatment, and cancer evolution [2–4]. These theoretical models can quantify how a biological system works. They allow us to abstract biological phenomena to a desired degree, explore the influence of certain mechanistic or technical assumptions on a system's kinetics or dynamics, train model parameters, and enable validation or testing of models in the light of available experimental or clinical data [5,6]. Note that there is a distinction between kinetics and dynamics. Kinetics describes the system over time without knowing the underlying forces or mechanisms. In contrast, the term dynamics implies that some knowledge or assumptions cause the observed temporal behaviour in a mechanistic way, e.g. within a theoretical model. Finally, theoretical models allow comparison of underlying theories.

One example of comparing underlying theories is the analogy between predator–prey systems in theoretical ecology and immune-cell–target-cell systems in theoretical immunology [7–9]. The adaptive immune system provides an effective defence against infectious agents, such as viruses, bacteria, and fungi [10–12], and plays a major role in discriminating self from non-self [13]. It has long been recognized that the immune status changes with cancer development [14,15]. While there are commonalities in how adaptive immune cells target and eliminate infected or cancer cells, there are also key differences. On the one hand, stimulated immune cells act as predators that identify, bind to, and kill target cells. On the other hand, the life cycles, evolutionary dynamics, and collective dynamics of predators and immune cells, or of prey and target cells, can be very distinct [8]. Thus, existing predator–prey dynamics rarely directly describe the dynamics of cancer–immune-system interactions (**Figure 29.1**A), and novel or adapted theories and models are needed.

A holistic understanding of the evolutionary dynamics of cancer in the context of an ageing immune system is still in its early stages. The need for deeper understanding is fuelled by the recent rise and diversification of immunotherapies and the increase in the fundamental importance of immuno-oncology. Mathematical modelling has succeeded in contextualizing and abstracting some existing knowledge in logical frameworks and is thus key to formulating novel hypotheses. It has been understood that cancer follows a non-linear evolutionary process that mainly occurs during somatic evolution in multicellular organisms [16,17]. The importance of immune system interactions has increasingly been recognized in cancer and cellular therapy [18]. During organismic development and somatic evolution, tissue generation and maintenance are not error-free, but stochastic events lead to potentially heritable aberrations that can manifest in cell populations [19]. Hereby, multiple processes impose strong negative selection against aberrant cells that naturally emerge due to genetic or epigenetic alterations. The immune system provides tight control mechanisms; innate and adaptive immune systems play a pivotal role in cancer elimination. In addition, spatial tissue structure and hierarchical tissue organization ensure tissue function and serve as checks and balances on somatic abnormalities

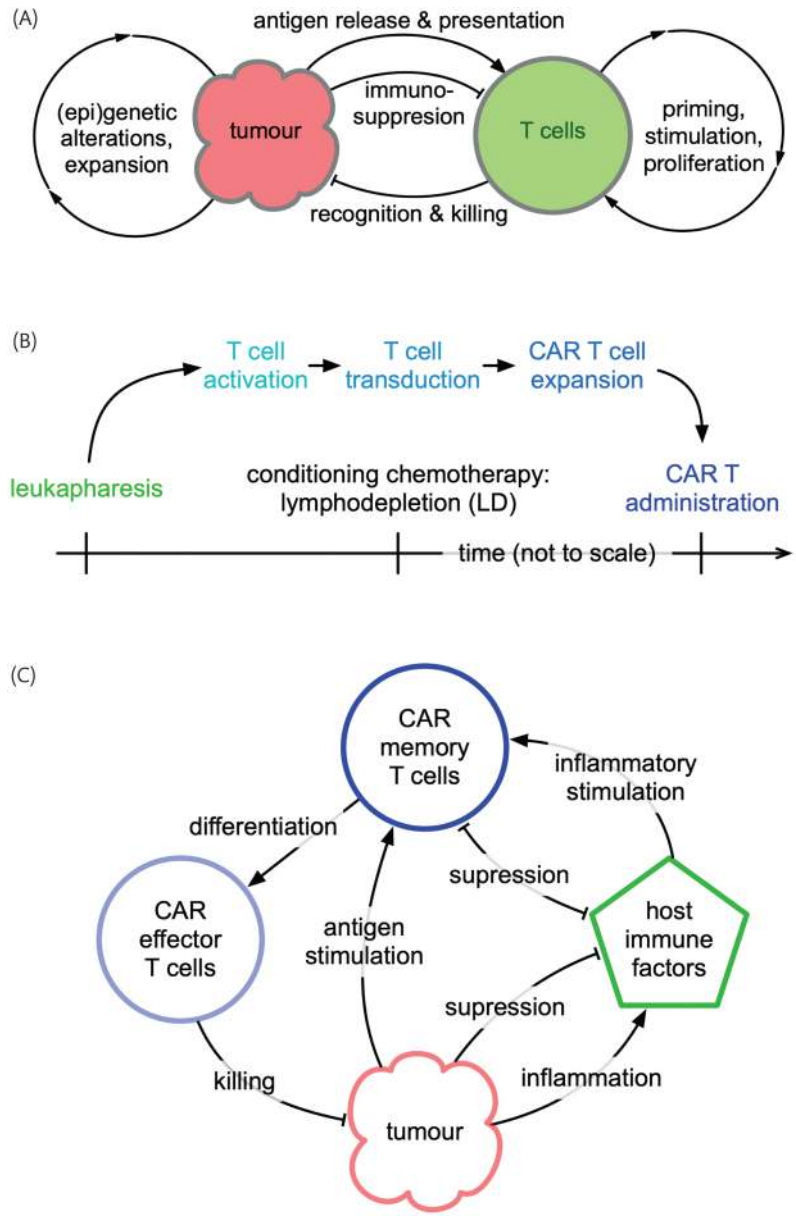


Figure 29.1. From T-cells to CAR T-cells to fight tumours. (A) Interactions of tumours and T-cells of the host immune system. (B) CAR T-cell timelines of the patient (bottom) and *ex vivo* manufacturing (top). (C) An example conceptual modelling framework involving CAR T-cell phenotypes (central memory, called 'memory' and effector memory, called 'effector'), tumour, and host immune factors (e.g. normal host lymphocytes) [38].

such as cancer [1]. As a result, cancer is a complex dynamical system that evolves through multiple bottlenecks and adaptations, resulting in substantial within- and across-cancer heterogeneity [20,21]. This heterogeneity can be decisive for cancer evolution and treatment response [22–24]. Current efforts in mathematical modelling seek to quantify the underlying non-linear and stochastic processes of cancer evolution and progression and their roles at different disease stages in the light of a temporally changing complex immune environment. In this complex ecological-evolutionary context, we attempt to understand the failures and successes of the rapidly developing field of immunotherapy.

Chimeric antigen receptor (CAR) T-cell therapy has greatly advanced personalized anti-cancer immunotherapy [25]. In this adoptive immune cell therapy approach, a patient's own (or a donor's) T-cells are extracted by leukapheresis and genetically engineered to express a synthetic receptor that binds a tumour antigen. After this manufacturing process, the CAR T-cells are expanded and infused back into the patient. They can then attack and destroy cancer cells (**Figure 29.1** B). Some of the most remarkable clinical improvements have been observed in the setting of CAR T-cell therapy of B-cell malignancies, such as acute lymphoblastic leukaemia and diffuse large B-cell lymphoma [26–31]. CAR T-cells 'are one of the first successful examples of synthetic biology and personalized cellular cancer therapy to become commercially available' [25].

Complex ecological processes drive cancer dynamics; the immune system heavily influences this ecology. At the cellular level, variation is provided by phenotypic plasticity and genotypic

diversification [32,33]. The selection emerging from this variation can be cell-autonomous (cell-intrinsic factors) [16] or non-cell-autonomous [34]. In the case of autonomous mechanisms, fitter subtypes experience positive selection independent of the (changing) composition of the diverse cancer cell population and potentially independent of the complex environment [35,36]. Non-autonomous mechanisms imply variable selective forces depending on the population composition and the environment. These mechanisms can emerge from cell-to-cell interactions, paracrine (short distance), or endocrine (long distance) interactions (via soluble factors) not only among cell cancer types but also between cancer and stromal (fibroblasts, myeloid-derived, or lymphoid-derived adaptive cells) [37]. The heterogeneous responses to adoptive cancer immunotherapies must be understood in the context of patient-specific complex tumour ecosystems. This 'ecology' gives rise to complex interactions that must be accounted for to model the therapy-disease dynamics (**Figure 29.1** C). The ecological predator–prey dynamics that drive tumour killing, CAR T expansion, and exhaustion occur in the context of the complex tumour-immune microenvironment.

29.2. Pharmacokinetic and pharmacodynamic modelling of adoptive T-cell therapies

Pharmacometrics traditionally relies on the principle of mass balance [39,40], in which the output of a drug is a linear function of the input (dose) and accumulation. Adoptive CAR T kinetics do not necessarily follow this principle; increasing the dose of the cellular therapy does not necessarily increase the exposure to the drug. Pharmacokinetics (PK) and pharmacodynamics (PD) are the traditional disciplines that are concerned with the kinetics of drugs or other substances once they enter the body (e.g. by absorption, distribution, and metabolic changes) and the dynamics of a drug's actions and effects on or off target (e.g. acting cytostatically or cytotoxically).

PK/PD modelling has been the cornerstone of quantitative modelling to translate pre-clinical evidence into clinical practice and characterize drug dosing, efficacy, and safety based on available biomarkers [41]. For adoptive immune cell therapies, the drug is a diverse set of cells that has the potential to proliferate, differentiate, and expand. Classical PK/PD descriptions cannot capture the mechanisms underlying these processes because these processes do not directly follow the classical PK schematics of absorption, distribution, metabolism, and excretion [42,43]. At the same time, the PD of adoptive T-cell therapies involves immune stimulation and especially targeted cell killing, e.g. cytotoxicity via perforin [44]. The dynamical processes involved in the action of cellular drugs require new mechanistic modelling approaches of complex immune-cell–target-cell interactions [45] and immune predation [46]. The field of mathematical models based on these observations has rapidly expanded in recent years [5,43]. There is a current sparsity of studies that have been able to compare data from different trials, institutions, or countries and an absence of centre-overarching data standards. Thus, qualitative and quantitative assessments of pan-study model performance and pan-cohort ranking of the predictive capabilities and mechanistic assumptions of specific models result in modelling lagging behind the rate of innovation in molecular drug design.

29.3. Descriptive modelling of CAR T-cell population kinetics in patients

Chaudhury et al. have nicely summarized the mathematical modelling of cellular kinetics and pharmacodynamics up to 2020 [43]. Patient conditioning before CAR generally leads to a temporary decline in the absolute lymphocyte count (**Figure 29.2** A). Near the nadir of this dip in the adaptive host immune cell density, the CAR T-cells are injected, after which they may expand, peak, and contract (**Figure 29.2** B). Consequently, the tumour burden declines due to the CAR's cancer-directed cytotoxicity (**Figure 29.2** C). **Figure 29.3** summarizes a few essential ways to model these non-linear dynamics mathematically, e.g. involving CAR T expansion, r_y, tumour killing by CAR T-cells, k, CAR T exhaustion, e. In the following, we

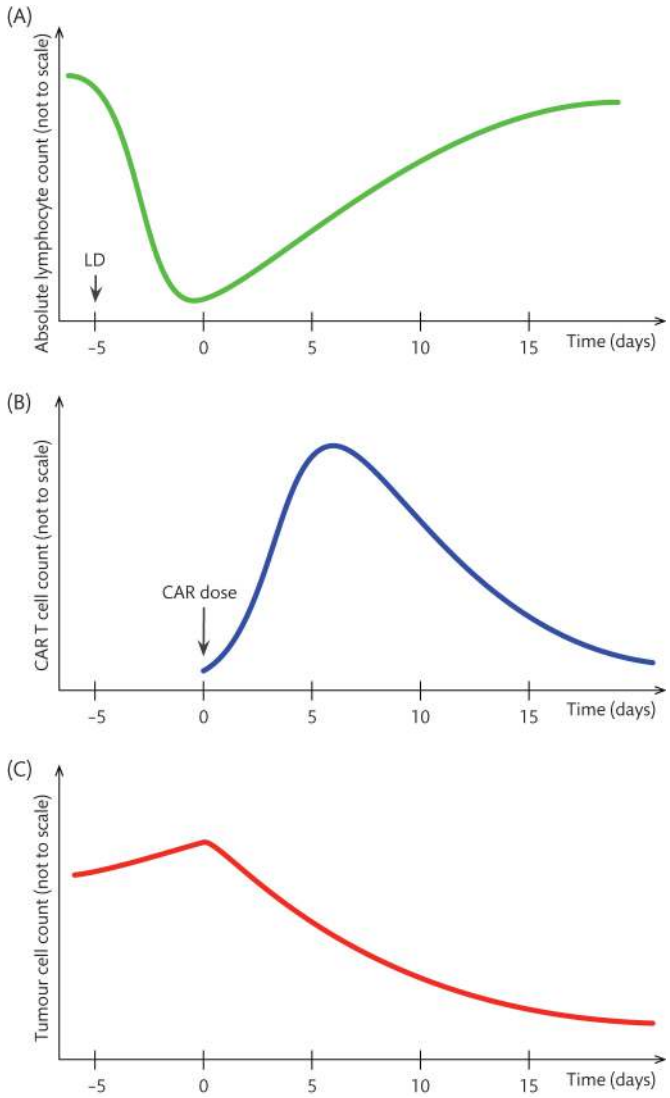


Figure 29.2. Qualitative dynamics. (A) Absolute lymphocyte count as a reaction to pre-CAR T conditioning chemotherapy (lymphodepletion, LD). (B) Typical qualitative dynamics of CAR T with expansion, peak, early fast contraction, and later slow contraction. (C) Assumed tumour dynamics as a consequence of CAR T killing.

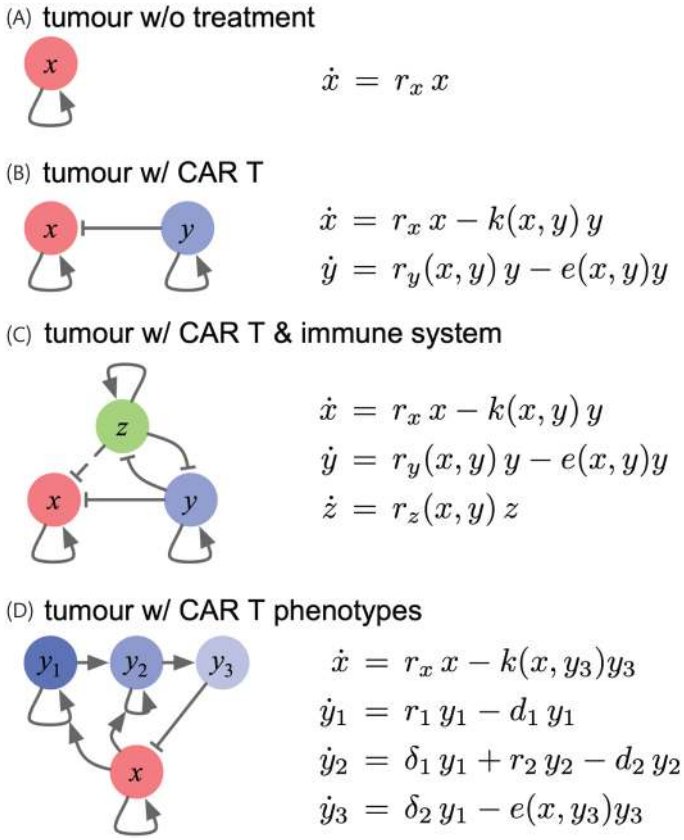


Figure 29.3. Capturing the ecological dynamics of CAR T–tumour interactions in mathematical models. Overview of a few typical modelling approaches with schematic on the left and ordinary differential equations on the right. By $\dot{x}$ we denote the derivative (change of) of the function $x(t)$, $\dot{x} = dx / dt$. (A) Tumour without any treatment. (B) Tumour and CAR T, where tumour killing and CAR T exhaustion can be a non-linear function of cancer and CAR T-cell densities [47,48]. (C) Adding host immune cells to the model of (B). (D) Accounting for CAR T phenotypic heterogeneity [49].

focus on three studies highlighting the descriptive abilities of mathematical models of these processes.

What basic features of cellular drug kinetics are important? One of the first published analyses of the kinetics of CAR T-cells in patients was developed by Stein et al. [50], which focused on the kinetics and expansion of CAR T-cell therapy against B-cell acute lymphatic leukaemia (ALL) and the impact of therapies for treating cytokine release syndrome (CRS). Next to immune effector cell-associated neurotoxicity, CRS not only comprises acute symptoms, such as chills, tiredness, nausea, pain, and fever persisting for multiple days but also manifests with other features of a systemic inflammatory response, such as organ dysfunction that can result from the direct effects of cytokine release. The effects of CRS are preventable or reversible in most patients if caught early [51]. Stein et al. addressed the three characteristic and clinically necessary phases: expansion, initial decline at a fast rate, and terminal decline at a slow rate. These three phases can be explained by effector and memory CAR T-cell compartments with a dynamic conversion between them; expansion is driven by effector CAR T-cell proliferation, initial decline by a phasing out of this expansion due to activation-induced cell death, and slow long-term decline by conversion from effector to memory CAR T-cells. Using a random effects statistical modelling approach based on CAR transgene copy concentration in the periphery (90 patients, over 800 time points in total [31]), this study was able to relate important CAR T-cell product characteristics, such as maximum concentration and area under the curve (AUC) to the relation of rapid contraction due to programmed apoptosis of activated effector CAR T-cells and the gradual decline of CAR T memory cells. Memory CAR T-cells may persist for years or even decades [30], especially in CAR T-cell therapies approved for paediatric and adult relapse refractory ALL. This pioneering modelling study confirmed the importance of peak expansion and AUC in a semi-mechanistic fashion. Yet, it did not show that treating CRS significantly impacts outcomes, but that mathematical modelling based on longitudinal data collected from blood samples is generally possible and useful. Such modelling can, e.g., leverage the dynamic number of CAR gene copies or the estimated number of CAR T-cells per volume in concert with other clinical covariates to characterize and stratify the complex drug kinetics across patients and address open questions related to efficacy, toxicity, and long-term response.

What is the role of the T-cell population structure, often called the CAR T-cell phenotype? A modelling approach of early survival prediction characterized the dynamics of four different CAR T-cell subtypes and tumour burden in 19 B-cell lymphoma patients [49], using a quantitative systems pharmacology approach that combines differential equations with broader statistical modelling involving clinical covariates (characteristics related to patients, therapies, or the CAR T-cell product). This study highlighted the importance of the differential kinetics of the various T-cell phenotypes concerning patient survival (see also Figure 29.3D). Based on a non-linear mixed-effects approach, the 'most important' modelling parameters describing product expansion were translated into a clinical composite score. This score led to a survival prediction based on the drug kinetics and the baseline metabolic tumour volume. In this way, complex characteristics of the cellular drug and intra-patient variability can be recapitulated using retrospective analyses based on a few patients. Such descriptive models rely on longitudinal measurements and, e.g., granular CAR T-cell phenotype information or tumour entity-dependent differences regarding CAR T-cell phenotypes and densities at the periphery and at the (solid) tumour site. Importantly, such modelling frameworks should adhere to standards and quality criteria that are only currently being defined for the rapidly developing field [52].

What is the role of patient-intrinsic fluctuations in the T- and tumour cell kinetics? To address the question of whether patient variability or intrinsic fluctuations due to small cell population size are important in treatment evasion [1,53], Kimmel et al. [47] devised a hybrid compartment modelling approach that switches gears from a deterministic (ordinary differential equation based) mode into a stochastic mode (e.g. numerically solved using the Gillespie algorithm [54]). This approach could simultaneously address the non-linear dynamics of T- and CAR T-cells and the stochastic nature of small tumour dynamics (**Figure 29.4**A and B). The benefit of such hybrid modelling lies in describing a probability of cancer elimination as a probabilistic event without needing a long-term stable state of elimination by persisting CAR T-cells. Thus, a tumour can be below detection levels for a long time or even go extinct, only driven by a temporary spike of CAR T-cells and the resulting killing. The nature of the probabilistic outcome is that a

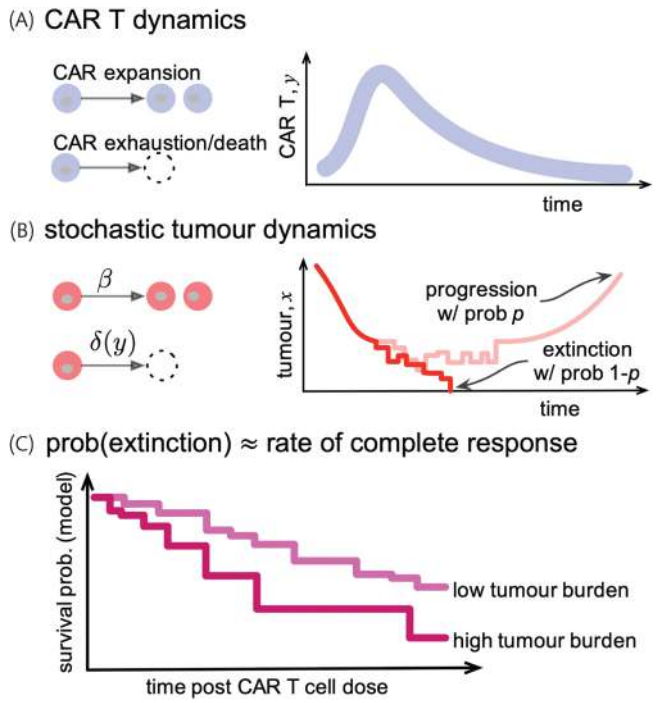


Figure 29.4. Leveraging stochastic dynamics of tumour extinction to describe patient population statistics. The long-term success of single-dose CAR T (y, (A)) can be used to model stochastic dynamics in the tumour (B) that grows at a rate β and is eliminated at a CAR T-dependent rate $\delta(y)$, giving rise to a branching process [1,47]. An individual tumour can probabilistically go extinct or relapse even with the same conditions and parameters (note that the probability of relapse, p, can be very different from the random value of 0.5). As a result, one can model survival as a probabilistic outcome (C).

long-term response with little to no tumour is highly probable. On the other hand, the waiting time to relapse can be exceedingly long. Another aspect of stochastic process modelling in response to drug kinetics is the introduction of inevitable outcome variation, even in patients comparable in almost any aspect (Figure 29.4C). Kimmel et al. [47] asked about the importance of this intrinsic variation by controlling for exogenous parameter variance. Interestingly, this variance, e.g. between-patient variability in tumour growth rate or CAR T effector-tumour target killing, played a more prominent role in the response and efficacy of the cellular drug. Thus, intrinsic stochasticity does not overrule inter-patient variability but may be an important determinant of the ability to define curative events, especially for single-dose cellular therapies that do not persist indefinitely.

These three examples highlight important aspects of CAR T-cell therapy modelling efforts. Firstly, we must understand non-linear and time-dependent properties in the expansion of CAR T-cells in the disease-specific context and the conversion to memory cells. Secondly, the CAR T-cell phenotype, defined by the abundance of naive memory (stem cell memory and central/effector memory) cells [55], can have an important role in the dynamics but is often only measured once. Thirdly, dynamic mechanistic assumptions about CAR T–resident immune system interactions may be crucial in understanding the drug's behaviour in patients. Lastly, probabilistic theories can be included to develop a more comprehensive understanding of long-term responses or cure events. These needs arise due to the unique circumstances of adoptive immune cell therapies that are typically only given once.

29.4. CAR T-cell therapy as an inflammatory process: the role of cytokines and toxicity

CAR T-cell therapy can cause severe CRS or a characteristic immune cell-associated neurologic syndrome (ICANS). Both events are associated with significant co-morbidity and mortality [29,56,57]. They are characterized by high levels of inflammatory cytokines and high numbers of CAR T-cells that spill into the blood [28]. Both forms of toxicity occur in predictable distributions across patient populations [58], thus shaping clinical outcomes.

Multiple processes can be identified as potential drivers of patient outcomes: the effects of lymphodepleting chemotherapy [47,48], the composition of the CAR T-cell product (e.g. phenotypic markers of T-cell differentiation into killer effector cells) [49], CAR T-cell expansion and durability [50], tumour killing [59,60], and tumour burden [27,47]. The use of inflammatory or homeostatic markers, such as cytokines, may be considered in modelling if needed for improved prediction. The explicit utility of inflammatory cytokines in treatment efficacy or side-effects models has yet to be shown satisfactorily. However, they play essential roles in the inflammatory process accompanying CAR T-cell therapy [61,62]. A key assumption is that a return of the wild-type lymphocytes to homeostatic levels can serve as an adequate proxy for the aberrant expression of inflammatory cytokines [47]. The up-regulation of homeostatic cytokines, such as IL-7 or IL-15, likely impacts CAR expansion and tumour killing [63]. An interesting addition to modelling could be the metabolic interactions between T-cells and tumour cells, requiring additional model compartments, e.g. glucose and cytokine levels. The concentrations of these molecules could then be integrated into existing ODE model frameworks, with a potential concentration-dependent effect on CAR T-cell activity and proliferation, sometimes called 'paradoxical' signalling [64]. Unfortunately, in our experience, the cytokine concentrations in patients fluctuate highly, complicating the reliable inference of cytokine dynamics.

Elevating key inflammatory cytokines (IFN-γ, IL-6, and IL-1) is also associated with severe CRS and ICANS. One could hypothesize that CAR T-cell expansion and the occurrence of toxicities can be at least partially explained by dynamic cytokine/biomarker levels in the blood. At the same time, homeostatic (IL-15) and inflammatory (IL-6, IFN-γ, and IL-1) cytokines may benefit CAR T-cell expansion. Computational, statistical, and mathematical models like the ones already developed [47,49] could be adapted to this problem to assess parameter sensitivities and establish clinical covariates. The modelling of cellular immunotherapies can benefit from an ecological perspective to develop a mechanistic understanding of treatment dynamics, failure and success.

29.5. Lymphodepletion before CAR T-cell therapy: necessary but not sufficient

Sufficient lymphodepletion (LD) is a pre-CAR T administration conditioning of the patient using a combination of chemotherapy agents [65]. LD is important in determining a durable response and

may also involve some tumour removal before CAR T administration. Kimmel et al. [47] considered that LD eliminates normal T-cells, giving rise to temporarily reduced competition among normal and engineered (CAR) T-cells. Normal and CAR T-cell populations then grow towards their respective carrying capacities, which gives rise to selection that is detrimental to the CAR T-cell populations in the long term but typically overshadowed by the CAR Ts' early expansion advantage. Specifically, the CAR T-cell-carrying capacity is lower than normal T-cells, resulting in their eventual removal. Immune recapitulation after LD must be driven by stem and progenitor cells [66,67], mediated by a complex signalling cascade [68]. Normal T-cells and CAR T-cells use this changing environment to proliferate and expand at different rates.

In the long term, CAR T-cells are disfavoured against normal T-cells [69]. In the context of mathematical modelling [47], coexistence is not possible. Thus, the CAR T-cell disadvantage ultimately leads to their decline and removal over time. However, CAR T-cell persistence due to effector memory cells could play a role in the long-term dynamics, especially in treating childhood acute lymphoblastic leukaemia. Still, it is unclear whether the associated mechanisms play a role in diffuse large B-cell lymphoma treatment in adults [26]. The decline of CAR T-cells may occur in a time frame longer than the expected survival time of the patient. Therefore, models that theoretically do not include eternal persistence may capture the clinical definition of CAR T persistence.

The expansion of CAR T-cells rapidly occurs within the first two weeks post-infusion, after which first rapidly declines and then decays slowly. The overall CAR T-cell turnover is expected to change during immune reconstitution, allowing us to hypothesize that additional levels of feedback exist between the immune system, the tumour, and CAR T-cells. Owens and Bozic explored the impact of different LD conditioning regimens on CAR-T-cell treatment outcomes [48]. The optimal LD regimen would depend on the CAR T dose and the tumour burden. The time between LD chemotherapy and CAR T-cell infusion (about 5 days in the clinic) was observed to be depending on the tumour growth rates. As this growth rate is often unknown, this model is a powerful example of how mathematical modelling can contribute to formulating novel testable hypotheses or treatment modifications. There might be room for improvement in selecting the timing and strength of LD, possibly on a personalized level.

So far, most models have assumed that tumour cell proliferation is independent of tumour burden. However, the cancer's net growth rate might depend on the tumour burden relative to a (patient-specific) maximum possible tumour size. In this context, one could explore the impact of other sources of variability that originate from a logistic dependence of tumour cell proliferation on tumour volume, called proliferation saturation [70,71]. Future studies should carefully consider non-linear relationships between tumour size and growth as a potential dynamic biomarker for the success of CAR T-cell therapy.

29.6. Modelling feedback from the tumour and predator–prey dynamics

Clinical outcomes of CAR T cell therapy have been shown to be impacted by cellular interactions between CAR T cells and patient tumours [72]. The interactions between cancer and immune cells can be understood as predator–prey, but this analogy has limitations [8]. Unlike predatory animals, immune cells do not feed on their prey (the target or cancer cells). Kareva et al. have also noted that immune–target cell systems are not expected to exhibit the oscillatory behaviour often seen in nature [8,73]. On the one hand, both target and immune cells compete for space and resources in the tumour microenvironment. On the other hand, increased tumour size often correlates with decreased immune cell density; these two key differences jointly create a new predator (immune cell)–prey (cancer cell) relationship. This relationship is driven by cell–cell signalling, immune editing, immunosuppression, resource availability, and temporal complexity rooted in healthy immune regulation. Together, these mechanisms determine tumour-immune feedback and alter the predatory character of the CAR T cells [59].

Due to its complex ecological character, in which the whole is more than the sum of its parts, there is a critical need to better understand the tumour microenvironment within which immune cells, such as CAR T cells, function. Modelling aims to predict how this environment can be perturbed to inhibit tumour growth more efficiently, e.g. relying on adapted predator–prey theory and methods from modelling in the field of ecology [46]. For example, Sahoo et al. observed *in vitro* in glioblastoma that CAR T-cell dose correlated inversely with the killing rate but correlated directly with the net rate of CAR T proliferation and exhaustion. This relationship suggests that at lower doses CAR T-cell killing is improved. Yet, the cells become more exhausted, all the while CAR T-cell exhaustion increases with antigen density and tumour growth rate [59]. This example shows that adoptive cellular therapies are an emerging and exciting application area for modelling personalized biological control strategies in ecological settings and the evolution of escape mechanisms. In this context, Hamilton et al. have noted that there are clear collective and resource competition effects involved in understanding CAR T cells as predators [46], possibly in connection to the handling time that predatory immune cells need to find and kill their targets. In addition, immune cell dynamics may be self-limiting, independent of the cancer cells [64] or other immune cells, impeding T cells from reaching and killing tumour cells [74]. Unlike 'conventional' predators, immune cell populations can 'work cooperatively to detect and destroy cancer cells' [46]. How these mechanisms collectively can improve CAR T cells' ability to control tumours of various sizes and limit the number of escape strategies remains to be seen.

29.7. Conclusions

Insights from ecology and evolution, especially mathematical modelling, often loosely called 'evolutionary dynamics' [2], are of great value to the emerging and fast-paced field of adoptive cellular immunotherapies. The mechanistic description of the cell population dynamics is crucial for understanding and further innovating these therapies. Ecology and evolution offer compelling frameworks for understanding complex, non-linear, and possibly stochastic systems, such as CAR T-cancer cell interplay. It is said that immuno-oncology 'is now in a position to reciprocate' [46], as more and more precise data about cellular dynamics [75], tumour burden [76], and inflammatory processes [77] becomes available, creating possible feedback that critically influences clinical outcomes. Advanced

immuno-oncology and modern mathematical modelling can help revolutionize cancer care using CAR T-cell therapies.

REFERENCES

1. Altrock PM, Liu LL, Michor F. The mathematics of cancer: integrating quantitative models. Nat Rev Cancer. 2015;**15**(12):730–45.
2. Nowak MA. Evolutionary dynamics: exploring the equations of life. Cambridge, MA: Harvard University Press; 2006.
3. Wodarz D, Komarova N. Dynamics of cancer: mathematical foundations of oncology. River Edge, NJ: World Scientific Publishing; 2014.
4. Rockne RC, Hawkins-Daarud A, Swanson KR, Sluka JP, Glazier JA, Macklin P, et al. The 2019 mathematical oncology roadmap. Phys Biol. 2019;**16**(4):041005.
5. Butner JD, Dogra P, Chung C, Pasqualini R, Arap W, Lowengrub J, et al. Mathematical modeling of cancer immunotherapy for personalized clinical translation. Nat Comput Sci. 2022;**2**(12):785–96.
6. Brady R, Enderling H. Mathematical models of cancer: when to predict novel therapies, and when not to. Bull Mathe Biol. 2019;**81**(10):3722–31.
7. Bell GI. Predator-prey equations simulating an immune response. Math Biosci. 1973;**16**(3-4):291–314.
8. Kareva I, Luddy KA, O'Farrelly C, Gatenby RA, Brown JS. Predator-prey in tumor-immune interactions: a wrong model or just an incomplete one? Front Immunol. 2021;**12**:668221.
9. Bekker RA, Zahid MU, Binning JM, Spring BQ, Hwu P, Pilon-Thomas S, et al. Rethinking the immunotherapy numbers game. J Immunother Cancer. 2022;**10**(e005107):1–5.
10. Germain RN. The art of the probable: system control in the adaptive immune system. Science. 2001;**293**(5528):240–5.
11. Deem MW, Hejazi P. Theoretical aspects of immunity. Annu Rev Chem Biomol Eng. 2010;**1**:247–76.
12. Wuthrich M, Deepe GS, Jr., Klein B. Adaptive immunity to fungi. Annu Rev Immunol. 2012;**30**:115–48.
13. Janeway CA, Jr. The immune system evolved to discriminate infectious nonself from noninfectious self. Immunol Today. 1992;**13**(1):11–6.
14. Oiseth SJ, Aziz MS. Cancer immunotherapy: a brief review of the history, possibilities, and challenges ahead. J Cancer Metastasis Treat. 2017;**3**(10):250–61.
15. Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. Nat Rev Immunol. 2020;**20**(11):651–68.
16. Michor F, Iwasa Y, Nowak MA. Dynamics of cancer progression. Nat Rev Cancer. 2004;**4**(3):197–205.
17. Scott J, Marusyk A. Somatic clonal evolution: a selection-centric perspective. Biochim Biophys Acta Rev Cancer. 2017;**1867**(2):139–50.
18. Cancer Research U. The immune system and cancer 2020 [updated 2020. https://web.archive.org/web/*/https://www.cancerresearchuk.org/about-cancer/what-is-cancer/body-systems-and-cancer/the-immune-system-and-cancer*.
19. Podlaha O, Riester M, De S, Michor F. Evolution of the cancer genome. Trends Genet. 2012;**28**(4):155–63.
20. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012;**12**(5):323–34.
21. Tabassum DP, Polyak K. Tumorigenesis: it takes a village. Nat Rev Cancer. 2015;**15**(8):473–83.
22. Raatz M, Shah S, Chitadze G, Bruggemann M, Traulsen A. The impact of phenotypic heterogeneity of tumour cells on treatment and relapse dynamics. PLoS Comput Biol. 2021;**17**(2):e1008702.
23. Opasic L, Zhou D, Werner B, Dingli D, Traulsen A. How many samples are needed to infer truly clonal mutations from heterogenous tumours? BMC Cancer. 2019;**19**(1):403.
24. Rocken C, Amallraja A, Halske C, Opasic L, Traulsen A, Behrens HM, et al. Multiscale heterogeneity in gastric adenocarcinoma evolution is an obstacle to precision medicine. Genome Med. 2021;**13**(1):177.
25. Feins S, Kong W, Williams EF, Milone MC, Fraietta JA. An introduction to chimeric antigen receptor (CAR) T-cell immunotherapy for human cancer. Am J Hematol. 2019;**94**(S1):S3–S9.
26. Locke FL, Ghobadi A, Jacobson CA, Miklos DB, Lekakis LJ, Oluwole OO, et al. Long-term safety and activity of axicabtagene ciloleucel in refractory large B-cell lymphoma (ZUMA-1): a single-arm, multicentre, phase 1-2 trial. Lancet Oncol. 2019;**20**(1):31–42.
27. Locke FL, Rossi JM, Neelapu SS, Jacobson CA, Miklos DB, Ghobadi A, et al. Tumor burden, inflammation, and product attributes determine outcomes of axicabtagene ciloleucel in large B-cell lymphoma. Blood Adv. 2020;**4**(19):4898–911.
28. Neelapu SS, Locke FL, Bartlett NL, Lekakis LJ, Miklos DB, Jacobson CA, et al. Axicabtagene ciloleucel CAR T-Cell therapy in refractory large B-cell lymphoma. New Engl J Med. 2017;**377**(26):2531–44.
29. Neelapu SS, Tummala S, Kebriaei P, Wierda W, Gutierrez C, Locke FL, et al. Chimeric antigen receptor T-cell therapy—assessment and management of toxicities. Nat Rev Clin Oncol. 2018;**15**(1):47–62.
30. Grupp SA, Kalos M, Barrett D, Aplenc R, Porter DL, Rheingold SR, et al. Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. New Engl J Med. 2013;**368**(16):1509–18.
31. Maude SL, Frey N, Shaw PA, Aplenc R, Barrett DM, Bunin NJ, et al. Chimeric antigen receptor T cells for sustained remissions in leukemia. New Engl J Med. 2014;**371**(16):1507–17.
32. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. Nature. 2013;**501**(7467):328–37.
33. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta. 2010;**1805**(1):105–17.
34. Marusyk A, Tabassum DP, Altrock PM, Almendro V, Michor F, Polyak K. Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. Nature. 2014;**514**(7520):54–8.
35. Foo J, Leder K. Dynamics of cancer recurrence. Ann Appl Probab. 2013;**23**(4):1437–68.
36. Nicholson MD, Cheek D, Antal T. Sequential mutations in exponentially growing populations. PLoS Comput Biol. 2023;**19**(7):e1011289.
37. Valkenburg KC, de Groot AE, Pienta KJ. Targeting the tumour stroma to improve cancer therapy. Nat Rev Clin Oncol. 2018;**15**(6):366–81.
38. Kimmel GJ, Locke FL, Altrock PM. Response to CAR T cell therapy can be explained by ecological cell dynamics and stochastic extinction events bioRXivorg. **2020**;2020:717074.
39. Beumer JH, Beijnen JH, Schellens JH. Mass balance studies, with a focus on anticancer drugs. Clin Pharmacokinet. 2006;**45**(1):33–58.
40. Center for Drugs and Biologics FaDA. Guideline for the format and content of the human pharmacokinetics and bioavailability section of an application. In: Research USDoHaHSCfDEa, editor. 1997.

41. Rowland M, Tozer TN. Clinical pharmacokinetics and pharmacodynamics: concepts and applications. 4th ed. Philadelphia, PA: Wolters Kluwer Health/Lippincott William & Wilkins; 2011.
42. Doogue MP, Polasek TM. The ABCD of clinical pharmacokinetics. Ther Adv Drug Saf. 2013;**4**(1):5–7.
43. Chaudhury A, Zhu X, Chu L, Goliaei A, June CH, Kearns JD, et al. Chimeric antigen receptor T cell therapies: a review of cellular kinetic-pharmacodynamic modeling approaches. J Clin Pharmacol. 2020;**60** Suppl 1:S147–59.
44. Benmebarek MR, Karches CH, Cadilha BL, Lesch S, Endres S, Kobold S. Killing mechanisms of chimeric antigen receptor (CAR) T cells. Int J Mol Sci. 2019;**20**(1283):1–21.
45. Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell-cell interactions and communication from gene expression. Nat Rev Genet. 2021;**22**(2):71–88.
46. Hamilton PT, Anholt BR, Nelson BH. Tumour immunotherapy: lessons from predator-prey theory. Nat Rev Immunol. 2022;**22**(12):765–75.
47. Kimmel GJ, Locke FL, Altrock PM. The roles of T cell competition and stochastic extinction events in chimeric antigen receptor T cell therapy. Proc R Soc B. 2021;**288**(20210229):1–9.
48. Owens K, Bozic I. Modeling CAR T-cell therapy with patient preconditioning. Bull Math Biol. 2021;**83**(42):1-36.
49. Mueller-Schoell A, Puebla-Osorio N, Michelet R, Green MR, Kunkele A, Huisinga W, et al. Early survival prediction framework in CD19-specific CAR-T cell immunotherapy using a quantitative systems pharmacology model. Cancers (Basel). 2021;**13**(11).
50. Stein AM, Grupp SA, Levine JE, Laetsch TW, Pulsipher MA, Boyer MW, et al. Tisagenlecleucel model-based cellular kinetic analysis of chimeric antigen receptor-T cells. CPT Pharmacometrics Syst Pharmacol. 2019;**8**(5):285–95.
51. Morris EC, Neelapu SS, Giavridis T, Sadelain M. Cytokine release syndrome and associated neurotoxicity in cancer immunotherapy. Nat Rev Immunol. 2022;**22**(2):85–96.
52. Mc Laughlin AM, Yee C. Letter to the editor: quality criteria for computational models predicting individual outcomes in CAR-T cell therapy. J Immunother Cancer. 2023;**11**(e006990):1–3.
53. Lenaerts T, Pacheco JM, Traulsen A, Dingli D. Tyrosine kinase inhibitor therapy can cure chronic myeloid leukemia without hitting leukemic stem cells. Haematologica. 2010;**95**(6):900–7.
54. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem. 1977;**81**(25):2340–61.
55. Lopez-Cantillo G, Uruena C, Camacho BA, Ramirez-Segura C. CAR-T cell performance: how to improve their persistence? Front Immunol. 2022;**13**:878209.
56. Lee DW, Santomasso BD, Locke FL, Ghobadi A, Turtle CJ, Brudno JN, et al. ASTCT consensus grading for cytokine release syndrome and neurologic toxicity associated with immune effector cells. Biol Blood Marrow Transplant. 2019;**25**(4):625–38.
57. Dholaria BR, Bachmeier CA, Locke F. Mechanisms and management of chimeric antigen receptor T-cell therapy-related toxicities. BioDrugs. 2019;**33**(1):45–60.
58. Jacobson CA. CD19 chimeric antigen receptor therapy for refractory aggressive B-cell lymphoma. J Clin Oncol. 2019;**37**(4):328–35.
59. Sahoo P, Yang X, Abler D, Maestrini D, Adhikarla V, Frankhouser D, et al. Mathematical deconvolution of CAR T-cell proliferation and exhaustion from real-time killing assay data. J R Soc, Interface/ R Soc. 2020;**17**(162):20190734.
60. Brummer AB, Xella A, Woodall R, Adhikarla V, Cho H, Gutova M, et al. Data driven model discovery and interpretation for CAR T-cell killing using sparse identification and latent variables. Front Immunol. 2023;**14**:1115536.
61. Santomasso B, Bachier C, Westin J, Rezvani K, Shpall EJ. The other side of CAR T-cell therapy: cytokine release syndrome, neurologic toxicity, and financial burden. Am Soc Clin Oncol Educ Book. 2019;**39**:433–44.
62. Silveira CRF, Corveloni AC, Caruso SR, Macedo NA, Brussolo NM, Haddad F, et al. Cytokines as an important player in the context of CAR-T cell therapy for cancer: their role in tumor immunomodulation, manufacture, and clinical implications. Front Immunol. 2022;**13**:947648.
63. Kochenderfer JN, Somerville RPT, Lu T, Shi V, Bot A, Rossi J, et al. Lymphoma remissions caused by anti-CD19 chimeric antigen receptor T cells are associated with high serum interleukin-15 levels. J Clin Oncol. 2017;**35**(16):1803–13.
64. Hart Y, Reich-Zeliger S, Antebi YE, Zaretsky I, Mayo AE, Alon U, et al. Paradoxical signaling by a secreted molecule leads to homeostasis of cell levels. Cell. 2014;**158**(5):1022–32.
65. Neelapu SS. CAR-T efficacy: is conditioning the key? Blood. 2019;**133**(17):1799–800.
66. Muranski P, Boni A, Wrzesinski C, Citrin DE, Rosenberg SA, Childs R, et al. Increased intensity lymphodepletion and adoptive immunotherapy—how far can we go? Nat Clin Pract Oncol. 2006;**3**(12):668–81.
67. Williams KM, Hakim FT, Gress RE. T cell immune reconstitution following lymphodepletion. Semin Immunol. 2007;**19**(5):318–30.
68. Turtle CJ, Hanafi LA, Berger C, Gooley TA, Cherian S, Hudecek M, et al. CD19 CAR–T cells of defined CD4+: CD8+ composition in adult B cell ALL patients. J Clin Invest. 2016;**126**:2123–38.
69. Wu L, Wei Q, Brzostek J, Gascoigne NRJ. Signaling from T cell receptors (TCRs) and chimeric antigen receptors (CARs) on T cells. Cell Mol Immunol. 2020;**17**(6):600–12.
70. Prokopiou S, Moros E, Poleszczuk J, Caudell J, Torres-Roca J, Latifi K, et al. A proliferation saturation index to predict radiation response and personalize radiotherapy fractionation. Radiat Oncol. 2015;**10**:159.
71. Poleszczuk J, Walker R, Moros E, Latifi K, Caudell J, Enderling H. Predicting patient-specific radiotherapy protocols based on mathematical model choice for Proliferation Saturation Index. Bull Math Biol. 2017;**80**:1195–206.
72. Kirouac DC, Zmurchok C, Deyati A, Sicherman J, Bond C, Zandstra PW. Deconvolution of clinical variance in CAR-T cell pharmacology and response. Nat Biotechnol. 2023 Nov;**41**(11):1606–17. Erratum in: Nat Biotechnol. 2023 Nov;41(11):1655.
73. Murray JD. Mathematical biology: I. An introduction. 3rd ed. New York: Springer; 2002.
74. Peranzoni E, Lemoine J, Vimeux L, Feuillet V, Barrin S, Kantari-Mimoun C, et al. Macrophages impede CD8 T cells from reaching tumor cells and limit the efficacy of anti-PD-1 treatment. Proc Natl Acad Sci U S A. 2018;**115**(17):E4041–50.
75. Haradhvala NJ, Leick MB, Maurer K, Gohil SH, Larson RC, Yao N, et al. Distinct cellular dynamics associated with response to CAR-T therapy for refractory B cell lymphoma. Nat Med. 2022;**28**(9):1848–59.
76. Dean EA, Mhaskar RS, Lu H, Mousa MS, Krivenko GS, Lazaryan A, et al. High metabolic tumor volume is associated with decreased efficacy of axicabtagene ciloleucel in large B-cell lymphoma. Blood Adv. 2020;**4**(14):3268–76.
77. Dean EA, Kimmel GJ, Frank MJ, Bukhari A, Hossain NM, Jain MD, et al. Circulating tumor DNA adds specificity to PET after axicabtagene ciloleucel in large B-cell lymphoma. Blood Adv. 2023;**7**(16):4608–18.

30

Modelling small cell lung cancer biology through deterministic and stochastic mathematical models

Srisairam Achuthan, Rishov Chatterjee, and Atish Mohanty

30.1. Introduction

Mathematical constructs that depict self-similarity, i.e. repeating patterns over multiple scales and non-integer dimensions, are referred to as fractals. Highly complex biological as well as universal entities have been described based on fractals [1]. Manifestations of fragments of repeating patterns are often observed in nature. In nature, structural patterns are often conserved and expressed across organisms including those from disparate realms. A noteworthy case is the analogy between the structure of a tree and that of a human lung. In both, the fractal design (self-similarity) and function are conserved as well as elegantly manifested. The central trunk of a tree, like the trachea in the human body, divides into wider branches, like the bronchi. These branches then bifurcate into increasingly smaller branches, like the bronchioles, and ultimately lead to the 'leaves', or alveoli, where gas exchange occurs. This similarity in structure also applies to function, and maintaining this natural order is crucial for biological equilibrium and overall health. By analysing changes in these patterns, it may be possible to identify and diagnose abnormal tissue at an early stage, leading to more effective and personalized cancer treatments [2]. Mathematical models are vital tools to study biological entity like the human lung that exhibits fractalness.

The objective of this chapter is to provide an overview of some of the mathematical models and principles reported in the literature underlying the growth, genetic diversity, and spread of small cell lung cancer (SCLC). This knowledge can serve as a framework for further analysing and utilizing the vast amount of genomic and clinical data available related to SCLC. By sharing this mathematical understanding with physicians and cancer researchers, more precise treatment plans could be created to improve patient outcomes. It is believed that cancer is triggered by a chaotic imbalance through random processes of genetic mutations and drift. However, cancer progression seems to restore balance, albeit in an abnormal state, through a predictable process of clonal selection. To model this, both deterministic equations, such as ordinary and partial differential equations, to describe tumour growth and surrounding tissue, as well as stochastic models, such as cellular automata (CA) and stochastic partial differential equations (PDEs), have been applied, to represent cancer initiation and progression. Principles from group theory and game theory could be applicable to describe the tumour cellular dynamics as well [3]. However, we have not covered these principles in this chapter. Together, these methods are crucial as they may provide a starting point for an integrated systems approach to developing new cancer treatments.

30.2. SCLC clinical states

SCLC is an aggressive form of lung cancer that often presents at advanced stages and is believed to originate from neuroendocrine cells found in the lung epithelium [4]. The cancer cells are small and round, and they rapidly grow and spread throughout the body. These cells retain markers such as CD56 [5, 6], chromagranin, synaptophysin [4], and the neuroendocrine transcription factor ASCL1 (achaete-scute complex homolog-like 1) [7]. SCLC is highly metastatic and is characterized by a large number of circulating tumour cells (CTCs) in the periphery, which contributes to relapse and poor prognosis [8, 9]. CTCs can have stem cell properties and may be used as diagnostic markers through liquid biopsy methods [10, 11, 12]. Majority of SCLC patients have a history of smoking, and the disease may present itself many years after smoking cessation [13, 14]. Treatment options for SCLC include chemotherapy, radiation therapy, and surgery; however, the prognosis for SCLC is generally poor. SCLC is characterized by its rapid and early metastatic growth, often spreading as clusters of cancer cells that may contribute to chemoresistance [15]. This suggests that molecules regulating cell–cell interactions, cell–matrix interactions, or cytoskeletal functions may be potential therapeutic targets for SCLC. Often, SCLC is classified into two stages: limited disease and extensive disease. Additional information about these two stages is provided in [2].

30.3. SCLC cells represent a diverse population

SCLC's intrinsic and ongoing genomic instability is a major factor in the diversity and evolution of the disease. With an average of 175 mutations per tumour, SCLC has one of the highest rates of genomic changes among solid tumours, ranging from 5.5 to 7.4 mutations per megabase [16, 17]. G-to-T transversions, a tobacco carcinogenesis signature, are frequently seen among SCLC patients and are consistent with a smoking history [16]. Even though SCLC is believed to originate from neuroendocrine cells found in lung epithelium [4], the cancer stem cell (CSC) population of SCLC is not clearly defined and appears to be diverse in terms of phenotype and genome. This may be due to a high number of mutations and/or epigenetic regulation. The phenotype of SCLC CSCs is not well understood and may express several markers, such as SOX2, CD44, CD56 (NCAM), CD90, CD105, CD133, Sall4, Oct4, nestin, S100β, or vimentin [18, 19]. The ASCL1 transcription factor plays a role in regulating neuroendocrine features and works together with Notch signalling in the normal differentiation of airway stem cells [20]. However, in SCLC, this pathway is altered, and the overexpression of E2F3 and loss of function of RB drive disease progression, likely due to mutations in the tumour suppressor gene TP53 [21]. TP53 (100%) and RB1 (93%) are the most commonly mutated genes in SCLC cases without chromotripsis [22]. Additionally, expression of the histone-lysine methyltransferase enhancer of zeste homolog 2 (EZH2) gene is strongly linked to disruption of the E2F transcription factors and RB1 pathway, found in 96% of SCLC cases [23]. More information about the diverse population of SCLC cells is provided in [2].

30.4. SCLC metastasis and the tumour microenvironment

The process of metastasis involves several distinct phases, beginning with the invasion of tumour cells into surrounding tissue and eventually the bloodstream. As CTCs, they can reach distant sites and grow if they could survive and interact with various tissues, such as extravasation through the endothelial lining of blood vessels [24]. The success of this process depends on specific cellular properties that may vary and are not only determined by genomic changes and their cellular consequences, but also by the effects that the tumour has on its microenvironment and how it interacts or responds to them. Both extravasation and the subsequent intravasation to establish metastatic sites can be regulated at multiple levels, involving ligands within the extra-cellular matrix (ECM), their receptors, including selectins, integrins, cadherins, CD44 and others, or chemokines and cytokines and their receptors. Additional interaction with immune cells or stromal cells further determines metastatic function [25]. A retrospective study analysed 251 SCLC patients diagnosed between 1999 and 2000 and found 152 (60.6%) with distant metastases. Target organ involvement included 20.3% liver, 18.3% bone, 15.5% brain, 10.0% lung, and 6.0% of adrenal gland [26]. In Figure 30.1,

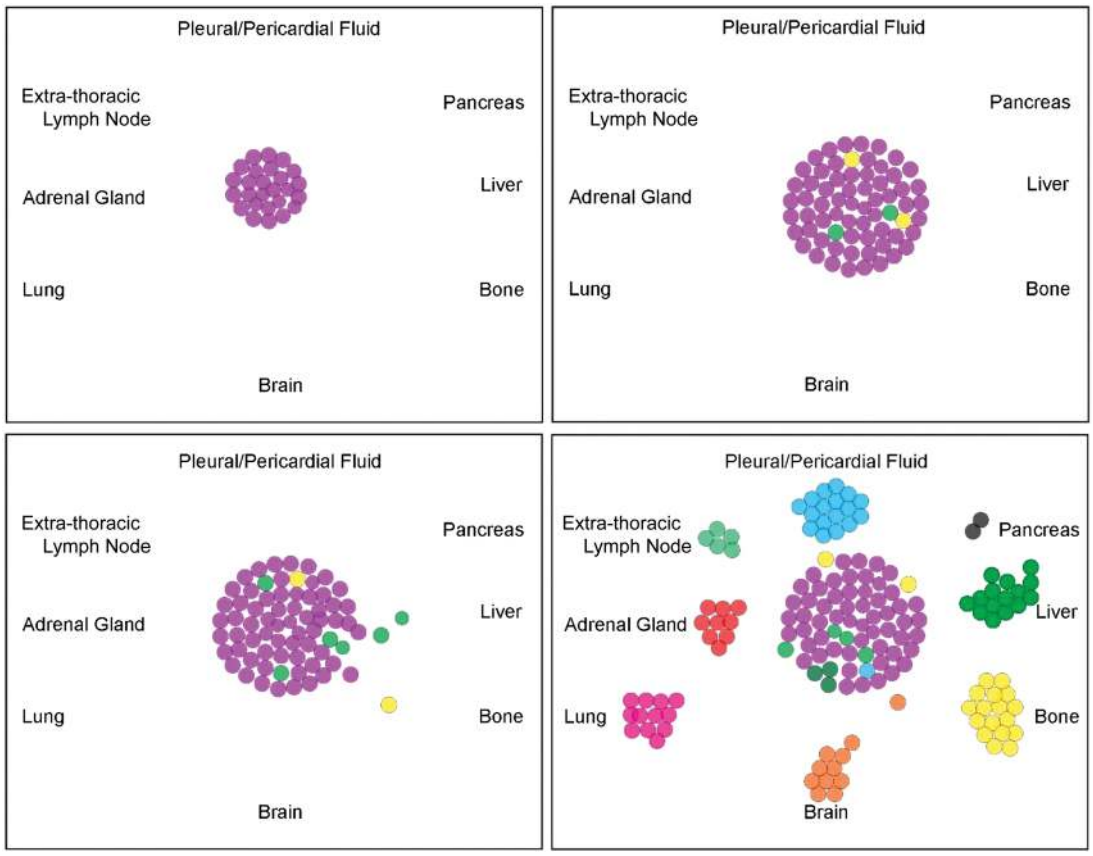


Figure 30.1. Most common metastatic sites of SCLC. Simulation of early time points and end points (from top to bottom right) of tumour cells (purple) metastasizing to different organs. *Source:* Adapted from Salgia et al., 2018.

a model was generated by JavaScript bubble chart, where the initial state is fixed with all cancer cells positioned at the primary site, and the final state is quantitatively fixed based on the metastasis sites population reported in Nakazawa et al. [26]. Cells with different metastatic phenotypes first appear at the primary site, and then cells with similar metastatic phenotypes cluster together and eventually metastasize to different sites. Metastasis is a multi-step process in the cancer model that includes steps, such as mitogenesis, morphogenesis, and motogenesis [27]. The mitogenesis gives the proliferation that then also affects morpho- and motogenesis. The process of metastasis is dependent on genetic regulation, protein network, and tumour–stroma interaction. As an example, it has been shown that PAX5 transcription factor is highly expressed in SCLC [28]. This then regulates chemokine receptor CXCR4 and RTKs, such as MET and RON. The receptors themselves cause a plethora of signal transduction events, such as activation of the focal adhesion protein FAK and Actin cytoskeleton. Ultimately, this leads to increased motility, invasion, and metastasis.

EMT is a crucial process for metastasis formation in cancer progression and development. The significance of the tumour stromal microenvironment for tumorigenicity and metastasis in SCLC has been well established, and some of the mechanisms involved have been already unraveled. To learn more about EMT, ECM, phenotypic plasticity of SCLC being a dynamic chaotic system where the initial conditions in the tumour microenvironment plays a crucial role in determining the final states, and other aspects of SCLC metastasis and tumour microenvironment, refer to [2].

30.5. Computational modelling

Deterministic and stochastic mathematical models have been used to study cancer. Deterministic models are mathematical models that describe the time evolution of a system using a set of differential equations. They describe the rate of change of variables in the system. The solution to the system is obtained when the value of the variables of the system of equations is computed at any given time. Deterministic models are useful for understanding the underlying mechanisms of cancer growth and spread, and for making predictions about the behaviour of a population of cancer cells. Examples of deterministic models used in cancer research include CA models and compartmental models.

Stochastic models, on the other hand, consider the randomness and uncertainty associated with biological processes, such as the growth and spread of cancer cells. These models use probability distributions and random variables to describe the behaviour of the system. Stochastic models are useful for studying the population dynamics of cancer cells as well as for understanding the effects of genetic and environmental factors on cancer progression. Examples of stochastic models used in cancer research include agent-based models (ABMs), Markov processes, and branching processes.

Single-cell events, like mutations, play a vital role in the development of cancer. To understand their impact, researchers have used various discrete methods that include both deterministic and stochastic approaches, such as CA [29], ABMs, and hybrid continuum-discrete approaches [30]. The CA models allow researchers to study the behaviour of individual cells and the interactions between cells, leading to a better understanding of cellular processes. While the ABMs focus on the behaviour of individual cells, the hybrid continuum-discrete approaches, on the other hand, combine the strengths of both CA and ABMs to provide a more comprehensive view of cellular processes. In recent years, multiscale modelling has emerged as a promising tool for studying cancer and other complex diseases. This approach links ordinary differential equations (ODEs) to cellular-level parameters [31], allowing researchers to study the dynamics of cancer progression and treatment response at multiple scales. The integration of discrete models with ODEs provides a comprehensive view of cancer development, from the molecular to the tissue level, and has led to new insights into the underlying mechanisms of cancer progression.

30.6. Cellular automata models

CA models are a type of deterministic mathematical model that are used to simulate the growth and spread of cancer cells. These models divide the space in which the cancer cells are growing into a regular grid of cells, with each cell representing a small volume of tissue. The state of each cell in the grid is determined by a set of rules that govern the behaviour of the cells in the neighbourhood. Cancer models based on differential equations address a continuum of cells at the tissue scale where the effect of individual cells is averaged. On the other hand, discrete models of tumour growth based on CA capture the response of individual cells as they interact with one another as well as with their microenvironment. CA models have been used to simulate a variety of different types of cancers at the cellular scale [32, 33, 34, 35, 36] and sub-cellular scale [37, 38], depending on the specific rules and initial conditions used. For example, in some CA models of solid tumours, cells could be in one of several different states, such as healthy, normal, or cancerous, and the rules govern the interactions between these cells and the rate at which they transition between states. An example of a CA model used to study cancer is the so-called 'Game of Life' model, which is based on the two-dimensional cellular automaton of the same name, invented by John Horton Conway [39]. This model represents cells as 'alive' or 'dead', and the rules determine how the cells interact with their neighbours to reproduce, die, or remain the same. This model could be used to simulate the growth of a tumour and its interactions with the surrounding tissue, by applying appropriate rule set. However, these models have limitations and need to be validated with experimental data to be useful in actual clinical scenarios.

CA was used to study the invasion of cancerous cells in a population of normal cells by Qi et al. [35]. In this context, a lattice cell represented a single biological cell. The state of each of the cells of the CA was assumed to be normal, cancerous, complex (cancerous and bound by white blood cell), or dead cancerous. Probabilistic rules were then applied to study the dynamics of the cellular states. However, this model did not explicitly consider growth promoting factors (such as the presence of blood vessels, nutrient supply, and oxygen) and growth inhibiting factors (such as toxic metabolites) for tumours that 'motivate' them to move far away from primary sites.

Macklin et al. [30] modelled the microenvironment-dependent birth rate b_i and death rate d_i as shown below:

$$b_i(t) = \begin{cases} b_{i,P}\left(1 - \eta_i R_i(t)\right) & \text{if } pO_{2,p} < pO_2 \\ b_{i,P}\left(\dfrac{pO_2 - pO_{2,N}}{pO_{2,p} - pO_{2,N}}\right)\left(1 - \eta_i R_i(t)\right) & \text{if } pO_{2,N} < pO_2 \le pO_{2,P} \\ 0 & \text{if } pO_2 \le pO_{2,N} \end{cases}$$

$$d_i(t) = \begin{cases} 0 & \text{if } pO_{2,N} < pO_2 \\ d^*_{i,N} & \text{if } pO_2 \le pO_{2,N} \end{cases}$$

where $d^*_{i,N}$ is a constant [40]. The microenvironment apoptosis rate $d_{i,A}$ is modelled as

$$d_{i,A}(t) = d^*_{i,A} + \left(d^{\max}_{i,A} - d^*_{i,A}\right) R_i(t)$$

where $d^{\max}$ is the maximum rate of apoptosis.

For the time interval $[t, t + \Delta t]$, each viable tumour cell has the chance to divide, undergo apoptotic death, or reach necrotic death. The probability for a live cell to perform one of the three actions in each time step until death is described by the below equations:

$$P(\text{cell division}) = 1 - \exp\left(d_{i,A}(t)\Delta t\right)$$
$$P(\text{apoptotic death}) = 1 - \exp\left(b_i(t)\Delta t\right)$$
$$P(\text{necrotic death}) = 1 - \exp\left(-d_{i,N}(t)\Delta t\right)$$

Each dead cell then has the probability that it will become lytic and rupture, leaving an empty space in the automaton model, based on the average duration the cell death. The probability of the dead cell to reach lysis is expressed in the below equation, where T_D is the duration of cell death:

$$P(\text{cell lysis}) = 1 - \exp\left(-\frac{1}{T_{i,D}}\Delta t\right)$$

30.7. Simulation example

To do this, by increasing the birth rate by 0.025 and increasing the viable cell Hill coefficient to 4 to simulate our growth rate in Figure 30.2, modified the viable lifespan of the tumour cells to increase the probability of apoptosis over time and decrease the time between apoptosis and necrosis due to hypoxia, and introduced a 3% rate of necrosis for peri necrotic tumour cells to fit our carrying capacity obtained in Figure 30.2. The remaining parameters of the simulation remain similar, where the duration of necrosis and cell death are the same, the system begins with a single cell at time = 0, and the drug parameters are unchanged. The first action performed by the BioFVM program as it creates the simulation is to fill any unoccupied spaces with fluid and consider any dead cells as an empty space [40]. The model then uses the following reaction–diffusion equations for the oxygen and drug, respectively, to apply the uptake rate across each cell:

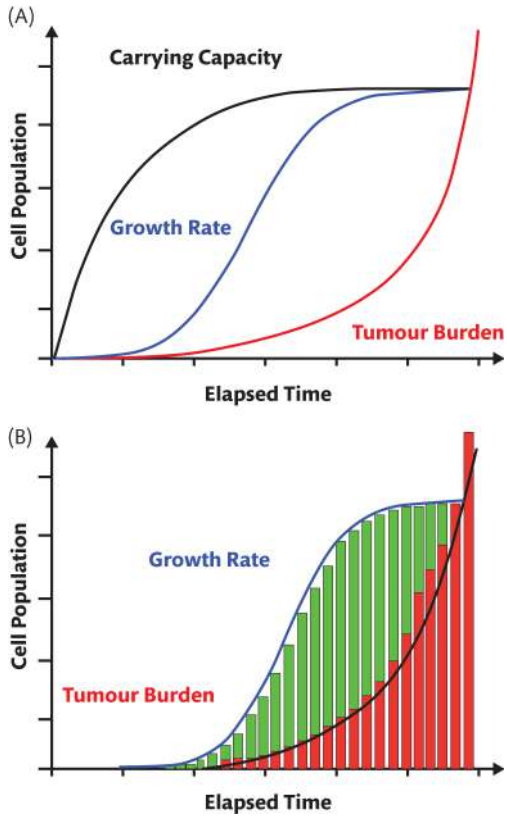


Figure 30.2. Tumour growth models. (A) Tumour growth model with dynamic carrying capacity. (B) Histogram of tumour evolution (green) alongside the histogram (red) of the metastatic growth. As the primary tumour reaches the maximum carrying capacity, the metastatic burden will increase exponentially from the detached primary cells travelling and growing at distant sites. The metastatic cell population grows beyond the original carrying capacity until it reaches a new carrying capacity due to the model describing the growth of cells in all metastatic sites rather than just the primary sites.
Source: Adapted from Salgia et al., 2018.

$$\frac{\partial pO_2}{\partial t} = D_{oxy}\nabla^2 pO_2 - \lambda_{\text{oxy}^\circ}\, pO_2 - \sum_{\text{cells}^\circ i} U_{i,oxy}\, pO_2$$
$$\frac{\partial c}{\partial t} = D_c\nabla^2 c - \lambda_c c - \sum_{\text{cells}\, i} U_{i,c}\, c$$

where the treatment by the drug is set at 5 μM at time $t = 528$ (day 22). When observing the two-dimensional image of the tumour over time, the cells become more hypoxic the closer they are to the middle. The oxygen concentration in the tumour begins to decrease in steps as the tumour begins to grow, leaving the population of hypoxic cells to grow rapidly.

With the presence of the drug at time $t = 528$, the live tumour cell's exposure to the drug E and its response to the drug R (having a Hill coefficient $h = 1$) are given by the below equations:

$$E_i(t) = \int_0^t c(s)ds$$
$$R_i(t) = \frac{E_i(t)}{\alpha_i + E_i(t)}$$

where α is the exposure for a half-maximum effect [40].

The BioFVM simulations may be captured after the two-dimensional tumour profile; cell automata model, basic agent model, and MATLAB model scripts have been edited and saved; and the simulations have been run through the command prompt [30]. Once the simulations have finished running, MATLAB is automatically prompted and a series of images, or visual interpretations of the system at each recorded time interval, are then opened. These images may then be saved as image files. This spatiotemporal model could be utilized to predict and personalize patient response to drug therapy using organoids, spheroids, mouse models, and zebrafish models.

30.8. Agent-based models

The cellular Potts model [41] is a more generalized CA that uses lattice dynamics to study interactions among biological cells. The cellular Potts model was used to study the formation of cell clusters because of assuming configurations of minimal adhesive free energy [42]. The CA model and the cellular Potts model fall under ABMs. The variables in ABM are individuals. These individuals are considered as agents, and a set of prescribed rules govern the behaviour of these agents. In the case of CA, the individual lattice cells are the agents.

ABMs can be used to study the evolution of a cancer population over time and to investigate the impact of various therapeutic interventions on the progression of the disease. ABMs have been applied to various types of cancers, including breast cancer, prostate cancer, and lung cancer. These models can provide valuable insights into the mechanisms underlying cancer growth and progression and can be used to evaluate potential therapeutic strategies and predict their efficacy. One of the advantages of ABMs is their ability to capture the complexity of the cancer microenvironment and the interactions between cancer cells and other cellular and non-cellular components. Additionally, ABMs can be used to study the impact of genetic and epigenetic heterogeneity in cancer populations and to incorporate information on specific cellular pathways and signalling networks. However, ABMs can also be computationally intensive and may require a large amount of data and computational resources to be effectively implemented. ABMs require significant amounts of data to initialize and parameterize the models, including data on cell properties, signalling pathways, and interactions between cancer cells and the microenvironment. The complexity of ABMs can make it challenging to interpret and understand the results, particularly when studying large, complex systems. ABMs can be difficult to validate due to the lack of experimental data or the difficulty in obtaining experimental data that directly supports the model predictions.

A discrete agent-based spatiotemporal model that incorporates the effects of nutrient supply, mechanical confinement that represents the tissue resistance against tumour cell movement, and toxicity of metabolites in the context of brain tumour progression was developed by Mansury [43]. They simulated the complex dynamic self-organizing and adaptive processes observed in tumours, namely spatial aggregation of tumour cells as clusters and their migration in search of suitable survival conditions.

Hybrid agent-based models combine the unusual effectiveness of continuum deterministic models to capture tumour dynamics at the tissue scale with discrete CA models at cellular and sub-cellular scales [44]. Tumour invasion of stroma and surrounding tissue are modelled as coupled non-linear PDEs. The PDEs are discretized to model cell migration and form the basis of the hybrid discrete-continuum model. This model enables specific properties of cells to be described, such as proliferation, death, cell–cell adhesion, and mutation.

30.9. Compartmental models

Compartmental models are a type of deterministic mathematical model that are used to study the population dynamics of cancer cells and the progression of cancer. These models divide the population of cancer cells into different compartments, each representing a different stage or state of the cancer. The compartments can represent different stages of cancer progression, such as normal, precancerous, or malignant cells, or different stages of treatment response, such as responsive, resistant, or dormant cells.

The compartmental models use a set of differential equations to describe the rate of change of the number of cells in each compartment over time, considering various processes such as cell growth, death, and migration. The behaviour of the system is represented by the overall dynamics of the population. Compartmental models can be used to simulate a variety of different types of cancers, depending on the specific compartments and interactions included in the model. One example of a compartmental model is the Gompertz model [45], which is a classic model for describing the growth of tumours and has been used to study various types of cancers. The model assumes that the growth rate of a tumour is initially slow, then accelerates, and eventually slows down again as the tumour reaches a maximum size:

$$V(t) = V_{inj}\left(\exp^{\frac{\alpha}{\beta}\left(1-\exp^{-\beta t}\right)}\right)$$

where $V(t)$ is the tumour size at time t, V_{inj} is the initial size with tumour parameters α and β.

Compartmental models are useful in simulating and studying the complex dynamics of cancer progression and can be used to test hypotheses about the mechanisms of cancer progression and the effects of various treatments. However, these models also have limitations and need to be validated with experimental data to be useful in actual clinical scenarios. Compartmental models do not consider the spatial distribution of cancer cells, which is particularly important for understanding solid tumours. It can be difficult to parameterize the compartmental models based on actual data, which can lead to unrealistic or misleading results. The models are deterministic in nature and do not consider the randomness and uncertainty associated with biological processes.

30.10. Deterministic continuous models

For a given set of initial conditions, models that produce the same results each time they are solved are known as deterministic models. These differ from stochastic or probabilistic models in that the

model results change each time they are solved even though the initial conditions do not. Deterministic models with one independent variable ('time') and one or more dependent variables (such as 'substrates' or 'metabolites') and represented by ODEs are ideal to capture dynamical processes. For example, Enderling and Chaplain [46] studied the rate of tumour growth cells. Utilizing parameters α (fraction of dividing tumour cells) and β (fraction of tumour dying cells), they showed that tumour cells could either be in (1) quiescence ($\alpha - \beta = 0$), (2) proliferative ($\alpha > \beta$), or (3) depleting ($\alpha < \beta$). Since tumours do not grow indefinitely in size, a more realistic representation of rate of tumour growth should take into account the carrying capacity constraint, K, representing the maximum population of cells also known as carrying capacity of the host cell. Hahnfeldt et al. [47] modelled K as a function of time and tumour size as follows:

$$\frac{dK}{dt} = \phi C - \varphi C^{\frac{2}{3}}$$

where ϕ and φ represent constant positive rates of angiogenesis stimulation and inhibition, respectively (**Figure 30.2**).

Mathematical models solely based on ODEs describe the total number of tumour cells over time but do not consider any spatial variables. It is essential to model the spatial variables along with time since processes such as cancer invasion and metastases are more potent killers than local tumour growth and are inherently spatial in nature. Models based on PDEs such as reaction diffusion are apt for quantitative substances of interest in cancer modelling (such as nutrients or oxygen) at a specific position (space) and time (t). PDE-based models are also referred to as continuum models since they are solved for continuously in space and time variables.

For example, Gatenby and Gawlinski [48] were one of the earliest to model cancer invasion as a spatiotemporal evolution of tumour cells (C), enzymes with H^+ ions (m), and ECM (v) as follows:

$$\frac{\partial C}{\partial t} = \nabla \cdot \left(D_C(1-v)\nabla C\right) + \rho C(1-C)$$
$$\frac{\partial m}{\partial t} = \nabla^2 m + \delta(C-m)$$
$$\frac{\partial v}{\partial t} = v(1-v) - \gamma m v$$

where D_C is the diffusion coefficient constant, ρ is the tumour cell proliferation rate constant, δ is the H^+ ion's production and decay rate constant and γ is the ECM degradation rate. Appropriate initial conditions and spatial region need to be specified to solve $\frac{\partial m}{\partial t} = \nabla^2 m + \delta(C-m)$, where tumour cells are assumed to proliferate and undergo non-linear diffusion and secrete H^+ ions that diffuse and degrade the normal tissue. The $H+$ ions are assumed to undergo linear decay with logistic growth for normal tissue in the absence of any cancer cells. Cancer cell migration processes, such as haptotaxis (i.e. directional cell migration in response to gradients of cellular adhesion molecules in the ECM or gradients of the ECM density), were modelled using a modified version of $\frac{\partial m}{\partial t} = \nabla^2 m + \delta(C-m)$ by Anderson [49].

The PDE-based models can be discretized using finite-difference approximations. To study individual cell movement, Anderson et al. investigated the discrete form of the continuous version built to study haptotaxis [50]. Spatial variables were discretized retaining time, t, to be continuous. Stochastic movement rules were incorporated to derive a biased random walk governing the motion of a single tumour cell. Dynamical models of cancer growth leading to chaotic behaviour have also been reported [51]. Itik and Banks were able to explicitly show the existence of deterministic chaotic dynamics by modelling the interactions and competitions between tumour cells and other cells of the body, such as healthy host cells and activated immune system cells. Based on ideas from Lie algebra [52], the control of chaotic dynamics of cancer growth has been recently formulated [53] in a three-dimensional cancer model for tumour growth. This spatiotemporal heterogeneity model could be utilized to understand the tumour evolution over time as well as attempt to predict the genetic phenotype that may correlate with metastasis and cancer progression. As we go forward from here, we must be able to incorporate mathematical modelling in SCLC. We can envision the utilization of various models in the behaviour of cell lines/three-dimensional models, organoids/spheroids along with PDX/CDX models, as well as tumour behaviour in natural progression and/or therapeutic response. We should be able to study the potential for mechanisms of resistance.

30.11. Application of neural networks to ODEs and PDEs

Mathematical modelling of SCLC from a deterministic or stochastic perspective often makes use of ODEs or PDEs to better simulate the disease's intricate mechanisms. Drawing from examples such as emulating the tumour microenvironment [30] or studying cell movement [51], these methods tend to rely on many strict constraints to find closed-form or efficient numerical solutions. Since various systems of SCLC holistically are complex with many confounding variables that influence behaviour of entities (e.g., cells or tissue), flexible mathematical models in the form of differential equations tend to be more challenging to solve in a closed-form manner, and numerical approximation tends to stop at local minima quite frequently. To lead to faster convergence while avoiding a cold start, deep neural networks can be leveraged for solving systems of differential equations that model phenomena of SCLC as optimization problems when limited data is present or when the dimensionality is large [54].

To frame the optimization problem that a neural network will minimize loss using gradient descent and backpropagation, the universal approximator property of neural networks is leveraged to transform the existing differential equation as a set of learnable parameters. The loss function itself is defined in a way that measures the deviation of the neural network's output from the true solution to the differential equation. Afterwards, the forward and backward passes of the network are utilized to minimize this loss. The iterative update process of the coefficients or unknowns of the differential equations is essentially weights that are adjusted to minimize the specified loss function. As an example, metastatic tumour growth and metastatic spreading can be modelled in a generalized way using an ODE and transport PDE model [55]:

$$\begin{cases} \dfrac{\partial}{\partial t}\rho(\upsilon,t)+\dfrac{\partial}{\partial \upsilon}\left[g_m(\upsilon)\rho(\upsilon,t)\right]=0, & \upsilon\in\left[\upsilon_{m,0},b\right),t\geq 0, \\ g_m(\upsilon_{m,0})\rho(\upsilon_{m,0},t)=\beta_p\left(\upsilon_p(t)\right)+\int_{\upsilon_{m,0}}^{b}\beta_m(\upsilon)\rho(\upsilon,t)d\upsilon, & t\in(0,+\infty), \\ \rho(\upsilon,0)=0, & \upsilon\in\left[\upsilon_{m,0},b\right), \end{cases}$$

where v represents the volume of the tumour, $b > v_{m,0}$ is the volume of the tumour at the saturated level, and $v_p(t)$ represents the volume of the primary tumour at time t. By discretizing the domain of the hybrid DE system, the partial derivatives of the model can be approximated using finite differences. By introducing a supervised training paradigm where the model can have input–output pairs from a numerical solver, training data can be generated from an artificial neural network to adjust weights of the input variables via backpropagation. The inputs would refer to initial/boundary conditions, and the outputs would refer to expected outputs. To frame the optimization problem, a loss function is necessary. Common loss functions include mean squared error (MSE) or mean absolute error (MAE) where the intent is to minimize loss to reach a global minimum during model training over several iterations.

Leveraging the weights obtained from the limited training data, the trained artificial neural network (ANN) model could then be used to make predictions on new boundary conditions defined by the model system. For further refinement, a physics-informed neural network (PINN) [56] can be leveraged to add more constraints on the defined loss function to converge to a minimum that more closely would approximate the true dynamics of the differential equation system. A PINN modifies the loss function by adding terms to the loss function which are associated with the rules described by the differential equation system. A general way of describing this modification is by introducing a term lambda that balances the contribution of the data and physics terms to the loss. Another loss term is attributed to the differential equation system independent of the ANN. Lambda is multiplied to this new loss term. The sum of loss terms is seen as the new loss function to optimize. Another approach to using ANNs for solving differential equations is to not only have examples of pairs of boundary conditions with outputs but to also introduce the predicted error of each pair from the numerical solver as an additional feature to the ANN. This integration is accentuated by the interactive mathematical modelling-artificial neural network (IMANN) framework [57].

In Figure 30.3, the numerical ODE model known as SOFC's anode numerical model is used to receive the inputs of an ODE with temperature and density as input parameters. The numerical model will predict overpotential, η, which has its squared difference computed from the experimental evidence for the specific boundary condition [57]. This result is fed into the ANN along with the input parameters that went into the numerical model to have a more calibrated, universal approximation of alpha and beta terms that come back into the numerical model for reinforcing the numerical model in future iterations until the squared difference of predicted overpotential and experimental overpotential converges to a global minimum. The flexibility of using ANNs for better approximation of ODE/PDE solutions can also be applied to equation systems that are high dimensional. This new development is presented by reformulating non-linear, parabolic differential equations as backward stochastic differential equations (BSDEs) [54]. The gradients of BSDEs

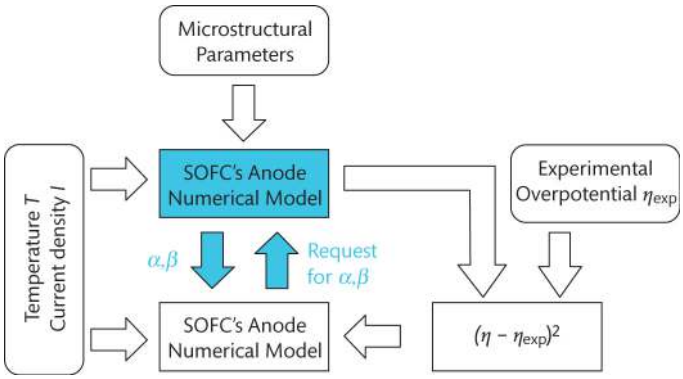


Figure 30.3. The interactive mathematical modelling-artificial neural network block diagram. *Source:* Adapted from Buchaniec et al. 2021.

can be much more easily approximated by ANNs invariant to the scale of the number of dimensions that makes it a robust approach when having limited input and output pairs from a numerical solver. Current solvers for very high dimensional PDEs are limited to the curse of dimensionality; therefore, simulation methods such as Monte Carlo and Feynman–Kac are used. ANNs that make predictions from approximating the gradients of BSDEs ultimately converge to lower loss values compared to Monte Carlo simulation for high-dimensional PDEs.

Although there are not many practical examples of high-dimensional PDEs currently in the computational oncology domain, there is potential for modelling tumour dynamics in the perspective of multi-omics that would lead to a significant number of dimensions when PDEs are developed. Reformulation of proposed PDEs to BSDEs while using ANNs for approximation will lead to more efficient solvers for large-scale systems in computational oncology.

30.12. Conclusions

Throughout this chapter, we have demonstrated that mathematical modelling is critical for advancing our understanding of small cell lung cancer (SCLC), a particularly aggressive and rapidly metastasizing form of lung cancer. We reviewed and explored various mathematical frameworks that shed light on the growth patterns, genetic diversity, and metastasis mechanisms of SCLC, highlighting the value of both deterministic and stochastic approaches. These models are vital in examining the intricate fractal nature of biological systems, particularly the lung's unique geometry. By utilizing these mathematical tools, researchers can develop algorithms that effectively analyse changes in disease phenotypes. We have emphasized the significance of deterministic models, such as differential equations, and stochastic models, like cellular automata, in comprehending the dynamics and progression of cancer.

Deterministic models that employ differential equations are essential for describing tumour growth and dynamics, while stochastic models capture the inherent randomness in biological processes. Notable examples of computational approaches, such as cellular automata and agent-based models, simulate cancer cell behaviour and interactions. The development of hybrid models, which integrate both deterministic and stochastic elements, provides a comprehensive perspective on cancer dynamics, effectively encapsulating

processes at both cellular and tissue levels. Furthermore, recent advancements in neural network architectures have significantly enhanced mathematical modelling, allowing for more accurate simulations of SCLC dynamics. These sophisticated models contribute to optimizing predictions regarding tumour behaviour and treatment responses, thereby supporting the move towards personalized medicine. The incorporation of physics-informed neural networks (PINNs) further improves the modelling of complex systems by offering enhanced approximations of the differential equations that govern SCLC.

In conclusion, the integration of mathematical models, computational techniques, and neural networks is crucial in deepening our understanding of SCLC biology and in developing innovative therapeutic strategies. This comprehensive approach aims to improve patient outcomes by facilitating the creation of more precise treatment plans tailored to the unique characteristics of the disease.

REFERENCES

1. Losa G, Merlini D, Nonnenmacher T, Weibel E. Eds. Fractals in Biology and Medicine. 2005. Birkhäuser Basel. https://link.springer.com/book/10.1007/3-7643-7412-8
2. Salgia R, Mambetsariev I, Hewelt B, Achuthan S, Li H, Poroyko V, et al. Modeling small cell lung cancer (SCLC) biology through deterministic and stochastic mathematical models. Oncotarget. 2018;**9**(40):26226–42 https://doi.org/10.18632/oncotarget.25360
3. Wölfl B, Te Rietmole H, Salvioli M, Kaznatcheev A, Thuijsman F, Brown JS, et al. The contribution of evolutionary game theory to understanding and treating cancer. Dyn Games Appl. 2022;**12**(2):313–42. https://doi.org/10.1007/s13235-021-00397-w
4. Park KS, Liang MC, Raiser DM, Zamponi R, Roach RR, Curtis SJ, et al. Characterization of the cell of origin for small cell lung cancer. Cell Cycle. 2011;**10**:2806–15. https://doi.org/10.4161/cc.10.16.17012
5. Lantuejoul S, Moro D, Michalides RJ, Brambilla C, Brambilla E. Neural cell adhesion molecules (NCAM) and NCAM-PSA expression in neuroendocrine lung tumors. Am J Surg Pathol. 1998;**22**:1267–76. doi: 10.1097/00000478-199810000-00012
6. Patel K, Moore SE, Dickson G, Rossell RJ, Beverley PC, Kemshead JT, et al. Neural cell adhesion molecule (NCAM) is the antigen recognized by monoclonal antibodies of similar specificity in small-cell lung carcinoma and neuroblastoma. Int J Cancer. 1989;**44**(4):573–8. doi: 10.1002/ijc.2910440402
7. Borges M, Linnoila RI, van de Velde HJ, Chen H, Nelkin BD, Mabry M, et al. An achaete-scute homologue essential for neuroendocrine differentiation in the lung. Nature. 1997;**386**:852–5. https://doi.org/10.1038/386852a0
8. Hodgkinson CL, Morrow CJ, Li Y, Metcalf RL, Rothwell DG, Trapani F, et al. Tumorigenicity and genetic profiling of circulating tumor cells in small cell lung cancer. Nat Med. 2014;**20**:897–903. https://doi.org/10.1038/nm.3600
9. Yu N, Zhou J, Cui F, Tang X. Circulating tumor cells in lung cancer: detection methods and clinical applications. Lung. 2015;**193**:157–71. doi: 10.1007/s00408-015-9697-7
10. Carter L, Rothwell DG, Mesquita B, Smowton C, Leong HS, Fernandez-Gutierrez F, et al. Molecular analysis of circulating tumor cells identifies distinct copy-number profiles in patients with chemosensitive and chemorefractory small-cell lung cancer. Nat Med. 2017;**23**:114–9. https://doi.org/10.1038/nm.4239.
11. Hou JM, Krebs MG, Lancashire L, Sloane R, Backen A, Swain RK, et al. Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer. J Clin Oncol. 2012;**30**:525–32. https://ascopubs.org/doi/abs/10.1200/JCO.2010.33.3716
12. Tanaka F, Yoneda K, Kondo N, Hashimoto M, Takuwa T, Matsumoto S, et al. Circulating tumor cell as a diagnostic marker in primary lung cancer. Clin Cancer Res. 2009;**15**:6980–86. doi: 10.1158/1078-0432.CCR-09-1095
13. Ou SH, Ziogas A, Zell JA. Prognostic factors for survival in extensive stage small cell lung cancer (ED-SCLC): the importance of smoking history, socioeconomic and marital statuses, and ethnicity. J Thorac Oncol. 2009;**4**:37–43. https://doi.org/10.1097/JTO.0b013e31819140fb.
14. Varghese AM, Zakowski MF, Yu HA, Won HH, Riely GJ, Krug LM, et al. Small-cell lung cancers in patients who never smoked cigarettes. J Thorac Oncol. 2014;**9**:8926. https://doi.org/10.1097/JTO.0000000000000142.
15. Klameth L, Rath B, Hochmaier M, Moser D, Redl M, Mungenast F, et al. Small cell lung cancer: model of circulating tumor cell tumorospheres in chemoresistance. Sci Rep. 2017;**7**:5337. https://doi.org/10.1038/s41598-017-05562-z.
16. Peifer M, Fernandez-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, et al. Integrative genome analyses identify key somatic driver mutations of small cell lung cancer. Nat Genet. 2012;**44**:1104–10. doi: 10.1038/ng.2396
17. Rudin CM, Durinck S, Stawiski EW, Poirier JT, Modrusan Z, Shames DS, et al. Comprehensive genomic analysis identifies SOX2 as a frequently amplified gene in small-cell lung cancer. Nat Genet. 2012;**44**:1111–6. https:// doi.org/10.1038/ng.2405
18. Codony-Servat J, Verlicchi A, Rosell R. Cancer stem cells in small cell lung cancer. Transl Lung Cancer Res. 2016;**5**:1625. doi: 10.3978/j.issn.2218-6751.2016.01.01
19. Zhang Z, Zhou Y, Qian H, Shao G, Lu X, Chen Q, et al. Stemness and inducing differentiation of small cell lung cancer NCI-H446 cells. Cell Death Dis. 2013;**4**:e633. https://doi.org/10.1038/cddis.2013.152
20. Ball DW. Achaete-scute homolog-1 and Notch in lung neuroendocrine development and cancer. Cancer Lett. 2004;**204**:159–69. doi: 10.1016/S0304-3835(03)00452-X
21. Kitamura H, Yazawa T, Sato H, Okudela K, Shimoyamada H. Small cell lung cancer: significance of *RB* alterations and TTF-1 expression in its carcinogenesis, phenotype, and biology. Endocr Pathol. 2009;**20**:101–7. doi: 10.1007/s12022-009-9072-4
22. George J, Lim JS, Jang SJ, Cun Y, Ozretic L, Kong G, et al. Comprehensive genomic profiles of small cell lung cancer. Nature. 2015;**524**:47–53. https://doi.org/10.1038/nature14664
23. Coe BP, Thu KL, Aviel-Ronen S, Vucic EA, Gazdar AF, Lam S, et al. Genomic deregulation of the E2 F/Rb pathway leads to activation of the oncogene EZH2 in small cell lung cancer. PLoS ONE. 2013;**8**:e71670. doi: 10.1371/journal.pone.0071670
24. Micalizzi DS, Maheswaran S, Haber DA. A conduit to metastasis: circulating tumor cell biology. Genes Dev. 2017;**31**:1827–40. https://doi.org/10.1101/gad.305805.117.
25. Reymond N, d'Agua BB, Ridley AJ. Crossing the endothelial barrier during metastasis. Nat Rev Cancer. 2013;**13**:858–70. doi: 10.1038/nrc3628
26. Nakazawa K, Kurishima K, Tamura T, Kagohashi K, Ishikawa H, Satoh H, et al. Specific organ metastases and survival in small cell lung cancer. Oncol Lett. 2012;**4**:617–20. https://doi.org/10.3892/ol.2012.792.

27. Lukas RV, Gondi V, Kamson DO, Kumthekar P, Salgia R. State-of-the-art considerations in small cell lung cancer brain metastases. Oncotarget. 2017;**8**:71223–33. https://doi.org/10.18632/oncotarget.19333
28. Kanteti R, Nallasura V, Loganathan S, Tretiakova M, Kroll T, Krishnaswamy S, et al. PAX5 is expressed in small-cell lung cancer and positively regulates c-Met transcription. Lab Invest. 2009;**89**:301–14. doi: 10.1038/labinvest.2008.168
29. Hogeweg P. Multilevel cellular automata as a tool for studying bioinformatic processes. Heidelberg: Springer-Verlag; 2010.
30. Macklin P, Edgerton M. Multiscale modeling of cancer—an integrated experimental and mathematical modeling approach. Cambridge, UK: Cambridge University Press; 2010.
31. Anderson A, Rejniak K, Gerlee P, Quaranta V. Modelling of cancer growth, evolution and invasion: bridging scales and models. Math Model Nat Phenom. 2007;**2**:1–29.
32. Dormann S, Deutsch A. Modeling of self-organized avascular tumor growth with a hybrid cellular automaton. In Silico Biol. 2002;**2**:393–406.
33. Kansal AR, Torquato S, Harsh GI, Chiocca EA, Deisboeck TS. Simulated brain tumor growth dynamics using a three-dimensional cellular automaton. J Theor Biol. 2000;**203**:367–82. https://doi.org/10.1006/jtbi.2000.2000.
34. Kimmel M, Axelrod D. Unequal cell division, growth regulation and colony size of mammalian cells: a mathematical model and analysis of experimental data. J Theor Biol. 1991;**153**:153–7. doi: 10.1016/s0022-5193(05)80420-5.
35. Qi AS, Zheng X, Du CY, An BS. A cellular automaton model of cancerous growth. J Theor Biol. 1993;**161**:1–12. https://doi.org/10.1006/jtbi.1993.1035.
36. Smolle J, Stettner H. Computer simulation of tumour cell invasion by a stochastic growth model. J Theor Biol. 1993;**160**:63–72. https://doi.org/10.1006/jtbi.1993.1004.
37. Düchting W. Tumor growth simulation. Comput Graph. 1990;**14**:505. doi: 10.1016/0097-8493(90)90073-7.
38. Düchting W, Ulmer W, Ginsberg T. Cancer: a challenge for control theory and computer modelling. Eur J Cancer. 1996;**32**(A):1283–92. doi: 10.1016/0959-8049(96)00075-5
39. Gardner M. Mathematical games—the fantastic combinations of John Conway's new solitaire game 'life'. Sci Am. 1970;**223**(4):120–23.
40. Ghaffarizadeh A, Friedman SH, Macklin P. BioFVM: an efficient, parallelized diffusive transport solver for 3-D biological simulations. Bioinformatics. 2016;**32**:1256–8. https://doi.org/10.1093/bioinformatics/btv730.
41. Wu F. The Potts model. Rev Mod Phys. 1982;**54**:235.
42. Steinberg M. Adhesion-guided multicellular assembly: a commentary upon the postulates, real and imagined, of the differential adhesion hypothesis, with special attention to computer simulations of cell sorting. J Theor Biol. 1975;**55**:431–43.
43. Mansury Y, Kimura M, Lobo J, Deisboeck T. Emerging patterns in tumor systems: simulating the dynamics of multicellular clusters with an agent-based spatial agglomeration model. J Theor Biol. 2002;**219**:343–70.
44. Anderson A. A hybrid mathematical model of solid tumor invasion: the importance of cell adhesion Math Med Biol. 2005;**22**:163–186
45. Gompertz, B. On the nature of the function expressive of the law of human mortality, and a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq. F. R. S. &c. By Benjamin Gompertz, Esq. F. R. S. Proc R Soc Lond. 1815;**2**:252–3. http://doi.org/10.1098/rspl.1815.0271
46. Enderling H, Chaplain MA. Mathematical Modeling of Tumor Growth and Treatment. Curr Pharm Des. 2014;**20**:1–7. doi: 10.2174/13816128196661311251504 34
47. Hahnfeldt P, Panigrahy D, Folkman J, Hlatky L. Tumor development under angiogenic signaling: a dynamical theory of tumor growth, treatment response, and postvascular dormancy. Cancer Res. 1999;**59**:4770–05.
48. Gatenby R, Gawlinski E. A reaction-diffusion model of cancer invasion. Cancer Res. 1996;**56**:5745–53.
49. Anderson A, Chaplain M, Newman E, Steele R, Thompson A. Mathematical modelling of tumour invasion and metastasis. Comput Math Methods Med. 2000;**2**:129–54. https://doi.org/10.1080/10273660008833042
50. Anderson A, Chaplain M. Continuous and discrete mathematical models of tumor-induced angiogenesis. Bull Math Biol. 1998;**60**:857–900. doi: 10.1006/bulm.1998.0042
51. Itik M, Banks S. Chaos in a three-dimensional Cancer Model. Int J Bifurcat Chaos. 2010;**20**(1): 71–79. https://doi.org/10.1142/S0218127410025417
52. Rietman E, Karp R, Tuszynski J. Review and application of group theory to molecular systems biology. Theor Biol Med Model. 2011;**8**(21):1–29. https://doi.org/10.1186/1742-4682-8-21
53. Shahzad M. Chaos control in three dimensional cancer model by state space exact linearization based on Lie. Algebra Math. 2016;**4**(2):33. https://doi.org/10.3390/math4020033
54. Han J, Jentzen A, Weinan E. Solving high-dimensional partial differential equations using deep learning. Proc Natl Acad Sci USA. 2018;**115**(34):8505–510. https://www.pnas.org/doi/10.1073/pnas.1718942115
55. Bulai IM, De Bonis MC, Laurita C, Sagaria V. Modeling metastatic tumor evolution, numerical resolution and growth prediction. Math Comput Simul. 2023;**203**:721–40. https://doi.org/10.1016/j.matcom.2022.07.002
56. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys. 2019;**378**:686–707. https://doi.org/10.1016/j.jcp.2018.10.045
57. Buchaniec S, Gnatowski M, Brus G. Integration of classical mathematical modeling with an artificial neural network for the problems with limited dataset. Energies. 2021;**14**(16):5127. https://doi.org/10.3390/en14165127

31

Mathematical models of resistance evolution under continuous and pulsed anti-cancer therapies

Einar Bjarki Gunnarsson and Jasmine Foo

31.1. Introduction

Cancer chemotherapy, the use of chemicals to treat cancer, became a standard treatment option following its successful application to childhood leukaemia and advanced Hodgkin's disease in the 1960s and 1970s [1]. Although patients often respond well initially, treatment usually fails due to the emergence of drug resistance. Resistance can develop through intrinsic mechanisms, such as poor absorption or rapid metabolism, which reduce the concentration of drug within the body [2]. It can also develop due to tumour cells acquiring specific genetic or epigenetic modifications that allow them to escape treatment. These modifications can cause resistance to many structurally and functionally unrelated drugs, e.g., via enhanced drug metabolism or drug efflux, which limits the accumulations of drug within the cell [2]. With the advent of cancer genomics and epigenomics, it has become clear that every tumour harbours cells with a variety of distinct mutational profiles and epigenetic states, some of which can confer resistance to treatment [3,4]. More recently, it has been observed that epigenetic states can be dynamic, causing cells to phenotypically switch between drug-sensitive and drug-tolerant states [5–7]. Epigenetic modifications usually occur at a much faster rate than genetic mutations, which can significantly increase the probability of treatment resistance, compared to resistance arising due to genetic mechanisms alone [8].

Chemotherapy usually attacks all rapidly dividing cells, both healthy and cancerous. It is applied with the premise that cancer cells are more susceptible due to their uncontrolled growth. An improved molecular understanding of cancer has paved the way for targeted therapeutics that inhibit specific proteins or biological pathways driving tumour evolution. Due to their specificity, targeted therapies are often better tolerated than conventional chemotherapy, but they nevertheless have their unique profile of adverse effects [9,10]. Targeted therapies can fail due to the same generic resistance mechanisms as chemotherapy, or due to drug-specific mechanisms, such as alterations of drug targets or activation of parallel or alternative biological pathways [2,11]. For example, in lung cancer, resistance to the tyrosine kinase inhibitor erlotinib can be acquired via the T790M point mutation in the epidermal growth factor receptor, which prevents the drug from binding to its target, or through MET amplification, which reactivates the PI3K/Akt signalling pathway otherwise suppressed by the drug [11].

A critical element of administering anti-cancer therapy is to determine the appropriate dosing schedule. Chemotherapy has traditionally been administered under the 'maximum tolerated dose' (MTD) paradigm. This involves giving a concentrated dose of the drug, often intravenously, followed by a prolonged treatment break of two or more weeks, to allow the patient to recover from drug toxicity [12,13]. The goal is usually to deliver as large a cumulative dose in as short a period as possible, with the premise that this will lead to maximal tumour reduction [14]. Conversely, the less toxic targeted drugs are often administered orally on a daily basis, in part due to their short half-lives [9]. Similar to chemotherapy, these drugs are commonly given at the maximum daily dose tolerated by the patient [15]. As our biological, chemical, mathematical, and medical understanding of cancer has advanced, these established treatment regimens have become increasingly challenged. Alternative dosing strategies that have been explored include low-dose continuous application of chemotherapy [16,13], and either pulsed application [17,18] or lowered daily dosing [19] for targeted therapies. As testing a large number of possible dosing strategies in pre-clinical and clinical studies is both resource-intensive and unethical, mathematical modelling offers an attractive method to search for better treatment strategies and to narrow down promising options for subsequent testing [20].

In this chapter, we discuss recent mathematical investigations of resistance evolution under continuous and pulsed anti-cancer therapies, focusing on the effect of treatment-induced resistance and phenotypic switching. In the following, continuous treatment refers to a treatment that applies a constant dose without treatment breaks.

Pulsed treatment, which can also be referred to as intermittent or periodic, is defined as a treatment that alternates (usually elevated) drug doses with drug holidays. We only consider treatment with a single anti-cancer agent, such as chemotherapy or targeted therapy. We note that this chapter is not intended to be a comprehensive review. Rather, our goals are to outline the key ingredients of mathematical models applied recently to study resistance evolution under continuous and pulsed therapies, and to extract insights into the conditions under which each strategy is preferable and under which lower cumulative doses are superior to higher doses. We conclude with a brief summary and discussion of important considerations for future work.

31.2. Key ingredients of mathematical models

We begin by outlining the main ingredients of mathematical models used recently to investigate continuous and pulsed anti-cancer therapies. A schematic overview is given in Figure 31.1.

31.2.1. Population dynamics in the absence of drug

The tumour population is usually assumed to consist of two types of cells: drug-sensitive and drug-resistant or drug-tolerant. Let $S(t)$ and $R(t)$ denote the number of sensitive and resistant (or tolerant) cells at time t, respectively. In the absence of the anti-cancer drug, we assume that sensitive (respectively resistant) cells divide at rate b_S (respectively b_R) and die at rate d_S (respectively d_R). For a stochastic model, this usually means that during an infinitesimally small time interval of length Δt, a sensitive cell divides with probability $b_S\Delta t$ and dies with probability $d_S\Delta t$. For a deterministic model, $b_S\Delta t$ (respectively $d_S\Delta t$) is the constant proportion of sensitive cells dividing (respectively dying) during a small interval of length Δt. We define $\lambda_S := b_S - d_S$ and $\lambda_R := b_R - d_R$ as the net division rates for each cell type. Resistant cells are usually assumed to proliferate more slowly, or at most as fast, as sensitive cells in the absence of drug, $\lambda_R \le \lambda_S$. This is often referred to as a 'cost of resistance', which is the notion that resistance mechanisms, such as enhanced efflux or activation of alternative signalling pathways, consume energy that would otherwise be devoted to proliferation [21,22].

We assume that sensitive cells transition to become resistant at rate μ_{SR}, and resistant cells transition back at rate μ_{RS} (Figure 31.1A). If the transition to resistance is due to a genetic mutation, it is assumed irreversible, in which case $\mu_{SR} > 0$ and $\mu_{RS} = 0$. If the transition is due to phenotypic switching, then $\mu_{SR} > 0$ and $\mu_{RS} > 0$. In many works, the transition between sensitivity and resistance is assumed to be directly influenced by the anti-cancer drug, see Section 31.2.2.

For the deterministic version of this model, the time evolution of the number of sensitive and resistant cells in the absence of drug can be described by the following system of differential equations:

$$\frac{dS}{dt} = \lambda_S \cdot S - \mu_{SR} \cdot S + \mu_{RS} \cdot R, \qquad (31.1)$$

$$\frac{dR}{dt} = \lambda_R \cdot R + \mu_{SR} \cdot S - \mu_{RS} \cdot R.$$

If $\mu_{SR} = 0$ and $\mu_{RS} = 0$, these equations lead to exponential growth at rates λ_S and λ_R per unit time for sensitive and resistant cells, respectively. If $\mu_{SR} > 0$ and $\mu_{RS} > 0$, the populations of sensitive and resistant cells eventually grow at a common exponential rate σ, given by

$$\sigma = \frac{(\lambda_S - \mu_{SR}) + (\lambda_R - \mu_{RS}) + \sqrt{((\lambda_S - \mu_{SR}) - (\lambda_R - \mu_{RS}))^2 + 4\mu_{SR}\mu_{RS}}}{2}$$

See [8] for the details. For a stochastic model, the differential equations in (31.1) describe the time evolution of the *average* number of sensitive and resistant cells.

The deterministic model in (31.1) can be modified to account for competition between sensitive and resistant cells for space and resources. To this end, it is common to consider paired logistic equations of the following or similar forms:

$$\frac{dS}{dt} = \lambda_S\left(1 - \frac{S+R}{K}\right)S - \mu_{SR} \cdot S + \mu_{RS} \cdot R, \qquad (31.2)$$

$$\frac{dR}{dt} = \lambda_R\left(1 - \frac{S+R}{K}\right)R + \mu_{SR} \cdot S - \mu_{RS} \cdot R.$$

For $\mu_{SR} = 0$ and $\mu_{RS} = 0$, these equations describe a Lotka–Volterra competition model with carrying capacity K. In this model, the net growth rates for sensitive and resistant cells are assumed to decrease linearly from λ_S and λ_R, respectively, to 0 as the total population size increases from 0 to K [23]. Sometimes, only the division rates are assumed to decrease with population size, in which case $\lambda_S\left(1 - \frac{S+R}{K}\right)S$ in (31.2) is replaced by $b_S\left(1 - \frac{S+R}{K}\right)S - d_S$, and $\lambda_R\left(1 - \frac{S+R}{K}\right)R$ by $b_R\left(1 - \frac{S+R}{K}\right)R - d_R$[22,24]. More generally, $\lambda_S\left(1 - \frac{S+R}{K}\right)S$ and $\lambda_R\left(1 - \frac{S+R}{K}\right)R$ in (31.2) can be replaced by terms of the form $f_S(S,R)\cdot S$ and $f_R(S,R)\cdot R$, where $f_S(S,R)$ and $f_R(S,R)$ are density-dependent growth rate functions. By choosing these functions appropriately, the logistic growth model in (31.2) can be replaced by other well-known tumour growth models, such as the Gompertz model or the von Bertalanffy model [25,26].

31.2.2. Dosing strategies and pharmacodynamics

To begin modelling the effect of anti-cancer therapy, we introduce the drug concentration $c(t) \ge 0$ as a function of time. Time is commonly measured in days, and $c(t)$ is often discretized so that on day i a given drug dose leads to a constant concentration c_i throughout the day, meaning that $c(t) = c_i$ for all $t \in [i-1,i)$ and $i \ge 1$. In many works, 'dose' and 'concentration' are used interchangeably, with the implication that the drug concentration reaching the tumour is proportional to the dose administered to the patient. We follow the same convention here. We note however that these are simplifications of the actual pharmacokinetics of an anti-cancer drug within the body, referring to its absorption, distribution, metabolism, and excretion [27]. In reality, the drug concentration in the body reaches a peak shortly after administration and decays over time, as is explicitly modelled in some works [28,24]. In addition, the serum concentration of the drug and the concentration reaching the tumour are not necessarily proportional to the dose given, meaning that it is not sufficient to know the dose to accurately model the response of the tumour to treatment [28,29].

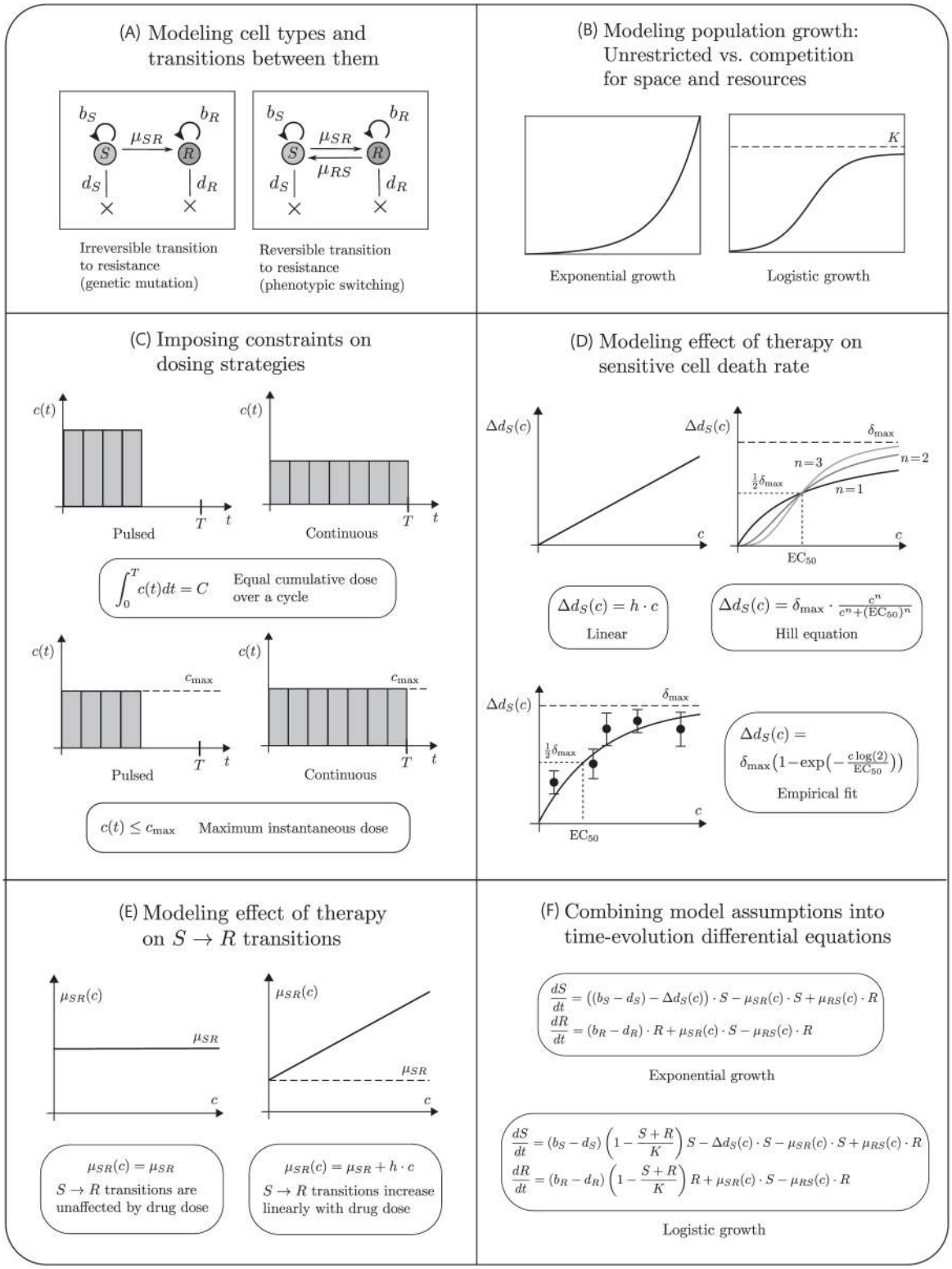


Figure 31.1. Main ingredients of recent mathematical models used to study resistance evolution under continuous and pulsed anti-cancer therapies. (A) First, the model structure must be determined, which delineates the cell types, their growth rates, and possible transitions between types. For example, in models of drug-sensitive and drug-resistant cells, genetic mutations from sensitivity to resistance are modelled by an irreversible transition, while phenotypic switching is modelled by a reversible transition. (B) Depending on the context, the model may assume unrestricted exponential growth of the cell population, or it may account for competition between cells for space and nutrients, producing a growth curve with a carrying capacity such as a logistic curve. (C) When comparing continuous and pulsed dosing strategies, some form of toxicity constraint should be imposed to establish a relationship between the size of a drug pulse and the number of days it can be tolerated. A common constraint is that of equal area under the curve across a treatment cycle, meaning that the same cumulative dose is applied over the cycle under all strategies. In some works, a maximum daily dose is considered instead, which allows a larger cumulative dose under continuous than pulsed strategies. (D) To model the effect of treatment, it is common to assume that the anti-cancer drug increases the death rate of drug-sensitive cells. The relationship $\Delta d_S(c)$ between the drug dose and the increase in sensitive cell death rate is modelled using a linear or non-linear equation such as the Hill equation or an exponential equation. The Hill equation can have either a concave or a sigmoidal shape, depending on the value of the Hill coefficient n. In some works, $\Delta d_S(c)$ is empirically determined by fitting a suitable curve to data from *in vitro* experiments. (E) For the case of drug-

To compare treatment outcomes under continuous and pulsed strategies, we must specify

(i) some form of toxicity constraint on the dosing strategy, which establishes a relationship between the size of a drug pulse and the number of days it can be tolerated,
(ii) the pharmacodynamic relationship between drug dose and cell proliferation, and
(iii) whether and how treatment affects transitions between sensitivity and resistance.

For (i), it is common to assume that treatment is conducted in cycles of T days, and that all dosing strategies have the same area under the curve, or the same cumulative dose, $\int_0^T c(t)dt = C$. This enables the application of larger daily doses for pulsed therapies than for continuous therapy. It is also common to assume a maximal instantaneous dose c_{max} so that $c(t) \le c_{max}$ for all $0 \le t \le T$. In this case, the cumulative dose allowed under continuous application is larger than under pulsed application (**Figure 31.1**C). We note that both in the pre-clinical and clinical setting, the cumulative dose tolerated under pulsed application can be significantly larger than for continuous application [17,18,30], and vice versa [16,28,31]. Thus, while the two simple assumptions presented here are mathematically convenient and useful to derive general insights, for the purposes of clinical translation, a mathematical model should be informed by an actual toxicity profile derived from pre-clinical or clinical data.

For (ii), it is common to assume that the anti-cancer drug affects the death rate of sensitive cells, according to one of the following relationships between the increase in death rate $\Delta d_S(c)$ under treatment and the dose applied c:

$$\Delta d_S(c) = h \cdot c, \tag{31.3}$$

$$\Delta d_S(c) = \delta_{max} \cdot \frac{c^n}{c^n + (EC_{50})^n}, \tag{31.4}$$

for some constants h, δ_{max}, $EC_{50} > 0$, and $n \ge 1$. In the linear model (31.3), the death rate under treatment is assumed to increase proportionally to the dose applied. In the Hill model (31.4), the drug effect levels off as the dose increases, approaching a maximum effect δ_{max} as $c \to \infty$ [32]. For $n = 1$, (31.4) takes the form of the Michaelis–Menten equation for enzyme kinetics [33]:

$$\Delta d_S(c) = \delta_{max} \cdot \frac{c}{c + EC_{50}}. \tag{31.5}$$

Here, $\Delta d_S(c)$ is concave. For $n > 1$, the Hill model (31.4) has a sigmoidal shape with an inflection point at $c = EC_{50}$, the dose at which the drug has half the maximal effect (**Figure 31.1**D).

In some works, the relationship between Δd_S and c is empirically determined by fitting a suitable curve to data from *in vitro* experiments [28,34]. In this case, an exponential curve of the following or similar form is often employed [35,36]:

$$\Delta d_S(c) = \delta_{max}\left(1 - \exp\left(-c \cdot \frac{\log(2)}{EC_{50}}\right)\right). \tag{31.6}$$

We note that for $c \ll EC_{50}$, using Taylor expansion, the non-linear equations (31.5) and (31.6) can be approximated by $\Delta d_S(c) = h \cdot c$, with $h = \delta_{max}/EC_{50}$ for (31.5) and $h = \delta_{max}\log(2)/EC_{50}$ for (31.6). Thus, the linear relationship (31.3) can be a reasonable approximation for the non-linear relationships (31.5) and (31.6) if all possible daily doses are significantly below the EC_{50} dose. Conversely, if all possible doses are significantly above EC_{50} dose, the constant function $\Delta d_S(c) = \delta_{max}$ can be a reasonable approximation for (31.5) and (31.6).

For (iii), many works assume that the transition from sensitivity to resistance is unaffected by the drug, meaning that $\Delta\mu_{SR}(c) = 0$ i.e., $\mu_{SR}(c) = \mu_{SR}$ for all $c \ge 0$. However, both chemotherapy and targeted therapies have been observed to induce mutations in cancer driver genes [37–40] and to drive phenotypic switches to drug-tolerant or drug-resistant states [6,35,41]. This is analogous to stress-induced mutagenesis observed in bacterial populations in response to antibiotics [42]. To incorporate this form of treatment-induced resistance, it is common to assume that transitions from sensitivity to resistance linearly increase with drug dose, $\Delta\mu_{SR}(c) = h \cdot c$. In some works, the drug is assumed to increase transitions from sensitivity to resistance independently of the dose, meaning that $\Delta\mu_{SR}(c) = k$ for some constant $k > 0$. As before, we note that non-linear relationships of the forms (31.5) and (31.6) can incorporate both these extremes. For the case of phenotypic switching, the rate of switching from resistance to sensitivity $\mu_{RS}(c)$ can also be assumed to depend on the drug dose.

Finally, in most of the works we consider, resistant cells are assumed to be 'fully resistant', meaning that they are not affected by the drug at any concentration. Otherwise, the effect of drug on resistant cells can be modelled using the relationships (31.3)–(31.6) or similar forms, with the constants h, δ_{max}, EC_{50}, and n taking different values for resistant cells than for sensitive cells.

31.2.3. Comparing time evolution across different strategies

Gathering all of the aforementioned ingredients, and assuming a deterministic model, the model dynamics for exponential growth can be captured by the following differential equations:

$$\begin{aligned} \frac{dS}{dt} &= (\lambda_S - \Delta d_S(c)) \cdot S - \mu_{SR}(c) \cdot S + \mu_{RS}(c) \cdot R, \\ \frac{dR}{dt} &= \lambda_R \cdot R + \mu_{SR}(c) \cdot S - \mu_{RS}(c) \cdot R. \end{aligned} \tag{31.7}$$

For logistic growth, the treatment effect $\Delta d_S(c)$ is usually treated as an exponential decay rate. This assumes that a fixed dose of the drug kills a certain proportion of tumour cells, independently of the size of the tumour, which is the 'log-kill model' [13,14]:

$$\begin{aligned} \frac{dS}{dt} &= \lambda_S\left(1 - \frac{S+R}{K}\right)S - \Delta d_S(c) \cdot S - \mu_{SR}(c) \cdot S + \mu_{RS}(c) \cdot R, \\ \frac{dR}{dt} &= \lambda_R\left(1 - \frac{S+R}{K}\right)R + \mu_{SR}(c) \cdot S - \mu_{RS}(c) \cdot R. \end{aligned} \tag{31.8}$$

When combined with a dosing function $c(t)$, these differential equations can be used to compute the response of the tumour to any dosing strategy and to compare their efficacy.

31.3. Review of previous work

In this section, we discuss recent mathematical investigations of resistance evolution under continuous and pulsed anti-cancer therapies. As stated in the introduction, we focus on the effects of treatment-induced resistance, as discussed in Section 31.3.2, and phenotypic switching, as discussed in Section 31.3.3. We begin by briefly discussing in Section 31.3.1 a set of investigations involving spontaneous resistance evolution through genetic mutation, which led to the first phase 1 clinical trial testing a mathematically optimized targeted treatment [29].

31.3.1. Earlier work on spontaneous resistance evolution

In [43], Foo and Michor study continuous and pulsed administration for targeted therapy, using the stochastic version of the exponential model (31.7). They propose a methodology for determining the best treatment schedule, where the objective is to either minimize the probability of resistance or the expected size of the resistant population. They suggest using clinical data to determine the relationship $t_{on}(c)$ between a daily dose c and the number of days t_{on} the dose can be tolerated during each treatment cycle, and to determine $\Delta d_S(c)$ empirically using *in vitro* experiments. Together, these functions give rise to a relationship $t_{on}(\Delta d_S)$ between the effect of drug on sensitive cells and the length of drug application during each cycle. The authors find that the preferability of high-dose pulsed schedules to more even schedules critically depends on the degree of convexity of the curve $\Delta d_S \mapsto t_{on}(\Delta d_S)$.

Using the same modelling framework, Chmielecki et al. [34] propose a novel strategy for treating EGFR-mutant non-small-cell lung cancer with erlotinib, where a large weekly pulse (20 μM) is applied in conjunction with a small continuous dose (1 μM) (see also Stein et al. [28]). The small continuous dose is applied to control the sensitive cell population. The large weekly pulse is applied since the division rate of resistant cells is observed to linearly decrease with drug dose, meaning that resistant cells are only partially resistant. Based on the results of [34], a phase 1 clinical trial was conducted, where twice-weekly pulses of erlotinib were combined with daily low doses [29]. The MTD was determined to be 1,200 mg on days 1–2 and 50 mg on days 3–7. The schedule was well tolerated, but it did not improve progression-free survival or prevent the emergence of EGFR T790M resistance compared to the standard daily dosing schedule. One possible explanation is that the median peak concentration under the drug pulses did not reach the 20 μM concentration studied in the pre-clinical model [34]. In fact, the pharmacokinetic data in [29] indicates that increasing the pulse dose from 600 to 1,350 mg does not lead to a significant increase in peak plasma concentration. It also shows significant inter-patient variability in plasma concentration, which suggests the importance of patient-specific treatment optimization and evaluating larger cohorts for efficacy.

31.3.2. The effect of treatment-induced resistance

We next discuss works which assume that transitions from sensitivity to resistance are elevated by the anti-cancer drug. To set the stage, we begin by noting that, in a recent work by Russo et al. [35], the authors investigate the responses of two colorectal cancer cell lines to increasing doses of targeted therapies. Using a mathematical model of drug-sensitive and drug-tolerant 'persister' cells, they conclude that transitions from sensitivity to tolerance are predominantly induced by the targeted therapies. For one of the cell lines (WiDr cells), the rate of transition linearly increases with the drug dose, while for the other cell line (DiFi cells), the rate is constant as a function of the dose, according to the best model fit under their framework. This work constitutes recent indication that the modelling assumptions discussed in Section 31.2.2 and applied in some of the following works are reasonable.

Greene et al. [44] consider a logistic competition model of the form (31.8), where both $\Delta d_S(c)$ and $\Delta \mu_{SR}(c)$ are assumed linear in c. They show an example where under equal daily dosing, a pulsed schedule outperforms a continuous schedule, while the opposite is true when transitions to resistance are unaffected by the drug, $\Delta \mu_{SR}(c) = 0$. They conclude that the level of resistance induction by the drug is a clinically significant parameter. It should be noted that Greene et al. only consider one value for the daily dose in this comparison, and this dose is later shown to be suboptimal for continuous application. In other words, they do not compare the optimal continuous and pulsed schedules (see [45] for a later work by the same authors which applies optimal control theory to the problem). In a more recent work, Kuosmanen et al. [46] study the same model as in [44], except $\Delta d_S(c)$ is now assumed to follow the Hill function (31.4) with $n = 2.3$, which has a sigmoidal shape. Starting from a population of sensitive cells only, the authors formulate an optimal control problem that aims to minimize the probability of resistance evolution or the size of the resistant population at a fixed time. The only constraint on the dosing strategy is a maximum instantaneous dose, $c(t) \leq c_{max}$. For the objective of minimizing the probability of resistance, the authors argue that the optimal strategy is often close to a continuous schedule, which leads them to study the simpler problem of finding the optimal continuous dose. They show that under even a small level of treatment-induced resistance, a dose significantly below c_{max} becomes optimal. This result is at least partly driven by the fact that in their model, the effect of drug on sensitive cells $\Delta d_S(c)$ levels off for large drug doses, while the transition rate effect $\Delta \mu_{SR}(c)$ is linear in c.

Kuosmanen et al. [46] note that if the objective is to minimize the probability of resistance, then in the absence of drug-induced resistance, $\Delta \mu_{SR}(c) = 0$, the optimal continuous treatment is the maximal dose $c(t) = c_{max}$. As indicated by both [44] and [46], this is not necessarily the case under a different objective, such as maximizing the time until the tumour reaches a certain size N_0. Indeed, under the logistic model (31.8), if N_0 is close to the carrying capacity of the tumour, the growth of resistant cells can be significantly suppressed by maintaining the sensitive cell population close to N_0. In this case, applying the largest dose, which maximally reduces the size of the sensitive population, is counterproductive even when the drug does not induce resistance. An example is shown in **Figure 31.2**. We note that the idea of maintaining the sensitive cell population to promote competition with resistant cells forms the basis of adaptive therapy strategies proposed by Gatenby [21]. We refer to the review [47] and the recent survey of open questions [48] for further discussion of these strategies.

In Angelini et al. [49], the authors investigate an exponential growth model of the form (31.7), where both $\Delta d_S(c)$ and $\Delta \mu_{SR}(c)$ are assumed to be logistic functions of the form $c \mapsto k(1 + e^{-r(c-s)})^{-1}$,

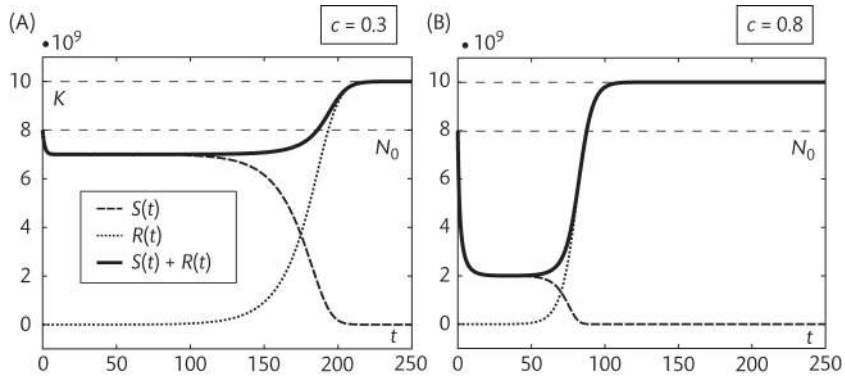


Figure 31.2. When sensitive and resistant cells compete for space and resources, maintaining a large number of sensitive cells under treatment can delay resistance evolution. Here, treatment is initiated when the tumour is at size $N_0 = 8 \times 10^9$, which is close to the carrying capacity of the tumour, $K = 10^{10}$. The time evolution of the model follows the logistic equations (31.8) with $\lambda_S = 1$, $\lambda_R = 0.2$, $\Delta d_S(c) = c$, $\mu_{SR}(c) = 10^{-6}$, and $\mu_{RS}(c) = 0$ for $c \geq 0$. In other words, the death rate of sensitive cells linearly increases with drug dose, and transitions from sensitivity to resistance are unaffected by the drug. In (A), a constant drug dose $c = 0.3$ is applied, and in (B), the constant dose $c = 0.8$ is applied. The time at which the tumour returns to its original size N_0 is significantly longer under the lower drug dose, which is due to the fact that the large sensitive population suppresses the growth rate $\lambda_R\left(1-\frac{S+R}{K}\right)$ of resistant cells.

which has a sigmoidal shape. The authors only consider continuous treatment, and they measure treatment success in terms of the time at which the tumour rebounds to its pretreatment size. Their results indicate that when the EC_{50} concentration of the drug is similar for $\Delta d_S(c)$ and $\Delta \mu_{SR}(c)$, the maximal continuous dose is optimal, although a wide range of intermediate doses may be almost as effective as the optimal dose. However, if the EC_{50} concentration is significantly larger for $\Delta \mu_{SR}(c)$ than for $\Delta d_S(c)$, meaning that drug-induced transitions to resistance become significant only at doses that are already effective at killing drug-sensitive cells, an intermediate dose becomes optimal. These results highlight the need to understand whether the dose range over which drug-induced transitions to resistance increase coincides with the dose range of increased cell kill. We note that since Angelini et al. use an exponential growth model, the competitive dynamics that contribute to the optimality of intermediate doses in Greene et al. [44] do not play a role.

We conclude this section by discussing a recent work by Mathur et al. [50], which considers both single-agent and combination anticancer therapies. For single-agent therapy, the authors use an exponential model of the form (31.7). They assume an equal cumulative dose across schedules and that $\Delta d_S(c)$ is linear in c. This means that if continuous treatment increases the death rate of sensitive cells by δ, a pulsed treatment applied for t_{on} days during a cycle of length T increases the death rate by $\delta T/t_{on}$. The authors also allow treatment to increase transitions from sensitivity to resistance, which occurs in a dose-independent manner as far as we can tell, $\Delta \mu_{SR}(c) = k$ with $k \geq 0$. Mathur et al. randomly generate parameters for 50,000 synthetic tumours, and for each tumour, they randomly sample 50 sets of values for t_{on} and T/t_{on}, generating 50 pulsed schedules. By comparing continuous therapy with the best pulsed therapy for each synthetic tumour, they find that pulsed application outperforms continuous application in all cases, delaying tumour recurrence by up to >140×. This substantial preference for pulsed schedules is likely driven by the assumptions that the effect of drug on sensitive cells does not level off at large doses and that treatment-induced resistance is dose independent, the latter of which puts a penalty on the length of drug application as opposed to the dose size.

31.3.3. The effect of phenotypic switching

We conclude by discussing works that incorporate phenotypic switching between sensitivity and resistance or tolerance. In [51], Akhmetzhanov et al. study a model of resistance in BRAF-mutant melanoma, where cell phenotypes are determined by the activities of two mutually inhibitory biological pathways. The activation dynamics of each pathway within each cell is modelled by a particle undergoing Brownian motion inside a double-well potential. A targeted anti-cancer agent changes the structure of the double-well potential associated with the main pathway. This inhibits the proliferation of main-pathway-active cells and facilitates activity transitions from the main to the alternative pathway. After making some simplifying assumptions, the authors arrive at a model with two cell states, sensitive and resistant, and they derive mean-field differential equations that have the same form as (31.7) (see expressions (3) and (4) and the paragraph following them in [51]). However, the functions $\Delta d_S(c)$, $\Delta \mu_{SR}(c)$, and $\Delta \mu_{RS}(c)$ have more complex forms than those discussed in Section 2.2 since they emerge from the underlying model of particles moving inside potentials.

Using the tumour size six months after treatment initiation as the objective, Akhmetzhanov et al. compare continuous and pulsed treatments with a treatment derived from optimal control theory. They find that the best continuous strategy applies an intermediate dose but do not discuss the underlying cause. We note that by the results of Angelini et al. [49], the fact that an anti-cancer drug induces transitions from sensitivity to resistance does not necessarily lead to optimality of an intermediate dose. According to expression (4) and Figure S2 of [51], the effect of drug on sensitive cells levels off at higher doses. In addition, the drug both encourages transitions to resistance and discourages transitions back in a dose-dependent manner, each of which increases the time spent in the resistant state relative to the sensitive state. In combination, these factors lead to the optimality of an intermediate dose. The relative importance of each modelling assumption in producing this result is an interesting avenue for further study.

Akhmetzhanov et al. also find that the best pulsed treatment outperforms the best continuous treatment by 12.2%, and the

treatment derived from optimal control theory leads to a further improvement of 4.6%. The optimal treatment is applied without breaks and involves two dose changes over the six-month period. Assuming that there is no resistant cell at the beginning of treatment, the maximal dose c_{max} is applied for a short initial period. This increases the proportion of resistant cells in the population to a certain level. Then, a significantly lower dose is used to maintain the resistant cells at the same proportion, which is possible due to the phenotypic switching between types. Finally, c_{max} is applied again for a short period since dosing at c_{max} is maximally effective at killing sensitive cells and reducing the tumour burden in the short term. Thus, in the model of [51] there appears to be an optimal population composition that can be maintained by a low drug dose, and the best treatment shifts the original population to this composition as quickly as possible and subsequently maintains it. We finally note that the best pulsed schedule, which performs similarly to the optimal schedule, induces the proportion of resistant cells to oscillate around the optimal level.

Cassidy et al. [24] have recently studied a model of phenotypic switching between drug sensitivity and drug tolerance, where switches occur on cell divisions, and the probability of switching depends on the dividing cell's age. Individually, each drug-tolerant cell has a negative net division rate, even in the absence of drug. However, drug-tolerant cells cooperate to divide faster as their frequency in the population increases. More precisely, the division rate of tolerant cells is assumed to increase as a function of the ratio θ between tolerant and sensitive cells according to the Hill equation (31.4). The anti-cancer drug affects sensitive cells according to the Michaelis–Menten relationship (31.5), while drug-tolerant cells are unaffected by it. Under this model, an aggressive treatment strategy that maximally reduces the sensitive cell population triggers the cooperative growth of drug-tolerant cells. Cassidy et al. therefore propose an adaptive strategy that aims at controlling the proportion of tolerant cells in the population. Under this strategy, every T days a fixed drug dose is applied if and only if the proportion between tolerant and sensitive cells is below a certain threshold. The threshold is chosen as the largest proportion θ^* for which the net division rate of tolerant cells is nonpositive. The authors parametrize their model using experimental data on non-small-cell lung cancer tumour spheroids treated with chemotherapy. They show that in this context the adaptive therapy leads to sustained tumour decay, whereas pulsed therapy applying a fixed dose every seven days drives the population to drug tolerance and eventually loses its effectiveness. We note that the two treatments compared are not necessarily the optimal versions of each strategy.

The previous two works suggest that under phenotypic switching, the best treatment strategy may involve maintaining the population at a certain desirable composition. As acknowledged by Akhmetzhanov et al. [51], implementing a strategy that necessitates frequent monitoring of the population composition requires the existence of non-genetic biomarkers that can be tracked non-invasively and inexpensively, which may impede clinical translation. As such, the identification of simple fixed continuous or pulsed strategies that perform comparably to the optimal schedule can be clinically useful [51]. In addition, we note that a model of switching between two types does not consider the evolution of more permanent and irreversible resistance mechanisms, whether due to genetic mutations or epigenetic reprogramming [8].

Finally, in the context of metastatic colorectal cancer treated with the anti-EGFR inhibitor panitumumab, Yin et al. [52] consider a three-type model of sensitive cells, cells harbouring a KRAS mutation rendering them resistant to panitumumab, and cells harbouring a secondary mutation rendering them resistant to a hypothetical second treatment targeting KRAS-mutated cells. They also model the release of DNA fragments into the bloodstream by each of the two mutated populations, creating circulating tumour DNA (ctDNA). They derive differential equations of the form (31.7), with three cell types, where the dosing function $c(t)$ for each drug only takes the values 0 or 1, meaning that only a single dose is considered for each drug. Transitions from sensitivity to resistance for each drug are assumed to be only active in the presence of the drug, and transitions back are only active in its absence. After parametrizing the model using patient data, the authors find that a pulsed treatment applying panitumumab for eight weeks out of a 12-week cycle prolongs the time until the tumour reaches its pretreatment size from 52 to 60 weeks compared to continuous application. They also show that an adaptive therapy that alternately applies panitumumab and the hypothetical second drug based on measured ctDNA levels further extends the rebounding time to 114–132 weeks.

31.4. Conclusions

The examples discussed in Section 31.3 demonstrate how the relative effectiveness of continuous versus pulsed schedules, and lower versus higher doses, crucially depends on the modelling assumptions outlined in Section 31.2. For example, if the effect of treatment on the sensitive cell death rate $\Delta d_S(c)$ levels off at higher doses, while the effect on transitions from sensitivity to resistance $\Delta\mu_{SR}(c)$ is linear, a continuous low-dose treatment may become optimal as in Kuosmanen et al. [46]. Conversely, if $\Delta d_S(c)$ is linear in c, while the drug increases transitions to resistance independently of the dose, $\Delta\mu_{SR}(c) = k$, then short, elevated pulses followed by treatment breaks may become optimal as in Mathur et al. [50]. As indicated in Section 31.2.2, we note that these two regimes can emerge if $\Delta d_S(c)$ and $\Delta\mu_{SR}(c)$ both follow a non-linear curve such as (31.5) or (31.6), where the EC_{50}'s for each curve are of different orders of magnitude.

Using the logistic growth model (31.8) creates competitive dynamics where it can be beneficial to maintain a large population of sensitive cells. This can create a preference for lower doses, even in the absence of drug-induced resistance, if the objective is to maximally delay tumour recurrence as opposed to preventing resistance evolution. For the case of phenotypic switching, applying treatment breaks or varying the dose over time may become beneficial, e.g., if the drug both increases transitions from sensitivity to resistance and inhibits transitions back in a dose-dependent manner, if resistant cells cooperate to grow faster at higher densities, or if transitions to resistance only occur in the presence of drug, while transitions back only occur in its absence. We also note that under phenotypic switching, it may become optimal to maintain a certain balance between sensitive and resistant cells that maximally inhibits long-term tumour growth.

The dose dependence of transitions between sensitivity and resistance remains poorly understood; this currently limits the clinical relevance of the mathematical studies discussed in Sections 31.3.2 and 31.3.3. The work of Russo et al. [35] is a recent example which

infers $\Delta\mu_{SR}(c)$ under a Bayesian framework using experimental data for two colorectal cancer cell lines treated with targeted therapies. In our opinion, mathematical modelers can make an important contribution to further research on therapy-induced resistance by investigating optimal experimental designs to jointly infer $\Delta d_S(c)$, $\Delta\mu_{SR}(c)$, and $\Delta\mu_{RS}(c)$ in a robust manner.

Most mathematical investigations to date have assumed mathematically convenient forms for $\Delta\mu_{SR}(c)$ and $\Delta\mu_{RS}(c)$, and they have generally considered one specific parameter regime. It is common to conduct sensitivity analysis to show how results change when the values of key model parameters are changed. However, it is equally important to tease apart how individual components of the mathematical model, including the ones depicted in Figure 31.1, affect qualitative results, such as whether continuous or pulsed schedules and whether low or high doses are preferred. Ideally, this calls for deriving theoretical results under a variety of different modelling assumptions. Such analysis is crucial for developing intuition for how the nature of interactions between cells in a particular cancer and the pharmacodynamics of a particular drug lead to specific treatment recommendations, and for understanding why different treatment recommendations emerge from different mathematical investigations. This is especially important in light of inter-patient heterogeneity, which suggests viewing each patient as harbouring a unique disease involving unique pharmacokinetics and pharmacodynamics.

REFERENCES

1. DeVita VT, Jr, Chu E. A history of cancer chemotherapy. Cancer Res. 2008;**68**:8643–53.
2. Foo J, Michor F. Evolution of acquired resistance to anti-cancer therapy. J Theor Biol. 2014;**355**:10–20.
3. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;**481**:306–13.
4. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. Cell. 2017;**168**:613–28.
5. Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell. 2010;**141**:69–80.
6. Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature. 2017;**546**:431–5.
7. Brown R, Curry E, Magnani L, Wilhelm-Benartzi CS, Borley J. Poised epigenetic states and acquired drug resistance in cancer. Nat Rev Cancer. 2014;**14**:747–53.
8. Gunnarsson EB, De S, Leder K, Foo J. Understanding the role of phenotypic switching in cancer drug resistance. J Theor Biol. 2020;**490**:110162.
9. Gerber DE. Targeted therapies: a new generation of cancer treatments. Am Fam Physician. 2008;**77**:311–9.
10. Liu S, Kurzrock R. Toxicity of targeted therapy: implications for response and impact of genetic polymorphisms. Cancer Treat Rev. 2014;**40**:883–91.
11. Neel DS, Bivona TG. Resistance is futile: overcoming resistance to targeted therapies in lung adenocarcinoma. NPJ Precis Oncol. 2017;**1**:1–6.
12. Kerbel RS, Kamen BA. The anti-angiogenic basis of metronomic chemotherapy. Nat Rev Cancer. 2004;**4**:423–36.
13. Benzekry S, Pasquier E, Barbolosi D, Lacarelle B, Barl´esi F, Andr´e N, et al. Metronomic reloaded: theoretical models bringing chemotherapy into the era of precision medicine. Semin. Cancer Biol. Elsevier. 2015;**35**:53–61.
14. Norton L. The Norton-Simon hypothesis revisited. Cancer Treat Rep. 1986;**70**:163–9.
15. Yesilkanal AE, Johnson GL, Ramos AF, Rosner MR. New strategies for targeting kinase networks in cancer. J Biol Chem. 2021;**297**:101128.
16. Lien K, Georgsdottir S, Sivanathan L, Chan K, Emmenegger U. Low-dose metronomic chemotherapy: a systematic literature analysis. Eur J Cancer. 2013;**49**:3387–95.
17. Milton DT, Azzoli CG, Heelan RT, Venkatraman E, Gomez JE, Kris MG, et al. A phase I/II study of weekly high-dose erlotinib in previously treated patients with nonsmall cell lung cancer. Cancer: Interdisciplinary Int J Am Cancer Soc. 2006;**107**:1034–41.
18. Grommes C, Oxnard GR, Kris MG, Miller VA, Pao W, Holodny AI, et al. 'Pulsatile' high-dose weekly erlotinib for CNS metastases from EGFR mutant non-small cell lung cancer. Neuro-oncology 2011; **13**:1364–9.
19. Yeo WL, Riely GJ, Yeap BY, Lau MW, Warner JL, Bodio K, et al. Erlotinib at a dose of 25 mg daily for non-small cell lung cancers with EGFR mutations. J Thorac Oncol. 2010;**5**:1048–53.
20. Michor F, Beal K. Improving cancer treatment via mathematical modeling: surmounting the challenges is worth the effort. Cell. 2015;**163**:1059–63.
21. Gatenby RA. A change of strategy in the war on cancer. Nature. 2009;**459**:508–9.
22. Strobl MA, West J, Viossat Y, Damaghi M, Robertson-Tessi M, Brown JS, et al. Turnover modulates resistance costs in adaptive therapy. Cancer Res. 2021;**81**:1135–47.
23. Wangersky PJ. Lotka-Volterra population models. Annu Rev Ecol Syst. 1978;**9**:189–218.
24. Cassidy T, Nichol D, Robertson-Tessi M, Craig M, Anderson AR. The role of memory in non-genetic inheritance and its impact on cancer treatment resistance. PLoS Comput Biol. 2021; **17**:e1009348.
25. Benzekry S, Lamont C, Beheshti A, Tracz A, Ebos JM, Hlatky L, et al. Classical mathematical models for description and prediction of experimental tumor growth. PLoS Comput Biol. 2014;**10**:e1003800.
26. Viossat Y, Noble R. A theoretical analysis of tumour containment. Nat Ecol Evol. 2021;**5**:826–35.
27. Garattini S. Pharmacokinetics in cancer chemotherapy. Eur J Cancer. 2007;**43**:271–82.
28. Stein S, Zhao R, Haeno H, Vivanco I, Michor F. Mathematical modeling identifies optimum lapatinib dosing schedules for the treatment of glioblastoma patients. PLoS Comput Biol. 2018;**14**:e1005924.
29. Yu H, Sima C, Feldman D, Liu L, Vaitheesvaran B, Cross J, et al. Phase 1 study of twice weekly pulse dose and daily low-dose erlotinib as initial treatment for patients with EGFR-mutant lung cancers. Ann Oncol. 2017;**28**:278–84.
30. Amin DN, Sergina N, Ahuja D, McMahon M, Blair JA, Wang D, et al. Resiliency and vulnerability in the HER2-HER3 tumorigenic driver. Sci Transl Med. 2010;**2**:16ra7.
31. West J, Newton PK. Chemotherapeutic dose scheduling based on tumor growth rates provides a case for low-dose metronomic high-entropy therapies. Cancer Res. 2017;**77**:6717–28.

32. Goutelle S, Maurin M, Rougier F, Barbaut X, Bourguignon L, Ducher M, et al. The Hill equation: a review of its capabilities in pharmacological modelling. Fundam Clin Pharmacol. 2008;**22**:633–48.
33. Michaelis L, Menten ML. Die kinetik der invertinwirkung. Biochem Zeitschrift. 1913;**49**:333–69.
34. Chmielecki J, Foo J, Oxnard GR, Hutchinson K, Ohashi K, Somwar R, et al. Optimization of dosing for EGFR-mutant non–small cell lung cancer with evolutionary cancer modeling. Sci Transl Med. 2011; **3**:90ra59.
35. Russo M, Pompei S, Sogari A, Corigliano M, Crisafulli G, Puliafito A, et al. A modified fluctuation-test framework characterizes the population dynamics and mutation rate of colorectal cancer persister cells. Nat Genet. 2022;**54**:976–84.
36. Yang EY, Howard GR, Brock A, Yankeelov TE, Lorenzo G. Mathematical characterization of population dynamics in breast cancer cells treated with doxorubicin. Front Mol Biosci. 2022; **9**:972146.
37. Szikriszt B, Póti Á, Pipek O, Krzystanek M, Kanu N, Molnár J, et al. A comprehensive survey of the mutagenic impact of′ common cancer cytotoxics. Genome Biol. 2016;**17**:1–16.
38. Venkatesan S, Swanton C, Taylor BS, Costello JF. Treatment-induced mutagenesis and selective pressures sculpt cancer evolution. Cold Spring Harbor Perspect Med. 2017;**7**:a026617.
39. Russo M, Crisafulli G, Sogari A, Reilly NM, Arena S, Lamba S, et al. Adaptive mutability of colorectal cancers in response to targeted therapies. Science. 2019;**366**:1473–80.
40. Cipponi A, Goode DL, Bedo J, McCabe MJ, Pajic M, Croucher DR, et al. MTOR signaling orchestrates stress-induced mutagenesis, facilitating adaptive evolution in cancer. Science. 2020;**368**:1127–31.
41. Goldman A, Majumder B, Dhawan A, Ravi S, Goldman D, Kohandel M, et al. Temporally sequenced anticancer drugs overcome adaptive resistance by targeting a vulnerable chemotherapy-induced phenotypic transition. Nat Commun. 2015;**6**:6139.
42. Long H, Miller SF, Strauss C, Zhao C, Cheng L, Ye Z, et al. Antibiotic treatment enhances the genome-wide mutation rate of target cells. Proc Natl Acad Sci. 2016;**113**:E2498–505.
43. Foo J, Michor F. Evolution of resistance to targeted anti-cancer therapies during continuous and pulsed administration strategies. PLoS Comput Biol. 2009;**5**:e1000557.
44. Greene JM, Gevertz JL, Sontag ED. Mathematical approach to differentiate spontaneous and induced evolution to drug resistance during cancer treatment. JCO Clin Cancer Inform. 2019;**3**:1–20.
45. Greene JM, Sanchez-Tapia C, Sontag ED. Mathematical details on a cancer resistance model. Front Bioeng Biotechnol. 2020;**8**:501.
46. Kuosmanen T, Cairns J, Noble R, Beerenwinkel N, Mononen T, Mustonen V. Drug-induced resistance evolution necessitates less aggressive treatment. PLoS Comput Biol. 2021;**17**:e1009418.
47. Belkhir S, Thomas F, Roche B. Darwinian approaches for cancer treatment: benefits of mathematical modeling. Cancers. 2021;**13**:4448.
48. West J, Adler F, Gallaher J, Strobl M, Brady-Nicholls R, Brown JS, et al. A survey of open questions in adaptive therapy: bridging mathematics and clinical translation. arXiv preprint arXiv:2210.12062 2022.
49. Angelini E, Wang Y, Zhou JX, Qian H, Huang S. A model for the intrinsic limit of cancer therapy: duality of treatment-induced cell death and treatment-induced stemness. PLOS Comput Biol. 2022;**18**:e1010319.
50. Mathur D, Taylor BP, Chatila WK, Scher HI, Schultz N, Razavi P, et al. Optimal strategy and benefit of pulsed therapy depend on tumor heterogeneity and aggressiveness at time of treatment initiation early tumor properties predict optimal treatment strategy. Mol Cancer Ther. 2022;**21**:831–43.
51. Akhmetzhanov AR, Kim JW, Sullivan R, Beckman RA, Tamayo P, Yeang CH. Modelling bistable tumour population dynamics to design effective treatment strategies. J Theor Biol. 2019; 474:88–102.
52. Yin A, Hasselt JG van, Guchelaar HJ, Friberg LE, Moes DJA. Anti-cancer treatment schedule optimization based on tumor dynamics modelling incorporating evolving resistance. Sci Rep. 2022;**12**:1–14.

32

Integrating *in silico* models with *ex vivo* data for designing better combinatorial therapies in cancer

Cameron Meaney, Dorsa Mohammadrezaei, and Mohammad Kohandel

32.1. Introduction

Historically, cancer therapy has primarily consisted of crude surgeries and singly administered compounds with severe side effects. In some ways, this description has changed little over the years, with many cancers remaining incurable and the standard of care treatments improving only slightly. In other ways, however, cancer medicine has changed dramatically, spurred on by progress in adjacent fields, such as medical imaging, engineering, and genomics. Modern cancer treatment plans utilize complex, sophisticated strategies involving combinations of therapies and relying on advanced technologies, which have led to earlier detection, better tumour targeting, and increased survival times. Much of this progression is owed to the ever-growing incorporation of mathematical and computational modelling into cancer medicine, both in better understanding the disease and improving its treatment. Indeed, the use of numerical methods in medical science is only expected to increase with time, with mathematical oncology poised to benefit particularly.

However, a key challenge in mathematical oncology is that models are built on data, which is often hard to come by in a medical context. Preclinical research typically begins with *in vitro* studies, followed by standard animal testing. However, as many of these models rely on cells with undermined immune systems or cells of a non-human origin, clinical translation is rare, with some studies putting it as low as 8% [1]. In the case of more advanced techniques, such as genome sequencing of tumour cells, these often similarly fail to accurately predict or describe phenotypes since many other biological factors are at play, such as protein expression, genetic mutations, microenvironment complexities, oncogenic amplifications, and post-transcriptional alteration that must be considered [2]. Simultaneous factoring in these non-genomic biological factors with existing experimental techniques is difficult, and methods for explaining interactions between them are lacking [3, 4]. In the case of human data specifically, financial, technological, and ethical considerations often prevent collecting enough data to inform models as much as would be desired. This can leave model elements, such as type, form, initial condition, parameter values, and validation data, needing to be artificially generated or estimated from insufficient observations. This poses a significant challenge for modern models of cancer systems. As a result, the development of a precise experimental method, presenting tumour phenotypes and closely reflecting the *in vivo* situation systems, is required for more precise disease modelling and drug screening.

Ex vivo models can provide a better experimental framework for understanding the influence of the tumour microenvironment and heterogeneity on cancer cell functions and have accordingly caught the attention of researchers in recent years. *Ex vivo methods* are defined as any technique that involves performing a drug screening on tumour cells or solid tumour tissue directly obtained from patients [5]. Although only a few results derived from *ex vivo* experiments have made their way to cancer clinics, these few results have shown remarkable outcomes demonstrating that there are alternatives when conventional treatment strategies have been exhausted [3]. The motivation behind *ex vivo* cultures of patient-derived tumour samples is to maintain the original tissue matrix as much as possible in experiment so that inter-patient and intra-patient variation can be included in the initial phases of drug development [6, 7]. *Ex vivo* methods were first used in haematologic malignancies by taking malignant cells from a blood sample, and *ex vivo* methods for solid tumours have proven more challenging since they frequently necessitate an invasive biopsy or surgical resection, and the acquired sample has limits in terms of accurately representing the complete range of heterogeneity. However, some *ex vivo* techniques have tried to replicate solid tumour heterogeneity in the experiments taking place outside of the patient body, with varying degrees of success, e.g. patient-derived cell lines [8, 9, 10], patient-derived cultures

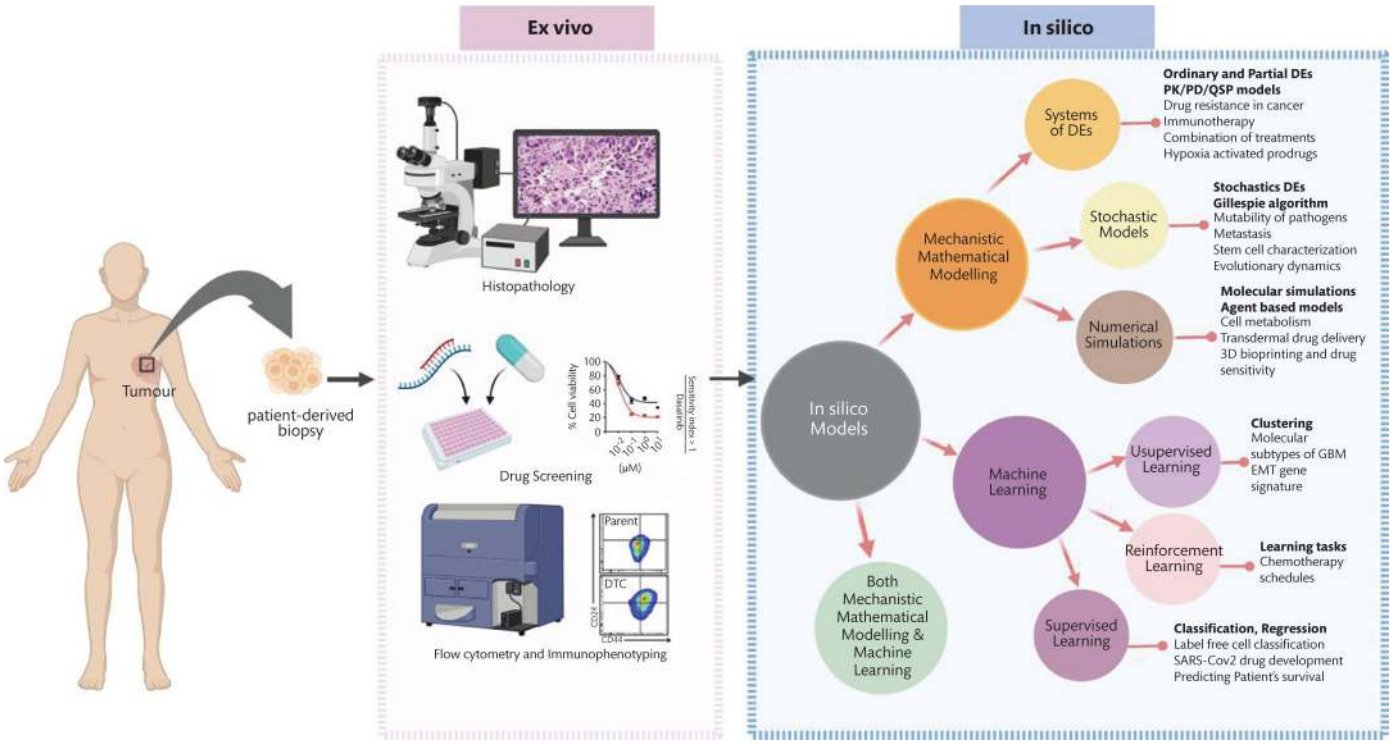


Figure 32.1. Integration of *ex vivo* data with *in silico* models for designing more efficient combinatorial therapies in cancer. The left image shows typical steps in *ex vivo* experiments. Data such as histopathological images, drug screening, cell viability, and immunophenotype of tumour cells determined by flow cytometry are collected in these studies that can be used for training and validation *in silico* simulations. The right image presents different *in silico* models, including mechanistic mathematical modelling, ML, and a combination of mechanistic mathematical modelling and ML. *Source:* The images are created with BioRender.com.

[11,12], patient-derived organoids [13,14,15,16], organoids on chips [17,18,19,20,21], assembloids [22,23,24], patient-derived explants [25,26], and tumour xenografts [27,28].

Similarly increasing in their utility and popularity in cancer research, *in silico* techniques have proven their immeasurable worth to modern medical research. Indeed, countless examples exist throughout the literature of works in which quantitative, *in silico* analysis has furthered our understanding of tumour development, improved an existing therapy, or aided in drug development, to name a few. In cancer medicine, *in silico* methods are typically employed for goals, such as modelling the spatiotemporal dynamics of a tumour system, predicting random events throughout tumour progression, or classifying patients into various groups. With these different goals, naturally, different *in silico* methods can be useful additions to analysis. These can be loosely broken into two categories: mechanistic modelling and machine learning (ML). Ordinary differential equations (ODEs), partial differential equations (PDEs), stochastic and agent-based models, and discrete models can all be classified as mechanistic models. And similarly, tree classifiers, random forests, support vector machines (SVM), and deep neural networks can all be classified as ML models.

In this chapter, we spotlight several works that combine *ex vivo* data with *in silico* models (Figure 32.1) to improve combination therapies for cancer. We focus on some works that exclusively rely on mechanistic models, some that exclusively rely on ML models, and one that details a hybrid mechanistic–ML model.

32.2. *In silico* models of triple-negative breast cancer *ex vivo* data

This section describes a series of papers that use mechanistic modelling to improve combination therapies of triple-negative breast cancer. The papers build on each other and utilize data taken from *ex vivo* experiments to inform their models.

Breast cancer remains one of the deadliest cancers despite extensive research into understanding and treating the disease. Its high mortality rate is owed primarily to the development of late-stage treatment resistance in many patients [29]. Breast tumours, like many other tumour types, may initially respond well to a given treatment but lose this positive response after some time. This is especially true in a particular type of breast cancer called *triple-negative breast cancer* (TNBC), termed by being negative for the production of the three important cell surface receptors HER2, oestrogen, and progesterone [30]. Treatment resistance occurs at a significantly higher rate in TNBC than in other breast cancers, leading to shorter survival times [31, 32]. Importantly, this development of late-stage treatment resistance in TNBC is observed in both chemotherapies and radiotherapies [33]. It is even observed in the emerging modality of immunotherapy, in which compounds are administered to a patient to better utilize the immune system to target the tumour [34]. In light of this, it is generally acknowledged that the major hurdle in bettering our treatment of breast cancer, particularly TNBC, is in overcoming the development of treatment resistance.

A full understanding of the biological mechanisms underlying treatment resistance remains elusive, although several theories exist. Until recently, the primary theory explaining acquired chemoresistance relied on the principles of simple Darwinian evolution, where advantageous inheritable traits are iteratively passed to daughter cells in a non-reversible fashion [35]. Specifically, in this case, the presence of a drug in the tumour creates a selective pressure towards cells that contain genetic mutations that confer a degree of resistance to the given treatment. For example, a cell may have an up-regulation of efflux transporters for which the drug is a substrate [36]. Then, with prolonged exposure to the chemotherapeutic, the minority of cells possessing the resistant phenotype become the likeliest survivors and eventually proliferate to become the majority, creating a chemoresistant tumour. However, through experimental observations, it has been shown that this is not the only mechanism at play in acquired resistance. For example, patients treated with chemotherapy can exhibit drug resistance even when no resistance-conferring genetic mutations can be identified—and indeed, similar observations have been made in studies of bacteria exposed to antibiotics [37, 38, 39]. Additionally, the simple Darwinian theory of adaptively acquired resistance is apparently refuted by the resensitization of cells after ceasing drug administration before restarting [40]. Hence, other mechanisms must be at play which drive acquired drug resistance.

Other theories have rested upon tumour heterogeneity, specifically with respect to cancer stem cells (CSCs). CSCs are a subpopulation of cells within a tumour which drive growth and differentiation. These cells typically have high proliferation rates and a higher degree of chemoresistance than differentiated cancer cells [41]. However, as our understanding of CSCs has improved, it has become clear that the binary classification of CSC vs. non-CSC is far too simplistic. Studies have demonstrated plasticity between CSCs and non-CSCs: in other words, cells may switch categories over time based on external factors [41]. Worse yet, the notion of splitting cells into distinct categories itself is perhaps overly simplistic, with the more correct interpretation being that cells exist on a continuous spectrum of genetic variability—though this is decidedly harder to characterize experimentally or model mathematically.

Recent studies, however, have revealed a new contributing factor explaining drug-induced chemoresistance: phenotypic switching. Specifically, evidence suggests that cancer cells can switch their phenotypic level of chemotolerance in response to the presence of a drug, even when an underlying genotypic change does not accompany this change [42, 43]. This drug-induced resistance can arise via protein expressions, kinase scaffolding, and signalling activations [44]. Though less well understood, this description of acquired chemotolerance does have experimental evidence to back it, as discussed below. Furthermore, a better understanding of it provides an opportunity to design therapies that bypass the resistance or perhaps even take advantage of drug-sensitive phenotypes. Naturally, when addressing such questions, *in silico* techniques can be of tremendous value.

Goldman et al. [45], for example, performed explant studies to examine the effect of drugs on the phenotypic expression of breast cancer cells. Tumour explants are a particularly useful experimental technique here as they preserve the cellular heterogeneity of the original tumour, which is crucial for examining the sizes of the subpopulations within. They examined key subpopulations, distinguished by their expression of two key molecules: CD44 and CD24. CD44 molecules are important because their expression can help to distinguish between CSCs and non-CSCs, or more generally, cells with a stem-like phenotype and cells with a non-stem-like phenotype. Specifically, cells that are CD44HiCD24Lo are considered stem-like and typically exhibit strong resistance to chemotherapy, as well as a host of other mesenchymal, stem-like characteristics. Conversely, cells which are CD44LoCD24Hi are considered non-stem-like and typically exhibit a degree of radiosensitivity and other traits typical of differentiated cells. Using their explant studies, they observed that after treating the cells with docetaxel, there was an increase in their expression of both CD44 and CD24. This suggested that chemotherapy induced a transition in the phenotype of treated cells towards a CD44HiCD24Hi state, termed the 'induced' state. These cells were characterized by low levels of apoptosis and exhibited a higher level of drug tolerance, including to other chemotherapeutics not used to induce the state.

To add theoretical backing to these observations, the authors created a compartmental mathematical model to analyse the phenotypic switching description of acquired drug resistance. Their model consisted of a system of three ODEs that described the time evolution of the size of the three different subpopulations within the tumour. In their model, cells in each compartment were assumed to proliferate and have a degree of plasticity—an ability to switch between different compartments. Using fluorescence-activated cell sorting, the authors were able to measure the population dynamics of the tumour explants. The authors then fit the model to tumour explants that had and had not been exposed to prior chemotherapy and observed the results. Based on the values of the fitted parameters in each case, they could make predictions about the mechanics underlying the phenotypic switching. For example, in the cells that had been previously exposed to chemotherapy, the proliferation rates of CSCs and induced cells were significantly increased, whereas the proliferation of non-CSCs became practically nonexistent. Furthermore, the fitted parameters suggested that treatment did not impact the rate at which CSCs differentiated into non-CSCs, but that bidirectional plasticity was significantly less likely in the cells that had been exposed to treatment. Transition rates were balanced between the induced and non-CSC compartments in drug-exposed cells, whereas in treatment-naive cells, this transition strongly favoured the non-CSCs. Finally, in the treatment-naive cells, CSCs and induced cells freely transition between compartments, though this transition was only in the direction of CSCs. Most importantly, the fitted parameter values for the two systems were distinct, meaning that the steady-state solutions of the ODE system were different.

The authors continued their experiments, again confirming several of these fitted results using cell sorting into the three compartments. These results led them to the idea of incorporating temporality into the administration of drugs that could take advantage of this phenotypic switching in response to chemotherapy. Specifically, they showed that treatments using a taxane-based therapy (docetaxel) followed by SFK inhibitors (dasatinib) resulted in an improvement of treatment effect. The rationale behind this combination is that docetaxel induced a phenotypic transition from non-CSC compartment to the induced state, which activates SFK signalling, making the cells sensitive to the effect of an SFK

inhibitor. To corroborate this, the authors generated additional tumour explants from tumours resistant to docetaxel. These explants were treated with either docetaxel alone or docetaxel followed by an SFK inhibitor. They found that the sequenced treatment resulted in a significant increase in tumour cell death compared to the single compound. Additionally, that administration of docetaxel followed by dasatinib was superior to simultaneous administration and administration in the reverse order.

Taken together, the *in silico* model of phenotypic switching allowed a scaffolding upon which a testable hypothesis could be generated. And indeed, the results of several *ex vivo* experiments agreed with the predictions of the ODE model. Most importantly, the combination of the *ex vivo* data and *in silico* modelling suggests a method for improving treatment of TNBC, which is notorious for its acquired resistance, by utilizing a novel combination chemotherapy. These results were also consistent with clinical findings showing similar results [46, 47, 48].

Another hallmark characteristic of tumours, including TNBC tumours, is dysregulated metabolism. In the absence of sufficient oxygen supply, normal cells resort to glycolysis rather than full oxidative respiration, a phenomenon termed *anaerobic glycolysis* [49, 50, 51]. Tumours often display high levels of hypoxia or anoxia, leading to cells utilizing anaerobic glycolysis as their primary metabolic pathway. Interestingly however, it is frequently observed that cancer cells, even in the presence of sufficient oxygen, will still resort to glycolysis for metabolism rather than complete full oxidative phosphorylation. This phenomenon is termed *aerobic glycolysis* or is more commonly referred to as the Warburg effect [49, 50, 51]. The outcome of this dysregulated metabolism within tumours is an acidic microenvironment that supports tumour progression. Evidence also shows that the utilization of this less efficient metabolic pathway contributes to treatment resistance although this is less well understood [52, 53].

Goldman et al. [45] investigated this altered metabolism to understand whether it contributed to conferring cross-drug resistance, that is, whether the administration of one drug can confer resistance to a different drug. The authors began by analysing cells that were either (1) pre-treated with docetaxel or were (2) naive to treatment, noting that the utilization of glycolytic metabolism was higher in the pre-treated cells, as evidenced by measuring the differences in extracellular acidification rate, oxygen consumption rate, and overall ATP abundance. After establishing the altered metabolic phenotype in pre-treated cells, the authors tested if the cells were resistant to doxorubicin using *in vitro* cultures. They administered doxorubicin along with lonidamine—an inhibitor of glucose metabolism—to isolate the effect of altered metabolism that minimally affects cell viability. When administering doxorubicin with lonidamide, cell death increased significantly compared to administration of doxorubicin alone, implicating the altered metabolism in drug–cross-drug tolerance. This effect was observed in both pre-treated and treatment-naive cells though the effect was more pronounced in pre-treated cells.

The authors performed further experiments to assess the temporal dynamics of the biological network underlying the switching in phenotypic plasticity and metabolism. Specifically, they applied sublethal doses of docetaxel which altered the relevant phenotypes without killing the cells. Using cell sorting flow cytometry, they observed that CD44 expression increased within 4 h of drug exposure, indicating that the phenotypic plasticity had been altered. On the other hand, the change in glucose uptake was more delayed and did not increase until 24 h after drug exposure, suggesting that a change in metabolic phenotype occurs after a change in plasticity phenotype. Given these temporal dynamics, the authors hypothesized that a combination chemotherapy with appropriate scheduling could overcome the resistance observed as a consequence of the altered phenotypes. Specifically, they considered a three-drug combination, including docetaxel, anthracycline, and a G6PD inhibitor.

A systems biology model was created to model the drug effect and signalling dynamics, with the values of all system parameters being taken from other previously published studies. The model consisted of four ODEs that tracked the amount of CD44, hypoxia-inducible factor (HIF1α), Glut1 (which encodes a glucose transporter), and reaction oxygen species. They assumed that cells transitioned into a drug-tolerant state after docetaxel exposure and modelled the tumour-killing effect of a glucose metabolism inhibitor. The simulation then determined how the addition of the glucose metabolism inhibitor resulted in the reduced glucose flux through the pentose phosphate pathway and used this as a proxy for cell death. The authors used the model to test different temporal sequencing of the chemotherapeutic combination. In all cases, the model predicted that simultaneous administration of lonidamine and doxorubicin after docetaxel treatment resulted in the best anti-tumour outcome.

The authors then used *ex vivo* experiments to study the potential clinical impacts of these results by generating tumour explants from fresh tumour biopsies. They tested the combination therapy of docetaxel, doxorubicin, and lonidamine, and observed that coadministration of doxorubicin and lonidamine after docetaxel provided the largest anti-tumour outcome, both in terms of cell death and reduction of the drug-tolerant phenotype. This was in agreement with the prediction of the *in silico* model, which also predicted this drug sequence as being the most efficacious. Taken altogether, these authors' results show that cancer cells are able to alter their metabolic pathway phenotype in response to chemotherapy. Furthermore, this alteration of the metabolism can confer a degree of tolerance to the administered drug as well as, importantly, a cross-tolerance to other drugs to which the cells are naive. And finally, they demonstrated that this reprogramming of the cancer cells' metabolic phenotype presents an opportunity that careful temporal sequencing can exploit to bypass the chemoresistance and yield greater anti-tumour efficacy. Of note, and just as in the discussion of the phenotypic switching model above, the *in silico* modelling provided a testable hypothesis that *ex vivo* experimentation could investigate and confirm. The result is a clinically relevant outcome that provides an opportunity for a novel combination therapy that could be useful in treating TNBC.

Heat shock protein 90 (Hsp90) plays many roles in cellular signalling, and accordingly, Hsp90 inhibitors have been studied both alone and in combination with other chemotherapies [54, 55]. Unfortunately, Hsp90-targeting compounds have yet to return results as efficacious as hoped, though hope still remains that it can still be used in the right combination and within the right treatment plan [56]. Cancer immunotherapies have emerged as a novel treatment option for tumours resistant to other conventional chemotherapeutics.

Unfortunately, tumours similarly commonly find ways to mitigate or bypass the action of CD8+ cytotoxic T cells and natural killer

(NK) cells. Even with strategies to prime these immune cells for tumour rejection, they often suffer from exhaustion, hampering their use as a long-term solution [57]. One idea, however, is to increase cancer cell surface receptors that can reinvigorate immune cells and enhance anti-tumour efficacy.

In Smalley et al. [58], the authors designed an engineered chemotherapy approach through utilization of *ex vivo* data, *in silico* modelling, and cancer nanomedicines that aims to reinvigorate NK cells to combat drug-induced resistance phenotypes. To begin, the authors used an *in vitro* coculture model with cells that were either treatment-naive or pre-treated with docetaxel and were therefore in a drug-tolerant state. They observed that the pre-treated drug-tolerant cells were also resistant to NK cells, whereas the treatment-naive cells remained more sensitive to the NK cells' effect. Through a series of *in vitro* experiments on pre-treated, drug-tolerant cells, combined with existing evidence from the literature, the authors created a list of protein families that are implicated in inducing drug tolerance following drug administration. These were Hsp90, Src, Akt, Casp-3, ERK, STAT3, HSF1, MICA, and NKG2D—all of which were included in a systems biology model. Hsp90 was an important node of the system that regulated the relationship between NK cell anti-tumour action and the tumour prosurvival pathways.

As evidenced by the results of the works described above, the precise ordering and timing of chemotherapies is an important consideration when planning cancer therapy, especially when attempting to overcome drug tolerance and resistance. The authors investigated the sequencing of Hsp90 inhibitors and docetaxel *in vitro* by examining the anticancer effects and NK cell action via NKG2D receptor ligand expression. They found that administering docetaxel followed by radicicol resulted in a synergistic treatment effect that maximized anti-tumour effect. To validate these experimental results, the ODE model was used to investigate the drug sequencing analogously. They used a genetic algorithm to fit the model to the *in vitro* data and then used the resulting model to investigate the scheduling. The simulation results agreed with the *in vitro* sequencing results, specifically that docetaxel first, radicicol second, was the optimal administration sequence in terms of anticancer effect. In addition to the anticancer effect, the authors also tested the model predictions related to Hsp90 disruption and NK cell recognition. They found that, as expected, Hsp90 inhibition sensitized tumour cells, significantly increasing the expression of MICA in the cells.

In summary, the combined results of the experimental and *in silico* models suggested two key findings: first, that the order of administration between docetaxel and radicicol was an important factor for improving anticancer effect; and second, that cells treated with docetaxel have suppressed NK cells that can be reversed by inhibition of Hsp90. Given these findings, the authors observed an opportunity to combine these drugs into a nanoparticle formulation. Since drugs *in vivo* tend to have shorter half-lives than *in vitro*, a nanoparticle formulation of the two drugs would theoretically allow the drug synergy window to be better exploited. This docetaxel–radicicol nanoparticle would have a fast release of docetaxel to eliminate the sensitive, treatment-naive cells and then the subsequent release of the radicicol would boost NK cell activity against the remaining, drug-tolerant cells. Using their *in silico* model, the authors predicted that this formulation would result in an improved anticancer effect compared to even the optimal sequencing of separate administration.

Once again, this study—which builds upon the results of the previous two studies in this section—incorporated *ex vivo* data with *in silico* techniques to arrive at a potential improvement for combination therapies with important clinical implications.

32.3. Combining *ex vivo* data and ML approaches

Thanks to the adoption of artificial intelligence (AI) and ML throughout the past decade, a new age in medicine has begun. AI has revolutionized how we process information and fundamentally changed how healthcare is delivered. ML techniques are being used to address challenges in current clinical and preclinical trials due to their capacity to automate essential procedures and practices [59, 60]. One of the challenges of *ex vivo* methods is the complexities in analysing patient-derived samples [61]. Hence, incorporating ML and deep-learning approaches in *ex vivo* studies can help overcome such complexities and improve the accuracy of findings while decreasing the time required for the analysis. In *ex vivo* experimental studies in oncology, ML has been mainly used for various applications, including tumour tissue detection, image analysis, patient survival prediction, clinical response anticipation, and therapy monitoring [62, 63, 64, 65]. In this section, we aim to provide an overview of the recent cancer treatment advances made by integrating ML and *ex vivo* techniques.

Ex vivo confocal laser scanning microscopy (CLSM) is a diagnostic technique for cutaneous squamous cell carcinoma (cSCC) detection. Clinical diagnosis of cSCC is complicated because of the overlap in clinical characteristics between cSCC and other skin neoplasms such as keratoacanthoma or basal cell carcinoma [66]. As a result, surgical excision and subsequent histopathologic analysis are essential to make an accurate diagnosis and determine the best course of treatment [67]. Unfortunately, the standard method of pathological evaluation, known as *frozen sectioning*, is based on labour-intensive and time-consuming processes. In addition to this, they present other issues, such as a partial breakdown in the cellular network of the tissue, difficulties in the cutting process, and poor quality of staining [68]. *Ex vivo* CLSM is one of the screening methods designed to solve existing problems with conventional pathology [69, 70]. Even with this, the evaluation of collected CLSM pictures is difficult, particularly when time is restricted, as it requires specialized knowledge and training [71].

To overcome this issue, Ruini et al. [71] showed the high potential of *ex vivo* CLSM to take advantages from incorporating ML algorithms into the interpretation and decision-making procedure by developing convolutional neural networks (CNNs) for automated tumour tissue diagnosis in *ex vivo* CLSM images of cSCC. In Ruini's study, 34 fresh tissue samples were extracted and tested shortly following excision. Then, following the histologically approved *ex vivo* CLSM detection, scientists annotated the tumour cells for classification for the purpose of training the CNN algorithm. To accomplish this, a MobileNet [72], a lightweight deep CNN, was employed. MobileNet is well known for its utilization of depth-wise separable convolutions, which reduces the amount of time required for processing. Compared to the expert assessment, CNN's overall sensitivity and specificity for distinguishing cSCC were 0.76 and 0.91, respectively. These findings showcase the capability of deep-learning

algorithms in recognizing cSCC locations on coloured *ex vivo* CLSM images and in differentiating those regions from tumour-free regions with a high level of sensitivity and specificity [65]. There are already published studies suggesting the potential use of *ex vivo* CLSM in detecting additional tumour forms, such as melanoma, prostate, and breast cancer. Consequently, the results of Ruini's study can serve as a guideline for the development of new standard deep-learning models for the automatic diagnosis of additional kinds of cancer cells or tissue on *ex vivo* CLSM pictures [71, 73, 74, 75, 76]. This new approach may be more effective and economical than the standard operating procedure present in use.

Incorporating ML into a combination of clinical and preclinical procedures can also assist in predicting patient survival [64]. One of the prevalent clinical methods today is magnetic resonance imaging (MRI) that is used to characterize tumours, particularly gliomas [77]. However, as a substitute or supplement to MRI-only techniques, radiolabelled amino acid Positron Emission Tomography (PET) tracer, such as L-S-methyl-11C-methionine (11C-MET), is regarded due to their high sensitivity and specificity as a promising diagnostic strategy for tumour characterization and prolonged treatment monitoring. L-S-methyl-(11C-MET) PET imaging is used to grade gliomas, determine tumour scope, determine brain biopsy sites, schedule radiotherapy, and monitor treatment [78, 79, 80]. Gliomas are the most prevalent type of brain tumour, which make up 81% of all cerebral malignancies. Anticipated patient survival differs by glioma kind, with glioblastoma multiforme—the most prevalent and deadly—having the lowest five-year survival rate at around 5%. In spite of the fact that *ex vivo* histopathologic and molecular evaluation is frequently used to make the definitive diagnosis, imaging is typically the main tool used in the process of diagnosing patients who may have gliomas. However, early detection, improved tumour therapy, and longer survival rates can be achieved by complex, sophisticated techniques involving the integration of multiple therapeutic techniques, including in *ex vivo*, *in vivo*, and *in silico* approaches [64].

For example, Papp et al. [64] developed ML-based survival prediction models for glioma utilizing *ex vivo* data, *in vivo* 11C-MET PET, as well as patient characteristics. This research involved a total of 70 participants, all of whom had treatment-naive gliomas that tested positive for 11C-MET and had *ex vivo* features extracted from histopathologies, such as tumour stage and histology. ML algorithms were employed to define relevant *in vivo*, *ex vivo*, and patient characteristics (such as age, height, and weight), as well as their respective weights in the ML model, with the purpose of predicting survival at 36 months. Four prediction models were developed through the utilization of the generated feature weights. The first of these models depended on *in vivo* data, *ex vivo* characterization, and patient features; the second of these models depended on *in vivo* information and patient characterization; the third of these models solely depended on *in vivo* features; and the fourth model depended on *ex vivo* data and patient information. Across all analysed characteristics, in patients with amino-acid–positive gliomas, grey-level cooccurrence matrix characteristics, such as entropy, angular second moment, and intensity characteristics like maximal tumour-to-background ratio (TBR), proved to be of modest value in predicting survival. However, some other features, such as the age of patients, *ex vivo* features, including isocitrate dehydrogenase 1 R132H mutational status, and the *in vivo* features, such as TBR total and spheric dice coefficient, found to be crucial for predicting survival. Furthermore, according to the results of a Monte Carlo cross-validation, the developed models that had the highest area under the curve (AUC) were *in vivo*-, *ex vivo*-, and patient-based model, as well as *ex vivo*- and patient-based one with 0.9 and 0.87, respectively. Accordingly, the most dominant features selected by the ML-chosen and ML-weighted features were dependent on patient and *ex vivo* data, followed by *in vivo* characterization as the next most important factor. These results recommended that the combination of patient characteristics, *ex vivo*, *in vivo* data, and ML methods, can result in the highest accurate survival predictions for amino acid PET–positive glioma patients. The proposed ML methods and predictive model are extremely general because they do not take into account any previous information regarding the inputs or derived characteristics. Consequently, it is feasible to investigate these ML techniques for various types of cancers.

The application of ML in *ex vivo* studies is not confined to image analysis and the detection of tumour tissue and patient survival; it also includes the prediction of early clinical response. Indeed, significant variation in treatment responses across diverse clinical settings necessitates accurate prediction of treatment response [62]. As an example, lymphoma is one of the most widespread haematological malignancies as well as one of the most prevalent forms of cancer in both animals and humans [81]. About three-fourths of canine lymphoma cases are multicentric lymphoma, while other types occur less frequently [82]. Multi-agent chemotherapy has been shown to achieve the highest response rates and the longest remission when treating canine lymphoma [83, 84]. However, treatment outcomes vary according to the administered drug combinations and lymphoma subsets. T cell subgroups of canine lymphoma, for instance, have a lower therapeutic efficiency than B-cell subgroups [85]. Therefore, it is necessary to precisely anticipate therapy response, particularly for individuals having lymphoma subgroups that are infrequent or have low response rates. By meeting this need, clinicians will be able to find the most effective medications for each patient and exclude those that are ineffective, hence enhancing treatment outcomes [62].

To this end, Bohannan et al. [62] integrated *ex vivo* chemosensitivity measurements with ML methods to assess the likelihood of chemotherapy drug efficiency for canine lymphoma. *Ex vivo* cell-based drug sensitivity tests are investigated extensively as a personalized medicine technique to simulate the tumour microenvironment *in vitro* and anticipate responses in patient lymphoproliferative diseases [86, 87, 88]. For developing the predictive system, Bohannan et al. [62] extracted active cancer cells from biopsies of diseased lymph nodes and acquired clinical feedback from 261 canine lymphoma patients who were expected to undergo at least one of five commonly used chemotherapeutic drugs (doxorubicin, vincristine, cyclophosphamide, lomustine, and rabacfosadine). In order to strengthen the probabilistic models of treatment response, they additionally included the immunophenotype of cancer cells derived from patients as assessed by flow cytometry. This is because characteristics collected from the findings of just drug sensitivity screening alone may not be sufficient for anticipating *in vivo* responding [50]. Then, 70% of treated patients for each drug were

randomly selected to train a random forest algorithm to determine the likelihood of a positive drug response. The remaining 30% of the dataset was used as a testing set to evaluate the accuracy of the model. The variables included the drug response output values and the proportion of the collected sample cells expressing a wide range of markers, such as lymphocytes, and large lymphocytes, and percentages of the cell population expressing CD21, class II MHC, CD3, CD8, and CD34. The results showed that the ROC–AUC for all drugs was consistently greater than 0.95, achieving a high distinction between positive and negative feedback in the dataset. Therefore, it showed that flow cytometry markers for immunophenotyping and drug sensitivity parameters are adequate for developing ML-based prediction models of chemotherapy response. Nonetheless, larger sample sizes and the addition of extra variables, such as genomic sequencing information or terms characterizing interactions between drugs, are anticipated to improve the prediction capacity of the given models.

To more accurately anticipate the response to chemotherapeutic treatments, it is crucial to preserve the clinical heterogeneity of cancer and stromal cells, tumour matrix, and tumour anatomy. However, the present standard *in vitro* and *ex vivo* preclinical techniques, which utilize cells and spheroids or *ex vivo* cancer models, are hampered by their poor ability to properly simulate the biology of the original tumour, resulting in poor clinical result prediction [89, 90, 91]. Majumder et al. [63] attempted to develop a clinically useful ML- and *ex vivo* based predictive model by designing and improving an *ex vivo* tumour ecosystem for colorectal cancer (CRC) and head and neck squamous cell carcinoma (HNSCC). To accomplish this, they cultured narrow cancerous explants with preserved cellular, microenvironmental, and architectural variability in culture wells covered with grade-matched tumour matrix support with personalized serum providing endogenous ligands. They combined the developed tumour ecosystems with an ML algorithm to create a platform that accurately anticipated the treatment effectiveness of anticancer drugs in patients. This platform was developed using tumour biopsy samples from 109 patients with CRC and HNSCC that had been treated with the same drugs, such as docetaxel, cisplatin, and 5-fluorouracil. The results of survivability, growth, histopathology, and apoptosis derived from these samples, along with the therapeutic outcome seen in the corresponding patients, were categorized as non-response, partial response, or full response, and this data were then employed as the training set for the ML algorithm. A linear prediction model was used for classification that was trained to employ SVMpAUC, a structural support vector machine classifier for improving partial AUC. This algorithm maintained an acceptable level of specificity while achieving 87.27% accuracy and 100% sensitivity on the testing dataset. Therefore, the reported platform can be considered an effective predictive tool that can be applied to various cancer types and therapeutic strategies, as evidenced by the overall feedback rates of HNSCC and CRC tumours to chemotherapeutic treatments.

Taken together, on the basis of all of these studies, we can infer that diverse combinations of *in vitro*, *ex vivo*, *in vivo*, and patient data combined with ML tools can improve personalized medicine tumour tissue diagnosis as well clinical practice prediction to an acknowledged high degree of accuracy.

32.4. Integrated systems biology and deep learning with *ex vivo* data

In this section, we provide a case study on the application of mechanistic models and deep-learning approaches to interrogate the response dynamics of an immunotherapy checkpoint inhibitor using *ex vivo* system with human HNSCC.

Immunotherapy has emerged as a promising new cancer treatment strategy and is widely considered to now be the fourth 'pillar' of modern cancer medicine. The most successful immunotherapy used in practice today is anti-PD-1 (programmed cell death protein) therapy, which works by blocking the binding of PD-1 to PD-L1 (programmed death ligand 1), preventing cancer cells from suppressing the activation of T cells, and evading the anti-tumour immune response [92]. While anti-PD-1 therapy is the most successful immunotherapy to date, patient response to the treatment is unpredictable—some patients respond well, and others have no response. This variability in response is thought to be attributable to underlying patient-specific biological traits though these traits are not well understood.

In a recent work, Smalley et al [93]. developed a systems biology model to investigate how the blockade of PD-1 affects the behaviour of relevant immune cells and how such therapy can elicit different response dynamics. Their model was based upon, and calibrated by, results from human *ex vivo* experiments and allowed for a framework through which immune checkpoint inhibitors could be tested *in silico*. This detailed and experimentally calibrated mathematical model is useful; however, an important goal that the model struggles to attain on its own is to identify patient-specific traits and biomarkers that could predict whether a patient is likely to be a responder to anti-PD-1 therapy. One way to answer such a question is through using ML techniques as described above. For example, one could amass a dataset of patients who have received anti-PD-1 immunotherapy, classify them according to their treatment response, train a neural network with this data, and then use this to predict which patients are likely to respond well to therapy. Unfortunately, this plan suffers from two key problems. Firstly, doing so may generate an accurate classification, but the classification algorithm is a black box, leaving researchers without interpretability, and no further insight into what the key traits which determine response actually are. And secondly, the network, to be properly trained, would require an amount of data that would be difficult or simply impossible to collect. This makes an exclusively ML-based approach impractical, meaning that a more sophisticated solution is necessary in order to predict the anti-PD-1 response from patient-specific traits.

In Przedborski et al. [34], the authors extend their systems biology model by creating a hybrid systems biology and deep-learning approach which allowed them to tackle these questions. Specifically, the systems biology model, calibrated by human *ex vivo* data, can generate an amount of data limited only by computational run time. Rather than using the small amount of available clinical data to train the anti-PD-1 response prediction neural network, they generated a much larger dataset using the systems biology model and then trained the neural network with this data. This technique bypasses one of the common shortcomings of exclusively ML-based models in medicine of the difficulty in obtaining enough data for training.

Furthermore, it also alleviates the lack of model explainability since results can be interpreted through the lens of the coupled systems biology model. The authors refer to this technique as a systems biology informed neural network.

A reasonable question to ask is whether the synthetically generated clinical dataset is a reasonable representation of the full clinical variability. The authors address this question by using feature selection techniques to show that the simulated dataset captures the key dynamics and variability seen in previous experimental and clinical investigations of anti-PD-1 therapy. From this analysis came other notable observations. For example, several of the parameters that feature selection identified as being crucial for predicting response to anti-PD-1 treatment (e.g., baseline levels of cytotoxic and naive T cells) can be experimentally measured or estimated. This means that patients could be pre-screened to estimate their likelihood of positive treatment response. Furthermore, knowledge of these key traits could give insight into mechanisms of drug resistance, allow for more efficient screening of patients, and perhaps even allow for development of treatments that could alter these values and improve response. Additionally, the simulated data was not used alone, but rather in concert with the *ex vivo* clinical data. Specifically, the network was first trained using the simulated data, then the *ex vivo* was incorporated into the training to fine-tune the network. This transfer approach greatly increased the accuracy of the network when evaluated on the *ex vivo* dataset. The reasoning for this is that since the *ex vivo* dataset is small, it does not capture the full range of possible dynamics that the much larger simulated dataset is able to. Therefore, pretraining on the larger dataset allows the network to generalize to previously unseen cases more accurately.

Of particular relevance, the results of these simulations also led to a rationale for a combination therapy that theoretically improves patient response to anti-PD-1 immunotherapy. In particular, the baseline population of cytotoxic T cells was identified as a primary driver of response to anti-PD-1 treatment. So, theoretically, increasing the T cell population prior to the stat of anti-PD-1 therapy would improve the patient response. Analysis of the systems biology model revealed that administration of IL-6 inhibitors and recombinant IL-12 to the patient would ultimately yield increased numbers of cytotoxic T cells, hence providing a rationale for a novel combination therapy that could improve anti-PD-1 therapy. Indeed, in previous studies, both IL-6 inhibitors and recombinant IL-12 individually showed the potential to improve the anti-tumour effects of anti-PD-1 therapy. Using the systems biology model, the authors showed that administering recombinant IL-12, followed by an IL-6 inhibitor second, prior to the anti-PD-1 agent (nivolumab, in this case), significantly increased the number of responders in their simulated clinical dataset.

Altogether, this study showed how systems biology models and ML models, in concert with *ex vivo* data, can be combined in such a way that the shortcomings of both styles can be mitigated. Most notably, the results suggested a novel triple combination therapy that theoretically improves the patient response to anti-PD-1 immunotherapy.

32.5. Conclusions

Ex vivo experimental methods have proven their utility in helping to alleviate the existing challenges of common *in vitro* and *in vivo* models in preclinical trials by being able to closely reflect the real situation systems through representing the original tumour phenotypes and heterogeneity. However, because of the complexities analysing results in patient-derived samples, more improved analytical techniques, such as mathematical and computational approaches, are required. Incorporation of different mathematical and computational modelling into cancer medicine can further our understanding of the disease and improve its treatment. In this work, we showcased several works that did just this. We divided these works by the methods of quantitative analyses they used: those that employ mechanistic mathematical modelling exclusively, those that employ AI exclusively, and those that employ both. Reviewing all of these works demonstrates the potential applicability of different mathematical methods to the successful evaluating of *ex vivo* data with the purpose of tumour tissue detection, image analysis, patient survival prediction, clinical response anticipation, and therapy monitoring.

The combination of *ex vivo* and *in silico* models has two benefits: first, mathematical and computational techniques can assist in more accurately analysing the outcomes originating from patient-derived samples and overcoming the complexities associated with *ex vivo* models; and second, using *ex vivo* models can support creating the dataset required for developing and validating mathematical models. For example, the challenge of obtaining adequate human data greatly restricts the use of ML and deep learning in cancer medicine. This is because it can be difficult to obtain this information in a medical setting due to financial, technological, and ethical constraints. However, by applying different anti-tumour therapies on *ex vivo* samples, researchers can provide results closer to clinical outcomes and supplement the required datasets for ML algorithms.

It is also important to note that in building predictive mathematical models, parameters acquired from patient-derived samples exclusively may not be sufficient, and the addition of other information, such as *in vitro* features, *in vivo* features, or patient characterization, may increase the predictive value of *ex vivo* evaluation. Hence, it can be concluded that personalized cancer treatments should include more complex and advanced strategies involving combinations of preclinical and clinical methods, such as *ex vivo*, *in vitro*, and *in silico* models to improve earlier detection, tumour targeting, and survival rates.

REFERENCES

1. Mak IWY, Evaniew N, Ghert M. Lost in translation: animal models and clinical trials in cancer treatment. Am J Transl Res. 2014;**6**(2):114.
2. Nairon KG, Skardal A. Biofabrication of advanced *in vitro* and *ex vivo* cancer models for disease modeling and drug screening. Future Drug Discov. 2021;**3**(3):Article FDD62.
3. Williams ST, Wells G, Conroy S, Gagg H, Allen R, Rominiyi O, et al. Precision oncology using ex vivo technology: a step towards individualised cancer care? Exp Rev Mol Med. 2022;**24**:e39.
4. Zitter R, Chugh RM, Saha S. Patient derived ex-vivo cancer models in drug development, personalized medicine, and radiotherapy. Cancers. 2022;**14**(12):3006.
5. Hwang TJ, Carpenter D, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of investigational drugs in late-stage clinical development and publication of trial results. JAMA Int Med. 2016;**176**(12):1826–33.

6. Bruna A, Rueda OM, Greenwood W, Batra AS, Callari M, Batra RN, et al. A biobank of breast cancer explants with preserved intra-tumor heterogeneity to screen anticancer compounds. Cell. 2016;**167**(1):260–74.e22.
7. Letai A, Bhola P, Welm AL. Functional precision oncology: testing tumors with drugs to identify vulnerabilities and novel combinations. Cancer Cell. 2022;**40**(1):26–35.
8. Kettunen K, Boström PJ, Lamminen T, Heinosalo T, West G, Saarinen I, et al. Personalized drug sensitivity screening for bladder cancer using conditionally reprogrammed patient-derived cells. Eur Urol. 2019 Oct;**76**(4):430–34.
9. Kim SY, Lee JY, Kim DH, Joo HS, Yun MR, Jung D, et al. Patient-derived cells to guide targeted therapy for advanced lung adenocarcinoma. Sci Rep. 2019 Dec;**9**(1):1–12.
10. Stringer BW, Day BW, D'Souza RCJ, Jamieson PR, Ensbey KS, Bruce ZC, et al. A reference collection of patient-derived cell line and xenograft models of proneural, classical and mesenchymal glioblastoma. Sci Rep. 2019 Mar;**9**(1):1–14.
11. Arjonen A, Mäkelä R, Härmä V, Rintanen N, Kuopio T, Kononen J, et al. Image-based ex vivo drug screen to assess targeted therapies in recurrent thymoma. Lung Cancer. 2020 Jul;**145**:27–32.
12. Mäkelä R, Arjonen A, Suryo Rahmanto A, Härmä V, Lehtiö J, Kuopio T, et al. *Ex vivo* assessment of targeted therapies in a rare metastatic epithelial–myoepithelial carcinoma. Neoplasia. 2020 Sep;**22**(9):390–98.
13. Boretto M, Maenhoudt N, Luo X, Hennes A, Boeckx B, Bui B, et al. Patient-derived organoids from endometrial disease capture clinical heterogeneity and are amenable to drug screening. Nat Cell Biol. 2019 Aug;**21**(8):1041–51.
14. Driehuis E, Kolders S, Spelier S, Lõhmussaar K, Willems SM, Devriese LA, et al. Oral mucosal organoids as a potential platform for personalized cancer therapy. Cancer Discov. 2019 Jul; **9**(7):852–71.
15. Ivanova E, Kuraguchi M, Xu M, Portell AJ, Taus L, Diala I, et al. Use of *ex vivo* patient-derived tumor organotypic spheroids to identify combination therapies for HER2 mutant non–small cell lung cancer. Clin Cancer Res. 2020 May;**26**(10):2393–403.
16. Yoshida T, Sopko NA, Kates M, Liu X, Joice G, McConkey DJ, et al. Impact of spheroid culture on molecular and functional characteristics of bladder cancer cell lines. Oncol Lett. 2019 Nov;**18**(5):4923–29.
17. Jin Jang K, Otieno MA, Ronxhi J, Keang Lim H, Ewart L, Kodella KR, et al. Reproducing human and cross-species drug toxicities using a Liver-Chip. Sci Transl Med. 2019 Nov;**11**(517):eaax5516.
18. Mao M, Bei HP, Lam CH, Chen P, Wang S, Chen Y, et al. Human-on-leaf-chip: a biomimetic vascular system integrated with chamber-specific organs. Small. 2020 Jun;**16**(22):2000546.
19. Mazzocchi AR, Rajan SAP, Votanopoulos KI, Hall AR, Skardal A. *In vitro* patient-derived 3D mesothelioma tumor organoids facilitate patient-centric therapeutic screening. Sci Rep 2018 Feb; **8**(1):1–12.
20. Schutgens F, Rookmaaker MB, Margaritis T, Rios A, Ammerlaan C, Jansen J, et al. Tubuloids derived from human adult kidney and urine for personalized disease modeling. Nat Biotechnol. 2019 Mar;**37**(3):303–13.
21. Vatine GD, Barrile R, Workman MJ, Sances S, Barriga BK, Rahnama M, et al. Human iPSC-derived blood-brain barrier chips enable disease modeling and personalized medicine applications. Cell Stem Cell. 2019 Jun;**24**(6):995–1005.e6.
22. Birey F, Andersen J, Makinson CD, Islam S, Wei W, Huber N, et al. Assembly of functionally integrated human forebrain spheroids. Nature. 2017 Apr;**545**(7652):54–9.
23. Kim E, Choi S, Kang B, Kong JH, Kim Y, Yoon WH, et al. Creation of bladder assembloids mimicking tissue regeneration and cancer. Nature. 2020 Dec;**588**(7839):664–69.
24. Song L, Yuan X, Jones Z, Griffin K, Zhou Y, Ma T, et al. Assembly of human stem cell-derived cortical spheroids and vascular spheroids to model 3-D brain-like tissues. Sci Rep 2019 Apr;**9**(1):1–16.
25. Bolenz C, Ikinger EM, Ströbel P, Trojan L, Steidler A, Fernández MI, et al. Topical chemotherapy in human urothelial carcinoma explants: a novel translational tool for preclinical evaluation of experimental intravesical therapies. Eur Urol. 2009 Sep;**56**(3):504–11.
26. Kelly JD, Williamson KE, Weir HP, McManus DT, Hamilton PW, Keane PF, et al. Induction of apoptosis by mitomycin-C in an ex vivo model of bladder cancer. BJU Int. 2000 May;**85**(7):911–17.
27. Pierre Gillet J, Maria Calcagno A, Varma S, Marino M, Green LJ, Vora MI, et al. Re- defining the relevance of established cancer cell lines to the study of mechanisms of clinical anti-cancer drug resistance. Proc Natl Acad Sci U S A. 2011 Nov;**108**(46):18708–13.
28. Karkampouna S, La Manna F, Benjak A, Kiener M, De Menna M, Zoni E, et al. Patient-derived xenografts and organoids model therapy response in prostate cancer. Nat Commun. 2021 Feb; **12**(1):1–13.
29. DeSantis CE, Ma J, Sauer AG, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. CA: Cancer J Clin. 2017;**67**(6):439–48. eprint. https://onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21412
30. Dent R, Trudeau M, Pritchard KI, Hanna WM, Kahn HK, Sawka CA, et al. Triple-negative breast cancer: clinical features and patterns of recurrence. Clin Cancer Res: An Off J Am Assoc Cancer Res. 2007 Aug;**13**(15 Pt 1):4429–34.
31. Pal SK, Childs BH, Pegram M. Triple negative breast cancer: unmet medical needs. Breast Cancer Res Treat. 2011 Feb;**125**(3):627–36.
32. Yao H, He G, Yan S, Chen C, Song L, Rosol TJ, et al. Triple-negative breast cancer: is there a treatment on the horizon? Oncotarget. 2016 Sep;**8**(1):1913–24.
33. Wende L, Li S, Chen IX, Liu Y, Ramjiawan RR, Leung C-H, et al. Combining losartan with radiotherapy increases tumor control and inhibits lung metastases from a HER2/neu-positive orthotopic breast cancer model. Radiat Oncol. (London, England). 2021 Mar;**16**:48.
34. Przedborski M, Smalley M, Thiyagarajan S, Goldman A, Kohandel M. Systems biology informed neural networks (SBINN) predict response and novel combinations for PD-1 checkpoint blockade. Commun Biol. 2021 July;**4**(1):1–15. Publisher: Nature Publishing Group.
35. Cairns J. Mutation selection and the natural history of cancer. Nature. 1975 May;**255**(5505):197–200.
36. Redmond KM, Wilson TR, Johnston PG, Longley DB. Resistance mechanisms to cancer chemotherapy. Front Biosci: J Virt Lib. 2008 May;**13**:5138–54.
37. Berrieman HK, Lind MJ, Cawkwell L. Do beta-tubulin mutations have a role in resistance to chemotherapy? Lancet Oncol. 2004 Mar;**5**(3):158–64.
38. Dawson CC, Intapa C, Jabra-Rizk MA. 'Persisters': Survival at the cellular level. PLOS Pathog. 2011 July;**7**(7):e1002121. Publisher: Public Library of Science.
39. Talpaz M, Silver RT, Druker BJ, Goldman JM, Gambacorti-Passerini C, Guilhot F, et al. Imatinib induces durable hematologic and cytogenetic responses in patients with

accelerated phase chronic myeloid leukemia: results of a phase 2 study. Blood. 2002 Mar;**99**(6):1928–37.
40. Cara S, Tannock IF. Retreatment of patients with the same chemotherapy: implications for clinical mechanisms of drug resistance. Ann Oncol: Off J Eur Soc Med Oncol. 2001 Jan;**12**(1):23–7.
41. Batlle E, Clevers H. Cancer stem cells revisited. Nat Med. 2017 Oct;**23**(10):1124–34. Publisher: Nature Publishing Group.
42. Marusyk, A, Almendro, V, Polyak, K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012 April;**12**(5):323–34.
43. Gupta PB, Fillmore CM, Jiang G, Shapira SD, Tao K, Kuperwasser C, et al. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. Cell. 2011 Aug; **146**(4):633–44.
44. Goldman A, Majumder B, Dhawan A, Ravi S, Goldman D, Kohandel M, et al. Temporally sequenced anticancer drugs overcome adaptive resistance by targeting a vulnerable chemotherapy-induced phenotypic transition. Nat Commun. 2015 Feb;**6**(1):6139. Publisher: Nature Publishing Group.
45. Goldman A, Khiste S, Freinkman E, Dhawan A, Majumder B, Mondal J, et al. Targeting tumor phenotypic plasticity and metabolic remodeling in adaptive cross-drug tolerance. Sci Signal. 2019 Aug;**12**(595):eaas8779.
46. Araujo JC, Mathew P, Armstrong AJ, Braud EL, Posadas E, Lonberg M, et al. Dasatinib combined with docetaxel for castration-resistant prostate cancer: results from a phase 1-2 study. Cancer. 2012 Jan;**118**(1):63–71.
47. Finn RS, Bengala C, Ibrahim N, Roché H, Sparano J, Strauss LC, et al. Dasatinib as a single agent in triple-negative breast cancer: results of an open-label phase 2 study. Clin Cancer Res: An Off J Am Assoc Cancer Res. 2011 Nov;**17**(21):6905–13.
48. Herold CI, Chadaram V, Peterson BL, Kelly Marcom P, Hopkins J, Kimmick GG, et al. Phase II trial of dasatinib in patients with metastatic breast cancer using real-time pharma-codynamic tissue biomarkers of Src inhibition to escalate dosing. Clin Cancer Res: An Off J Am Assoc Cancer Res. 2011 Sep;**17**(18):6061–70.
49. Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. Nat Rev Cancer. 2011 Feb 11(2):85–95.
50. Cantor JR, Sabatini DM. Cancer cell metabolism: one hallmark, many faces. Cancer Discov. 2012 Oct;**2**(10):881–98.
51. Warburg O. On the origin of cancer cells. Science (New York, NY). 1956 Feb;**123**(3191):309–14.
52. Bosc C, Selak MA, Sarry J-E. Resistance is futile: targeting mitochondrial energetics and metabolism to overcome drug resistance in cancer treatment. Cell Metab. 2017 Nov;**26**(5):705–707.
53. Rahman M, Hasan MR. Cancer metabolism and drug resistance. Metabolites. 2015 Sep:**5**(4):571–600.
54. Lu X, Xiao L, Wang L, Ruden DM. Hsp90 inhibitors and drug resistance in cancer: the potential benefits of combination therapies of Hsp90 inhibitors and other anti-cancer drugs. Biochem Pharmacol. 2012 Apr;**83**(8):995–1004.
55. Zhao R, Davey M, Hsu Y-C, Kaplanek P, Tong A, Parsons AB, et al. Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the Hsp90 chaperone. Cell. 2005 Mar;**120**(5):715–27.
56. Wang H, Lu M, Yao M, Zhu W. Effects of treatment with an Hsp90 inhibitor in tumors based on 15 phase II clinical trials. Mol Clin Oncol. 2016 Sep;**5**(3):326–34.
57. Vivier E, Tomasello E, Baratin M, Walzer T, Ugolini S. Functions of natural killer cells. Nat Immunol. 2008 May;**9**(5):503–10.
58. Smalley M, Kumar Natarajan S, Mondal J, Best D, Goldman D, Shanthappa B, et al. Nanoengineered disruption of heat shock protein 90 targets drug-induced resistance and relieves natural killer cell suppression in breast cancer. Cancer Res. 2020 Dec;**80**(23):5355–66.
59. Benjamens S, Dhunnoo P, Mesk´o B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. NPJ Digital Med. 2020 Sep;**3**(1):1–8.
60. Muehlematter UJ, Daniore P, Vokinger KN. Approval of artificial intelligence and machine learning-based medical devices in the USA and Europe (2015–20): a comparative analysis. Lancet Digit Health. 2021 Mar;**3**(3):e195–203.
61. Graça G, Lau CHE, Gonçalves LG. Exploring cancer metabolism: applications of metabolomics and metabolic phenotyping in cancer research and diagnostics. Advances in Experimental Medicine and Biology. 2020; **1219**, 367–85.
62. Bohannan Z, Pudupakam RS, Koo J, Horwitz H, Tsang J, Polley A, et al. Predicting likelihood of in vivo chemotherapy response in canine lymphoma using ex vivo drug sensitivity and immunophenotyping data in a machine learning model. Vet Comp Oncol. 2021;**19**(1):160–71.
63. Majumder B, Baraneedharan U, Thiyagarajan S, Radhakrishnan P, Narasimhan H, Dhandapani M, et al. Predicting clinical response to anticancer drugs using an *ex vivo* platform that captures tumour heterogeneity. Nat Commun. 2015;**6**, Article 6169.
64. Papp L, Pötsch N, Grahovac M, Schmidbauer V, Woehrer A, Preusser M, et al. Glioma survival prediction with combined analysis of *in vivo* 11 C-MET PET features, ex vivo features, and patient features by supervised machine learning. J Nucl Med. 2018;**59**(6):892–9.
65. Ruini C, Schlingmann S, Jonke Z, Avci P, Padrón-Laso V, Neumeier F, et al. Machine learning based prediction of squamous cell carcinoma in ex vivo confocal laser scanning microscopy. Cancers. 2021;**13**(21):1–13.
66. Hyung Ryu T, Kye H, Eun Choi J, Hyun Ahn H, Chul Kye Y, Hong Seo S. Features causing confusion between basal cell carcinoma and squamous cell carcinoma in clinical diagnosis. Ann Dermatol. 2018 Feb;**30**(1):64–70.
67. Alam, M, Armstrong, A, Baum, C, Bordeaux, JS, Brown, M, Busam, KJ. et al. Guidelines of care for the management of cutaneous squamous cell carcinoma. J Am Acad Dermatol. 2018 Mar; **78**(3):560–78.
68. Desciak EB, Maloney ME. Artifacts in frozen section preparation. Dermatol Surg. 2000;**26**(5):500–504.
69. Ragazzi M, Longo C, Piana S. *Ex vivo* (fluorescence) confocal microscopy in surgical pathology: state of the art. Adv Anat Pathol. 2016;**23**(3):159–69.
70. Rajadhyaksha M, Menaker G, Flotte T, Dwyer PJ, González S. Confocal examination of nonmelanoma cancers in thick skin excisions to potentially guide mohs micrographic surgery without frozen histopathology. J Invest Dermatol. 2001 Nov;**117**(5):1137–43.
71. Hartmann D, Krammer S, Bachmann MR, Mathemeier L, Ruzicka T, von Braunmühl T. Simple 3-criteria-based ex vivo confocal diagnosis of basal cell carcinoma. J Biophoton. 2018 Jul;**11**(7):e201800062.
72. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. ArXiv. 2017 Apr.
73. Bertoni L, Puliatti S, Reggiani Bonetti L, Maiorana A, Eissa A, Azzoni P, et al. *Ex vivo* fluorescence confocal

microscopy: prostatic and periprostatic tissues atlas and evaluation of the learning curve. Virchows Archiv: Int J Pathol. 2020 Apr;**476**(4):511–20.
74. Krishnamurthy S, Cortes A, Lopez M, Wallace M, Sabir S, Shaw K, et al. *Ex vivo* confocal fluorescence microscopy for rapid evaluation of tissues in surgical pathology practice. Arch Pathol Lab Med. 2018 Mar;**142**(3):396–401.
75. Puliatti S, Bertoni L, Pirola GM, Azzoni P, Bevilacqua L, Eissa A, et al. *Ex vivo* fluorescence confocal microscopy: the first application for real-time pathological examination of prostatic tissue. BJU Int. 2019;**124**(3):469–76.
76. Rezende RM, Lopes ME, Menezes GB, Weiner HL. Visualizing lymph node structure and cellular localization using ex-vivo confocal microscopy. J Vis Exp: JoVE. 2019 Aug; **2019**(150):e59335.
77. Jost Fouke S, Benzinger T, Gibson D, Ryken TC, Kalkanis SN, Olson JJ. The role of imaging in the management of adults with diffuse low grade glioma: a systematic review and evidence-based clinical practice guideline. J Neuro-Oncol. 2015 Dec;**125**(3):457–79.
78. Galldiks N, Law I, Pope WB, Arbizu J, Langen KJ. The use of amino acid PET and conventional MRI for monitoring of brain tumor therapy. NeuroImage: Clin. 2017 Jan;**13**:386–94.
79. Yamane T, Sakamoto S, Senda M. Clinical impact of 11C-methionine PET on expected management of patients with brain neoplasm. Eur J Nucl Med Mol Imaging. 2010 Nov; **37**(4):685–90.
80. Young Yoo M, Chul Paeng J, Jeong Cheon G, Soo Lee D, Key Chung J, Edmund Kim E, et al. Prognostic value of metabolic tumor volume on 11C-methionine PET in predicting progression-free survival in high-grade glioma. Nucl Med Mol Imaging. 2015 Dec;**49**(4):291–7.
81. Zandvliet M. Canine lymphoma: a review. Veterinary Quarterly. 2016 Apr;**36**(2):76–104. https://doi.org/10.1080/01652176.2016.1152633
82. Ponce F, Marchal T, Magnol JP, Turinelli V, Ledieu D, Bonnefont C, et al. A morphological study of 608 cases of canine malignant lymphoma in France with a focus on comparative similarities between canine and human lymphoma morphology. Vet Pathol. 2010 May;**47**(3):414–33.
83. Curran K, Thamm DH. Retrospective analysis for treatment of na¨ıve canine multicentric lymphoma with a 15-week, maintenance-free CHOP protocol. Vet Comp Oncol. 2016 Aug;**14**:147–55.
84. Rao S, Lana S, Eickhoff J, Marcus E, Avery PR, Morley PS, et al. Class II major histocompatibility complex expression and cell size independently predict survival in canine B-cell lymphoma. J Vet Inter Med. 2011 Sep;**25**(5):1097–105.
85. Valli VE, Kass PH, San Myint M, Scott F. Canine lymphomas: association of classification type, disease stage, tumor subtype, mitotic rate, and treatment with survival. Vet Pathol. 2013 Feb;**50**(5):738–48. http://dx.doi.org/10.1177/0300985813478210.
86. Blom K, Nygren P, Larsson R, Andersson CR. Predictive value of ex vivo chemosensitivity assays for individualized cancer chemotherapy: a meta-analysis. SLAS Technol. 2017 Jan;**22**(3):306–14. https://doi.org/10.1177/2472630316686297.
87. Dietrich S, Oleś M, Lu J, Sellner L, Anders S, Velten B, et al. Drug-perturbation-based stratification of blood cancer. J Clin Invest. 2018 Jan;**128**(1):427–45.
88. Frismantas V, Pamela Dobay M, Rinaldi A, Tchinda J, Dunn SH, Kunz J, et al. *Ex vivo* drug response profiling detects recurrent sensitivity patterns in drug-resistant acute lymphoblastic leukemia. Blood. 2017 Mar;**129**(11):e26–37.
89. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Wai Lau K, et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature. 2012 Mar;**483**(7391):570–75.
90. Sharma SV, Haber DA, Settleman J. Cell line-based platforms to evaluate the therapeutic efficacy of candidate anticancer agents. Nat Rev Cancer 2010 Mar;**10**(4):241–53.
91. Vaira V, Fedele G, Pyne S, Fasoli E, Zadra G, Bailey D, et al. Preclinical model of organotypic culture for pharmacodynamic profiling of human tumors. Proc Natl Acad Sci U S A. 2010 May; **107**(18):8352–56.
92. Sznol M, Chen L. Antagonist antibodies to PD-1 and B7-H1 (PD-L1) in the treatment of advanced human cancer. Clin Cancer Res: Off J Am Assoc Cancer Res. 2013 Mar;**19**(5):1021–34.
93. Smalley M, Przedborski M, Thiyagarajan S, Pellowe M, Verma A, Brijwani N, et al. Integrating systems biology and an ex vivo human tumor model elucidates PD-1 blockade response dynamics. iScience. 2020 June;**23**(6):101229.

33

Tumour-immune co-evolution dynamics and its impact on immunotherapy optimization

Annice Najafi and Jason George

33.1. Historical aspect of immunotherapy

Tumour immunology and cancer immunotherapy have recently become tremendously active areas of research, but the foundations of this exciting field have been developed over centuries (**Figure 33.1**) [1]. The idea of immunity can be traced back to 430 BC when Thucydides discovered that Athenians infected with the plague did not become infected a second time [2]. This was followed by the famous thirteenth-century story of Peregrine Laziosi, a cancer patient who, on the day of his scheduled leg amputation, had exhibited significant tumour regression. At the time, the immune system's role in cancer elimination was unknown; subsequently, Laziosi became the patron saint for people living with cancer [3]. About four centuries later, a cascade of events led to the advent of the first vaccine. As smallpox spread across Europe during the seventeenth century, Lady Mary Montagu, the wife of the British ambassador in Istanbul, was intrigued by the Ottoman practice of inoculation and attempted to return it to England by granting Charles Maitland a royal licence to test the variolation method on willing prisoners, all of whom survived smallpox. Maitland then inoculated the two daughters of the Princess of Wales, resulting in the public's general approval. In 1796, Dr Edward Jenner inoculated an eight-year-old patient against smallpox with cowpox viral material collected from an infected milkmaid. The patient experienced mild infection and healed shortly thereafter, which later led to the widespread use of vaccination in England [4,5]. By 1878, Louis Pasteur successfully vaccinated lab chickens against cholera and extended the same immunization method for anthrax and rabies [6].

After the advent of the first vaccines, a series of key observations of tumour regression occurring in the setting of bacterial infections were made in Europe. In 1725, Deidier first noted that syphilis patients developed fewer tumours, and Sir James Paget observed more generally that infections could cause tumour regression. This latter observation led to subsequent attempts to induce tumour regression via bacterial infection. In 1869, Wilhelm Busch deliberately inoculated a cancer patient with erysipelas and observed tumour regression, and at the time, the underlying causative was unknown [7]. Shortly thereafter, Friedrich Fehleisen repeated the same experiment and ultimately identified *Streptococcus pyogenes* as the primary contributor to this process [8].

In New York, Dr. William Coley also noted sarcoma regression following infection with erysipelas. He attributed these miraculous post-surgical regressions to the stimulation of the immune system. Coley began experimenting with his hypothesis by injecting a solution known as Coley's toxins, made of two dead bacteria: *Streptococcus pyogenes* and *Serratia marcescens* [7,8]. Coley is now known as the 'father of immunotherapy', but his approach was met with substantial scepticism by the scientific community. Variability in the site and method of inoculation, and even questionable cancer diagnoses in some patients, complicated the validity of his results. As a result, the earliest form of immunotherapy was definitively abandoned [7,8].

The next significant discoveries occurred in the 1920s at Johns Hopkins when Raymond Pearl noted an inverse relationship between cancer and tuberculosis (TB) [9]. Holmgren noted the lack of allergy to tuberculin in cancer patients and exploited this observation to diagnose cancer. He later attempted to treat gastric cancer patients with the available vaccine against TB, bacillus Calmette–Guerin (BCG). Although Holmgren mentions immune stimulation in a rectal cancer patient, he more generally notes 'no remarkable' effect on cancer growth [10]. In the decades to follow, several doctors noted the effectiveness of this vaccine in fighting against melanoma and bladder tumours [11–13], with this vaccine eventually receiving FDA approval for the treatment of carcinomas in 1990 [14]. The underlying mechanism is still unclear, although it is known that cytotoxic T-cell-mediated immunity, antigen presentation, and pro-inflammatory responses are augmented through the utilization of BCG [14].

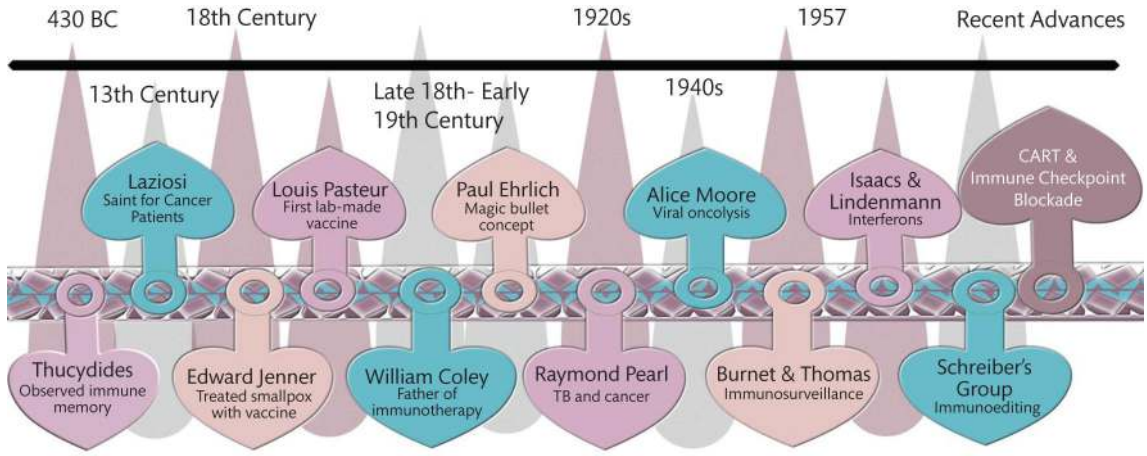


Figure 33.1. The historical aspect of immunotherapy, demonstrating the key historical events of the field of immunotherapy.

In addition to tumour-specific immune regression, independent effects of tumour targeting were emerging. George Beatson discovered a connection between sex steroid hormones and hormone-sensitive cancers. Beatson began treating breast cancer patients through the surgical removal of ovaries (oophorectomy). Although this method later became widely accepted and is still in use, the underlying mechanisms linking hormones, inflammation, and tumour growth were unknown at the time [15]. Several conceptual advances followed this: Rudolph Virchow noted an association between cancer and inflammation [16], and the Noble Prize laureate Paul Ehrlich conceptualized the existence of cell surface receptors capable of recognition and binding to specific molecules. This led Ehrlich to work on developing preferential chemical targets for these receptors. Ehrlich also postulated that the immune system could repress several carcinomas [17,18].

In the 1940s, Alice Moore performed the first viral oncolysis [19]. Moore noted that both slow-growing and rapidly proliferating tumours were unaffected by the Russian Encephalitis virus in the study suggesting a higher potential for therapeutic escape for highly proliferating and slowly proliferating tumours [20]. This observation was rediscovered in the setting of tumour-immune interactions about a decade later when George Klein defined the rejection of medium-sized inocula over small-sized and large-sized inocula referred to as the 'sneaking-through' phenomenon of tumours [21].

This was followed by the discovery of interferons (IFNs) in 1957 by Isaacs and Lindenmann, named for their ability to interfere with viral infections [22]. IFNs were later purified and administered to patients and yielded encouraging results against nearly every type of cancer [23]. Additional leukocyte-derived cytokines known as 'interleukins' (IL) [24] were later discovered [25] and have been implicated in mediating tumour-immune infiltration [26]. The clinical use of interleukin-2 (IL-2) immunotherapy in 1987 was followed by the inoculation of murine models by IL-2 expanded tumour infiltrating leukocytes (TILs) resulting in the total eradication of tumours [27]. Collectively, the above discoveries established a resurgence in interest for developing immunotherapies in the first half of the twentieth century. This work laid the scientific foundation for subsequent development of the modern principles of cancer immunotherapy.

33.2. Immunosurveillance and immunoediting

In 1957, Frank Macfarlane Burnet and Lewis Thomas hypothesized the cancer immunosurveillance concept that thymus-dependent immune cells interrogate host cells to detect inappropriately transformed cells [28,29]. Osias Stutman argued that immunosurveillance implies hosts with defective immune systems should have higher rates of occurrences of either chemically induced or spontaneous tumours and proceeded to test this empirically [30]. Unfortunately, his analysis was limited to CBA/H nude immunocompromised mice at the time, which had many shortcomings, including the presence of detectable αβ T cells and fully functioning tumour suppressor systems such as p53. These limitations confounded his ultimate finding of no statistical difference in tumour incidence between immunocompromised and control groups. As a result, the field experienced waning scientific interest in the immunosurveillance hypothesis for several decades [31–33].

Renewed interest in immunosurveillance occurred when scientists repeated Stutman's experiment and found that nude BALB/c mice died more often of cancer [34]. The protective role of interferon γ(IFN-γ) was discovered when 129/SvEv mice, without the IFN-γ receptor, and STAT1 were shown to be more sensitive to the tumour-inducing ability of 3-methylcholanthrene (MCA) [35]. The immunoediting concept was then born by the Schreiber group. They utilized recombinant activating gene knockout mice, which results in complete immunodeficiency of T cells, natural killer (NK) cells, and B cells. Their work showed that the 129/SvEv *RAG2*$^{-/-}$ mice were more susceptible to tumours than the control highlighting the role of immune system in recognizing tumours. Intriguingly, wild-type tumours' reduced immunogenicity compared to *RAG2*$^{-/-}$ tumours in the absence of a functioning immune response suggested the immune system's role in refining the tumour's immunogenicity [31–33,36].

Collectively, these findings demonstrated the key role of immunosurveillance on tumour progression and elimination, and the more general concept that immunity not only prevents but also sculpts the immunogenicity of developing tumours became known as the immunoediting hypothesis [31,32]. Schreiber's group further defined three associated dynamical phases to this

process: elimination, equilibrium, and escape [31,32]. During the elimination phase, tumour cells that have lost their original tumour suppressive features draw the attention of the host immune system and become targeted. Both innate and adaptive immunity are at play with the infiltrating lymphocytes recognizing antigens and secreting IFN-γ. IFN-γ contributes direct apoptotic and anti-proliferative effects on cancer cells as well as indirect suppressive mechanisms, including angiogenesis inhibitors. Dendritic cells (DCs) follow cell debris, draining them into the lymphatic vessels. Meanwhile, chemokines provoke an immune response by recruiting NK cells and macrophages to the site. In some cases, tumours are completely destroyed in this phase. However, should the immune system fail to eradicate tumour cells completely, the residual population enters the equilibrium phase [31–33].

In the equilibrium phase, cells are kept in a dormant state which takes the longest period among the three phases of immunoediting [31–33]. However, tumour dormancy and its underlying mechanisms are poorly understood [37,38]. For example, it is unclear to what extent cancer cells become truly dormant, with zero growth and immune-mediated death versus a state of net zero growth where birth and death rates are matched. Some studies have shown that tumour cells can exhibit non-proliferating properties while halted in a cell cycle arrest in G1/G0 through the combination of the effects of the tumour necrosis factor (TNF) and IFN-γ [38,39]. While other studies emphasize population or angiogenic dormancy in which case the tumour population size remains constant, and its growth is counterbalanced with its death through apoptosis due to which the tumour fails to vascularize and may remain below a clinically detectable size, increasing the likelihood of immune evasion [37]. Koebel et al. showed that dormant tumours would only become malignant in the absence of T lymphocytes, IFN-γ or IL-2, while they failed to progress in the absence of innate immunity components, such as NK cells, thus highlighting the importance of the adaptive immunity in maintaining equilibrium. In addition, they showed that the dormant tumour allowed considerable infiltration of immune cells, such as T cells, macrophages, and $B220^{+}$ cells, resulting in a higher death rate [38,40]. The role of the adaptive immune-mediated modulation of tumour dormant states is gaining interest as a potential therapeutic avenue [31–33,36,38]. If cancers escape from the dormant state, the tumour cells form an immunosuppressive environment, and the immune cells can no longer act on inhibiting the growth of the tumour. This marks the transition into the final elimination phase [31–33].

33.3. Adoptive cellular therapy and checkpoint blockade

Adoptive cellular and checkpoint blockade therapies emerged based on the concept that immunotherapy is constricted by regulatory mechanisms that prevent immune activation and the existence of an immunosuppressive tumour microenvironment (TME). Adoptive cellular therapy utilizes immune cells to enhance anti-tumour immunity [41]. The earliest accounts of adoptive cellular therapies trace back to 1957, when Edward Thomas performed the first cellular therapy on a leukaemia patient. Given the immune ablative nature of early irradiation therapy, the patient was treated with bone marrow infusion from his identical twin brother [42]. It was later discovered that leukaemia remission rates were enhanced by a donor-derived T-cell-mediated graft-versus-host effect against the minimal residual cancer population [43,44]. HSCT is still in use today for treating haematological malignancies and is a curative option for many patients [45,46].

A more recently developed strategy, known as chimeric antigen receptor (CAR) T-cell therapy, was introduced in 1989 by Zelig Eshar and subsequently granted FDA approval for haematological malignancies in 2017 [47–49]. CAR T cells, composed of tumour-specific binding domains with T-cell intracellular signalling domains, kill target-bearing cells. Although clinical trials have demonstrated encouraging results, cancer progression and toxicities arise in some patients, necessitating further refinement and combination therapy for improved individualized treatment [47–49]. For example, CAR T-cell therapy may lead to cytokine release syndrome (CRS) [49–52], a serious side condition where large amounts of cytokines, including INF-γ, TNF-α, IL-6, and IL-2, are released. CRS can be mitigated by blocking the IL-6 signalling pathway or suppressing cytokine production with dexamethasone or tocilizumab [49,50]. Other treatments, such as corticosteroids, can reduce transduced T cells and thus reduce inflammation and graft-versus-host disease. At the same time, methotrexate can not only indirectly help with the treatment of CRS but also be effective in treating autoimmune diseases caused by off-target CAR T cells [50]. Furthermore, empirical evidence has shown that the body acquires resistance to CAR T-cell therapy through subsequent exposure reducing the effectiveness of this type of immunotherapy [50–52].

Although CAR T-cell therapy was initially designed for haematological malignancies, its utility has carried over to solid tumours through bispecific CAR T cells, which can target two antigens concurrently or through immune checkpoint inhibitors [53]. Other types of immune cells commonly used for adoptive cellular therapy are NK cells, cytokine-induced killer cells, lymphokine-activated killer cells, and DCs. While all but the latter activate in an antigen-independent manner against tumours, DCs have been used in immunotherapy due to their ability to regulate adaptive immunity [54].

Perhaps most significantly, the discovery of the immune checkpoint T-cell molecules. programmed death-1 (PD-1) and cytotoxic T lymphocyte antigen-4 (CTLA-4), by Nobel Laurates Tasuku Hanju and James Allison demonstrated an inhibition in T-cell proliferation that occurs upon binding with PD-L1 and B7, respectively. The discovery that these inhibitory molecules are often up-regulated in cancer populations led to checkpoint blockade strategies that successfully treat many cancer subtypes [55–56]. However, not all patients respond to treatment, and an immune-permissive TME and the proper immune recognition of tumours are required for effective therapeutic responses [57]. Studies of blockade therapies have shown that tumour mutational characteristics often correlate with response to therapy; thus, better therapeutic responses can be achieved by targeting mutant neoantigens [58,59]. Furthermore, studies focusing on non-small-cell lung cancer (NSCLC) samples have shown that better anti-tumour responses are achieved by targeting clonal neoantigens [59].

Given the early success of immune checkpoint blockade, there was an immediate question of how cancers evolve in a microenvironment following treatment. To address this, Riaz et al. studied a cohort of melanoma patients sub-categorized based on previous CTLA-4 inhibition using ipilimumab (IPI). These patients were

then given PD-1 blockade agent Nivolumab (Nivo) and assessed via whole exome and T-cell receptor (TCR) sequencing. Riaz et al. observed a reduction of mutations and neoantigen diversity post-treatment with Nivo in both IPI-treated and IPI-naive groups, which is consistent with active immunoediting taking place [61]. They also discovered an increase in the fraction of TILs [61], a common indicator of therapeutic response [60], upon Nivo treatment relative to that of (untreated) IPI-naive patients. However, this enrichment was significantly greater in IPI-naive patients than in IPI-treated ones. Although similar rates of therapeutic response were observed in both groups, overall survival was only associated with tumour mutational load in the IPI-naive group [61], suggesting TME dependence on the difference in effects between checkpoint blockade agents. Lack of association between therapeutic response and mutational load suggests an alternative, non-genetic driver of outcome.

33.4. Intra-tumoral heterogeneity

As established above, if immunoediting is the driving force that imposes a continuous selective pressure on cancer, then intra-tumoral heterogeneity (ITH) may be viewed as one evolutionary consequence that can lead to population-level diversification and impaired therapeutic response. The failure of immunoediting and immunotherapy is often attributed to ITH [63] that was first discovered by Rudolph Virchow. Virchow observed ITH in the form of variations in cancer cell shape and morphology [64]. Huxley further generalized this concept by categorizing features of ITH as genetic, epigenetic, and environmental [65]. ITH arises from several environmental pressures in the form of metabolic, immunological, pH changes, and therapeutic components. Mutations that randomly accrue throughout tumour development diversify the tumour population, resulting in competition for survival in the presence of selection [63, 66–68].

Furthermore, ITH may also manifest within the neighbouring cells of the tumour, affecting the stroma and components of the immune system. This heterogeneity often confers a survival advantage through phenotypic plasticity. Throughout development, the tumour population and corresponding microenvironment undergo adaptive alterations that contribute to the population-level heterogeneity. On the other hand, immunoediting selects tumour subclones with low antigenicity so that ITH drives tumour progression and immune escape. Mounting experimental evidence suggests that ITH results in worse therapeutic response suggesting ITH should be targeted for therapy [63]. However, although spatial ITH can be therapeutically targeted, temporal ITH may be an aggressive driver of tumour progression and, thus, more difficult to target [63]. ITH is complemented by heterogeneity in the immune response to tumours. In particular, the state of the T cells' response can vary widely in time and across patients. To complicate matters, the effects of cancer immunosurveillance prior to detection also appear to influence the likelihood of immunotherapy success. Tumour samples from CTL-infiltrated lung cancer patients demonstrate antigen depletion and MHC loss [69]. In solid tumours, such as hepatocellular carcinoma and lung cancer, tumour antigen-specific CTL enrichment appears to reflect spatial geography of malignant populations as well as whether or not antigens are ubiquitous or localized [70,71].

33.5. Unravelling patterns of tumour evolution

In the preceding section, we discussed ITH and its contribution to therapy resistance. Here, we provide a brief history and a description of the endeavours that revealed patterns of clonal evolution leading to ITH. In the early 1900s, Boveri postulated the somatic mutation theory, proposing that chromosomal rearrangements result in the loss of the segment of chromosomes responsible for inhibiting cell proliferation and thus lead to cancer [72]. This was followed by Whitman's postulate that cancer initiates from mutated cells [72]. Shortly thereafter, scientists discovered that carcinogenic ionizing radiation is also mutagenic, providing additional evidence supporting the relationship between somatic mutations and cancer [73].

Meanwhile, two opposing views emerged concerning whether mutant acquisition was the result of environmental changes, or if it was pre-existent and emerged as a result of environmental selective pressures. This dilemma remained inadequately addressed until 1943, when Max Delbrück and Salvador Luria investigated the origin of mutations in bacteria. In the Delbrück–Luria experiment, the viral agent infected five tubes of a bacterial strain susceptible to a specific virus. They demonstrated that the predicted fluctuation in the number of resistant colonies in the test tubes could resolve whether mutations were acquired versus pre-existent. This famous set of experiments showed that pre-existing mutations can arise prior to an inciting selective pressure [74].

By the end of the 1940s, Bauer's theory of a single mutated cell resulting in cancer received several objections from the scientific community regarding its failure to explain the relationship between age and cancer incidence. Although Bauer responded to the objections by stating that mutated cells remain latent in the body in a carcinogen-dependent manner, the scientific community soon followed up with the very first attempts at mathematical modelling of cancer [75]. With Muller's concept proposed in 1951 that the acquisition of multiple mutations by a single cell is required for the occurrence of cancer [76], three statistical projects investigated the relationship between age and cancer mortality [75,77,78].

33.6. Cancer-age incidence models

Fisher and Hollomon in 1951 and Nordling in 1953 used data from stomach cancer in US women and statistics from Europe and the United States about cancer-related mortality in men, respectively, to relate age and cancer incidence. They concluded that the logarithm of age and logarithm of cancer-related death rate are proportional, and the death rate increases six times as rapidly between the ages of 25 and 74 years. Data from young people were excluded because they believed that the underlying mechanisms behind cancer at early ages are not the same at later ages, and data from older people are unreliable, perhaps due to the low number of samples and high mortality rate. Fisher and Hollomon hypothesized the need for a critical number of cancer cells for malignant growth (critical size hypothesis), leading to the conclusion that cancer cells and cancer incidence are proportional to the fifth/sixth power of the concentration of carcinogen, which was contradictory to experimental data.

Armitage and Doll took the approach of assuming mutation rates to be rare but constant throughout the life of the patient [75,77,78]. From this, the rate r of cancer incidence could be explained by

$$r = kp_1p_2p_3p_4p_5p_6t^6$$

where r is the cancer incidence rate at age t, k is a constant, and p_i is the probability of occurrence of the i th successive mutation. That means the logarithm of the cancer incidence rate is directly proportional to the logarithm of the age [79]:

$$\log r = 6\log t + C$$

Armitage and Doll were determined to investigate this hypothesis further for cancers of different sexes and sites. As opposed to Nordling, Armitage and Doll did not restrict their description of successive events to mutations. They found that the data is consistent with the results of the previous two studies in cancers that are more likely to be independent of environmental factors for both sexes. Thus, the effects of carcinogens remain constant throughout life. On the other hand, the data deviated from the proposed model in cancers where developmental factors affect the development of the tumour variably, such as prostate and lung cancer. The model proposed by Armitage and Doll is often referred to as the multi-stage theory of carcinogenesis [75,77,78].

In other instances, tumour incidence was discovered to follow specific trends. In 1971, Alfred Knudson observed, among 48 retinoblastoma cases, that children with hereditary retinoblastoma typically developed the tumour in both of their eyes as opposed to others who developed the tumour in only one eye. Based on these observations, Knudson proposed the concept of a tumour suppressor gene and postulated that children with hereditary retinoblastoma have one defective copy of the gene, which was mutated in the germline. Because germline mutations impact all somatic cells, children with hereditary retinoblastoma develop more tumours. On the other hand, individuals with sporadic cases of retinoblastoma born with two normal alleles need to develop mutations in both copies in somatic cells. This concept became known as the 'two-hit' hypothesis. In his paper, Knudson showed that the relationship between age and the fraction of retinoblastoma cases that are not diagnosed is linearly proportional in cases of children with multiple cancers in both eyes, while it is better explained with a quadratic function of age in months for children with only one tumour. Knudson concluded from the data that tumours are distributed with a Poisson distribution, and he estimated the mean number of tumours per gene carrier [79]. This study was followed by probabilistic models proposed by Moolgavkar and Knudson describing the incidence of retinoblastoma and cancer initiation and progression [80].

33.7. Patterns of tumour evolution

During the 1960s and 1970s, cytogenetic and karyotyping studies gave insights into the contribution of polyploidy to cancer. Notably, the study by Felix Mitelman in 1971 showed heterogeneity in chromosomal numbers and clonal evolution in 50 albino Rous mice. In that study, a stepwise accumulation of chromosome numbers was shown to be associated with loss of differentiation and decreased collagen production in sarcoma cells [81]. A separate study investigated the inactivation of the X-chromosome at the glucose-6-phosphate dehydrogenase locus in women who were heterozygous for the gene. Although the existence of both isoenzymes was observed in normal tissue, only one phenotype was functional in tumour cells suggesting tumour clonality [82].

In 1975, Peter Nowell concluded the clonal pattern of tumour evolution from the studies mentioned above and suggested a unicellular origin for tumours. Nowell developed the first model of linear tumour evolution. In his model, a single cell acquires a selective fitness advantage over neighbouring cells via an induced change. This is followed by neoplastic progression either immediately or after a latent period. Later if a new genetic variant is induced, it will get eliminated due to metabolic disadvantages or immunological destruction. However, if this new variant provides further fitness advantage, it will soon dominate the population with its progeny. This process is commonly known as a selective sweep. The utility of this model diminished as higher resolution data was acquired, and its most successful application describes colorectal cancer evolution [83].

Branched evolution was the next alternative evolutionary framework applied to understand cancer progression whereby expansion of distinct clones can occur in parallel, each with a low likelihood of selective sweeps. One of the most famous models of this type is the multi-type branching model developed by Bozic et al., where tumour evolution was modelled as a stochastic discrete time branching process [84]. In this model, in each step, either a cell dies or gives birth to two daughter cells. The daughter cell can either be identical to the producing cell or acquire a mutation that changes the fitness advantage. The probability of giving birth in each generation is referred to as b_j, while the probability of dying is d_j. While a cell can either die or give birth at every time point, and thus $b_j + d_j = 1$, it may give birth to a mutated cell with probability u. The death rate is variable and changes as a function of generation j and may be explained via the function $d_j = 1/2(1-s)^j$ [84].

Two types of mutations were considered in this process, drivers, which conferred a fitness advantage, and passengers, which were neutral. Through this model, Bozic et al. found the number of passenger mutations as a function of driver mutations. They used published sequencing data and the Cancer-specific High-throughput Annotation of Somatic Mutations (CHASM) algorithm, a supervised learning method using a random forest model for identifying missense mutations, for conferring the fitness advantage of mutations. After fitting their model to the data, they identified the same average selective advantage as 0.4% for the two types of cancers considered, glioblastoma multiforme and pancreatic adenocarcinoma. These results suggest that the average selective advantage of drivers is not dependent on the cancer type and is universal. Additionally, the selective advantage of drivers is very low, which suggests that the rate of extinction of driver mutations is very high. In addition, Bozic considered the model when transitioning to a continuous-time process. Bozic noted that by making the death rate constant $d = 1$ while switching the birth rate to $b = 1 + sj$ (s is the selection parameter and j is the generation), the population becomes infinite at a finite time. On the other hand, the death parameter can be tweaked to $d = 1 - sj$ while the birth parameter remains constant as $b = 1$ to solve this problem [84].

Although the simplicity of this foundational mathematical model makes it amenable to analytic characterization and direct and universal experimental testing, both of which become nontrivial in

models with many parameters, this model assumes an exponential pattern of tumour growth. Thus, it loses accuracy in describing populations undergoing Gompertzian growth. Another issue of this model is that it does not consider the effects of deleterious passenger mutations. A more recent study by McFarland et al. utilized a stochastic population model and framed tumour progression in the context of a tug of war between driver mutations and deleterious passenger mutations, with the death rate increasing as a Gompertzian function. Interestingly, their model predicted a significantly different average selective advantage than the Bozic model [84,85].

The above model was extended to describe mutational events resulting in the emergence of resistant phenotypes. In this model, mutations are categorized as deleterious, neutral, or advantageous depending on their growth rate relative to that of the original population. Mutated clones arrive according to a Poisson process with parameter $\lambda = Mu$, where u is the mutation rate and M is the population size, from which Bozic et al. estimated that the expected number of clones is k when the population reaches k/u and thus the probability of k clones in a population of size M is $\lambda^k e^{-\lambda} / k!$. Intriguingly, their results suggest that the ratio of medians and means for the number of cells for the two subclones is parameter free, which they attributed to the fact that the branching process with a large size and slight mutation ($M \gg 1 and\, u \ll 1$) possibly can be approximated by a pure birth process. This parameter-free ratio was then confirmed from clonal abundances measured by liquid biopsy-derived circulating tumour DNA data. Subsequent work developed a mathematical framework for estimating the relative sizes of treatment-resistant clones and the probability of resistant clone acquisition based on the tumour size, thereby providing insight into the heterogeneity of the tumour population, which is important for therapeutic planning [86].

Multi-type branching frameworks can also be utilized in a more specific tumour-immune interaction setting. Iwasa et al. investigated the dynamics of immune escape utilizing a continuous-time branching process where a heterogenous quasi-species struggles to survive with a low reproductive ratio and needs to accrue mutations in several nucleotides under harsh selective pressures, such as an adaptive immune system, to successfully invade the host. Immune escape can then be modelled through a Galton–Watson branching process whereby a wild-type sequence with a subcritical reproductive ratio undergoes a number of successive forward mutations (which can happen simultaneously) to reach an escape sequence with a reproductive ratio exceeding 1. Their model shows that successful immune escape depends on population size, reproductive ratios, mutation rates, and initial fitness of the invading host. They also emphasize that the independence of lineages assumed in their model is a limitation and may fail in capturing the transfer of genetic information between the lineages [87].

Through the linear and branched patterns of tumour evolution mentioned above, the tumour originates from a single mutated cell and propagates in a stepwise fashion; however, some experimental evidence supports another pattern of tumour evolution called punctuated evolution. Through punctuated evolution, a plethora of mutations occur at the early stages of tumour evolution and only one or few of these clones expand afterwards. Thus, in contrast to branched and linear evolutions where the origin of cancer can be traced back to point mutations, in punctuated patterns of evolution, cancer is often caused by copy number variations and chromosomal abnormalities. The tumour's pattern of evolution has important therapeutic values as cancers propagating with a branched or linear pattern of evolution can easily be targeted for mutations or copy number variation since most of the population is uniform while punctuated evolution suffers from immense ITH and successful eradication of the tumour is contingent upon correctly targeting truncal mutations [66].

Although next-generation sequencing data has allowed us to investigate the tumour extensively at the genetic level, single-biopsy samples fail to capture the temporal and spatial heterogeneity needed to properly assess evolutionary models of cancer progression. Suitably specialized experimental studies have begun to emerge that can offer further insight. Perhaps, most notably, the TRACERx project investigates the spatial pattern of tumour evolution in NSCLC, clear-cell Renal Cell Carcinoma (ccRCC), longitudinal and multi-region samples with the goal of optimizing therapy based on the patient's genetic profile [88,89]. These studies predict slow cancer growth over prolonged periods that were then followed by immune escape. In the NSCLC data, heterogeneity of copy number variations, and not mutations, was associated with worse patient outcomes, while the ccRCC data was characterized by loss of chromosome 3p, associated with tumour development by ages of 30–50 years [88,89]. The TRACERx cohort, and similar approaches, should be championed for providing a suitable amount of data for comparison to predicted trajectories of sophisticated mathematical models.

33.8. Models of tumour-immune interaction

Unlike traditional treatments that target specific pathways in cancer cells, immunotherapy acts indirectly by stimulating the patient's immune system and in doing so imposes a distinct type of evolutionary pressure [90]. This results in significant variability in immune response and cancer progression. Mathematical models play a crucial role for predicting the complex behaviour of tumour-immune interactions where tumour, immune system, and immunotherapy are mutually affected by one another. Mathematical models that have been developed over the past 40–50 years have identified variables affecting the efficacy of therapy, thereby improving patient outcomes [91]. This section provides a brief review of some of the models used to study tumour-immune interactions. A majority of the available models can be sub-categorized into predator–prey models, deterministic differential equation models, and stochastic process models.

33.8.1. Predator–prey

The first use of the predator–prey models in oncology can be traced back to the early 1970s when Bell utilized a simplistic predator–prey framework to model and track the dynamics of tumour-immune population in different parameter regimes [92]. This was followed by 1977 with DeLisi and Rescigno, where they modelled the interaction between a population of lymphocytes and tumour cells and showed that the probability of immune evasion increases with tumour size [93]. A later predator–prey model by Bocharov et al. captured the sneak-through phenomenon for lymphocytic choriomeningitis viral infection, which, similar to tumour progression, may result in replication and immune evasion [94]. Although predator–prey models have been widely used to explain cell–cell interactions in the tumour-immune setting, their use is questionable due to their underlying assumptions [95–98]. In these models,

the existence of predators depends on the density of the prey population that corresponds to T cells' clonal expansion and contraction in the presence and absence of abundant antigenic signatures. However, when applied to study immune recognition of cancer, these models implicitly assume T-cell recognition of cancer, which in general is not guaranteed to be constant over the relevant time interval [98]. These models therefore omit a complete description of the stochastic tumour-immune recognition that is important for elimination.

33.8.2. Ordinary differential equations

Ordinary differential equations (ODEs) and partial differential equations (PDEs) are another type of tumour-immune interaction models commonly utilized. In these models, the dynamics of populations are tracked using ODE if time is the only independent variable, while they may be represented by PDE if multiple spatial dimensions are considered. Here, we provide two examples of ODE models.

33.8.3. ODE of macrophages: a double-edged sword

Macrophages are innate cells and known to possess both pro-tumorigenic and anti-tumorigenic effects. Highly influenced by stromal signalling, monocytes can potentially differentiate into an M1 state and display tumour inhibitory behaviour through the secretion of cytotoxic elements such as nitric oxide or they may differentiate into a tumorigenic M2 state and secrete growth promoting cytokines [99]. Early models interrogated the role of macrophages in eradicating tumour cells in avascular settings [100] followed by models investigating the ability of engineered macrophages in targeting tumour cells or delivering drugs that showed unusual sensitivity of these therapies to tumour and therapy parameters [101–102].

Using a differential equation model, den Breems and Eftimie investigated the interplay between macrophage polarization and tumour progression [103]. Through their model, the dynamics of M1 and M2 macrophages can be tracked via the differential equations below, which allow for two-state transitions and exogenous cellular and cytokine activation:

$$\frac{\mathrm{d}X_{\mathrm{M1}}}{\mathrm{d}t} = (a_s x_{\mathrm{Ts}} + a_{\mathrm{m1}} x_{\mathrm{Th1}}) x_{\mathrm{M1}} \left(1 - \frac{x_{\mathrm{M1}} + x_{\mathrm{M2}}}{\beta_{\mathrm{M}}}\right) - \delta_{\mathrm{m1}} x_{\mathrm{M1}} - k_{12} x_{\mathrm{M1}} x_{\mathrm{M2}} + k_{21} x_{\mathrm{M1}} x_{\mathrm{M2}}$$

$$\frac{\mathrm{d}X_{\mathrm{M2}}}{\mathrm{d}t} = (a_s x_{\mathrm{Tn}} + a_{\mathrm{m2}} x_{\mathrm{Th2}}) x_{\mathrm{M2}} \left(1 - \frac{x_{\mathrm{M1}} + x_{\mathrm{M2}}}{\beta_{\mathrm{M}}}\right) - \delta_{\mathrm{m2}} x_{\mathrm{M2}} + k_{12} x_{\mathrm{M1}} x_{\mathrm{M2}} - k_{21} x_{\mathrm{M1}} x_{\mathrm{M2}}$$

In the above equations, X_{M1} and X_{M2} represent the density of M1 and M2 macrophages; x_{Ts} is the density of immunogenic tumour cells, in the presence of which M1 macrophages get activated with an activation rate of a_s; x_{Tn} is the density of non-immunogenic tumour cells, in the presence of which M2 macrophages get activated with an activation rate of a_s; M1 and M2 macrophages get activated in the presence of Th1 and Th2 helper T cells (represented by x_{Th1} and x_{Th2}) with rates a_{m1} and a_{m2}, respectively. The death rates of the two types of macrophages are represented by δ_{m1} and δ_{m2} for M1 and M2 macrophages, respectively, where they transition from an M1 state to M2 and vice versa with rates k_{12} and k_{21}.

Using this model, Den Breems and Eftimie showed that type II immune responses correlate with tumour growth and the k_{12} and k_{21} are related to the delay in tumour growth and tumour size. Clinically, it has been observed that the prominent presence of M2 macrophages within the TME leads to worse patient outcomes [103]. This was further supported by another model that considered the effects of Tie2 Expressing Macrophages (TEMs) emanating from a separate precursor macrophage. Although phenotypically similar to M2 macrophages, these macrophages facilitate angiogenesis and tumour growth through both Angiopoeitin-2 and differentiation to M2 macrophages through interleukin-10. Interestingly, the results showed that the active presence of M2 macrophages regardless of TEM results in larger tumour growth highlighting the potential failure of TEM ablation in an M2-abundant TME [104].

33.8.4. Sneaking through with ODEs

Following up on Klein's discovery of the 'sneaking-through' phenomenon, in 1980 Grossman and Berke proposed an ODE model considering the interactions between immunogenic tumour cells and T lymphocytes. They tracked the dynamics of three types of lymphocyte populations, X, Y, and Z, representing precursor, proliferating, and killer cells, respectively [105].

In their model, the precursor lymphocyte population represented by X grows and decays at a rate of s and e', respectively. X cells differentiate into proliferating lymphocytes denoted as Y in the presence of antigenic tumour cells and at a rate that is proportional to their density and the antigen density denoted as Ag. Y cells grow for τ time until they reach a resting and immunocompetent state (memory cell), and as this state was considered functionally similar to the X state, the two states are considered the same. If further stimulation by antigens takes place, in that case Y cells divide at a rate of dy for a period of τ' until they become cytotoxic killer cells denoted as Z. The tumour cells on the other hand (density equivalent to Ag) replicate at a constant rate of b. Z cells then kill tumour cells at a fast pace and die at a rate of c. Another type of immunogenic antigen molecules, denoted by Ag′, are assumed to be generated by tumour cells at a rate of b' or by destroyed tumour cells at a rate proportional to the rate of killing. These cells are then removed with a rate of b''. Grossman and Berke considered the effects of blocking factors that are antigen-bound antibody complexes or free antigens capable of blocking the immune system from destroying tumour cells through the mass-action law. However, they did not consider the blocking of killer cells as their blocking was deemed unimportant in tumour immunology. The densities of the three types of lymphocytes can be tracked via the differential equations below where the constants (a, a', and a'') denote the strengths of antigen–lymphocyte interactions and Ag_T is the total amount of antigen ($\mathrm{Ag}_T = \mathrm{Ag}+\mathrm{Ag}'$):

$$\frac{dX}{dt} = s - e'X - a\mathrm{Ag}_T X + eY(t-\tau)$$

$$\frac{dY}{dt} = a\mathrm{Ag}X - a'\mathrm{Ag}_T Y - dY$$

$$\frac{dZ}{dt} = a''\mathrm{Ag}(t-\tau')Y(t-\tau') - cZ.$$

The dynamics of tumour cells were then tracked via the following equation where k is a proportionality factor:

$$\frac{dAg_T}{dt} = \left(b + b' + b''\right)Ag - b''Ag_T + \frac{(k-1)aAgZ}{1+\left(\frac{\alpha}{\beta}\right)Ag}$$

Next, they performed stability analysis on steady-state solutions and concluded that the existence of a stable steady state depends on antigen density in a manner that leads to minimal immune responses to slow and rapidly growing threats. Thus, Grossman and Berke successfully captured the 'sneak-through' phenomenon of slow-growth threats using an ODE [105].

Another application of ODE models is in the context of immune-mediated tumour dormancy to understand escape mechanisms. Through these models, whereby cancer cells grow as a function of immune cells, tumour cell resistance to the immune system's line of defence can be modelled via reduced immune predation strength and immune recruitment potential. This set of models can be utilized to interrogate mechanisms of escape that prolong the tumour's dormant state and thus improve patient therapeutic outcomes [106].

33.8.5. Stochastic models of tumour-immune coevolution

To address some of the limitations of deterministic models mentioned previously, George and Levine devised a stochastic model to explain the coevolution of an adaptive immune system and cancer [107]. Their model considered cancer as a dynamic threat that may acquire a fully immune-evasive phenotype, such as class-I MHC down-regulation. They assumed cancer immune detection, should it occur, is possible once the population has exceeded a lower threshold, which may be determined by total size or total net growth rate, the latter reflecting the growth-threshold conjecture of immune detection [107,108]. Immune recognition probability was parameterized by TCR repertoire diversity and turnover, and from this they calculated the likelihood and mean arrival time of an immune-evasive phenotype. Their results demonstrated that the dynamical behaviour resulting from the growth-threshold conjecture yields tumours with slow and fast growth rates to have larger escape probabilities, providing the first analytic argument of 'sneak-through' as an emergent phenomenon resulting from stochastic cancer immune evasion. Size-limited detection, on the other hand, generates evasion probabilities that vary monotonically with net growth. These dynamics, while less relevant for understanding passive tumour evolution, could be applied to understand therapies where immune activation may largely be independent of cancer growth rate, such as CAR T-cell therapy [107].

This mathematical framework was further applied to account for the fact that TCR repertoire diversity and thymic output decrease as a function of age to predict acute myeloid leukaemia (AML) incidence. There, they modelled T-cell turnover, and repertoire diversity declines as a function of age using available data on T-cell receptor excision circles [109]. Their results showed that cancer, which can either arise due to immune-evasive mutations or population escape of immune recognition, mostly occurs due to the rare arrival of immune-evasive mutations at an early age, slightly increases because of a decrease in thymic output, and eventually decreases in late age due to lower TCR repertoire diversity. Moreover, this model of immune-specific cancer incidence significantly outperforms the multi-stage carcinogenesis model in describing AML incidence [107]. This finding seems to be general to many cancer types for models of cancer incidence that focuses on immune function [110].

This model was later extended [107] to intermediate threats such as neoantigen down-regulation or mutation in the continual presence of adaptive immunity. Through this model, a parent clone grows as a pure birth process, may face immune recognition, and die through a pure death process. However, the evolution of the tumour may continue if the parent clone successfully acquires a resistant phenotype prior to extinction. This process repeatedly continues until either elimination or escape occurs. This model evolves through a Galton–Watson branching process [111]. Through fitting the model to empirical data of multiple cancer types, they discovered a significant correlation between early cancer incidence (ages 0–40 years) and per-cell evasion rate. Thus, their model highlighted the importance of immunosurveillance in the early stages of the disease. Next, they applied their model to TRACERx ccRCC and NSCLC data. From the ccRCC data, they found that the tumour undergoes 27 recognition cycles before immune escape. From the parameter estimation of their model, they identified that initiation events are significantly higher than incidence, highlighting the existence of an active immune system. While they found that squamous cell carcinoma and smoking were related to an immunosuppressed system, from their calculations of the mean–variance of clonal non-synonymous mutations, they observed an opposite behaviour for non-smokers and adenocarcinoma. Overall, their results demonstrate the immense need for cancer-type-specific treatment strategies. Their model successfully related immune evasion and the branched pattern of tumour evolution [111].

Recent work has explicitly modelled tumour evasion via antigen loss and considers the aggressiveness of the cancer evasion strategy as a key model parameter. Cancer optimal evasions strategies can be solved for using stochastic dynamic programming, and immune microenvironmental features, including T-cell recognition rates, are predicted to affect the immunogenicity, ultimately resulting in 'cold' or 'warm' tumours [112].

33.9. Conclusions

In this chapter, we reviewed the key concepts in the history of cancer immunotherapy and evolution, wherein mathematical modelling has played a central modern role in improving patient outcomes. With the current popularity of immunotherapy and the widespread availability of high-resolution data, there is an even larger need for data-driven and theoretical models to better understand individual therapeutic failure resulting from patient-specific intra-tumoral heterogeneity. Stochastic modelling is an attractive framework for advancing our understanding, given the inherent random nature of tumour progression along with significant uncertainty in every detailed parameter required to fully define the relevant problems.

REFERENCES

1. Couzin-Frankel J. Cancer immunotherapy. Science. 2013 Dec 20;**342**(6165):1432–3.
2. Littman RJ. The plague of Athens: epidemiology and paleopathology. Mt Sinai J Med: J Transl Pers Med. 2009 Oct;**76**(5):456–67.
3. Krone B, Kölmel KF, Grange JM. The biography of the immune system and the control of cancer: from St Peregrine to contemporary vaccination strategies. BMC Cancer. 2014 Dec;**14**(1):1–13
4. Riedel S. Edward Jenner and the history of smallpox and vaccination. In Baylor University Medical Center Proc. 2005 Jan 1;**18**(1): 21–5. Taylor & Francis.
5. Willis NJ. Edward Jenner and the eradication of smallpox. Scottish Med J. 1997 Aug;**42**(4):118–21.
6. Ullmann A. Pasteur-Koch: Distinctive ways of thinking about infectious diseases. Microbe–Am Soc Microbiol. 2007;**2**(8):383.
7. McCarthy EF. The toxins of William B. Coley and the treatment of bone and soft-tissue sarcomas. Iowa Orthop J. 2006;**26**:154.
8. Butterfield LH, Kaufman HL, Marincola FM, editors. Cancer immunotherapy principles and practice. New York: Demos Medical Publishing; 2021 Aug 25.
9. Pearl R. Cancer and tuberculosis. Am J Hyg. 1929;**9**:97–159.
10. Holmgren I. Employment of BCG, especially in intravenous injection. Acta Med Scand. 1936 Jan 12;**90**(S78):350–61.
11. Rosenthal SR. Surgery, recall antigens, immunity, and bacillus Calmette-Guerin vaccination. Am J Surg. 1988 Jul;**156**(1):1–3.
12. Old LJ, Clarke DA, Benacerraf B, Stockert E. Effect of prior splenectomy on the growth of sarcoma 180 in normal and Bacillus Calmette-Guerin infected mice. Experientia. 1962 Jul;**18**(7):335–6.
13. Morton DL, Eilber FR, Holmes EC, Hunt JS, Ketcham AS, Silverstein MJ, Sparks FC. BCG immunotherapy of malignant melanoma: summary of a seven-year experience. Ann Surg. 1974 Oct;**180**(4):635.
14. Morales A. BCG: A throwback from the stone age of vaccines opened the path for bladder cancer immunotherapy. Can J Urol. 2017 Jun 1;**24**(3):8788–93.
15. Beatson GT. On the treatment of inoperable cases of carcinoma of the mamma. Lancet. 1896;**2**.
16. Schmidt A, Weber OF. In memoriam of Rudolf Virchow: a historical retrospective including aspects of inflammation, infection and neoplasia. In: Dittmar T, Zaenker KS, Schmidt A, editors. Infection and inflammation: impacts on oncogenesis. Vol. **13**. Contributions to Microbiology. Karger Publications; 2006. p. 1–15.
17. QUE RU. Paul Ehrlich (1854-1915): man with the magic bullet. Singapore Med J. 2010;**51**:1.
18. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. Nat. Immunol. 2002 Nov 1;**3**(11):991–8.
19. Moore AE. The destructive effect of the virus of Russian Far East encephalitis on the transplantable mouse sarcoma 180. Cancer. 1949 May;**2**(3):525–34.
20. Moore AE. Viruses with oncolytic properties and their adaptation to tumors. Ann N Y Acad Sci. 1952 Jul;**54**(6):945–52.
21. Gatenby PA, Basten A, Creswick P. 'Sneaking through': a T-cell-dependent phenomenon. Br J Cancer. 1981 Nov;**44**(5):753.
22. Isaacs A, Lindenmann J. Virus interference. I. The interferon. Proc R Soc Lond. Ser B–Biol Sci. 1957 Sep 12;**147**(927):258–67.
23. Pestka S, Krause CD, Walter MR. Interferons, interferon-like cytokines, and their receptors. Immunol Rev. 2004 Dec;**202**(1):8–32.
24. Mizel SB, Farrar JJ. Revised nomenclature for antigen-nonspecific T-cell proliferation and helper factors. Cell Immunol. 1979 Dec 1;**48**(2):433–6.
25. Morgan DA, Ruscetti FW, Gallo R. Selective in vitro growth of T lymphocytes from normal human bone marrows. Science. 1976 Sep 10;**193**(4257):1007–8.
26. Baluna R, Vitetta ES. Vascular leak syndrome: a side effect of immunotherapy. Immunopharmacology. 1997 Oct 1;**37**(2-3):117–32.
27. Topalian SL, Muul LM, Solomon D, Rosenberg SA. Expansion of human tumor infiltrating lymphocytes for use in immunotherapy trials. J Immunol Methods. 1987 Aug 24;**102**(1):127–41.
28. Burnet M. Cancer—a biological approach: III. Viruses associated with neoplastic conditions. IV. Practical applications. Br Med J. 1957 Apr 4;**1**(5023):841.
29. Lawrence HS, editor. Cellular and humoral aspects of the hypersensitive states: a symposium of the New York Academy of Medicine. New York: Hoeber-Harper; 1959.
30. Stutman O. Tumor development after 3-methylcholanthrene in immunologically deficient athymic-nude mice. Science. 1974 Feb 8;**183**(4124):534–6.
31. Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. Annu. Rev. Immunol. 2004 Apr 23;**22**:329–60.
32. Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. Immunity. 2004 Aug 1;**21**(2):137–48.
33. Kim R. Cancer immunoediting: from immune surveillance to immune escape. Cancer Immunother. 2007 Jan 1:9–27.
34. Engel AM, Svane IM, Mouritsen S, Rygaard J, Clausen J, Werdelin O. Methylcholanthrene-induced sarcomas in nude mice have short induction times and relatively low levels of surface MHC class I expression. Apmis. 1996 Jul 8;**104**(7–8):629–39
35. Kaplan DH, Shankaran V, Dighe AS, Stockert E, Aguet M, Old LJ, et al. Demonstration of an interferon γ-dependent tumor surveillance system in immunocompetent mice. Proc Natl Acad Sci. 1998 Jun 23;**95**(13):7556–61.
36. Shankaran V, Ikeda H, Bruce AT, White JM, Swanson PE, Old LJ, et al. IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. Nature. 2001 Apr 26;**410**(6832):1107–11.
37. Mittal D, Gubin MM, Schreiber RD, Smyth MJ. New insights into cancer immunoediting and its three component phases—elimination, equilibrium and escape. Curr Opin Immunol. 2014 Apr 1;**27**:16–25.
38. Teng MW, Swann JB, Koebel CM, Schreiber RD, Smyth MJ. Immune-mediated dormancy: an equilibrium with cancer. J Leukocyte Biol. 2008 Oct;**84**(4):988–93.
39. Braumüller H, Wieder T, Brenner E, Aßmann S, Hahn M, Alkhaled M, et al. T-helper-1-cell cytokines drive cancer into senescence. Nature. 2013 Feb 21;**494**(7437):361–5.
40. Xue W, Zender L, Miething C, Dickins RA, Hernando E, Krizhanovsky V, et al. Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. Nature. 2007 Feb;**445**(7128):656–60.
41. Laskowski T, Rezvani K. Adoptive cell therapy: living drugs against cancer. J Exp Med. 2020 Dec 7;**217**(12):e20200377.
42. Thomas ED, Lochte HL, Jr, Lu WC, Ferrebee JW. Intravenous infusion of bone marrow in patients receiving radiation and chemotherapy. New Engl J Med. 1957 Sep 12;**257**(11):491–6.

43. Collins RH, Jr, Goldstein S, Giralt S, Levine J, Porter D, Drobyski W, et al. Donor leukocyte infusions in acute lymphocytic leukemia. Bone Marrow Transpl. 2000 Sep;**26**(5):511–6.
44. Ye Y, Yang L, Yuan X, Huang H, Luo Y. Optimization of donor lymphocyte infusion for AML relapse after Allo-HCT in the era of new drugs and cell engineering. Front Oncol. 2021;**11**.
45. Aljurf M, Weisdorf D, Alfraih F, Szer J, Müller C, Confer D, et al. Worldwide Network for Blood & Marrow Transplantation (WBMT) special article, challenges facing emerging alternate donor registries. Bone Marrow Transpl. 2019 Aug;**54**(8):1179–88.
46. Horowitz MM, Gale RP, Sondel PM, Goldman JM, Kersey J, Kolb HJ, et al. Graft-versus-leukemia reactions after bone marrow transplantation. Blood. 1990 Feb 1;**75**(3):555–62.
47. Gross G, Waks T, Eshhar Z. Expression of immunoglobulin-T-cell receptor chimeric molecules as functional receptors with antibody-type specificity. Proc Natl Acad Sci. 1989 Dec;**86**(24):10024–8.
48. Eshhar Z. From the mouse cage to human therapy: a personal perspective of the emergence of T-bodies/chimeric antigen receptor T cells. Hum Gene Ther. 2014 Sep 1;**25**(9):773–8.
49. Sengsayadeth S, Savani BN, Oluwole O, Dholaria B. Overview of approved CAR-T therapies, ongoing clinical trials, and its impact on clinical practice. EJHaem. 2022 Jan;**3**:6–10.
50. Smith AJ, Oertle J, Warren D, Prato D. Chimeric antigen receptor (CAR) T cell therapy for malignant cancers: summary and perspective. J Cell Immunother. 2016 Nov 1;**2**(2):59–68.
51. Srivastava S, Riddell SR. Chimeric antigen receptor T cell therapy: challenges to bench-to-bedside efficacy. J Immunol. 2018 Jan 15;**200**(2):459–68.
52. Zmievskaya E, Valiullina A, Ganeeva I, Petukhov A, Rizvanov A, Bulatov E. Application of CAR-T cell therapy beyond oncology: autoimmune diseases and viral infections. Biomedicines. 2021 Jan 9;**9**(1):59.
53. Wing A, Fajardo CA, Posey AD, Shaw C, Da T, Young RM, et al. Improving CART-cell therapy of solid tumors with oncolytic virus–driven production of a bispecific T-cell engager armed oncolytic adenovirus improves CART cell therapy. Cancer Immunol Res. 2018 May 1;**6**(5):605–16.
54. Galon J, Bruni D. Tumor immunology and tumor evolution: intertwined histories. Immunity. 2020 Jan 14;**52**(1):55–81.
55. Waldman AD, Fritz JM, Lenardo MJ. A guide to cancer immunotherapy: from T cell basic science to clinical practice. Nat Rev Immunol. 2020 Nov;**20**(11):651–68.
56. Darvin P, Toor SM, Sasidharan Nair V, Elkord E. Immune checkpoint inhibitors: recent progress and potential biomarkers. Exp Mol Med. 2018 Dec;**50**(12):1–1.
57. Daly RJ, Scott AM, Klein O, Ernst M. Enhancing therapeutic anti-cancer responses by combining immune checkpoint and tyrosine kinase inhibition. Mol Cancer. 2022 Dec;**21**(1):1–20.
58. Hugo W, Zaretsky JM, Sun LU, Song C, Moreno BH, Hu-Lieskovan S, et al. Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. Cell. 2016 Mar 24;**165**(1):35–44.
59. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. New Engl J Med. 2014 Dec 4;**371**(23):2189–99.
60. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science. 2016 Mar 25;**351**(6280):1463–9.
61. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. Cell. 2017 Nov 2;**171**(4):934–49.
62. Topalian SL, Taube JM, Anders RA, Pardoll DM. Mechanism-driven biomarkers to guide immune checkpoint blockade in cancer therapy. Nat Rev Cancer. 2016 May;**16**(5):275–87.
63. Vitale I, Shema E, Loi S, Galluzzi L. Intratumoral heterogeneity in cancer progression and response to immunotherapy. Nat Med. 2021 Feb;**27**(2):212–24.
64. Huxley J. Biological aspects of cancer. George Allen & Unwin; 1958.
65. David H. Rudolf Virchow and modern aspects of tumor pathology. Pathol–Res Pract. 1988 Jun 1;**183**(3):356–64.
66. Davis A, Gao R, Navin N. Tumor evolution: linear, branching, neutral or punctuated?. Biochim Biophys Acta (BBA)–Rev Cancer. 2017 Apr 1;**1867**(2):151–61.
67. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: past, present, and the future. Cell. 2017 Feb 9;**168**(4):613–28.
68. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. Biochim Biophys Acta (BBA)–Rev Cancer. 2010 Jan 1;**1805**(1):105–17.
69. Rosenthal R, Cadieux EL, Salgado R, Bakir MA, Moore DA, Hiley CT, et al. Neoantigen-directed immune escape in lung cancer evolution. Nature. 2019 Mar;**567**(7749):479–85.
70. Losic B, Craig AJ, Villacorta-Martin C, Martins-Filho SN, Akers N, Chen X, et al. Intratumoral heterogeneity and clonal evolution in liver cancer. Nat Commun. 2020 Jan 15;**11**(1):1–5.
71. Joshi K, de Massy MR, Ismail M, Reading JL, Uddin I, Woolston A, et al. Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. Nat Med. 2019 Oct;**25**(10):1549–59.
72. Soto AM, Sonnenschein C. One hundred years of somatic mutation theory of carcinogenesis: is it time to switch?. BioEssays. 2014 Jan;**36**(1):118–20.
73. Muller HJ. The effects of X-radiation on genes and chromosomes. Science. 1928;**67**(1728).
74. Luria SE, Delbrück M. Mutations of bacteria from virus sensitivity to virus resistance. Genetics. 1943 Nov;**28**(6):491.
75. Nordling C. A new theory on the cancer-inducing mechanism. Br J Cancer. 1953 Mar;**7**(1):68.
76. Muller HJ. Cancer biology. Sci. Prog. 1951;**7**:130.
77. Fisher JC, Hollomon JH. A hypothesis for the origin of cancer foci. Cancer. 1951 Sep;**4**(5):916–8.
78. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer. 2004 Dec;**91**(12):1983–9.
79. Knudson Jr AG. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci. 1971 Apr;**68**(4):820–3.
80. Moolgavkar SH, Knudson AG. Mutation and cancer: a model for human carcinogenesis. JNCI: J Natl Cancer Inst. 1981 Jun 1;**66**(6):1037–52.
81. Mitelman F. The chromosomes of fifty primary Rous rat sarcomas. Hereditas. 1971 Dec;**69**(2):155–86.
82. Linder D, Gartler SM. Glucose-6-phosphate dehydrogenase mosaicism: utilization as a cell marker in the study of leiomyomas. Science. 1965 Oct 1;**150**(3692):67–9.

83. Nowell PC. The clonal evolution of tumor cell populations: acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression. Science. 1976 Oct 1;**194**(4260):23–8.
84. Bozic I, Antal T, Ohtsuki H, Carter H, Kim D, Chen S, et al. Accumulation of driver and passenger mutations during tumor progression. Proc Natl Acad Sci. 2010 Oct 26;**107**(43):18545–50.
85. McFarland CD, Mirny LA, Korolev KS. Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. Proc Natl Acad Sci. 2014 Oct 21;**111**(42):15138–43.
86. Bozic I, Nowak MA. Timing and heterogeneity of mutations associated with drug resistance in metastatic cancers. Proc Natl Acad Sci. 2014 Nov 11;**111**(45):15964–8.
87. Iwasa Y, Michor F, Nowak MA. Evolutionary dynamics of invasion and escape. J Theor Biol. 2004 Jan 21;**226**(2):205–14.
88. Jamal-Hanjani M, Hackshaw A, Ngai Y, Shaw J, Dive C, Quezada S, et al. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. PLoS Biol. 2014 Jul 8;**12**(7):e1001906.
89. Turajlic S, Xu H, Litchfield K, Rowan A, Chambers T, Lopez JI, et al. Tracking cancer evolution reveals constrained routes to metastases: TRACERx renal. Cell. 2018 Apr 19;**173**(3):581–94.
90. George JT, Levine H. Implications of tumor–immune coevolution on cancer evasion and optimized immunotherapy. Trends Cancer. 2021 Apr 1;**7**(4):373–83.
91. Agur Z, Vuk-Pavlović S. Mathematical modeling in immunotherapy of cancer: personalizing clinical trials. Mol Ther. 2012 Jan 1;**20**(1):1–2.
92. Bell GI. Predator-prey equations simulating an immune response. Math Biosci. 1973 Apr 1;**16**(3–4):291–314.
93. DeLisi C, Rescigno A. Immune surveillance and neoplasia—1 a minimal mathematical model. Bull Math Biol. 1977 Mar;**39**(2):201–21.
94. Bocharov G, Ludewig B, Bertoletti A, Klenerman P, Junt T, Krebs P, et al. Underwhelming the immune response: effect of slow virus growth on CD8+-T-lymphocyte responses. J Virol. 2004 Mar 1;**78**(5):2247–54.
95. d'Onofrio A, Gatti F, Cerrai P, Freschi L. Delay-induced oscillatory dynamics of tumour–immune system interaction. Math Comput Model. 2010 Mar 1;**51**(5-6):572–91.
96. Rocha AM, Costa MF, Fernandes EM. On a multiobjective optimal control of a tumor growth model with immune response and drug therapies. Int Trans Oper Res. 2018 Jan;**25**(1):269–94.
97. Singh S, Sharma P, Singh P. Stability of tumor growth under immunotherapy: a computational study. Biophys Rev Lett. 2017 Jun 21;**12**(02):69–85.
98. Kareva I, Luddy KA, O'Farrelly C, Gatenby RA, Brown JS. Predator-prey in tumor-immune interactions: a wrong model or just an incomplete one?. Front Immunol. 2021; **12**:668221.
99. Chanmee T, Ontong P, Konno K, Itano N. Tumor-associated macrophages as major players in the tumor microenvironment. Cancers. 2014 Aug 13;**6**(3):1670–90.
100. Owen MR, Sherratt JA. Modelling the macrophage invasion of tumours: effects on growth and composition. Math Med Biol: J IMA. 1998 Jun;**15**(2):165–85.
101. Byrne HM, Cox SM, Kelly CE. Macrophage-tumour interactions: in vivo dynamics. Discrete Continuous Dyn Syst-B. 2004;**4**(1):81.
102. Owen MR, Byrne HM, Lewis CE. Mathematical modelling of the use of macrophages as vehicles for drug delivery to hypoxic tumour sites. J Theor Biol. 2004 Feb 21;**226**(4):377–91.
103. den Breems NY, Eftimie R. The re-polarisation of M2 and M1 macrophages and its role on cancer outcomes. J Theor Biol. 2016 Feb 7;**390**:23–39.
104. Mahlbacher G, Curtis LT, Lowengrub J, Frieboes HB. Mathematical modeling of tumor-associated macrophage interactions with the cancer microenvironment. J Immunother Cancer. 2018 Dec;**6**(1):1–7.
105. Grossman Z, Berke G. Tumor escape from immune elimination. J Theor Biol. 1980 Mar 21;**83**(2):267–96.
106. Wilkie KP, Hahnfeldt P. Mathematical models of immune-induced cancer dormancy and the emergence of immune evasion. Interface Focus. 2013 Aug 6;**3**(4):20130010.
107. George JT, Levine H. Stochastic modeling of tumor progression and immune evasion. J Theor Biol. 2018 Dec 7; **458**:148–55.
108. Arias CF, Herrero MA, Cuesta JA, Acosta FJ, Fernández-Arias C. The growth threshold conjecture: a theoretical framework for understanding T-cell tolerance. R Soc Open Sci. 2015 Jul 8;**2**(7):150016.
109. Naylor K, Li G, Vallejo AN, Lee WW, Koetz K, Bryl E, et al. The influence of age on T cell generation and TCR diversity. J Immunol. 2005 Jun 1;**174**(11):7446–52.
110. Palmer S, Albergante L, Blackburn CC, Newman TJ. Thymic involution and rising disease incidence with age. Proc Natl Acad Sci. 2018 Feb 20;**115**(8):1883–8.
111. George JT, Levine H. Sustained coevolution in a stochastic model of cancer–immune interaction. Cancer Res. 2020 Feb 15;**80**(4):811–9.
112. George JT, Levine H. Optimal cancer evasion in a dynamic immune microenvironment generates diverse post-escape tumor antigenicity profiles. Elife. 2023 Apr 25;**12**: e82786.

34

Mechanistic modelling and machine learning to establish structure–activity relationship of nanomaterials for improved tumour delivery

Maria Jose Peláez, Shreya Goel, Vittorio Cristini, Zhihui Wang, and Prashant Dogra

34.1. Introduction

Cancer continues to be among the leading causes of death worldwide, with systemic therapy (e.g. chemotherapy and immunotherapy) being an important treatment option for the majority of cancers [1]. The challenges of drug insolubility, fragility, or instability *in vivo*, and the collateral damage to healthy cells associated with non-targeted delivery of cytotoxic agents have prompted the need for innovative methods to improve the safety and efficacy of systemic cancer therapies. As a result, engineered nanomaterials have emerged as the putative 'magic bullets' capable of targeted tumour delivery of the encapsulated cargo. By preserving drug stability and improving its systemic pharmacokinetics, nanomaterials enhance the safety and efficacy of systemic cancer therapies [2].

The past two decades have witnessed a surge in preclinical research and development activities in cancer nanomedicine (**Figure 34.1**) [3,4]; however, the pace of clinical translation of these novel nanomaterials has failed to keep up, with only a handful of nanomaterials arriving to clinical trials or in the clinic [5]. As per a comprehensive meta-analysis of preclinical studies, the tumour delivery efficiency of nanoparticles (NPs) has stagnated at a median value of 0.7% of the injected dose, across a range of nanomaterial and tumour types [6]. This delivery problem has primarily contributed to the limited clinical success of cancer nanomedicine [7]. Additionally, the associated potency of nanomaterials to induce deleterious effects in blood and healthy tissues presents a major barrier to regulatory approval of cancer nanomedicines [8–10].

For this, experimental efforts to better understand the effect of NP physicochemical properties (e.g. size, shape, Zeta potential, surface chemistry, hardness, porosity, core material type, state of aggregation, and crystallinity) on the systemic pharmacokinetics, tumour delivery efficiency, and safety of NPs, referred to here as the structure–activity relationship (SAR), have been made [6,7,11–13]. Based on the divide and conquer strategy, experimental efforts to establish the SAR of NPs have been made in isolation [14–16], i.e. various physicochemical properties have been studied separately for their effect on the biological behaviour of NPs. Generalization of outcomes from such studies is useful, but the occurrence of nonlinear effects of particle properties on biological outcomes may preclude the elucidation of a comprehensive SAR of NPs *in vivo*. Also, it becomes an upheaval task to experimentally study parameter combinations in the vast, multidimensional parameter space of the NP-mediated drug delivery process to establish a universal SAR of NPs.

To this end, mathematical modelling has been a critical tool employed by scientists to unravel the mechanistic underpinnings of pharmacokinetics, safety, and tumour delivery of NPs [17]. Through data-driven and mechanistic modelling, various biochemical and biophysical nano–bio interactions, physiological processes, and tumour dynamics can be investigated for their nonlinear effects to obtain a comprehensive SAR map of NPs, which can generate design guidelines for developing NPs with optimized pharmacokinetics and improved safety and tumour delivery efficiency. In this chapter, we cover the fundamental nano–bio interactions and transport phenomena relevant to this problem and discuss the various mathematical modelling and machine-learning approaches that have been developed to study NP SAR pertaining to systemic pharmacokinetics, toxicity, and tumour-targeted delivery of NPs.

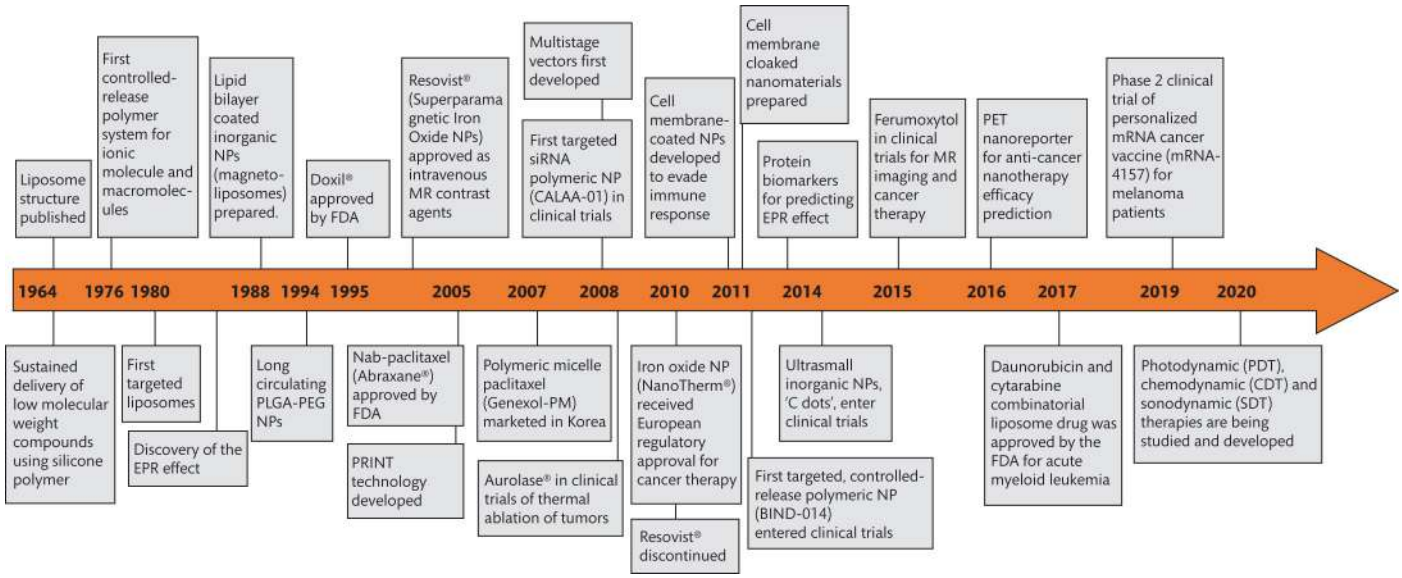


Figure 34.1. A historical timeline of the major advancements in cancer nanomedicine. *Source:* Adapted from Dogra et al. [17].

34.2. Nano–bio interactions and transport phenomena *in vivo*

To better appreciate the complexity of the delivery problem mentioned above, it is important to understand the relevant nano–bio interactions and transport phenomena that determine the *in vivo* fate of NPs. As shown in Figure 34.2, the passage of NPs from the site of injection to the site of action (i.e. cancerous cells) can be divided into three stages: (i) vascular, (ii) transvascular, and (iii) interstitial [17,18]. Along this path, NPs encounter several biochemical, biophysical, anatomical, and physiological challenges that ultimately govern their biological fate, and determine the toxicity and tumour delivery efficiency of NPs, as described below.

Following injection, as NPs enter the blood stream, they are exposed to a high concentration of plasma proteins (60–80 g/l), namely albumins, apolipoproteins, and opsonins that adsorb on the surface of NPs and form a (bio)molecular corona around the particles. Corona formation alters the synthetic identity of NPs and imparts them a new *in vivo* identity, thereby influencing other nano–bio interactions (e.g. NP interactions with cell surface receptors or cell membranes). The corona around the particles is a dynamic entity, and depending upon the physicochemical properties of the NPs, the relative affinity of plasma proteins for NP surface can vary, and thus the corona can have a varied composition at steady state. As a result, the effect of the biomolecular corona on microscopic nano–bio interactions can vary based upon the properties of the NP, which can manifest at the global scale in the form of altered pharmacokinetics

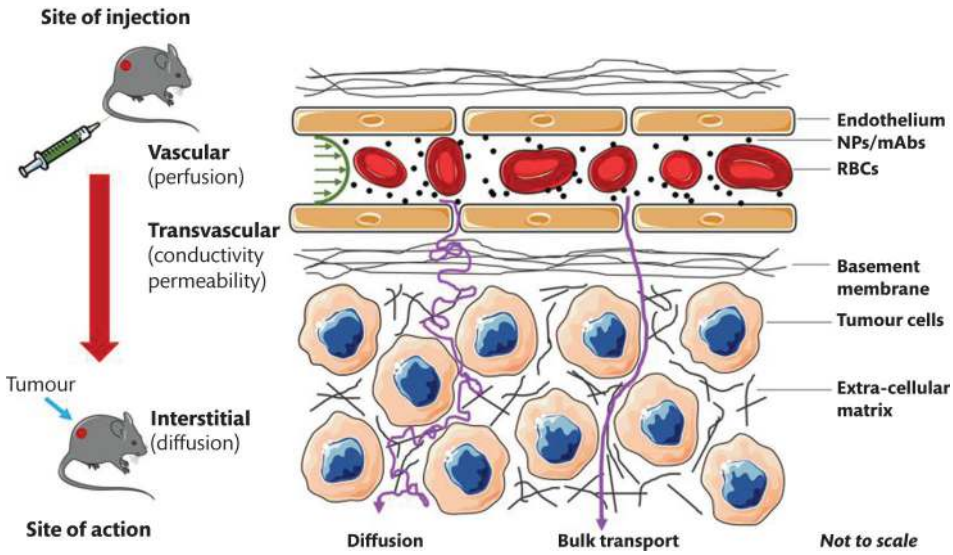


Figure 34.2. Nanoparticle transport in tumours. Biophysical barriers involved in the delivery of NPs to tumours via microcirculation. Source: Adapted from Dogra et al. [17,18].

and tumour delivery efficiency of NPs. To minimize such undesired NP interactions, surface coating of NPs with hydrophilic polymers, such as polyethylene glycol (PEGylation), is a commonly used strategy to sterically block plasma protein binding of NPs. However, optimization of the density of the grafted polymer is necessary to ensure protection from plasma proteins without compromising NP activity [3].

Further, perfusion of the tumour (i.e. passage of blood through the vasculature of tumour) carries the NPs from the systemic circulation into the tumour vascular space. Perfusion is quantified as the blood flow rate per unit mass of organ and is a property of the biological system, varying across tissues and species [19]. Blood flow rate Q is measured as the ratio of the pressure difference between arterial and venous ends of blood vessels (ΔP) and flow resistance (FR):

$$Q = \Delta P / \mathrm{FR}, \tag{34.1}$$

where flow resistance is the product of apparent viscosity (due to erythrocyte density) and geometrical resistance, which tends to be greater in tumours compared to healthy tissues, due to vessel deformations and abnormalities, thus lowering tumour perfusion. Therefore, the perfusion-dependent influx of NPs into the tumour vascular volume, as defined by the product of Q and NP concentration in the systemic circulation, largely determines the fraction of injected dose of NPs arriving in the tumour vasculature [20]. Simultaneously, NPs are also being transported through perfusion to the healthy tissues and organs of the body based on their respective blood flow rates, which partly governs the systemic pharmacokinetics of NPs.

Once NPs arrive at the finest blood vessels, (i.e. microcapillaries), they encounter a highly dynamic environment because of the engagement of the capillary network in the exchange of materials (i.e. nutrients and wastes) with tissue interstitium. This exchange is made feasible by the porous nature of the capillary wall, allowing easy exchange of fluid and solutes driven by a combination of diffusion and advection. The combined effect of diffusion and bulk transport that governs the flux J of molecules through the capillary wall into the interstitium can be expressed as

$$J = \overbrace{PS\left(C_\mathrm{p} - C_\mathrm{i}\right)}^{\text{Diffusion}} + \overbrace{LS(1-\sigma)\left[\left(p_\mathrm{v} - p_\mathrm{i}\right) - \sigma\left(\pi_\mathrm{v} - \pi_\mathrm{i}\right)\right]C_\mathrm{p}}^{\text{Advection}}, \tag{34.2}$$

where P represents permeability of solute across vascular wall, S is the vascular surface area, C_p and C_i are the molecular concentrations in systemic circulation and interstitium, respectively, L is the hydraulic conductivity of fluid across the endothelial barrier, σ is the reflection coefficient that determines the impedance to solute passage across the vascular wall, $p_\mathrm{v} - p_\mathrm{i}$ represents the difference between hydrostatic pressures between microvasculature and interstitial space, and $\pi_v - \pi_i$ is the difference in osmotic pressure across the wall. P, L, and σ are the critical transport parameters that govern the tendency of a substance to undergo transvascular extravasation and are a function of NP physicochemical properties (especially NP size and surface charge) and porosity of the vascular wall [20].

Note that based on the pore size of vessel walls, microcapillaries in healthy tissues are classified as (i) continuous, (ii) fenestrated, and (iii) sinusoidal. Continuous capillaries have pore sizes <5 nm (e.g. brain, lungs, muscles, and skin), fenestrated capillaries haves pores <15 nm (e.g. kidneys), and sinusoids in liver have pores <200 nm, while those in spleen are ~5 μm [21]. Thus, the NP to pore size ratio becomes a critical determinant (in addition to NP surface charge) of NP extravasation into tissue interstitium, which comprises the transvascular stage of NP transport [22]. Similarly, due to malformed neoangiogenic vessels, tumour vessel walls have larger pores (~1,700 nm) [23], which enhances their permeability, allowing passive accumulation of NPs in tumour interstitium. On the other hand, the enhanced permeability of tumour vessels also causes greater leakage of fluid into the interstitium, which builds up the interstitial fluid pressure in tumours that counters the hydrostatic pressure necessary for advection. This is further aggravated by dysfunctional lymphatic drainage caused by solid stress in a growing tumour, ultimately leading to diffusion being the primary means of NP extravasation in tumour interstitium [24].

Finally, once NPs are in the tumour intersitium, they diffuse through the dense extra-cellular matrix (ECM), with little to no support from advection (due to high interstitial fluid pressure) [25], to arrive in the vicinity of cancerous cells, described by

$$\frac{\partial C_\mathrm{i}}{\partial t} = D\nabla^2 C_\mathrm{i} \tag{34.3}$$

where C_i is the concentration of NPs in tumour interstitium and D is the diffusion coefficient of NPs (which is a function of NP size). Diffusion-mediated interstitial transport of NPs limits their penetration distance and thus delivery of cargo to cancerous cells distant from the tumour-feeding capillaries is challenging [26,27]. *Readers are referred to the following review for a detailed discussion on intratumoral transport barriers to drug delivery* [28]. Once in the proximity of cells, NPs maybe internalized through clathrin-dependent/independent endocytosis, caveolin-dependent/independent endocytosis, or receptor-mediated endocytosis, depending upon their size and surface characteristics [29–32].

The preferential accumulation of NPs in the tumour interstitium, as discussed above, is exploited to achieve passive tumour targeting of NPs, referred to as the enhanced permeability and retention (EPR) effect [33]. However, such preferential accumulation of NPs is also observed in mononuclear phagocytic system (MPS) organs (particularly, liver and spleen). This is driven by the high porosity of sinusoidal vessel walls and the occurrence of resident macrophages (such as Kupffer cells in liver and splenic macrophages in spleen) in the lumen of sinusoids, which causes immediate sequestration of circulating NPs in the intersitium of these organs [34,35]. Given the large physical dimension of MPS organs, compared to tumours, a major fraction of the injected dose of NPs is captured in the liver and spleen, thereby reducing NP circulation time and their availability for accumulation in the tumours. Prolonged retention of NPs in MPS organs also raises the possibility of NP-induced toxicity due to oxidative stress caused by the generation of reactive oxygen species, cell membrane damage, possible interactions of NPs with intracellular organelles, such as mitochondria and nuclei, to alter cellular metabolism and induce DNA damage [36,37]. *Thus, the optimization of NP design to reduce uptake in MPS organs, which can prevent potential toxicities and improve NP circulation time to improve tumour accumulation, is the holy grail for clinical translation of cancer nanomedicine.*

As evident in the description above, NP delivery to the tumour is a multiscale process accompanied by microscopic nano–bio interactions and transport processes (dependant on NP properties, MPS physiology, and tumour characteristics), which govern the systemic pharmacokinetics, safety, and tumour delivery efficiency of NPs. To effectively study NP-mediated tumour drug delivery, integration of experimental work with mechanistic mathematical modelling and machine learning is critical to enable exploration of the vast, multidimensional parameter space and establish universal SAR of NPs to guide rational design of NPs for improved tumour-targeted delivery.

34.3. Mechanistic modelling to characterize SAR of NPs

To support the preclinical development and clinical translation of NPs, the characterization and improvement of their systemic pharmacokinetics, safety, and tumour delivery efficiency is of utmost importance. To this end, several mathematical modelling approaches have been developed to study the SAR of NPs [17]. These models have been developed to study processes spanning across multiple spatiotemporal scales; depending upon the length and timescales to be modelled, a particular modelling approach is selected, keeping the computational efficiency and model complexity in context. Broadly, mechanistic modelling approaches can be classified as discrete, continuum, and hybrid (Table 34.1).

Discrete or agent-based models generally operate at the level of autonomous agents or entities, which in the current context can be cells or NPs. The high spatiotemporal resolution of such models allows the inclusion of detailed molecular or cellular level processes, which is particularly relevant to the investigation of microscopic nano–bio interactions. However, the computational cost associated with the modelling of multiple agents and their interactions limits the scalability of such models. To overcome this limitation, continuum models are used, which model the behaviour of agents as continuous mass rather than discrete particles. Assuming homogenous distribution of agents in the occupied volume (e.g. NPs in the tumour vasculature), these models study the bulk behaviour of the system by considering spatiotemporal averages of the underlying (smaller scale) processes and randomness. Thus, due to reduced spatiotemporal resolution, these models are computationally less expensive and thus can easily be extended to study large-scale systems, e.g. whole-body biodistribution dynamics of NPs. These models however limit the study of the effects of heterogeneity within the system, (e.g. polydispersity of NP samples or intercellular differences in expression of cancer cell surface receptors), which necessitates the hybridization of discrete and continuum modelling approaches, leading to hybrid models. Leveraging on the strengths of the two approaches, in hybrid models, some components are modelled as discrete entities, while the rest are modelled as continuous mass. In the following sections, we review selected mechanistic modelling methods (with a focus on physiologically based pharmacokinetic (PBPK) models) that are fundamental to the characterization of NP pharmacokinetics, safety, and tumour-targeted drug delivery for establishing the SAR of NPs.

34.3.1. Models pertaining to NP pharmacokinetic characterization

One of the most fundamental types of continuum models used for the whole-body pharmacokinetic characterization of drugs is PBPK models. PBPK models involve compartmentalization of the body into physiological volumes representing organs and tissues, connected in an anatomical fashion via physiological fluid (blood and lymph) flow (Figure 34.3). The model is formulated as a system of ordinary differential equations (ODEs) to describe the transport phenomena responsible for mass transfer between compartments and

Table 34.1. Common quantitative approaches to study SAR of NPs (systemic pharmacokinetics, safety, and tumour delivery efficiency).

Model type		Applications	Advantages	Disadvantages
Mechanistic models	Continuum (ODE/PDE)	Model the average spatiotemporal dynamics of NP biodistribution	-Low computational cost -Model centrally based large-scale processes -Allows data-driven empirical description or mechanistic formalism of the system -(ODE) has a high clinical applicability	-Hard to describe heterogeneity -Inclusion of phenomenological model parameters, which may not have a direct biological meaning, proving difficult to measure experimentally -Solutions are deterministic
	Discrete	Model the spatiotemporal dynamics on a cellular level, studying NP–cell, cell-to-cell, and cell–environment interactions	-Captures heterogeneity of biological entities and their underlying mechanisms -Solutions are non-deterministic	-High computational cost -Harder development (multiscale modelling) and implementation -Impractical for large-scale global systems
	Hybrid	Model the spatiotemporal global dynamics between populations, including cellular level interactions	-Captures heterogeneity of biological entities and their underlying mechanisms -Multiscale modelling technique -Model large-scale processes	-High computational cost -Harder to develop -Use of many numerical techniques -Difficult to do model parameterization
Artificial intelligence		Predict, classify, and find relationships regarding drug delivery and distribution, patient status, and treatment efficiency	-Can handle non-linear relationships -High predictive capability - Has a high clinical applicability	-High computational cost -Requires large amount of consistent data, which may be hard to find in medical databases -Use of complex algorithms

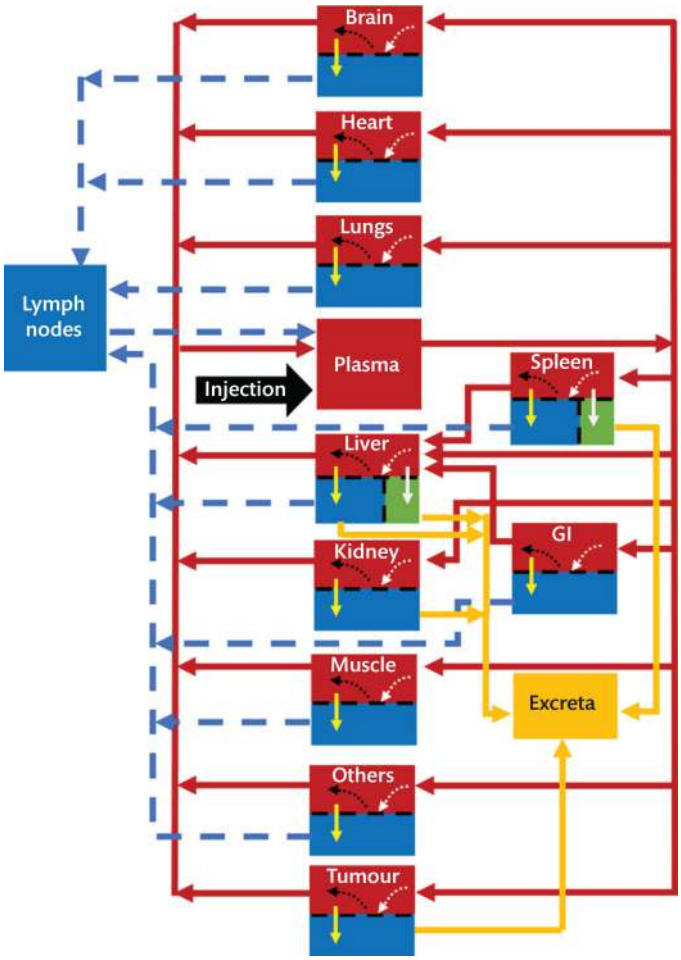


Figure 34.3. Structure of whole-body PBPK model. Notation: red arrows: plasma flow; dashed blue arrows: lymph flow; bright yellow arrows: extravasation; white arrows: phagocytosis; mustard arrows: excretion; dotted white arrows: NP deposition on the vascular wall; dotted black arrows: NP dislodging from the vascular wall. Source: Reproduced with permission from Dogra et al. [38].

sub-compartments, thereby characterizing the concentration kinetics of drugs across the body. Since the PBPK modelling framework is based on mechanistic and physiological parameters, it can help with interspecies extrapolation of drug PK, thereby allowing predictions for humans (based on physiological differences) from animals, to support investigational new drug applications. Additionally, these models allow the characterization of drug PK in patients or special populations (e.g. paediatrics, elderly, and pregnant women) based on the PK in healthy volunteers. Further, PBPK models can be integrated to drug–response (or pharmacodynamic) models to characterize drug safety and efficacy. Since the model incorporates physiological parameters (organs volumes, blood-, and lymph-flow rates), drug-related parameters (permeability and lipophilicity), and treatment regimen parameters (dose, route of administration, and treatment frequency), it can investigate the mechanistic basis of drug PK and support the optimization of control parameters to achieve a target PK profile for optimal treatment outcomes.

To accurately characterize the systemic PK of NPs, it is critical for a PBPK model to capture the key nano–bio interactions and transport processes (detailed in the previous section). Due to physical differences between small molecules and NPs, the adoption of traditional PBPK models to study NP dynamics *in vivo* requires modification of modelling assumptions and model reparameterization to account for the effect of NP properties (such as size and shape) on absorption, distribution, metabolism, and excretory processes. Thus, experimental studies that can quantify nano–bio interactions and systemic PK of NPs are critical to the development of well-calibrated PBPK models. For this purpose, non-invasive, whole-body imaging techniques to quantify the time-dependent biodistribution of NPs are commonly used [18]. The most used imaging modalities include magnetic imaging, nuclear imaging (PET and SPECT), and optical imaging (fluorescence and bioluminescence), which involve labelling of NP surface with small molecules (i.e. imaging agents) that upon stimulation with an input signal, or without, produce a reporter signal detectable by the receiver of the imaging device. This generates a time series of images, which upon quantification provide the necessary NP biodistribution kinetics data across the body for the development and calibration of a PBPK model. Also, key mechanistic and physiological parameters relevant to NP pharmacokinetics and tumour delivery, such as fluid (blood and lymph) flow rates, compartment volumes (vascular and extravascular), vascular permeability and surface area, and cellular internalization rates, can be quantified through the above imaging techniques. Note that the choice of the imaging modality is based upon the chemical limitations of NP labelling and the requirements of spatial resolution, tissue penetration depth, sensitivity, and cost effectiveness of the imaging modality. Over the past decade, several data-driven PBPK modelling efforts have helped characterize the systemic pharmacokinetics of novel NPs. *Readers are referred to a detailed review of NP PBPK models by Kumar et al.* [7].

As a representative example, here we discuss an imaging data-driven PBPK model developed by Dogra et al. to study the effect of NP physicochemical properties on the whole-body pharmacokinetics and tumour delivery efficiency of mesoporous silica NPs [38]. As shown in **Figure 34.3**, the model is composed of 12 compartments that represent major organs and tissues of interest. Each compartment comprises two sub-compartments that represent the vascular (red) and extravascular (blue) volumes of the organs. NPs upon injection into systemic circulation are transported via blood flow to the vascular sub-compartments of various organs, from where a fraction of freely circulating NPs extravasate into the extravascular sub-compartment (depending upon the corresponding lymph flow rate), while the remaining fraction exits the vascular sub-compartment to re-join the systemic circulation. Note that in the vascular sub-compartment, depending upon their physicochemical properties, NPs bind and unbind from the vascular endothelium, which affects their circulation time and the tendency to extravasate through vessel wall pores or be phagocytosed by macrophages. The extravasating NPs enter the lymph node compartment to re-join systemic circulation, which enables the conservation of mass. In this model, based upon experimental evidence, the liver, spleen, and kidneys are considered to participate in the excretion of NPs, with the liver and spleen also comprising an additional sub-compartment representing the macrophages of the MPS. The various nano–bio interactions (except plasma protein binding) and transport phenomena discussed in the sections before are captured in this model through a system of ODEs that are solved numerically, and the solution is compared for its accuracy against *in vivo* data

obtained through SPECT-CT imaging of radiolabelled mesoporous silica NPs [14]. The model predictions of whole-body pharmacokinetics showed excellent agreement with *in vivo* data, with predictions of tumour delivery also leading to physiologically plausible results. The model was then used for local and global sensitivity analyses that unravelled the key roles of NP degradation rate, tumour blood viscosity, NP size, tumour vascular fraction, and tumour vascular porosity in determining the tumour delivery efficiency of these NPs. The physiological structure of the model, along with the inclusion of mechanistic parameters to represent NP-specific *in vivo* interactions, allows the application of this model to NP types other than mesoporous silica, as demonstrated by the adaptation of this model to study ultrasmall porous silica NPs [15] and microparticles [39].

34.3.2. Models pertaining to NP-mediated tumour-targeted drug delivery

To evaluate the bulk accumulation of NPs in the tumour, based on the EPR effect, and their penetration into the tumour tissue, continuum or discrete modelling approaches have been employed. From a continuum modelling perspective, an obvious extension of whole-body PBPK models, as shown in Figure 34.3, involves the inclusion of an additional compartment representing the tumour [7,11,38]. NPs administered into the systemic circulation are fed into the tumour vasculature through perfusion, and as discussed before, from the tumour vascular sub-compartment, NPs exit into the tumour interstitial sub-compartment in a vascular permeability-limited fashion, i.e. based on the relative size of NPs and vessel wall pores. Such a modelling approach allows the understanding of the effect of NP parameters simultaneously on systemic pharmacokinetics and tumour delivery efficiency, which can be used for multiobjective optimization, i.e. simultaneously identifying NP design sub-space that minimizes NP accumulation in the MPS and maximizes accumulation in the tumour. Given that the timescale of NP clearance from the body (hours–days) is short compared to tumour growth timescales (weeks–months), therefore, to investigate the SAR of NPs following a single injection, a static tumour-integrated PBPK model is sufficient [38]. However, if required to understand the effect of treatment regimen parameters on tumour outcomes, it can become necessary to include a growth component in the tumour compartment to accurately simulate the time-dependant changes in the tumour parameters that can affect NP delivery efficiency over the course of treatment. Note that the healthy compartments in this scenario are modelled to be static, which can also be simplified and lumped into one or two compartments, based on the compartmental PK modelling approach [40]. As a representative example, a minimal modelling approach was developed by Dogra et al. [41,42], where they lumped the healthy organs into a single compartment representing the systemic circulation, which connected to a tumour compartment through blood flow. Since their modelling approach was oriented towards improving the understanding of tumour delivery efficiency of NPs and optimizing treatment outcomes of NP-mediated gene therapy, the tumour compartment was then sub-compartmentalized into vascular, interstitial, and cellular volumes, and tumour growth dynamics was modelled to evaluate the effect of changing tumour physiology on NP delivery and to include a pharmacodynamic component in the model. Their analysis identified novel combinations of NP-delivered gene therapy with standard-of-care drugs for breast cancer to improve treatment outcomes.

While the continuum modelling approaches mentioned above (which at times rely on phenomenological descriptors of biological interactions and process) allow the investigation of multiple parameters simultaneously (i.e. NP-related, physiology-related, tumour-related, and treatment-regimen-related parameters) for their effect on NP accumulation in the tumour, it is challenging to investigate intra-tumoral NP transport dynamics, interactions with the tumour microenvironment, and cellular or nuclear internalization of NPs from a detailed, mechanistic perspective with such multiscale models. As such, discrete or hybrid modelling approaches are often used for a detailed description of the tumour microenvironment and the associated transport phenomena to study intra-tumoral NP SAR. As an example, to understand the effects of NP size on tissue distribution and penetration efficiency, a multiscale model was developed by Islam et al. using a time-adaptive Brownian dynamics algorithm that accounted for advection, diffusion, and cellular uptake of NPs [43]. This model included particle–cell interactions, cell–surface adhesion, particle dispersion, particle capture, and penetration. The results showed that in a system that considers the interaction of NPs with the cell walls and advection, NP size (10 and 100 nm) effects on NP distribution and penetration were less pronounced than that in a cell-free *in vitro* system. Similarly, Sykes at al. performed Monte Carlo simulations to study diffusion of NPs through tumour ECM [44]. They modelled the tumour ECM in three dimensions as an anisotropically oriented network of collagen fibres to study the mobility of NPs through collagen matrices of varying densities. Collagen fibres were approximated as immobile cylinders, and NP–fibre collision was assumed to be elastic. NP movement was simulated as a discrete random walk following the Stokes–Einstein relation for the diffusion of spherical particles in a fluid with low Reynolds number. Their model helped elucidate the mechanisms underlying particle size-dependent retention of NPs in clinically relevant tumour conditions. Also, Wirthl et al. studied tumour delivery of NPs by adapting a multiphase tumour growth model to include tumour microenvironmental transport barriers (Figure 34.4) [45]. The model included fives phases, with the ECM being the solid phase, and tumour cells, host cells, and interstitial

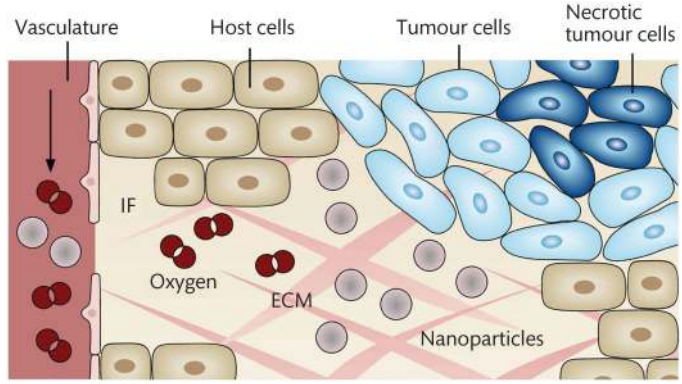


Figure 34.4. Components of the multiphase tumour growth model. The model comprises a solid phase, the ECM, three fluid phases, host cells, tumour cells and interstitial fluid (IF), and the vasculature that is modelled as an independent porous network. In addition, the phases transport species, namely necrotic tumour cells, oxygen, and nanoparticles. Source: Reproduced with permission from Wirthl et al. [45].

fluid representing fluid phases flowing through the pores of the solid phase. Additionally, the vasculature was modelled as an independent porous network with blood flow and as the source of injected NPs. Furthermore, the model assumed that oxygen and nutrients were mainly transported by the vasculature and interstitial fluid. To model the extravasation of the NPs to the interstitial fluid, both the transendothelial (diffusion through capillary vessel walls, dependent on particle size) and the interendothelial pathways (convective process influenced by pressure gradients) were included; additionally, lymphatic drainage (interstitial fluid to the lymph system) was also modelled. Simulation results showed that solid tumours developed a non-perfused core and increased interstitial pressure, which were identified as the two features of solid tumours that limit NP delivery. Only in the case of small NPs, the transport was dominated by diffusion, and NPs reached the entire tumour. The permeability of the blood-vessel endothelium had a major impact on the amount of delivered NPs. *Readers are referred to the following review for more such modelling examples* [17].

34.3.3. Models pertaining to NP safety assessment

While NP parameter optimization to achieve the target systemic PK profile and efficient tumour delivery are critical to support the clinical translation of cancer nanomedicines, it is equally important to ensure the safety associated with exposure of healthy tissues to the administered NPs. Modelling efforts to understand the effect of NP properties on their toxic potential include the application of continuum and hybrid modelling approaches, with representative examples discussed below.

Laomettachit et al. designed a hybrid model based on the PBPK framework to estimate NP concentration in the liver and used this value as input to a cell-response model, which predicted the extent of cell death in the liver due to the accumulation of the titanium dioxide NPs [46]. This model demonstrated the dose-dependent toxicity of titanium dioxide NPs in liver, i.e. the tissue damage from a low dose of NPs is negligible and reversible due to compensation by cell proliferation, while high exposure can cause irreversible tissue damage unless a large fraction of cells undergoes cell division to renew the damaged tissue mass. Also, Lin et al. developed a PBPK model of gold NPs that described the tissue distribution of NPs across different species to be integrated to *in vitro* or *in vivo* toxicity data for quantitative risk assessment [47]. The model consisted of seven compartments, which represented the major organs of the body (plasma, liver, spleen, kidneys, lung, brain, and others), including transcapillary membrane transport, endocytosis, different distribution coefficients for the plasma and the tissue, and biliary and urinary excretion. The model was validated using data from mice, rats, and pigs, and extrapolations made to predict behaviour in humans, with rats and pigs identified as the more appropriate species for human extrapolations.

34.4. Artificial intelligence to characterize SAR of NPs

In addition to mechanistic models discussed above, an increasing number of studies are now leveraging the strengths of artificial intelligence at handling large volumes of *in vitro* and *in vivo* data to identify correlations between NP properties and various aspects of NP SAR, including NP toxicity [48,49]. Machine-learning techniques use learning-based algorithms to establish relationships and patterns from the available data (high quantity of data of fixed structure and periodicity). These types of algorithms can be divided into supervised learning (task-driven), unsupervised learning (data-driven), and reinforcement learning (learning from mistakes). Now, we briefly discuss some important advances in this direction.

Lin et al. used various machine-learning algorithms (classic models, ensemble models, support vector machines, and neural networks) to determine the effect of NP physicochemical properties (e.g. size, zeta potential, shape, and core material), tumour models, and cancer types on tumour delivery efficiency [50]. The results from the analysis suggested that cancer type, zeta potential, and the core material of NPs have the greatest effect on tumour delivery efficiency. Further, to analyse NP transport and predict NP distribution, Stillman et al. developed a modelling platform called EVONANO (**Figure 34.5**), which included three central modules: (i) virtual tumour model (simulation of virtual tumour growth using agent-based model), (ii) virtual tissue model (simulation of NP–tissue interactions using stochastic reaction–diffusion equations), and (iii) genetic algorithm-based machine-learning module (for optimization of NP design) [51]. This platform generated different representative scenarios to model and optimize NP distribution in the tumour under different conditions. Results show that the ML algorithm can find a treatment solution that kills 95% of the cancer cells with a low dosage for the homogeneous tumour environment. Meanwhile, the best solution for the heterogeneous tumour environment showed that a combined treatment was able to kill 99% of cancer cells and 80% of cancer stem cells. Additionally, the study showed that NPs with a high diffusion coefficient and low binding affinity showed the best results for both types of tumours.

Nademi et al. used three different machine-learning algorithms (random forest, multilayer perceptron, and linear regression) to identify design insights to improve cellular uptake of short interfering RNA (siRNA) by hydrophobically modified polyethylenimine (PEI) NPs into various breast cancer cell lines [52]. For this, the complete dataset included 197 data points from which 75% was used for model training and the remaining 25% for model testing. The models identified that the most significant determinants of cellular uptake were PEI-to-siRNA weight ratio, type of hydrophobic substitution, total number of C, unsaturation level of C on the lipid substitution, and the total number of thioester groups on the modified PEIs.

Kingston et al. created an imaging (3D microscopy) and image analysis machine-learning workflow to study the interactions of NPs with cells in metastatic tumours, as well as analyse the physiology of micrometastases [53]. Due to their small size and sparse distribution, the interaction and distribution of NPs in micrometastases has been limited, but, due to their proximity to blood vessels, the authors hypothesized that they would be a good target for NP delivery. Thus, the authors combined tissue clearing, 3D microscopy, and machine learning (segmentation and image analysis algorithms) to assess their hypothesis. Results showed that on average the primary tumour cells were found at 16 μm from the closest blood vessel, while micrometastases were found at 8 μm. Additionally, an eight times higher density of NP-positive cells in micrometastases

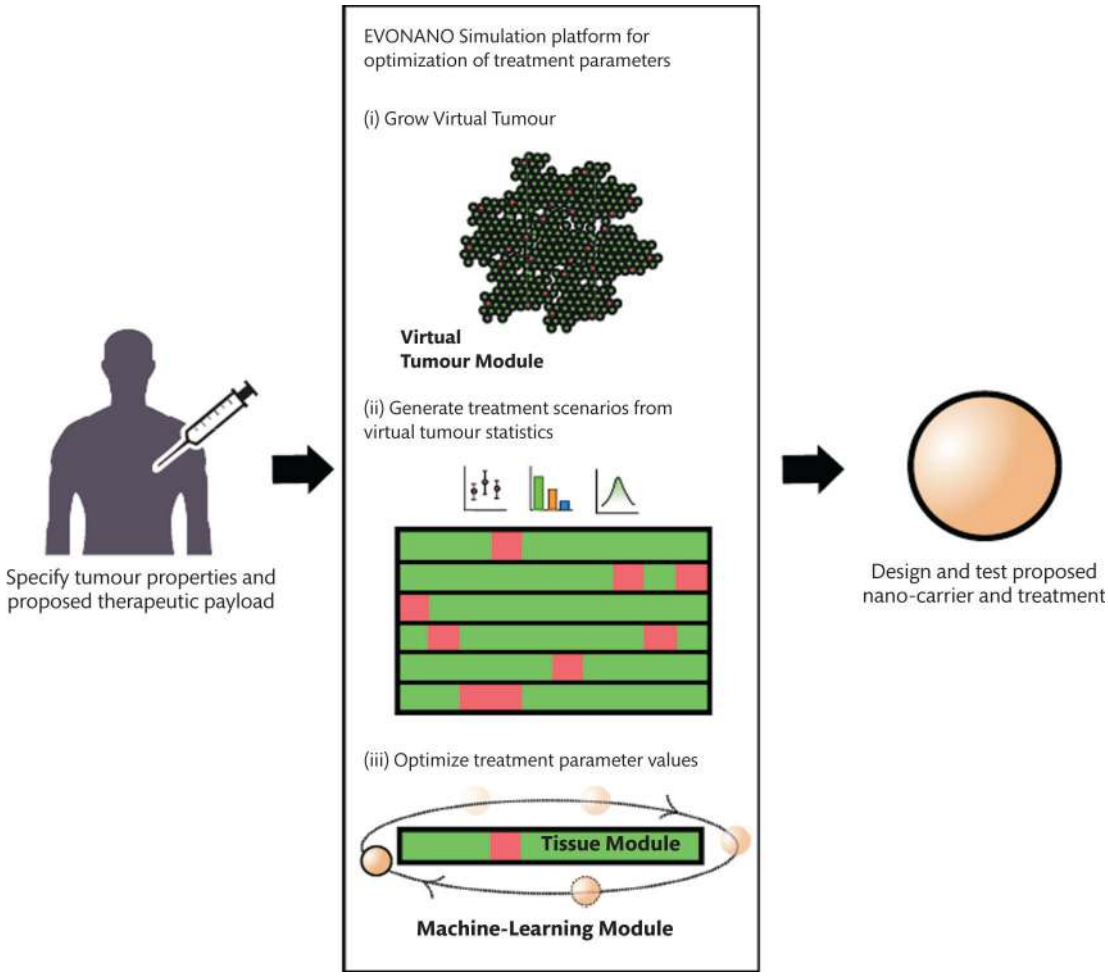


Figure 34.5. Schematic of the EVONANO platform. The EVONANO platform begins by specifying tumour and possible nanocarrier properties that are then used as assumptions within the EVONANO simulation platform. This proceeds as follows: (i) a virtual tumour is first grown to a sufficient size using the virtual tumour model; (ii) summary statistics, such as necessary penetration distance, are calculated from the virtual tumour and used to generate representative treatment scenarios; (iii) parameter values are then optimized using the tissue module and machine-learning module. The nanocarrier and treatment parameters can then be designed and tested using in vitro/in vivo methods. Source: Reproduced with permission from Stillman et al. [51].

compared to primary tumour cells was observed. Moreover, 24 h post injection, in micrometastases 50% of the cells were positive for NPs, whereas only 17% were positive in the primary tumour cells. Finally, in both cases, NP intensity decreased with higher distance from blood vessels.

To study the delivery and accumulation of anticancer drugs (DOX) transported by NPs, Goswami et al. developed a framework consisting of computer vision and a machine-learning module [54]. In the first module, images obtained from a fluorescence microscope were analysed and the average luminance (direct measure of fluorescence directly correlates with cell uptake) value was extracted. Later, this information was used to predict and observe general accumulation trends of DOX by using least-squares regression. Results showed that initially free DOX showed higher concentrations (~0.42 times higher at 4 h) than DOX-loaded NPs, whereas after 24 h, DOX-loaded NPs showed higher intensity values than free DOX in the cell cytoplasm (~1.98 times higher at 24 h) and at the whole cell level (~1.35 times higher at 24 h). Thus, concluding that free DOX accumulates fast at initial times, but also effluxes fast, while DOX-loaded NPs initiate with a slower accumulation, which later increases, and are retained in the cell for an extended period.

34.5. Conclusions

In the past two decades, preclinical research in cancer nanomedicine has focused on the development of novel nanoformulations for the optimization of tumour-targeted drug delivery. However, due to biological interactions and physiological processes *in vivo*, the tumour delivery efficiency of NPs has remained consistently low, which has manifested in the form of limited clinical translation of cancer nanomedicine. To this end, establishing the quantitative relationship between NP properties and their *in vivo* behaviour pertinent to systemic pharmacokinetics, safety, and tumour delivery efficiency is critical to optimize the design of NPs or identify treatment regimen strategies for improved treatment outcomes. Integration of experimental studies with mathematical modelling and/or machine learning enables efficient investigation of the vast

parameter space of this multiscale system. We presented the key challenges associated with NP-mediated drug delivery and the most common methods used for the investigation of NP SAR, with a focus on systemic pharmacokinetics, safety, and tumour delivery efficiency of NPs.

Acknowledgements

This work was supported in part by the National Institutes of Health (NIH) Grants 1R01EB035545 (P.D.) and 1R01CA253865 (Z.W.).

REFERENCES

1. Butner JD, Dogra P, Chung C, Pasqualini R, Arap W, Lowengrub J, et al. Mathematical modeling of cancer immunotherapy for personalized clinical translation. Nat Comput Sci. 2022;**2**(12):785–96.
2. Mitchell MJ, Billingsley MM, Haley RM, Wechsler ME, Peppas NA, Langer R. Engineering precision nanoparticles for drug delivery. Nat Rev Drug Discov. 2021;**20**(2):101–24.
3. Noureddine A, Butner JD, Zhu W, Naydenkov P, Peláez MJ, Goel S, et al. Emerging lipid-coated silica nanoparticles for cancer therapy. In: Saravanan M, Barabadi H, editors. Cancer nanotheranostics. Vol **1**. Cham: Springer International Publishing; 2021. p. 335–61.
4. Hosoya H, Dobroff AS, Driessen WH, Cristini V, Brinker LM, Staquicini FI, et al. Integrated nanotechnology platform for tumor-targeted multimodal imaging and therapeutic cargo release. Proc Natl Acad Sci. 2016;**113**(7):1877–82.
5. Shi J, Kantoff PW, Wooster R, Farokhzad OC. Cancer nanomedicine: progress, challenges and opportunities. Nat Rev Cancer. 2017;**17**(1):20.
6. Wilhelm S, Tavares AJ, Dai Q, Ohta S, Audet J, Dvorak HF, et al. Analysis of nanoparticle delivery to tumours. Nat Rev Mater. 2016;**1**(5):1–12.
7. Kumar M, Kulkarni P, Liu S, Chemuturi N, Shah DK. Nanoparticle biodistribution coefficients: a quantitative approach for understanding the tissue distribution of nanoparticles. Adv Drug Deliv Rev. 2023;**194**:114708.
8. Simak J, De Paoli S. The effects of nanomaterials on blood coagulation in hemostasis and thrombosis. WIREs Nanomed Nanobiotechnol. 2017;**9**:e1448.
9. Sharma S, Parveen R, Chatterji BP. Toxicology of nanoparticles in drug delivery. Curr Pathobiol Rep. 2021;**9**(4):133–44.
10. Yu T, Malugin A, Ghandehari H. Impact of silica nanoparticle design on cellular toxicity and hemolytic activity. ACS Nano. 2011;**5**(7):5717–28.
11. Cheng Y-H, He C, Riviere JE, Monteiro-Riviere NA, Lin Z. Meta-analysis of nanoparticle delivery to tumors using a physiologically based pharmacokinetic modeling and simulation approach. ACS Nano. 2020;**14**(3):3075–95.
12. Labouta HI, Asgarian N, Rinker K, Cramb DT. Meta-analysis of nanoparticle cytotoxicity via data-mining the literature. ACS Nano. 2019;**13**(2):1583–94.
13. Yu T, Greish K, McGill LD, Ray A, Ghandehari H. Influence of geometry, porosity, and surface characteristics of silica nanoparticles on acute toxicity: their vasculature effect and tolerance threshold. ACS Nano. 2012;**6**(3):2289–301.
14. Dogra P, Adolphi NL, Wang Z, Lin Y-S, Butler KS, Durfee PN, et al. Establishing the effects of mesoporous silica nanoparticle properties on in vivo disposition using imaging-based pharmacokinetics. Nat Commun. 2018;**9**(1):4551.
15. Goel S, Ferreira CA, Dogra P, Yu B, Kutyreff CJ, Siamof CM, et al. Size-optimized ultrasmall porous silica nanoparticles depict vasculature-based differential targeting in triple negative breast cancer. Small. 2019;**15**:1903747.
16. Shukla SK, Sarode A, Wang X, Mitragotri S, Gupta V. Particle shape engineering for improving safety and efficacy of doxorubicin—a case study of rod-shaped carriers in resistant small cell lung cancer. Biomater Adv. 2022;**137**:212850.
17. Dogra P, Butner JD, Chuang Y-l, Caserta S, Goel S, Brinker CJ, et al. Mathematical modeling in cancer nanomedicine: a review. Biomed Microdevices. 2019;**21**(2):40.
18. Dogra P, Butner JD, Nizzero S, Ruiz Ramirez J, Noureddine A, Pelaez MJ, et al. Image-guided mathematical modeling for pharmacological evaluation of nanomaterials and monoclonal antibodies. Wiley Interdiscip Rev Nanomed Nanobiotechnol. 2020:e1628.
19. Shah DK, Betts AM. Towards a platform PBPK model to characterize the plasma and tissue disposition of monoclonal antibodies in preclinical species and human. J Pharmacokinet Pharmacodyn. 2012;**39**(1):67–86.
20. Jain RK, Stylianopoulos T. Delivering nanomedicine to solid tumors. Nat Rev Clin Oncol. 2010;**7**(11):653–64.
21. Sarin H. Physiologic upper limits of pore size of different blood capillary types and another perspective on the dual pore theory of microvascular permeability. J Angiogenes Res. 2010;**2**(1):14.
22. Stylianopoulos T, Soteriou K, Fukumura D, Jain RK. Cationic nanoparticles have superior transvascular flux into solid tumors: insights from a mathematical model. Ann Biomed Eng. 2013;**41**(1):68–77.
23. Hashizume H, Baluk P, Morikawa S, McLean JW, Thurston G, Roberge S, et al. Openings between defective endothelial cells explain tumor vessel leakiness. Am J Pathol. 2000;**156**(4):1363–80.
24. Frieboes HB, Wu M, Lowengrub J, Decuzzi P, Cristini V. A computational model for predicting nanoparticle accumulation in tumor vasculature. PLos ONE. 2013;**8**(2):e56876.
25. Terracciano R, Carcamo-Bahena Y, Royal ALR, Messina L, Delk J, Butler EB, et al. Zonal intratumoral delivery of nanoparticles guided by surface functionalization. Langmuir. 2022;**38**(45):13983–94.
26. Wang Z, Kerketta R, Chuang Y-L, Dogra P, Butner JD, Brocato TA, et al. Theory and experimental validation of a spatio-temporal model of chemotherapy transport to enhance tumor cell kill. PLoS Comput Biol. 2016;**12**(6):e1004969.
27. Brachi G, Ruiz-Ramírez J, Dogra P, Wang Z, Cristini V, Ciardelli G, et al. Intratumoral injection of hydrogel-embedded nanoparticles enhances retention in glioblastoma. Nanoscale. 2020;**12**(46):23838–50.
28. Brocato T, Dogra P, Koay EJ, Day A, Chuang Y-L, Wang Z, et al. Understanding drug resistance in breast cancer with mathematical oncology. Curr Breast Cancer Rep. **2014**:1–11.
29. Perera RM, Zoncu R, Johns TG, Pypaert M, Lee F-T, Mellman I, et al. Internalization, intracellular trafficking, biodistribution of monoclonal antibody 806: a novel anti-epidermal growth factor receptor antibody. Neoplasia. 2007;**9**(12):1099–110.
30. Belleudi F, Marra E, Mazzetta F, Fattore L, Giovagnoli MR, Mancini R, et al. Monoclonal antibody-induced ErbB3 receptor internalization and degradation inhibits growth and migration of human melanoma cells. Cell Cycle. 2012;**11**(7):1455–67.

31. Zhang S, Gao H, Bao G. Physical principles of nanoparticle cellular endocytosis. ACS Nano. 2015;**9**(9):8655–71.
32. Haqqani AS, Bélanger K, Stanimirovic DB. Receptor-mediated endocytosis and brain delivery of therapeutic biologics. Front Drug Deliv. 2024;**4**:1360302.
33. Ikeda-Imafuku M, Wang LL-W, Rodrigues D, Shaha S, Zhao Z, Mitragotri S. Strategies to improve the EPR effect: a mechanistic perspective and clinical translation. J Control Release. 2022;**345**:512–36.
34. Tsoi KM, MacParland SA, Ma X-Z, Spetzler VN, Echeverri J, Ouyang B, et al. Mechanism of hard-nanomaterial clearance by the liver. Nat Mater. 2016;**15**(11):1212.
35. Cataldi M, Vigliotti C, Mosca T, Cammarota M, Capone D. Emerging role of the spleen in the pharmacokinetics of monoclonal antibodies, nanoparticles and exosomes. Int J Mol Sci. 2017;**18**(6):1249.
36. Yao Y, Zang Y, Qu J, Tang M, Zhang T. The toxicity of metallic nanoparticles on liver: the subcellular damages, mechanisms, and outcomes. Int J Nanomed. 2019;**14**:8787–804.
37. Sukhanova A, Bozrova S, Sokolov P, Berestovoy M, Karaulov A, Nabiev I. Dependence of nanoparticle toxicity on their physical and chemical properties. Nanoscale Res Lett. 2018;**13**(1):44.
38. Dogra P, Butner JD, Ramírez JR, Chuang Y-l, Noureddine A, Brinker CJ, et al. A mathematical model to predict nanomedicine pharmacokinetics and tumor delivery. Comput Struct Biotechnol J. 2020;**18**:518–31.
39. Goel S, Zhang G, Dogra P, Nizzero S, Cristini V, Wang Z, et al. Sequential deconstruction of composite drug transport in metastatic breast cancer. Sci Adv. 2020;**6**(26):eaba4498.
40. Hendriks BS, Reynolds JG, Klinz SG, Geretti E, Lee H, Leonard SC, et al. Multiscale kinetic modeling of liposomal Doxorubicin delivery quantifies the role of tumor and drug-specific parameters in local delivery to tumors. CPT Pharmacometrics Syst Pharmacol. 2012;**1**(11):e15.
41. Dogra P, Ramírez JR, Butner JD, Peláez MJ, Chung C, Hooda-Nehra A, et al. Translational modeling identifies synergy between nanoparticle-delivered miRNA-22 and standard-of-care drugs in triple-negative breast cancer. Pharm Res. 2022;**39**:511–28.
42. Dogra P, Ramírez JR, Butner JD, Peláez MJ, Cristini V, Wang Z, editors. A multiscale model to identify limiting factors in nanoparticle-based miRNA delivery for tumor inhibition. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico. 2021:4230–33.
43. Islam MA, Barua S, Barua D. A multiscale modeling study of particle size effects on the tissue penetration efficacy of drug-delivery nanoparticles. BMC Syst Biol. 2017;**11**(1):113.
44. Sykes EA, Dai Q, Sarsons CD, Chen J, Rocheleau JV, Hwang DM, et al. Tailoring nanoparticle designs to target cancer based on tumor pathophysiology. Proc Natl Acad Sci. 2016;**113**(9):E1142–E51.
45. Wirthl B, Kremheller J, Schrefler BA, Wall WA. Extension of a multiphase tumour growth model to study nanoparticle delivery to solid tumours. PLoS ONE. 2020;**15**(2):e0228443.
46. Laomettachit T, Puri IK, Liangruksa M. A two-step model of TiO(2) nanoparticle toxicity in human liver tissue. Toxicol Appl Pharmacol. 2017;**334**:47–54.
47. Lin Z, Monteiro-Riviere NA, Kannan R, Riviere JE. A computational framework for interspecies pharmacokinetics, exposure and toxicity assessment of gold nanoparticles. Nanomedicine (Lond). 2016;**11**(2):107–19.
48. Huang Y, Li X, Cao J, Wei X, Li Y, Wang Z, et al. Use of dissociation degree in lysosomes to predict metal oxide nanoparticle toxicity in immune cells: Machine learning boosts nano-safety assessment. Environ Int. 2022;**164**:107258.
49. Concu R, Kleandrova VV, Speck-Planche A, Cordeiro M. Probing the toxicity of nanoparticles: a unified in silico machine learning model based on perturbation theory. Nanotoxicology. 2017;**11**(7):891–906.
50. Lin Z, Chou W-C, Cheng Y-H, He C, Monteiro-Riviere NA, Riviere JE. Predicting nanoparticle delivery to tumors using machine learning and artificial intelligence approaches. Int J Nanomed. **2022**:1365–79.
51. Stillman NR, Balaz I, Tsompanas M-A, Kovacevic M, Azimi S, Lafond S, et al. Evolutionary computational platform for the automatic discovery of nanocarriers for cancer treatment. NPJ Comput Mater. 2021;**7**(1):150.
52. Nademi Y, Tang T, Uludağ H. Modeling uptake of polyethylenimine/short interfering RNA nanoparticles in breast cancer cells using machine learning. Adv NanoBiomed Res. 2021;**1**(10):2000106.
53. Kingston BR, Syed AM, Ngai J, Sindhwani S, Chan WCW. Assessing micrometastases as a target for nanoparticles using 3D microscopy and machine learning. Proc Natl Acad Sci. 2019;**116**(30):14937–46.
54. Goswami S, Dhobale KD, Wavhale RD, Goswami B, Banerjee SS. Computer vision and machine-learning techniques for quantification and predictive modeling of intracellular anticancer drug delivery by nanocarriers. Appl AI Lett. 2022;**3**(1):e50.

SECTION 8

Ecology, evolution, and cancer

35

Decoding cancer evolution through adaptive fitness landscapes

Rowan Barker-Clarke, Eshan S. King, Jeff Maltas, J. Arvid Ågren, Dagim Shiferaw Tadele, and Jacob G. Scott

35.1. Introduction

Cancer is a disease that starts and ends with evolution. In all non-transmissible forms, the tumour cells are directly descended from the healthy cells of the patient. It is host-derived cells undergoing mutation or modification that eventually evolve into a cancerous state. These expanding tumour cells outcompete healthy, non-malignant cells for space and resources.

The clinical treatment of cancer aims to successfully eradicate tumour cells without harming the host, yet total eradication via cancer treatment is still difficult to achieve across many cancers. This is in part due to heterogeneity within and between patients, alongside the nature of the similarity between healthy and malignant cells in a patient [1]. Unfortunately, the most difficult problem to address is the deadliest: the evolution of resistance in cancer cells, arising in response to the treatment designed to eradicate them.

The initial treatment-naive expansion of tumour cells often results in plasticity and diversity [1]; however, it is subsequent treatment pressure that promotes the expansion of resistant cancer cells [2]. Thus, the standard approach to cancer treatment, aimed at rapidly reducing initial tumour volume, can do little to prevent subsequent therapy-resistant tumours in the patient [3]. The intrinsic notion that treatment impacts the evolution of patient tumours has motivated research aimed at altering existing treatment paradigms [4–6]. Due to the nuance and mathematical precision required, the development of these novel therapeutic approaches often consists of interdisciplinary work with mathematical and computational biologists. One such approach aiming to better inform treatment regimens is 'adaptive therapy'. Adaptive therapy uses predictive models informed and parameterized by the patient's disease combined with the real-time response of a patient's tumour to therapy to improve patient response [6–8]. Although adaptive therapy leverages some aspects of tumour evolution, we introduce in this chapter the notion of 'evolutionary therapy' which aims to advance these ideas. The aim of evolutionary therapy is not only to predict the future evolution of a tumour but also to actively steer it. To achieve this goal, we require a suite of control approaches and a deep understanding of the different evolutionary pathways available to cancer cells over time. It is the object at the heart of this chapter, the adaptive fitness landscape, which encapsulates predictability, encoding the map of all possible evolutionary trajectories [9].

In this chapter, we examine the adaptive landscape, its history, and its successes. We illustrate how the adaptive landscape provides the road map for evolution and thus for successful cancer treatment. We introduce the models and mathematics used to study adaptive landscapes and review successful studies that assess how the underlying landscape structure impacts evolution. We also discuss the unique evolutionary mechanisms present in cancer and highlight some experimental methods that allow us to study cancer evolution. We consider and demonstrate how fitness landscapes may shed light on cancer initiation and treatment resistance and finally explore some of the exciting future directions available for study at the intersection of adaptive landscapes and cancer research.

35.1.1. What is an adaptive fitness landscape?

Fitness broadly refers to the ability of an organism to survive and reproduce. The term 'fitness landscape' was coined by Sewall Wright in 1932 [10] and has since been used to analyse and discuss so many systems that the exact meanings of landscape and fitness, even within biology, vary by context [11–15]. Fitness landscapes have been used to study various phenomena undergoing biological evolution, such as the evolution of antibiotic resistance in bacteria and the evolution of protein sequences and RNA folding [16–19].

In the context of cancer, fitness typically refers to the growth rate of the tumour cells. This can be genotype-specific, meaning that different genetic changes in tumour cells impart different fitness effects. In a traditional 3D representation of a fitness landscape, the x- and y-axes represent different genotypes (i.e. different combinations of genetic traits), and the z-axis represents the fitness of each genotype. The landscape is then a graphical representation of the relationship between genotype and fitness (see **Figure 35.1** for examples).

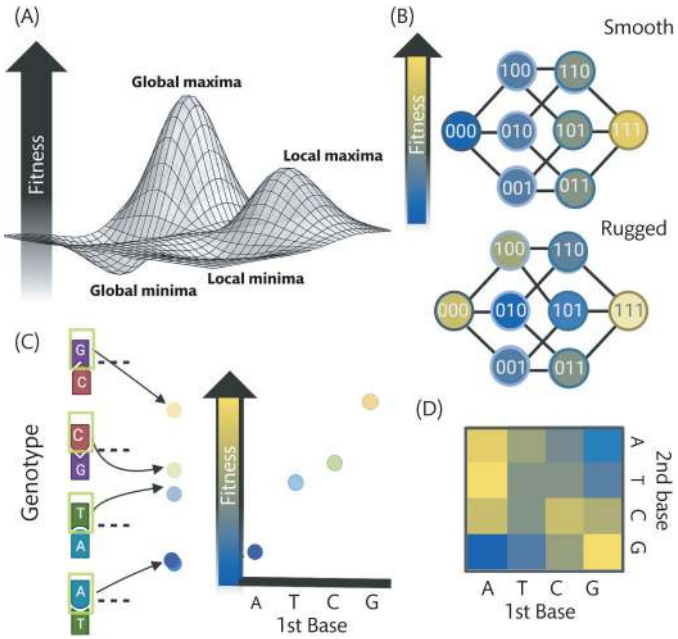


Figure 35.1. Different representations of fitness landscapes. (A) A generic and rugged 3D landscape with two valleys and two peaks is shown. In three dimensions, the *x*–*y* plane is the space of possible genetic combinations and fitness is elevation. Genotypes at local maxima are less fit than the genotype of the global maxima. Similarly, minima in the landscape also vary in depth. (B) Two comparative 2D examples of three allele landscapes with and without epistasis are shown. In the smooth landscape, mutations increase fitness in predictable independent ways. (C) Examples of how specific bases at certain positions within a genotype could convey differing fitness. Here, the third base in a sequence is mapped to a different fitness. (D) If more bases are incorporated, higher dimensional maps like this 2D heat map can represent fitness landscapes. Fitness landscapes can represent any kind of mutation, e.g. using alleles (01) or base pairs (ATCG).

One important concept in fitness landscapes is the 'peak' or 'optimal' genotype; the genotype that has the highest fitness in the landscape. In a smooth landscape, the peak genotype is directly accessible via a simple adaptive walk, and thus the fitness increase should be relatively smooth over time. In a rugged landscape, the peak genotype may be more difficult to reach, and the fitness increase may be very variable over time. Fitness landscapes can also be understood in terms of 'adaptive valleys' and 'adaptive peaks'. An adaptive valley is a region of the landscape where there are few genotypes with high fitness, while an adaptive peak is a region of the landscape where there are many genotypes with high fitness. Just as particles are drawn to attractive potential wells in classical physics, selection pressure upon mutating populations guides the evolving populations towards peaks in the landscape. The shape of the fitness landscape can influence the direction and speed of evolution as well as the probability of population extinction.

The ruggedness of a landscape is caused by a property called epistasis. Without epistasis, a fitness landscape is smooth with a gentle slope; small differences in genotype space correspond to set differences in fitness. This smoothness means that there are many viable paths towards high fitness such that the shortest paths towards global peaks are linear. On the other hand, a rugged fitness landscape has many peaks and valleys, with relatively small differences in genotype leading to large differences in fitness. This means that there may be a few viable paths towards high fitness and that adaptation may proceed slowly or unpredictably [20].

Fitness landscapes can also be used to understand how populations evolve. In a population, the distribution of genotypes may shift over time as different genotypes become more or less common. Both random fluctuations and various variable factors, such as the size of the population, the rate of reproduction, the mutation rate, and the presence of selective pressures, can influence this. Evolution via selective pressures can be modelled as an adaptive walk, which is a form of optimization on the fitness landscape. Survival of the fittest 'moves' the population genotype distribution across the landscape towards fitness peaks. Under selection, adaptive mutations become more frequent as a population evolves, approaching peaks in the landscape. Overall, fitness landscapes are a powerful tool for understanding the evolution of organisms and populations, providing a representation of the relationship between genotype and fitness and helping researchers to understand how different factors, such as interactions between genes, mutation and selection, influence the evolution of a population.

35.1.2. Mapping the vastness of evolutionary space

In the 1930s, prior to the discovery of the structure of DNA, Wright, among others, understood that information stored in chromosomes in the nucleus was inherited by offspring. Scientists discovered through experiments that the information within genes could be randomly mutated, and these mutations could impact the properties of a species and thus broadly its fitness. The landscape metaphor came into being partly because, at the time, scientists already appreciated that there were likely hundreds to thousands of possible mutations in each species. They reasonably concluded that the number of possible different combinations of mutations in a population was intractably large. The vastness of evolutionary space, which Wright described as 'an almost infinite field of possible variations', motivated the development of the fitness landscape to help explain evolution within it. His initial paper on the roles of mutation and selection in evolution was published in 1931 [21]. He soon followed this with a description of a metaphorical landscape complete with the aid of illustrative mutational hyper-cubes (higher dimensional punnet squares) and contour maps [10] (see Figure 35.1).

Today, we know that the human genome consists of approximately 20,000 genes and over three billion base pairs. A combinatorially complete landscape would consist of all possible combinations of mutant alleles in a genome. The resultant number of possible genomes has been described as hyper-astronomical [22]. For example, given just the possibility of only one type of mutation in each gene region allows for $2^{20,000}$ or $4 \times 10^{6,020}$ different possible combinations. This number dwarfs the number of atoms in the observable universe (approximately 7×10^{72}). Thus, the idea of mapping all possible versions of the human genome to their effects is infeasible.

This realization, rather than halting all attempts to understand this space, has driven many different scientific questions, aided by physicists, computer scientists, and abstract mathematicians about how the size of the theoretical genotype space impacts evolution. Focused experimental studies have analysed smaller sub landscapes in depth, while mathematical studies have used abstract mathematics and algebraic approaches to explore how species can adapt and evolve in such intractable spaces [20,23,24].

Box 35.1 Mathematics of fitness and epistasis

Fitness landscapes represent mappings from a basis to fitness,

$$\text{mutable basis} \rightarrow \text{fitness}. \tag{35.1}$$

Without epistasis, fitness is just a linear sum of the fitness contributions from each mutation or allele,

$$w = f(w_{1\ldots n}) = \sum_{i=1}^{n} w_i. \tag{35.2}$$

The degree of epistasis, ε, is defined as

$$\varepsilon \equiv \ln w_{a,b} - \ln w_a - \ln w_b. \tag{35.3}$$

35.1.3. Genetic, RNA, and protein landscapes

One advantage of fitness landscapes is that the mechanistic link between genetic information and fitness needs not be known such that the formalism lends itself to various proxies or analogies for adaptive fitness. Adaptive landscapes have been studied experimentally and theoretically to better understand both macroscopic and microscopic evolutionary scales. The same techniques for studying genotype–fitness landscapes have also been leveraged to explore genotype–phenotype mappings. Many of these systems have been studied theoretically within the framework of adaptive walks. As such, there is a large body of successful work aimed at probing the nature of evolution across genetic, transcriptomic, and protein spaces using similar techniques [16,17,24–28].

Proteins and RNAs, as shorter and simpler structures than the genome, have formed more tractable spaces for mechanistic theoretical and experimental landscape work. Combined with a good understanding of the underlying biophysics of RNA and protein folding, early RNA folding and protein folding models have been used to understand the power of genotype–phenotype–fitness mappings [16,25]. For example, the physical chemistry of base pair interactions has been used to build landscape studies on the evolution of RNA folds and protein binding domains [17]. Pitt and co-worker calculated the reaction rate constant for every point mutant of a catalytic RNA and were able to use biochemical activity measures and sequencing to rapidly construct the fitness landscape of 10^7 unique RNA sequences [17]. These works have supported the notion that fitness landscape objects are informative in understanding mechanistic aspects of evolution and have helped develop experimental landscape measuring techniques.

To date, genomic and protein adaptive landscapes have typically been reduced to a focal subsystem of interest for experimental studies. When measuring a landscape, experiments may be focused on mutations to a single gene, to a binding site of interest, or abstracted to a generalized mutation within a combination of genes related to a function or pathway. An important example is of a fitness landscape in HIV, where measurement of the protein landscape has aided the development of anti-viral compounds [26]. Examples of experimental genotype–fitness landscape exploration include work elucidating the response to antibiotics of bacterial populations containing all different possible mutations in β-lactamase [29]. Chen et al. expanded this to quantify all single mutational effects on fitness and EC50 of VIM-2 β-lactamase across a 64-fold range of ampicillin concentrations [30].

Fitness and adaptive landscapes attempt to describe phenomenally large genotypic spaces and encode our phenotypic understandings. These mappings have a long history and a demonstrated utility across many biological systems, including bacteria, protein, and RNA evolution. This solid basis has driven experimental and mathematical evolutionary study and provides an excellent framework with which to study the complex evolutionary dynamics present in cancer.

35.2. Modelling fitness landscapes

To the systems biologist or mathematician, the idea of mapping a genotype to a high-dimensional fitness surface is conceptually simple, yet the biological importance and implications of the shape of this surface are profound. The landscape model and mapping have formed the basis of many theoretical and empirical studies and have fortified and expanded upon traditional notions of the relationship between genotype and fitness. In this section, we discuss the mathematical modelling of landscapes, evolution, and the non-linearity of the fitness surface. We highlight how the shape of this landscape impacts theoretical population dynamics and evolvability, and how the non-linear interactions, or epistasis can be measured and modelled.

35.2.1. Theory of landscapes

The fitness landscape is a specific type of mathematical mapping, resulting in an n-dimensional surface where the fitness w depends on the position in the state space. The state could be, for example, the DNA sequence, RNA sequence, protein folding, or genes and their mutational state. Fitness, w, can be a direct or resultant phenotypic property, such as reproductive fitness, stability, or binding affinity. For each mapping there exists a mutable basis (e.g. genotype) and a resultant fitness property (e.g. division rate) to be selected upon.

In the simplest case, each gene or mutation is selected upon independently, and this linear or additive behaviour produces what is known as an additive landscape. Under this assumption, the fitness of a genome with n mutations is just a linear function, i.e. a sum of the fitness changes associated with individual mutations (Equation 35.2). The central concept that allows alteration of the shape of this fitness landscape is 'epistasis'. This refers to the interaction between different genes and how they affect each other's contributions to fitness. In some cases, one gene may have a strong effect on fitness, while in other cases, the effect of a gene may depend on the presence or absence of other genes. The degree to which these fitness effects are additive and the relative importance of epistasis has been a subject of discussion for many decades [10,31–34].

35.2.2. Epistasis in the genome

Long understood to be present through experimental evidence, genetic epistasis refers to the non-additive interaction between different genes and how they affect each other's contributions to an organism's traits. This interaction can be either positive, negative, or neutral, depending on the specific genes involved and the context in which they are expressed. Epistasis describes how genes interact with each other and explains the resultant shape of the high-dimensional surface of a fitness landscape.

Although epistasis has been described using various terms across biology, Weinreich et al. describe the following forms of epistasis relative to the genetic background [31]. These forms are sign and magnitude epistasis. Sign epistasis means that the mutation is beneficial (positive) on some genotypic backgrounds and deleterious (reciprocal) on others. In magnitude epistasis, the mutation has a fixed direction of impact on the fitness of an organism or cell but a magnitude dependent upon the background genome.

Genetic epistasis can also occur at different levels of the genetic hierarchy, such as between different genes, between different chromosomes, or between different genetic pathways. For example, one gene may affect the expression of another gene, or one chromosome may affect the expression of genes on another chromosome. This can have important implications for the evolution of populations. In some cases, it can lead to the evolution of complex traits that are composed of multiple genes rather than being controlled by a single gene.

The evolution of complex traits is common in cancer where it is often the case that multiple adaptive steps are required for tumour evolution [35]. In the human genome and within cancer, increased genetic diversity and mutation rate increase the importance of understanding epistatic effects. These epistatic effects can influence the direction and speed of a tumour's evolution, as well as the probability that a targetable mutation fixes in the population [36].

35.2.3. Evolution with epistasis

In the additive landscape, evolution is simple; there is a single global maximum and, as a result, an accessible evolutionary path to the maximum from any point. Wright and others had, however, observed and considered the possibility of epistasis. This non-linear interaction between alleles is mathematically defined in Equation 35.3, where w_i is the fitness of a genome with mutation at site i.

This value, ε, is positive in the presence of positive sign epistasis—a supplemental positive effect due to the interactions of two alleles or mutations such that $\varepsilon > 0$ and for reciprocal sign epistasis, $\varepsilon < 0$, refers to a more negative effect on fitness compared to the effects of these mutations alone.

Epistasis was a key motivation for the creation of adaptive landscapes, for if a mutation did not confer a constant and guaranteed benefit to a cell, the fitness landscape would not be smooth and additive. Epistasis is therefore an important factor that shapes the surface of fitness landscapes, which are representations of the relationship between genotypes (i.e. combinations of genetic traits) and fitness. In the presence of epistasis, a landscape is multi-peaked and the accessibility of trajectories to the global fitness maximum may be entirely dependent on the starting genotype, and the full impact of epistasis on evolution by natural selection is still an open question [31,37].

The adaptive fitness landscape helps illustrate how epistasis is critical in determining whether a population can quickly explore a landscape and access global peaks in fitness. As epistasis influences how smooth or rugged a landscape is, it affects the direction and speed of evolution. The nature of adaptation to environmental pressures, such as oxygen or drug concentration, would be far more complex in landscapes with strong epistatic effects. Understanding the nature of epistasis present in cancer and its effect on the fitness of cells can help scientists to better understand the evolution of tumours and their composite subpopulations.

35.2.4. Modelling evolution on landscapes

If the landscape is already defined experimentally or numerically, fitness landscapes can be used to simulate the evolution of populations over time and to test different hypotheses about evolution. In cancer, our aim is to use this surface and models on it to predict the evolution of a tumour. Many mathematical approaches have been developed to model cancer progression, but two types of mathematical models are commonly used to study resistance evolution in tandem with fitness landscapes: population genetics (typically continuous) models and individual-based stochastic models [38].

One popular model is the strong selection weak mutation model of evolution [39], which models evolution as a stochastic point process in which mutations are immediately fixed or eliminated; the evolving population is represented as a moving point in genotype space. In general, population genetics models are used to study the evolution of populations by considering the changes in the frequency of different genes over time and modelling the changing mean and variance of allele frequencies in the population. These models are often based on the Hardy–Weinberg equilibrium and Wright–Fisher processes, which describe the relationship between genotype frequency and gene frequency in a population.

Population-based models can be written in the form of ordinary or stochastic differential equations, and examples include models that successfully predict clinical drug synergies or patient outcomes [40,41]. These models can assess how different selective pressures might influence the time to cancer initiation or the evolution of resistance within a population [42–44].

Individual-based models are used to study the evolution of populations by simulating the behaviour of individual organisms and the interactions between them [45]. These cellular-automata models can incorporate intrinsic heterogeneity and can be used to understand how different factors, such as competition for resources or cell–cell interactions, might influence the evolution of a population [2,46,47]. Platforms, such as PhysiCell, and specific libraries, such as HAL, have enabled easier agent-based modelling of populations [48,49]. The use of agent-based models allows cells to explore complex evolutionary landscapes while encoding spatial dependencies, particularly allowing for dynamic environmental properties and ecological interactions.

35.2.5. Evolvability and epistasis

In order to assess the effect of epistasis on these evolutionary models, mathematicians have developed methods of both quantifying epistasis in landscapes and generating landscapes with varying epistasis [23,50]. Although the exact measure of epistasis between two genes can be quantified (Equation 35.3), other measures have been developed to quantify the magnitude of average epistatic effects over larger landscapes.

One notable landscape generation methodology produces the '*NK*' set of model landscapes [23], which were developed in 1989 by Kauffman. Initially applied to model the binding affinity of antibody molecules to antigens, *NK* landscapes were quickly adopted as a useful general model of epistasis. *NK* landscapes mimic the adaptive landscapes underlying the adaption of the immune system to novel pathogens, particularly in that the K parameter can be tuned to match the observed average number of mutations needed to reach a peak within the landscape. Generated statistically, N and K define

the landscape and epistasis where N represents the number of sites in the genotype or protein sequence and K is the degree of epistatic interaction. These model landscapes are popular for their 'tunably rugged' properties, in which the fitness distribution and the nature of epistasis are predetermined.

Many studies have used simulated landscapes, such as the *NK* landscapes, to reinforce the intuition that the navigability of a landscape depends on its structure and the idea that evolutionary plasticity is a mechanism of escape from local maxima and is required to traverse large spaces [51–54]. The exploration of high-dimensional landscape topology and its ruggedness or high-order epistasis is one of the most powerful uses of abstract mathematics in the study of evolutionary spaces [36,55,56]. More recently, these ideas have been expanded to discuss the nature of complexity in evolution, understanding complexity as a restriction to evolvability, and the implications on natural evolution and its mechanisms [20]. Numerical solutions derived through computation have also allowed higher dimensional systems to be studied numerically as well as allowing exhaustive simulation of evolution on combinatorially complete theoretical landscapes.

As the availability of experimentally derived landscapes has increased, studies such as those by Aguilar et al. have allowed the exploration of landscapes in the natural world and how these compare to the range of theoretical possibilities [57]. Studies in the β-lactamase gene have reinforced ideas around adaptability, demonstrating that epistatic interactions increase the number of indirect paths accessible to evolution, delaying commitment to a specific evolutionary trajectory [19]. The idea of epistasis that motivated the landscape description has been the study of many works including the emphasis on the role of epistasis in evolution [34]. Experimental work investigating the presence of epistasis has shown that it can increase evolvability, i.e. the ability of a species to reach a global fitness peak [50,53].

The structure of generic genotype–phenotype mappings has also been examined to further understand evolution. The structure and connectivity of a high-dimensional genotype–phenotype space can promote certain phenotypes through frequency alone [58]. Exhaustive explorations of relatively simple spaces, such as RNA folding, show fascinating insights including observations that the landscape places limits on naturally observed RNA folds and that common folds predicted by landscapes reflect the frequency of folds in nature [24,58,59].

35.3. Cancer evolution

Large quantities of experimental and patient data have aided the study and understanding of cancer and its evolution. Spatial tissue sampling, single-cell DNA and RNA sequencing, and techniques such as genetic 'barcoding' add unprecedented levels of detail to evolutionary experiments. Paired with CRISPR knockouts in mice and high-throughput drug assays in cell lines, these experiments allow a fine-grained approach to understanding treatment responses. The availability of this data has allowed for comparative phylogenetic, experimental, mathematical, and observational methods to be utilized in cancer studies.

In this section, we review the different evolutionary processes and their study in cancer. We illustrate how different evolutionary experiments aim to sample the adaptive landscapes underlying cancer evolution and discuss how evolutionary therapy can aim to control evolution within patient tumours.

35.3.1. Evolution of cancer cells

Cancer is a disease state characterized by common features including rapid and uncontrolled growth [60]. As cancer biologists, we know that cell-intrinsic mechanisms that prevent rapid and uncontrolled growth are eroded during cancer development while mutation rate and proliferation are increased [60]. Tumour growth comes ultimately at the host's expense and is then maintained under therapy when cancer cells access therapy-resistant states.

Understanding how and why cancer develops and proliferates has highlighted two particularly interesting features in the context of evolution, the increased rate of mutational supply and the manipulation of selection pressures. There exist several mechanisms by which mutational rate is increased and selection decreased, e.g., cancer commonly has reduced efficacy of DNA repair [61] and defects within the pathways that signal apoptosis in the event of genomic error [60]. In the absence of genome correction, resulting in higher mutation rates, individual genomes within a tumour population gain mutations that can be advantageous, neutral, or deleterious [62,63].

The ability of a tumour to manipulate selection pressure also involves the manipulation of the microenvironment [60,64,65]. Immune evasion through the modification of immune cells and antigen presentation reduces the negative selection pressure upon tumour cells, and the recruitment of macrophages also reduces the likelihood of the host's immune system eradicating cells via phagocytosis [66]. Tumour cells have also been shown to modify their microenvironment via the recruitment of cancer-associated fibroblasts in their surroundings, cells that promote tumour growth [67] or modify vasculature [64].

This accelerated evolution and modified selection permits cancer populations to alter their fitness landscape and more rapidly explore it. In this accelerated evolution regime, predictability and control become increasingly important factors in cancer treatment. By understanding which specific mutations are driving the growth and progression of cancer, we can construct a more realistic landscape and thus create better evolutionary models and design better control protocols.

35.3.2. Driver mutations in the landscape of cancer

Driver mutations are genetic changes that contribute to the development and progression of cancer by enabling cells to survive, grow, and divide in an uncontrolled manner. These mutations can occur in various genes, including oncogenes (genes that promote cancer) and tumour suppressor genes (genes that inhibit cancer).

Driver mutations reflect repeatability in cancer evolution, as the occurrence of specific driver mutations is seen across large numbers of distinct cancers and patients. Specific 'driver' mutations reflect strong peaks, or 'global epistasis', in the cancer landscape, promoting the gain of these mutations during adaptive evolution and increasing predictability [68]. In recent studies, such as those by Hosseini et al., the analysis of tumour genomes reveals predictable pathways inferred from phylogenetic analysis [69]. Understanding the many pathways available and involved in cancer evolution is equivalent to fully understanding the adaptive walks that are possible on a cancer fitness landscape. Knowledge of

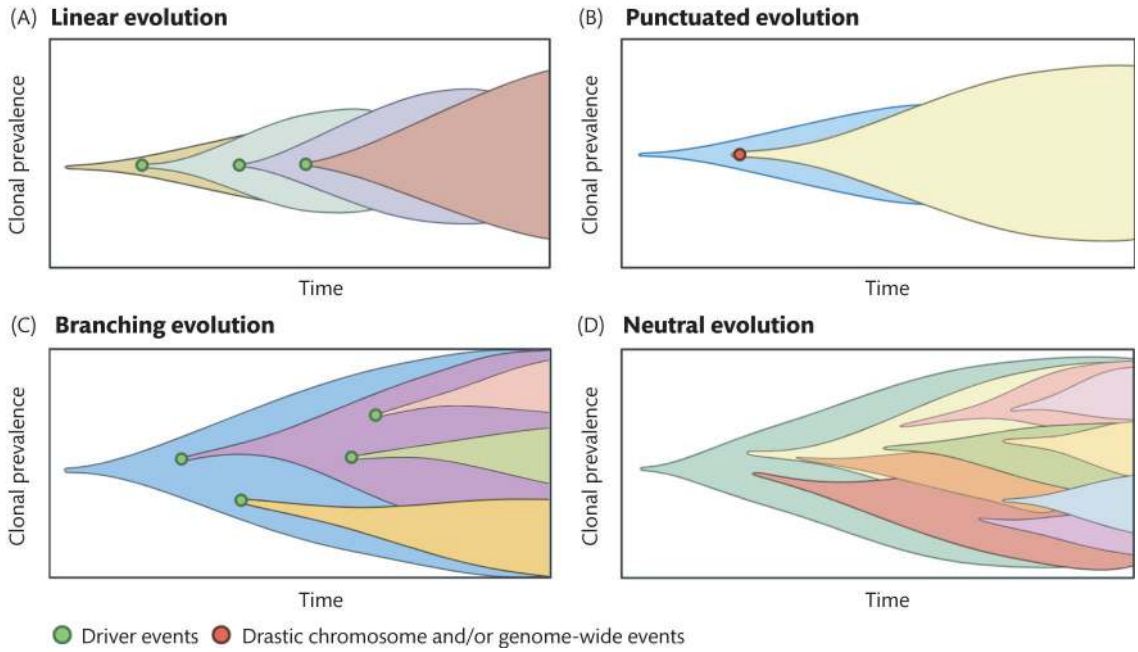


Figure 35.2. Possible evolutionary modes as reflections of the local landscape. The variable selection pressures and mutational accessibility in the landscapes, as well as the size of the population, can result in different 'modes' of evolution. These modes can be understood as reflections of the local fitness landscape to which the tumour or population has access. (A) Linear evolution involves the stepwise gradual movement of the entire population upwards in fitness through different genotypes. (B) Sudden or complex genome-wide events in punctuated evolution reflect a large single-step movement across landscape space resulting in subsequent diversification and expansion. (C) Branching evolution involves the divergence of a population resulting in a mixture of the fittest genotypes. (D) Neutral evolution involves an apparent lack of selection, resulting in populations acquiring and retaining random mutations with no sweeps or predictable clonal dominance.

driver mutations is intrinsic to the idea of the fitness landscape and the development of targeted therapies.

There exist many identified genes in cancer which are commonly mutated. *TP53, PI3KCA*, and *KRAS* are three such drivers, with *TP53* mutations found in almost 35% of all cancer cases in the United States [70]. One of the most commonly mutated genes, *TP53*, is a clear example of positive sign epistasis. Mutations in this gene are often deemed necessary, but not sufficient, for tumour growth, and the presence of *TP53* mutations in combination with other driver mutations is strongly associated with cancer progression. There is evidence for widespread epistatic interactions in the human genome in cancer [71].

While common mutations are repeatedly seen across cancer patients, similar cancer phenotypes can arise via distinct mechanisms. As a consequence, genomic studies in cancer provide strong evidence for both repeatable and convergent evolution [72]. By identifying both epistasis and robust peaks in fitness landscapes, we can elucidate whether specific mutations demonstrate epistasis and how strongly they are selected for [72].

35.3.3. Clonal selection and cancer landscapes

Due to the random nature of genetic mutation during cell division, mutations constantly occur within the genome of both normal and cancer cells. However, these mutations usually undergo negative selection and most mutations do not undergo expansion. In developing tumours, evolution typically occurs through soft or hard sweeps involving the selection and dominance of the fittest multiple or single sub-clones, respectively. However, the macroscopic tumour can often contain heterogeneous populations, and alternate types of selection of the sub-clones can be reflected in the phylogenetic tree of clones within tumours. Monitoring the strength of selection and frequency of mutation has led to four standard models of cancer evolution (Figure 35.2): linear, punctuated, branching, and neutral evolution [73]. These modes reflect differences in how the tumour population diversifies over time and contains implications about the underlying fitness landscape.

Linear evolution describes a stepwise increase in the fitness of tumour cells, and hard sweeps of the new fittest genome reflect a gradual increase in fitness. This mode of evolution reflects an additive genetic or epigenetic landscape, reflecting stepwise cell modification or mutation. Genome-wide punctuated evolution describes the evolution of a population via a hard sweep occurring via a specific genome-wide event from which the population rapidly diversifies. Such an event, by the nature of a hard sweep, involves a large genomic change arriving at a local maximum with strong global epistasis where the tumour expands and the genome of the tumour population remains. Branched evolution reflects the divergence of a population and the expansion of multiple clones. This would be more commonly seen in rugged or epistatic landscapes where multiple local peaks are present and branched evolution occurs as different subpopulations of cells evolve independently to nearby fitness maxima. In flat landscape regions, neutral evolution can also play a role in the development and progression of tumours [62] These neutral mutations (those which confer no benefit) can be amplified by the selection of associated driver mutations and contribute to the overall genetic diversity of the tumour, which can influence its behaviour and response to treatment [62].

Modern work aims at disentangling and estimating these modes of selection and their strength using temporal inference from the sequencing of clones with evolution and selection in mind [74]. Sequencing can reveal expected mutations within known pathways, epistasis in the presence of background mutations, as well the existence of multiple pathways to resistance [75]. The prevalence of mutations can inform us about the underlying landscapes and how treatment has altered selection pressures on cancer sub-clones [76,77].

35.3.4. Experimental and engineered landscapes

Evolutionary experiments validate or challenge our ideas of evolution and our understanding of the fitness landscape. The required approach to probe a landscape entirely depends upon the nature of the landscape of interest. Experimental evolution and engineered fitness landscapes provide two distinct methods aimed at understanding the underlying landscape. In the presence of epistasis, it is uncommon that an entire landscape is easily accessible under experimental evolution, and thus the direct engineering of cells with mutations associated with resistance can be used to probe specific landscapes of interest.

When engineering fitness landscapes, a complete genetic space can be measured whether it is accessible to adaptive evolution or not. In the analysis of some small protein landscapes, engineering allows fully exhaustive experimental construction [29]. In cancer cell lines, transfection and genome-editing techniques allow the insertion of specific genes into cell lines with specific point mutations. This allows us to measure the fitness of mutations and their combinations within specific regions of genotype space. Within cancer cells, once the stable expression of specific mutations of interest is established, growth or viability assays can be used to assess fitness.

Evolved resistance mirrors more closely the existing process within cancer patients and is useful in exploring larger regions of genotype space. Evolving resistance typically involves exposing cells to gradually increasing concentrations of a drug over time. In this way, the tumour cells studied undergo linear evolution via an adaptive walk, and the possible mutations to the genome are restricted only by increasing fitness. Sequencing the genomes present within the population over time allows experimental evolution to measure the trajectory of an evolving population upwards in a fitness landscape by measuring its genotypic coordinates over time. Although this method is less biased towards specific regions of the genome, measuring any low-fitness areas is unlikely under adaptive evolution. *In vivo* or *in vitro* evolutionary experiments in cancer are usually carried out using xenografts in mice or using cell lines, as illustrated in **Figure 35.3**A (adapted from [2]).

Combining cell engineering with measurements and sequencing from experimental evolution allows examination of the adaptive landscape while controlling for off-target mutations. Such studies have so far been carried out predominantly in bacterial populations. These successes include the complete landscape of mutations to genes in a malaria strain [78] and in *Escherichia coli*, the β-lactamase gene [29]. Although engineering stable gene mutations in cancer cell lines is more difficult than in bacteria, similar landscape construction methods are underway.

As tumours collect mutations, selection under treatment results in the expansion of drug-resistant clones, ultimately resulting in a drug-resistant tumour. For the clinical translation of fitness landscapes, there must be something actionable we can

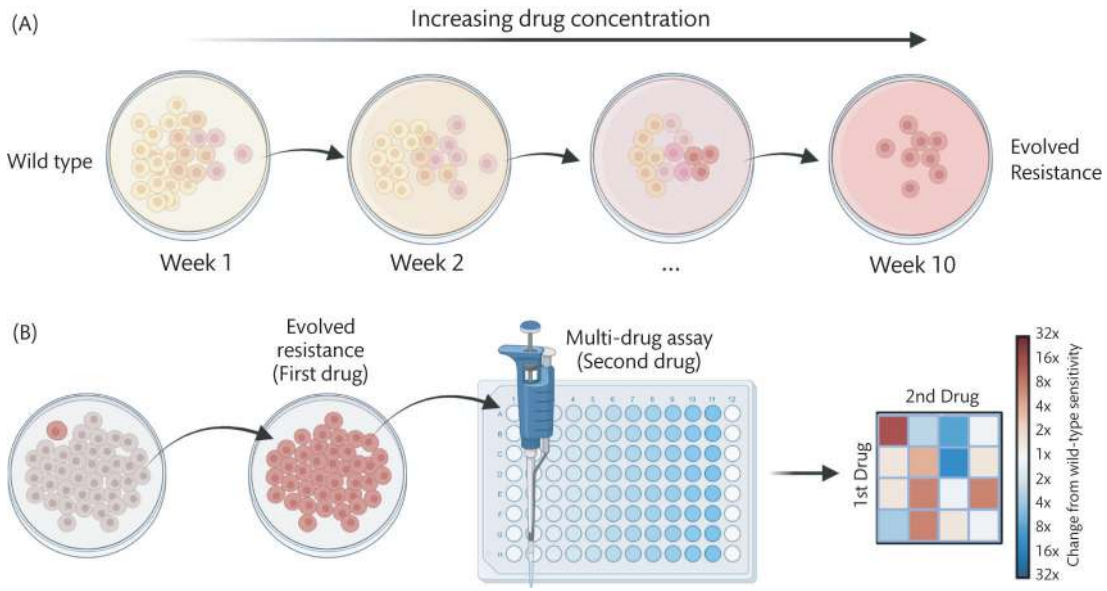


Figure 35.3. Experimental evolution of resistance informs us about the underlying fitness landscape. (A) An example experiment aimed at evolving resistance is shown. At each time point a population is plated on a dish of increasing drug concentration. Evolution is fundamentally a function of time and selection pressure. Evolutionary experiments implement selection pressures, and the expansion of clones within drug gradients represents restricted linear evolution. Experimental repeats and sequencing elucidate global epistatic effects and alternate resistance pathways. (B) In collateral sensitivity experiments, cells undergo evolutionary selection under a first drug. These resistant cells (red) undergo high-throughput drug assays used to measure their sensitivity to a selection of other drugs. By repeating this technique, initially evolving resistance to separate drugs within the entire drug panel, a collateral sensitivity map can be constructed. Collateral effects are revealed by comparing resistance in one drug to a second, with red in the heat map showing collateral (cross) resistance and blue showing collateral sensitivity.

gain from them. Having discussed the measurement of cancer fitness landscapes and experimental evolution, we examine multi-drug resistance and how fitness landscapes encode a potential clinical solution.

35.3.5. Collateral sensitivity

Collateral sensitivity occurs when the evolved mechanism of resistance to the therapeutic drug results in an increase in sensitivity to a different drug; experimental outline for this type of experiment is shown in Figure 35.3B. Recently, a number of studies have suggested exploiting the evolution of collateral sensitivity as a potential method to slow, or even reverse, the development of resistance [4,79,80]. In the fitness landscape paradigm, this is analogous to the population being driven towards a fitness maximum in the therapeutic fitness landscape, while that same evolutionary trajectory results in the population being driven towards a fitness minimum in the collaterally sensitive fitness landscape. The more anti-correlated the fitness landscapes are on this evolutionary trajectory, the more collaterally sensitive the population will be [81].

Unfortunately, collateral evolution can also result in *decreased* sensitivity, typically referred to as collateral resistance, cross-resistance, or multi-drug resistance. Further complicating the picture is the recent observation that these collateral effects are often transient, and ever-changing with continued therapy and evolution, giving only a small window to deliver effective treatments [4,82]. While these observations greatly increase the complexity of a potential optimal treatment, they also offer hope that properly judicious and thoughtful therapies may significantly prolong treatment efficacy or reverse resistance entirely.

35.3.6. Steering evolution

Modern treatment approaches involve adapting to the changing genetic characteristics of a tumour over time [83]. The goal of adaptive therapy is to optimize treatment in real time to maintain a sensitive tumour volume and extend survival. This treatment is based on the specific genetic, molecular, and phenotypic characteristics of the tumour rather than using a one-size-fits-all aggressive treatment plan.

There are several ways that adaptive therapy is currently envisaged and implemented, depending on the specific goals and objectives of the treatment. For example, adaptive therapy can involve:

- Dynamic treatment scheduling: This involves adjusting the timing, frequency, or intensity of treatment based on the response of the tumour to treatment. For example, if a tumour is responding well to treatment, the treatment may be continued or intensified. If the tumour is not responding, the treatment may be changed or stopped.
- Targeted therapy: The use of drugs or other therapies that are specifically designed to target the genetic changes that are driving the development and progression of the tumour.
- Combination therapy: Combinations of multiple drugs or therapies are used to target different aspects of the cancer cells or the tumour microenvironment at once.

Clinical trials of adaptive therapy and many *in silico* models have demonstrated strategies aimed at reducing and controlling tumour burden through the careful application of therapies aimed at reducing population size and thus resistance evolution [84,85].

Evolutionary therapy aims to leverage ideas from adaptive therapy along with fitness landscapes and growth dynamics to plan protocols directly aimed at preventing resistance or guiding a population to a collaterally sensitive region [4,86]. Collaterally sensitive genotypes allow evolution to be leveraged for patient benefit. Work like that of Yoon et al. specifically looks at scheduling to achieve maximum therapeutic results from collateral sensitivity [44]. In bacteria, experimental and theoretical combined work has also demonstrated the ability to steer populations both towards and away from resistance [27]. The use of therapies aimed at predicting, anticipating, and guiding tumour composition falls under the scope of evolutionary control. The fitness landscape formalism also allows for methods of steering evolution borrowed from more complex control theory [87].

Populations can be considered to traverse evolutionary spaces between genomic states in a probabilistic manner, and this interpretation has inspired applications derived from quantum control theory. The work of Kimura and others has previously shown how the evolution of allele frequencies in genetic populations can be described using equations traditionally used for describing fluids [88]. These equations are analogous in some ways to the Schrodinger equations describing quantum particles. Iram et al. built upon this to show that populations could be controlled *in silico* using parallels derived from quantum counter-diabatic control theory [89]. Control of tumour populations is ideally both rapid and accurate, in the same way that counter-diabatic driving moves quantum particles between states.

As understandings of both tumour populations and the fitness landscape develop mathematically, it is likely that more cross-disciplinary methods to decode and control complex cancer evolution will become available.

35.4. Future directions

The fitness landscape can easily be integrated with modern techniques or experiments in the pursuit of further understanding and control over clinical cancer evolution. In this section, we discuss some of the more recent developments aimed at integrating fitness landscapes into more realistic or complex models of patients.

35.4.1. Dose-dependent fitness landscapes

What is the translational potential of fitness landscapes? How might we leverage fitness landscapes to predict or control the evolution of cancer within a patient undergoing therapy? To address these questions, it is important to consider pharmacokinetic and pharmacodynamic effects. Importantly, the drug concentration experienced by a tumour *in vivo* will vary in time and space due to variables, such as drug diffusion, metabolism, and drug dosing schedules. This raises another important question: how does changing drug dose or the tumour environment shape the adaptive landscape?

Dose-dependent evolutionary trade-offs, or fitness costs to resistance, often arise as a consequence of resistance mechanisms [84,90–92]. Drug resistance mechanisms often impose metabolic burdens or impair vital functions of the organism. In other words, a mutant may 'trade' its drug-free growth rate in exchange for drug resistance. For instance, in an experiment with laboratory-evolved non-small-cell lung cancer subject to a targeted inhibitor, drug-resistant cells incurred a growth-rate penalty relative to drug-sensitive cells [93].

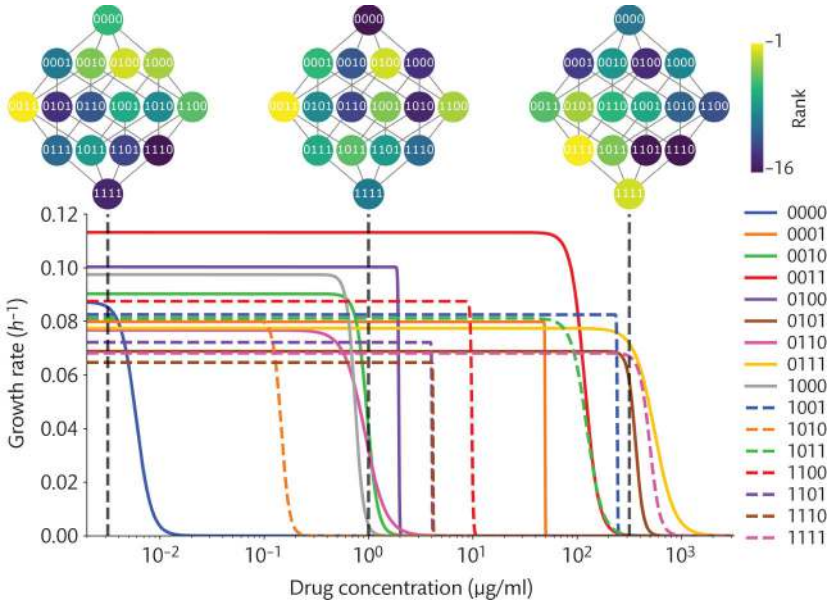


Figure 35.4. Seascapes as dose–response curves model evolutionary trade-offs. Dose–response curves with respect to drug concentration are shown for all 16 possible genotypes in a four-allele system. 0000 corresponds to the wild-type and 1111 to the presence of four mutant alleles. The wild type has a sharp reduction in growth rate at low drug concentrations. Traditional fitness landscapes coloured by rank fitness are vertical cross-sections of this seascape and are shown above the drug–response curves. These individual landscapes result from relative growth rate at three different drug concentrations. *Source:* Adapted from King et al. [97].

When considering trade-offs in growth rate, it is clear that the fitness landscape is context (or dose) dependent—the relative fitness of different mutants will change as the drug-selective pressure changes. To model these dynamics, fitness seascapes extend the fitness landscape model to include dose-dependent effects [12,94–96]. While others have previously used the term fitness seascapes to describe a time-varying fitness landscape, here we use fitness seascape to refer to a fitness landscape that changes due to any arbitrary variable (or variables), such as drug concentration, pH, or temperature.

A straightforward way of modelling fitness seascapes is with collections of genotype-specific dose–response curves; canonical fitness landscapes, as described earlier in this chapter, are then vertical cross-sections of this structure. This is illustrated in Figure 35.4 with data from bacteria (adapted from King et al [97]). Here, genotypes are represented by a string of 4 bits, with each bit representing the presence or absence of a specific drug-resistance-conferring mutation. Each genotype is associated with a dose–response curve, and this collection of dose–response curves forms a fitness seascape, encoding fitness trade-offs. For instance, in Figure 35.4, the four allele genotypes encoded by 0111 and 1111 (most mutated), which exhibit a high level of drug resistance, incur a growth-rate cost relative to other less resistant genotypes, including 0011 and 0000 (the wild type). Fitness landscapes, which are represented as the rank order of the growth rate at a given drug concentration, are then vertical cross-sections of this structure. We can use these cross-sections to visualize the stark changes in the fitness rank of different genotypes as a function of drug concentration.

Although the data within Figure 35.4 comes from an experiment in bacteria, dose–response curves in cancer often exhibit a similar pattern of dose-dependent trade-offs, resulting in fitness landscapes that vary as a function of drug concentration. Future work will involve characterizing more complete fitness seascapes in cancer and modelling multiple genotypes that may arise during the evolution of a tumour. Owing to these dose-dependent effects, it will be important to understand the pharmacokinetic dynamics within a tumour environment to gain a more realistic model of evolution. Furthermore, it is well known that tumour populations modify the microenvironment, affecting variables, such as pH and oxygen gradients [98]. These variables likely alter the adaptive landscape in a dose-dependent manner, resulting in distinct fitness seascapes. Understanding fitness landscapes as high-dimensional genotype-by-environment interactions will be important to predict, and ultimately control the evolution of drug resistance in cancer.

35.4.2. Landscapes and machine learning

Machine-learning techniques, with their ability to solve complex classification and prediction problems, are an important tool in understanding increasingly large biological datasets. Currently, deep neural net classifiers have been trained to predict the fitness of different genotypes based on their genetic characteristics. Using classifiers to predict drug response based on cell-intrinsic factors such as gene expression enables scientists to abstractly model the underlying fitness landscape [77,99,100].

Machine-learning approaches can also address the problem of immeasurably large landscapes via the imputation of missing fitness values [101,102]. Dimension reduction techniques provide possible solutions to address the high dimensionality of complete fitness landscapes. As demonstrated across bacterial and cancer studies, background mutations and ecological effects can also influence population distributions [103]. In these cases, higher dimensional landscapes become essential to model [104]. Certainly, the integration of cutting-edge image classification and convolutional and recursive neural network machine-learning techniques across pathological and experimental studies of evolution will facilitate

novel approaches to mapping and interpreting fitness landscapes and controlling tumour evolution.

35.4.3. Beyond genetic fitness landscapes

Although many existing cancer landscape studies focus on the presence and control of genetic mutations, additional layers of biological complexity and fluctuations require higher dimensional mappings and modified control theorems. Cancer's ability to modify its fitness landscape through intrinsic properties and environmental factors drives a desire to apply adaptive landscape principles across these dimensions. For example, it is clear that some cancers enormously benefit from structural changes and copy number alterations [105]. Epigenetic modifications also allow for cancer fitness plasticity and more evolvability on a landscape, potentially allowing the epigenetic landscape to combine with the genetic one permitting escape from a suboptimal local maximum [106].

Fitness landscape theory also has similar potential in understanding the evolution of metastasis, a critical factor in patient outcomes. In a metastatic landscape, fitness could be understood as the ability of a cell to metastasize or to establish itself at a different site. Work by Jolly et al. demonstrates that plasticity in the form of the epithelial–mesenchymal transition allows cancer cells to escape local maxima, priming cells for metastasis [107]. Metastatic potential remains an understudied measure of cell fitness and presents an axis to control in future studies. Although increasing the dimensions that require control, these wider studies provide an explanation for the evolutionary divergence between experiments and patients, and incorporating these dimensions brings us closer to accurately predicting evolution.

The ideas developed in fitness landscape theory translate into the predictability of other facets of tumour-relevant evolution, such as an understanding of the dynamics of immune adaptation [66,108], neo-antigens [109], or metabolism [110]. Successful long-term approaches will involve the modelling and control of the multiple interacting dimensions of these complex systems. The combination of wider landscape ideas with game theory models will develop more complex evolutionary models for an investigation into the co-evolution of different populations.

35.5. Conclusions

The allure of predictability in biology and the clinical necessity of evolutionary and ecological control in the face of cancer resistance will continue to drive the field [87,94]. Although the fitness landscape is sometimes understood as simply a useful metaphor, scientists and mathematicians across disciplines have proven it to be more powerful than this [111]. As both a measurable and a mathematical object, it allows for predictive simulations of evolution and provocative theoretical questions about evolution itself. As an effective framework that can help explain the observations of convergent and repeatable evolution, the fitness landscape provides a powerful theoretical basis for questions about the nature of evolution. The fitness landscape presents a figurative, analysable, and measurable map of evolution, and the expansion of fitness landscape ideas can help to explain how cancer evolves and evades treatment via the traversal and modification of its fitness surface.

Within cancer, the accelerated evolution of hyper-mutating immortal cells results in evolutionary dynamics that promote an accelerated exploration of the fitness landscape. In this regime of accelerated evolution, the existence of an underlying landscape has aided the inference of drug sensitivity and forms the basis for new therapeutic approaches and intelligently derived protocols aimed at steering evolution. Determined exploration of these landscapes across different tumour types and drug combinations will permit novel control methods aiming to target tumours using collaterally sensitive drug combinations.

Whether we choose to study cancer from an evolutionary perspective or not, evolutionary forces shape patient outcomes. The lens of fitness landscapes looks only to aid this perspective. The majority of the limitations and criticisms of landscape formalisms lie in the simplicity of some approaches and assumptions around the limits of their utility. These limitations are challenged when the fitness landscape is extended to multiple dimensions, and many approaches are as yet unexplored. For example, when models require the inclusion of background mutations or alternate evolutionary mechanisms, fitness landscapes can provide a starting point and context for these discussions. Furthermore, the fitness landscape framework is not limited to purely genetic evolution or evolution via single mutational steps. As a result, fitness landscapes permit many possibilities, including integrating various stochastic tools to model the probabilistic nature of evolution when background noise or epistatic effects are at play.

With increasing evidence for epistatic interactions with background mutations [104], there will be a need to lean on the theoretical tools within stochastic mathematics and physics to model non-linear surfaces and diverse populations as probability distributions [89]. Evolutionary landscape models encapsulating alternate mechanisms of evolution or this stochasticity will likely produce models more effective at explaining the scope of cancer outcomes and the possibility of control. As an understanding of tumour-specific evolution grows, so does the understanding of the tumour in its environment as a whole. The study and popularity of tumour microenvironments, through ecological interactions and environmental constraints, has highlighted the need for models of tumour evolution that incorporates the microenvironment. The fitness landscape provides a basic graph-like framework for analysing individual and community evolution, providing a mathematical and conceptual basis for an interesting exploration of cancer evolution and treatment. Fitness landscapes provide a broad enough framework within which the essential aspects of cancer ecology and community dynamics can eventually be integrated. The evolutionary understanding of the cancer landscape's multiple dimensions can be combined with treatment information and control theories to form predictive and clinically relevant approaches to cancer therapy.

REFERENCES

1. Marusyk A, Janiszewska M, Polyak K. Intratumor heterogeneity: The Rosetta Stone of therapy resistance. Cancer Cell. 2020;**37**:471–84.
2. Vander Velde R, Yoon N, Marusyk V, Durmaz A, Dhawan A, Miroshnychenko D, et al. Resistance to targeted therapies as a multifactorial, gradual adaptation to inhibitor specific selective pressures. Nat Commun. 2020;**11**:1–13.

3. Mansoori B, Mohammadi A, Davudian S, Shirjang S, Baradaran B. The different mechanisms of cancer drug resistance: a brief review. Adv Pharm Bull. 2017;**7**:339.
4. Scarborough JA, McClure E, Anderson P, Dhawan A, Durmaz A, Lessnick S, et al. Identifying states of collateral sensitivity during the evolution of therapeutic resistance in Ewing's sarcoma. iScience. 2020;**23**:101293.
5. Lin KH, Rutter JC, Xie A, Pardieu B, Winn ET, Bello RD, et al. Using antagonistic pleiotropy to design a chemotherapy-induced evolutionary trap to target drug resistance in cancer. Nat Genet. 2020;**52**:408–17.
6. Strobl MA, West J, Viossat Y, Damaghi M, Robertson-Tessi M, Brown JS, et al. Turnover modulates the need for a cost of resistance in adaptive therapy. Cancer Res. 2021;**81**:1135–47.
7. West J, You L, Zhang J, Gatenby RA, Brown JS, Newton PK, et al. Towards multidrug adaptive therapy. Cancer Res. 2020;**80**:1578–89.
8. Kim E, Brown JS, Eroglu Z, Anderson AR. Adaptive therapy for metastatic melanoma: predictions from patient calibrated mathematical models. Cancers. 2021;**13**:823.
9. Gabbutt C, Graham TA. Evolution's cartographer: mapping the fitness landscape in cancer. Cancer Cell. 2021;**39**:1311–3.
10. Wright S. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. Proc Sixth Int Congr Genet. 1932;**1**:355–66.
11. Pigliucci M. Sewall Wright's adaptive landscapes: 1932 vs. 1988. Biol Philos. 2008;**23**:591–603.
12. Mustonen V, Lässig M. From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. Trends Genet. 2009;**25**:111–9.
13. McCandlish DM. Visualizing fitness landscapes. Evolution. 2011;**65**:1544–58.
14. Pigliucci M. Landscapes, surfaces, and morphospaces: what are they good for. In: Svensson E, Calsbeek R, editors. The adaptive landscape in evolutionary biology [Internet]. Oxford University Press; 2013.
15. Srivastava M, Payne JL. On the incongruence of genotype-phenotype and fitness landscapes. PLoS Comput Biol. 2022;**18**:e1010524.
16. Van Nimwegen E, Crutchfield JP, Huynen M. Neutral evolution of mutational robustness. Proc Natl Acad Sci. 1999;**96**:9716–20.
17. Pitt JN, Ferré-D'Amaré AR. Rapid construction of empirical RNA fitness landscapes. Science 2010;**330**:376–9.
18. Salverda ML, Dellus E, Gorter FA, Debets AJ, Van Der Oost J, Hoekstra RF, et al. Initial mutations direct alternative pathways of protein evolution. PLoS Genet. 2011;**7**:e1001321.
19. Palmer AC, Toprak E, Baym M, Kim S, Veres A, Bershtein S, et al. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. Nat Commun. 2015;**6**:1–8.
20. Kaznatcheev A. Computational complexity as an ultimate constraint on evolution. Genetics. 2019;**212**:245–65.
21. Wright S. Evolution in Mendelian populations. Genetics. 1931;**16**:97.
22. Kauffman SA. At home in the universe: the search for laws of self-organization and complexity. USA: Oxford University Press; 1995.
23. Kauffman SA, Weinberger ED. The NK model of rugged fitness landscapes and its application to maturation of the immune response. J Theor Biol. 1989;**141**:211–45.
24. Dingle K, Ghaddar F, Šulc P, Louis AA. Phenotype bias determines how natural RNA structures occupy the morphospace of all possible shapes. Mol Biol Evol. 2022;**39**:msab280.
25. Bonhoeffer S, McCaskill JS, Stadler PF, Schuster P. RNA multi-structure landscapes. Eur Biophys J. 1993;**22**:13–24.
26. Kouyos RD, Leventhal GE, Hinkley T, Haddad M, Whitcomb JM, Petropoulos CJ, et al. Exploring the complexity of the HIV-1 fitness landscape. PLoS Genet. 2012;**8**:e1002551.
27. Nichol D, Jeavons P, Fletcher AG, Bonomo RA, Maini PK, Paul JL, et al. Steering evolution with sequential therapy to prevent the emergence of bacterial antibiotic resistance. PLoS Comput Biol. 2015;**11**:e1004493.
28. Das SG, Direito SO, Waclaw B, Allen RJ, Krug J. Predictable properties of fitness landscapes induced by adaptational tradeoffs. eLife. 2020;**9**:e55155.
29. Firnberg E, Labonte JW, Gray JJ, Ostermeier M. A comprehensive, high-resolution map of a gene's fitness landscape. Mol Biol Evol. 2014;**31**:1581–92.
30. Chen J, Fowler D, Tokuriki N. Environmental selection and epistasis in an empirical phenotype–environment–fitness landscape. Nat Ecol Evol. 2022;**6**:427–38.
31. Weinreich DM, Watson RA, Chao L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. Evolution. 2005;**59**:1165–74.
32. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB. Should evolutionary geneticists worry about higher-order epistasis? Curr Opin Genet Dev. 2013;**23**:700–7.
33. Blanquart F, Bataillon T. Epistasis and the structure of fitness landscapes: are experimental fitness landscapes compatible with Fisher's geometric model? Genetics. 2016;**203**:847–62.
34. Weinreich DM, Lan Y, Jaffe J, Heckendorn RB. The influence of higher-order epistasis on biological fitness landscape topography. J Stat Phys. 2018;**172**:208–25.
35. Dasari K, Somarelli JA, Kumar S, Townsend JP. The somatic molecular evolution of cancer: Mutation, selection, and epistasis. Prog. Biophys. Mol. Biol. 2021;**165**:56–65.
36. Sailer ZR, Harms MJ. High-order epistasis shapes evolutionary trajectories. PLoS Comput Biol. 2017;**13**:e1005541.
37. Poelwijk FJ, Kiviet DJ, Weinreich DM, Tans SJ. Empirical fitness landscapes reveal accessible evolutionary paths. Nature. 2007;**445**:383–6.
38. Anderson AR, Chaplain MA, Newman EL, Steele RJ, Thompson AM. Mathematical modelling of tumour invasion and metastasis. Comput. Math. Methods Med. 2000;**2**:129–54.
39. Gillespie JH. A simple stochastic gene substitution model. Theor Popul Biol. 1983;**23**:202–15.
40. Sun X, Bao J, Shao Y. Mathematical modeling of therapy-induced cancer drug resistance: connecting cancer mechanisms to population survival rates. Sci Rep. 2016;**6**:1–12.
41. Butner JD, Wang Z, Elganainy D, Al Feghali KA, Plodinec M, Calin GA, et al. A mathematical model for the quantification of a patient's sensitivity to checkpoint inhibitors and long-term tumour burden. Nat Biomed. Eng. 2021;**5**:297–308.
42. Moolgavkar SH, Knudson AG. Mutation and cancer: a model for human carcinogenesis. JNCI: J Natl Cancer Inst. 1981;**66**:1037–52.
43. Beerenwinkel N, Antal T, Dingli D, Traulsen A, Kinzler KW, Velculescu VE, et al. Genetic progression and the waiting time to cancer. PLoS Comput. Biol. 2007;**3**:e225.
44. Yoon N, Vander Velde R, Marusyk A, Scott JG. Optimal therapy scheduling based on a pair of collaterally sensitive drugs. Bull Math Biol. 2018;**80**:1776–809.
45. West J, Robertson-Tessi M, Anderson AR. Agent-based methods facilitate integrative science in cancer. Trends Cell Biol. 2022 Apr;**33**(4):300–11.

46. Gerlee P, Anderson AR. An evolutionary hybrid cellular automaton model of solid tumour growth. J Theor Biol. 2007;**246**:583–603.
47. Thomas DS, Cisneros LH, Anderson AR, Maley CC. In silico investigations of multi-drug adaptive therapy protocols. Cancers. 2022;**14**:2699.
48. Ghaffarizadeh A, Heiland R, Friedman SH, Mumenthaler SM, Macklin P. PhysiCell: an open source physics-based cell simulator for 3-D multicellular systems. PLoS Comput Biol. 2018;**14**:e1005991.
49. Bravo RR, Baratchart E, West J, Schenck RO, Miller AK, Gallaher J, et al. Hybrid automata library: a flexible platform for hybrid modeling with real-time visualization. PLoS Comput Biol. 2020;**16**:e1007635.
50. Ferretti L, Schmiegelt B, Weinreich D, Yamauchi A, Kobayashi Y, Tajima F, et al. Measuring epistasis in fitness landscapes: the correlation of fitness effects of mutations. J Theor Biol. 2016;**396**:132–43.
51. Flyvbjerg H, Lautrup B. Evolution in a rugged fitness landscape. Phys Rev A 1992;**46**:6714.
52. Hegarty P, Martinsson A. On the existence of accessible paths in various models of fitness landscapes. Ann Appl Probab. 2014;**24**:1375–95.
53. Greenbury SF, Schaper S, Ahnert SE, Louis AA. Genetic correlations greatly increase mutational robustness and can both reduce and enhance evolvability. PLoS Comput Biol. 2016;**12**:e1004773.
54. Greenbury SF, Louis AA, Ahnert SE. The structure of genotype-phenotype maps makes fitness landscapes navigable. Nat Ecol Evol. 2022;**6**:1742–52.
55. Gavrilets S. Evolution and speciation on holey adaptive landscapes. Trends Ecol Evol. 1997;**12**:307–12.
56. Fragata I, Blanckaert A, Louro MAD, Liberles DA, Bank C. Evolution in the light of fitness landscape theory. Trends Ecol Evol. 2019;**34**:69–82.
57. Aguilar-Rodrıguez J, Payne JL, Wagner A. A thousand empirical adaptive landscapes and their navigability. Nat Ecol Evol. 2017;**1**:1–9.
58. Schaper S, Louis AA. The arrival of the frequent: how bias in genotype-phenotype maps can steer populations to local optima. PLoS ONE. 2014;**9**:e86635.
59. Dingle K, Schaper S, Louis AA. The structure of the genotype–phenotype map strongly constrains the evolution of non-coding RNA. Interf Focus. 2015;**5**:20150053.
60. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;**100**:57–70.
61. Sancar A, Lindsey-Boltz LA, Ünsal-Kaçmaz K, Linn S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu Rev Biochem. 2004;**73**:39–85.
62. McFarland CD, Mirny LA, Korolev KS. Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. Proc Natl Acad Sci. 2014;**111**:15138–43.
63. Diaz-Uriarte R. Cancer progression models and fitness landscapes: a many-to-many relationship. Bioinformatics. 2018;**34**:836–44.
64. Anderson AR, Weaver AM, Cummings PT, Quaranta V. Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. Cell. 2006;**127**:905–15.
65. Kim SK and Cho SW. The evasion mechanisms of cancer immunity and drug intervention in the tumor microenvironment. Front Pharmacol. 2022 May;**13**:868695.
66. Gourmet L, Secrier M, Sottoriva A, Zapata L. Immune evasion impacts the selective landscape of driver genes during tumorigenesis. bioRxiv 2022.
67. Martinez-Outschoorn UE, Balliet RM, Rivadeneira D, Chiavarina B, Pavlides S, Wang C, et al. Oxidative stress in cancer associated fibroblasts drives tumor-stroma co-evolution: a new paradigm for understanding tumor metabolism, the field effect and genomic instability in cancer cells. Cell Cycle. 2010;**9**:3276–96.
68. Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science. 2014;**344**:1519–22.
69. Hosseini SR, Diaz-Uriarte R, Markowetz F, Beerenwinkel N. Estimating the predictability of cancer evolution. Bioinformatics. 2019;**35**:i389–97.
70. Mendiratta G, Ke E, Aziz M, Liarakos D, Tong M, Stites EC. Cancer gene mutation frequencies for the US population. Nat Commun. 2021;**12**:1–11.
71. Wang X, Fu AQ, McNerney ME, White KP. Widespread genetic epistasis among cancer genes. Nat Commun. 2014;**5**:1–10.
72. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, et al. The evolutionary history of 2,658 cancers. Nature. 2020;**578**:122–8.
73. Davis A, Gao R, Navin N. Tumor evolution: linear, branching, neutral or punctuated? Biochim. Biophys Acta (BBA)–Rev Cancer. 2017;**1867**:151–61.
74. Graham TA, Sottoriva A. Measuring cancer evolution from the genome. J Pathol. 2017;**241**:183–91.
75. Gopal P, Sarihan EI, Chie EK, Kuzmishin G, Doken S, Pennell NA, et al. Clonal selection confers distinct evolutionary trajectories in BRAF-driven cancers. Nat Commun. 2019;**10**:1–14.
76. Bolton KL, Ptashkin RN, Gao T, Braunstein L, Devlin SM, Kelly D, et al. Cancer therapy shapes the fitness landscape of clonal hematopoiesis. Nat Genet. 2020;**52**:1219–26.
77. Caravagna G, Heide T, Williams MJ, Zapata L, Nichol D, Chkhaidze K, et al. Subclonal reconstruction of tumors by using machine learning and population genetics. Nat Genet. 2020;**52**:898–907.
78. Jiang PP, Corbett-Detig RB, Hartl DL, Lozovsky ER. Accessible mutational trajectories for the evolution of pyrimethamine resistance in the malaria parasite Plasmodium vivax. J Mol Evol. 2013;**77**:81–91.
79. Maltas J, Wood KB. Pervasive and diverse collateral sensitivity profiles inform optimal strategies to limit antibiotic resistance. PLoS Biol. 2019;**17**:e3000515.
80. Maltas J, Krasnick B, Wood KB. Using selection by nonantibiotic stressors to sensitize bacteria to antibiotics. Mol Biol Evol. 2020;**37**:1394–406.
81. Maltas J, McNally DM, Wood KB. Evolution in alternating environments with tunable interlandscape correlations. Evolution. 2021;**75**:10–24.
82. Zhao B, Sedlak JC, Srinivas R, Creixell P, Pritchard JR, Tidor B, et al. Exploiting temporal collateral sensitivity in tumor clonal evolution. Cell. 2016;**165**:234–46.
83. Enriquez-Navas PM, Wojtkowiak JW, Gatenby RA. Application of evolutionary principles to cancer therapy. Cancer Res. 2015;**75**:4675–80.
84. Das Thakur M, Salangsang F, Landman AS, Sellers WR, Pryer NK, Levesque MP, et al. Modelling vemurafenib resistance in melanoma reveals a strategy to forestall drug resistance. Nature. 2013;**494**:251–5.

85. West JB, Dinh MN, Brown JS, Zhang J, Anderson AR, Gatenby RA. Multidrug cancer therapy in metastatic castrate-resistant prostate cancer: an evolution-based strategy. Clin Cancer Res. 2019;**25**:4413–21.
86. Acar A, Nichol D, Fernandez-Mateos J, Cresswell GD, Barozzi I, Hong SP, et al. Exploiting evolutionary steering to induce collateral drug sensitivity in cancer. Nat Commun. 2020;**11**:1–14.
87. Lässig M, Mustonen V. Eco-evolutionary control of pathogens. Proc Natl Acad Sci. 2020;**117**:19694–704.
88. Kimura M. On the probability of fixation of mutant genes in a population. Genetics. 1962;**47**:713.
89. Iram S, Dolson E, Chiel J, Pelesko J, Krishnan N, Güngör Ö, et al. Controlling the speed and trajectory of evolution with counterdiabatic driving. Nat Phys. 2021;**17**:135–42.
90. Jacqueline C, Biro PA, Beckmann C, Moller AP, Renaud F, Sorci G, et al. Cancer: a disease at the crossroads of trade-offs. Evol Appl. 2017;**10**:215–25.
91. Kaznatcheev A, Peacock J, Basanta D, Marusyk A, Scott JG. Fibroblasts and Alectinib switch the evolutionary games played by non-small cell lung cancer. Nat Ecol Evol. 2019;**3**:450–6.
92. Hausser J, Szekely P, Bar N, Zimmer A, Sheftel H, Caldas C, et al. Tumor diversity and the trade-off between universal cancer tasks. Nat Commun. 2019;**10**:5423.
93. Farrokhian N, Maltas J, Dinh M, Durmaz A, Ellsworth P, Hitomi M, et al. Measuring competitive exclusion in non–small cell lung cancer. Sci Adv. 2022;**8**:eabm7212.
94. Lässig M, Mustonen V, Walczak AM. Predicting evolution. Nat Ecol Evol. 2017;**1**:1–9.
95. Merrell DJ. The adaptive seascape: the mechanism of evolution. Minne ed. UK: University of Minnesota Press; 1994.
96. Agarwala A, Fisher DS. Adaptive walks on high-dimensional fitness landscapes and seascapes with distance-dependent statistics. Theor Popul Biol. 2019;**130**:13–49.
97. King ES, Pelesko J, Maltas J, Barker-Clarke R, Dolson E, and Scott JG. Fitness seascapes facilitate the prediction of therapy resistance under time-varying selection. bioRxiv. 2022.
98. Damaghi M, West J, Robertson-Tessi M, Xu L, Ferrall-Fairbanks MC, Stewart PA, et al. The harsh microenvironment in early breast cancer selects for a Warburg phenotype. Proc Natl Acad Sci. 2021;**118**:e2011342118.
99. Gerdes H, Casado P, Dokal A, Hijazi M, Akhtar N, Osuntola R, et al. Drug ranking using machine learning systematically predicts the efficacy of anti-cancer drugs. Nat Commun. 2021;**12**:1–15.
100. Chawla S, Rockstroh A, Lehman M, Ratther E, Jain A, Anand A, et al. Gene expression based inference of cancer drug sensitivity. Nat Commun. 2022;**13**:1–15.
101. Skums P, Tsyvina V, Zelikovsky A. Inference of clonal selection in cancer populations using single-cell sequencing data. Bioinformatics. 2019;**35**:i398–i407.
102. Fernandez-de-Cossio-Diaz J, Uguzzoni G, Pagnani A. Unsupervised inference of protein fitness landscape from deep mutational scan. Mol Biol Evol. 2021;**38**:318–28.
103. Card KJ, Thomas MD, Graves JL Jr, Barrick JE, Lenski RE. Genomic evolution of antibiotic resistance is contingent on genetic background following a long-term experiment with Escherichia coli. Proc Natl Acad Sci. 2021;**118**:e2016886118.
104. Ogbunugafor CB, Eppstein MJ. Genetic background modifies the topography of a fitness landscape, influencing the dynamics of adaptive evolution. IEEE Access. 2019;**7**:113675–83.
105. Martins FC, Couturier DL, De Santiago I, Sauer CM, Vias M, Angelova M, et al. Clonal somatic copy number altered driver events inform drug sensitivity in high-grade serous ovarian cancer. Nat Commun. 2022;**13**:1–14.
106. Huang S. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. Cancer Metastasis Rev. 2013;**32**:423–48.
107. Jolly MK, Mani SA, Levine H. Hybrid epithelial/mesenchymal phenotype(s): The 'fittest' for metastasis? Biochim. Biophys. Acta (BBA)–Rev Cancer. 2018;**1870**:151–7.
108. El-Kenawi A, Gatenbee C, Robertson-Tessi M, Bravo R, Dhillon J, Balagurunathan Y, et al. Acidity promotes tumour progression by altering macrophage phenotype in prostate cancer. Br J Cancer. 2019;**121**:556–66.
109. Lakatos E, Williams MJ, Schenck RO, Cross WC, Househam J, Zapata L, et al. Evolutionary dynamics of neoantigens in growing tumors. Nat Genet. 2020;**52**:1057–66.
110. Pinheiro F, Warsi O, Andersson DI, Lässig M. Metabolic fitness landscapes predict the evolution of antibiotic resistance. Nat Ecol Evol. 2021;**5**:677–87.
111. Hartl DL. What can we learn from fitness landscapes? Curr Opin Microbiol. 2014;**21**:51–7.

36

A case against causal reductionism in acquired therapy resistance

Andriy Marusyk

36.1. Introduction: how do tumours escape initially effective targeted therapies?

The reductionistic molecular biology studies identified many cancer-associated mutations and specific molecular mechanisms underlying malignant phenotypes. Many of these oncogenic mutations occur in cell signalling genes, resulting in constitutive context-independent signalling, enabling cells to ignore constraints to proliferation and survival. Cancer cells are often 'addicted' to this abnormal signalling, i.e. its suppression can halt cell proliferation and induce cell death, even though other genetic and epigenetic alterations associated with the oncogenic progression remain unaffected. This phenomenon paved the way for the development of highly specific pharmacological inhibitors that target different oncogenic addictions. Often, these targeted therapies induce strong and durable clinical responses while avoiding severe toxicities associated with less selective traditional cytotoxic therapies. Therefore, for cancers defined by the presence of strong 'druggable' oncogenic addictions (such as EML4–ALK fusion in a subset of lung cancers), targeted therapies became the preferred frontline therapeutic option. In turn, the success of many types of targeted therapies fuelled the development of precision oncology, i.e. the search for druggable oncogenic addictions in cancers that lack targeted therapy options. While precision oncology is getting overshadowed by immune therapies, it remains one of the main directions of cancer research.

Despite typically eliciting strong and durable responses, targeted therapies are not curative in advanced metastatic cancers. Some of the tumours with targetable mutations fail to respond (so-called intrinsic resistance). In tumours that do respond, regression eventually reaches a plateau, with tumour burden stabilized at a certain level. This stabilization, called minimal residual disease (MRD), reflects the ability of some of the tumour cells to resist elimination. Following the seminal paper that described the phenomenon in targetable EGFR mutant lung cancers, the residual cells are referred to as drug-tolerant persisters [1]. Stabilization of MRD burden under continuous treatment reflects a near-neutral proliferation/death balance in populations of persisters, with cell proliferation substantially suppressed by cytostatic effects of therapy. If the treatment is stopped, residual tumours typically resume fast growth. However, even under continuous treatment, the persistent populations evolve resistance over time, losing sensitivity to the cytostatic and cytotoxic effects of therapy. This leads to a transition from near-neutral to net-positive tumour growth relapse. While persistence and resistance are typically considered qualitatively distinct phenomena, a growing body of evidence suggests a substantial heterogeneity of persistent tumour cells and a lack of clear boundaries. Recent analyses of genetic and phenotypic progression from residual disease to relapse suggest a continuum of resistance, from barely surviving under therapy (categorized as persistence) to complete loss of drug sensitivity and everything in between [2,3].

36.2. How do we approach the challenges of persistence and resistance?

The success of targeted therapies paved the way for a near-consensus opinion that the solutions to the challenges of persistence and resistance could be found within the same precision oncology paradigm that produced the strong initial gains. That is, overcoming both the residual disease and relapse requires identifying specific molecular mechanisms responsible for persistence and resistance and developing specific inhibitors directed against these mechanisms. This paradigm led to the identification of a large spectrum of mechanisms, including both mutational (copy number and nucleotide-level changes) and epigenetic (expression-level changes without changes in the DNA) cell-intrinsic changes, as well as environmentally mediated resistance, where specific paracrine signals from local or systemic environment compensate for the suppression of the targeted addiction. These discoveries spurred the development of multiple pharmacological inhibitors of varying degrees of specificity and on-target activity and the launching of new clinical trials. However, these developments failed to meet the high expectations. The clinical introduction of more potent inhibitors, capable of negating common

on-target mutational resistance mechanisms (such as replacement of the first-generation ALK inhibitor crizotinib with second- and third-generation inhibitors alectinib and lorlatinib), enabled a substantial improvement in the rate and duration of remissions but still failed to resolve the inevitability of relapse.

New drugs directed against off-target resistance mechanisms with the purpose of being used in combination with the frontline inhibitor fared even worse. When not prohibited by increased systemic toxicities, targeting individual persistence/resistance mechanisms in combination therapy settings provided either no measurable effects or transient and modest improvement/recovery of tumour responses. At best, these new inhibitors enabled oncologists to play an ultimately doomed 'whack-a-mole' game, with shallower and less durable responses, until running out of options. Precision oncology strategies to deal with persistence failed to achieve cures as well. At best, adding a co-inhibitor that suppresses the ability of tumour cells to survive upon the shutdown of the main 'driver' decreases MRD size and prolongs remissions. An example of this relative success is the addition of MEK inhibitors to the backbone of BRAF-targeting therapies in melanomas, now being extended to other targetable contexts, such as ALK+ lung cancers. Some of the promising leads (such as adding cMET inhibitors to the backbone of EGFR inhibitors in EGFR mutant lung cancers) failed to show strong effects in clinical trials [4]. Very often, the ability to co-inhibit a different target over extended time frames is limited by systemic toxicities, which becomes especially relevant for the accumulation of side effects when added to the backbones of effective frontline inhibitors capable of achieving remissions that last for years (such as alectinib and osimertinib).

36.3. Why are we failing?

Why did the precision oncology paradigm in dealing with persistence/resistance fail to meet the high expectations? One of the most obvious challenges is the issue of inter-tumour variability. Each patient's cancer is unique, as it represents a unique 'evolutionary experiment' shaped by the random nature of mutational processes and a degree of stochasticity of clonal dynamics. Moreover, the effect of specific mutations can be modified by epistatic interactions with other alleles, mutational timing/order, life histories, sex, age, etc. [5]. Thus, cancers with the same main targetable mutational driver can have substantially different evolutionary trajectories, with no two tumours being identical in karyotypes, genetic mutations, and gene expression patterns. This variability can be even higher at relapse, as even near-identical experimental tumours can develop resistance through various molecular mechanisms.

The challenge of tumour-tumour variability is further exacerbated by the issue of mutational, phenotypic, and microenvironmental intra-tumour heterogeneity [6,7]. The existence of intra-tumour heterogeneity was noted well before the introduction of targeted therapies [8]. However, these observations were typically ignored in favour of a simplistic view of near-homogenous tumours arising through a series of clean clonal succession [9]. However, the frustrating inevitability of acquired resistance, as well as technical development enabling analyses at regional and even single-cell resolution, eventually forced the acceptance of the more complex reality. Now, tumours are commonly recognized as heterogeneous entities composed of phenotypically and genetically distinct neoplastic populations, undergoing continuous diversification through mutational processes and plasticity-mediated state transitions.

Intra-tumour heterogeneity extends to the differences in therapy sensitivity, with multiple persistent and resistant phenotypes coexisting and coevolving in the same tumours. The coexistence of different mutational resistance mechanisms has been described in multiple therapeutic contexts [7]. Successful targeting of any individual population leads to a competitive release of subpopulations with different resistance mechanisms. This mechanism could be underlying the 'whack-a-mole game', which is lost with the emergence of populations refractory to all of the available therapeutic options. In principle, intra-tumour heterogeneity should be less of an issue when dealing with persistent cells within the MRD, as smaller residual population sizes and more limited phenotypic diversity should, in principle, improve the odds of therapeutic success. However, despite identifying multiple persistence mechanisms, there have been no success stories of successfully eradicating the residual disease with targeted therapy combinations in advanced metastatic cancers.

36.4. The elephant in the room: multifactorial causation of resistance

Even though the issues of inter- and intra-tumour heterogeneity pose formidable challenges, one can still argue that the current framework for discovering and targeting individual persistence/resistance mechanisms is fundamentally sound. Given the continuous advances in resolution and precision of molecular diagnostics tools, it might soon be possible to detect subpopulations with different resistance mechanisms. The current limits of 'druggability' might be overcome by new technological developments, such as tagging the targeted proteins for proteasome degradation. Thus, it should be, in principle, possible to identify and eliminate all of the resistant subpopulations one by one (**Figure 36.1**A). The progress might be slow and incremental, but one could expect to eventually reach the goal of eradicating therapy-refractory cancers without redrawing current strategies for understanding and treating cancers.

The prospects of eventual success within the current paradigms of the war on cancer are tarnished, however, by another issue that is currently residing in a blind zone of the mainstream efforts. The current strategies of focusing on one (co)target at a time are based on an implicit assumption that, at the level of individual tumour cells, the phenomena of persistence, as well as intrinsic and acquired resistance, can be reduced to a single or at least a dominant cause. If this assumption is correct, disruption of this mechanism should lead to the inevitable death of the cell, at least as long as the mechanism was correctly identified and the cell has not changed during the treatment. If persistence/resistance is multifactorial, i.e. if more than one mechanism is responsible for the viability of persistent cells, or the ability of resistant cells to grow in the face of the initially effective drug, effective disruption of any single mechanism could only have a partial effect or no effect at all (**Figures 36.1B and Figures C and 36.2**). The issues of intra-tumour heterogeneity and the ability of neoplastic populations to create new variability in persistence/resistance mechanisms would further exacerbate the issue.

Despite its critical importance for the soundness of current precision oncology efforts, the assumption of a single-cause resistance

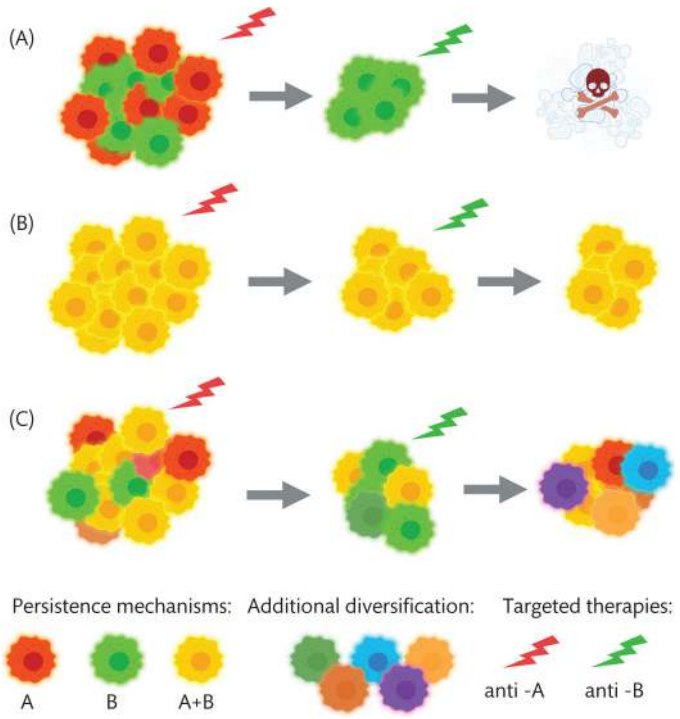


Figure 36.1. Therapeutic implications of multifactorial resistance. (A) Single-cause resistance: heterogeneous residual tumours. Tumours can be eliminated by sequentially targeting different populations. Multifactorial resistance: inhomogeneous (B) and heterogeneous (C) tumours. Sequential targeting of different resistance mechanisms cannot eliminate tumours. Additional diversification further complicates the challenge.

is rarely made explicit except in mathematical modelling studies [10,11]. Instead, it is implicit within the rationale for the search for the 'critical regulators' or 'drivers' of resistance phenotype. The terminology is sufficiently vague to leave room for multiple interpretations, thus avoiding the need to discuss or defend the subject. The single-cause assumption is also relevant for those cases when resistance is attributed to 'cancer stem cells'[12], 'stemness'[13], or epithelial to mesenchymal transition (EMT) [14]. These terms refer to cell phenotypes characterized by enhanced phenotypic plasticity, i.e. the ability to adaptively rewire and rearrange gene expression, signalling and metabolic networks in response to external stimuli. In principle, consideration of plasticity-mediated cell state transitions should imply complex, multifactorial expression changes affecting many genes and multifactorial causation. In practice, the link between stemness/EMT and resistance is attributed to enhanced expression of multidrug resistance genes, regardless of whether the drug in question serves as their substrate. Alternatively, stemness or EMT is described as proximal, specific resistance mechanisms analogous to mutational changes, essentially explaining away the resistance phenomenon and closing the door for deeper interrogation of the phenomenon.

36.5. What are the roots of the single-cause assumption?

The assumption of single-cause resistance appears to be a specific case of the more general issue described as causal reductionism or single-cause fallacy [15,16]. This fallacy is a well-described issue that impacts many areas of scientific inquiry where standard practices do not involve the need to adhere to robust quantitative reasoning standards. The prevalence of this fallacy in life sciences is quite understandable, given the dominance and success of reductionistic paradigms. The molecular biology revolution has provided the concepts, knowledge, and tools that enabled the identification of specific molecular mechanisms of proximal causation. The mechanistic understanding enabled numerous breakthroughs across various medical conditions, from diabetes to infectious diseases. The commercial success of tyrosine kinase inhibitors such as Gleevec led to the rush to develop new targeted therapies within the pharmacological industry, which in turn created demand and provided support for the search for targetable oncogenic addictions across all types of cancers. The clinical and commercial success of targetable therapies bolstered the case for the sufficiency of reductionistic approaches to the point where most research studies and grant submissions could be described by the 'Gene X is a critical regulator of hallmark Y in cancer Z' formula.

Is there solid empirical evidence for single-mechanism causation of resistance? To address this question, let us consider the criteria for defining a given gene or pathway as the cause of the resistance phenotype. In a typical target discovery pipeline, a candidate resistance mechanism is identified through studies that compare therapy sensitive/naive with therapy-resistant primary tumours or experimental models. Intermediate steps are omitted from considerations due to convenience or lack of access to relevant samples, such as the challenge of sampling MRD in clinics. The comparisons could involve searching for recurrent genetic mutations, comparing

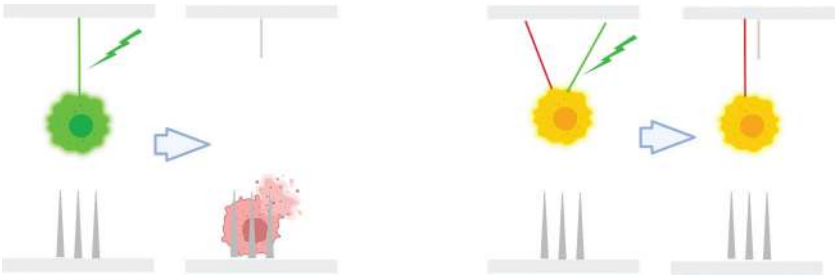


Figure 36.2. Consequences of single-cause fallacy. A tumour cell whose survival relies on a single molecular mechanism (depicted by a tethering line) can be eliminated by targeting this mechanism. However, if more than one mechanism underlies cell viability, targeting individual resistance mechanism is expected to have limited impact on cell viability.

gene expression through transcriptomics or proteomics analyses, or functional screens (genetic or pharmacological). These analyses/screens typically produce a list of candidate mechanisms. The list is then examined to find a few candidates that satisfy novelty, clinical relevance, and targeting potential criteria. Once the short list of potential candidates is compiled, functional validation studies are performed by assessing the impact of suppression and activation of the candidate mechanism. A candidate that passes the functional validation criteria is typically selected to become a central player in a paper reporting a novel mechanism. If multiple candidates pass the criteria, they often become seeds of multiple papers, each reporting a new 'critical regulator'.

A reasonable validation entails the demonstration that the putative mechanism is both *necessary* and *sufficient* to explain resistance. However, a few studies strictly adhere to this criterion. A typical validation pipeline entails demonstrating the statistical significance and reproducibility of the effects of the suppression and activation of a putative mechanism. As long as suppression of the putative mechanism enhances sensitivity while its activation reduces it, the boxes are checked, especially if this demonstration is performed in several experimental systems and with different assays. Whether the *magnitude* of the observed effect is sufficient to explain the clinically observed resistance levels is usually overlooked. To an extent, this omission reflects the limitations of available experimental systems. In many cases, the assays are performed in experimental models that are not directly relevant to the cell/cancer type being studied. For example, the impact of specific hotspot mutation on sensitivity to TKIs is commonly studied in an IL3-dependent Ba/F3 cell line [17]. Exogenous expression of kinases enables Ba/F3 cells to grow without IL3, while pharmacological inhibition of the kinases restores IL3 dependence. While the Ba/F3 system provides a highly convenient platform for experimental and screening studies, the genetic and epigenetic context of this murine Pro-B cell line is highly distinct from the contexts of cells that are clinically relevant targets. However, even when more relevant *in vitro* and *in vivo* experimental systems are available, the question of the sufficiency of the putative resistance mechanism to fully account for the ability of tumour cells to maintain net-positive growth in the presence of treatment is rarely asked.

Thus, considering therapy resistance as a single-cause phenomenon is not based on solid first principles or robust empirical evidence. But do we have any evidence to support the alternative scenario of the multifactorial nature of therapy persistence/resistance? The coexistence of multiple resistance mechanisms within the same tumour in clinical relapses is a well-documented phenomenon [18–20]. Since such a coexistence can reflect intra-tumour heterogeneity, where distinct resistant subpopulations contain distinct resistance mechanisms, this type of evidence is only circumstantial without further interrogation. However, in some cases, the circumstantial evidence is sufficiently strong to speculate that multiple cell-intrinsic resistance mechanisms might act within the same neoplastic cells [21]. Moreover, ample evidence exists for the existence of compound on-target mutations in the form of multiple nucleotide-level changes in the same target or a point mutation plus amplification of the DNA locus that is most likely acquired in multiple mutational steps. A growing body of direct experimental evidence documents the contribution of multiple specific expression-level changes in mediating reduced drug sensitivity, mediated by epigenetic mechanisms [2,22,23]. Notably, genetic and epigenetic mechanisms are not mutually exclusive and can co-occur in the same cells as they evolve resistance [23]; drug sensitivities of evolving lineages are further modulated by systemic and microenvironmental influences. Conceptually, scenarios where resistance can be fully attributed to a single mechanism can be viewed as a special, limiting case of multifactorial resistance. The opposite, however, is not true. Much like an observation of a single black swan is sufficient to refute the 'all swans are white' statement, a single case of multifactorial resistance should be sufficient to refute the single-cause fallacy in therapy resistance. Yet, the question of whether the resistance is reducible to a single molecular cause is rarely evoked, discussed, or challenged. Therefore, like the possibility of substantial intra-tumour heterogeneity at the dawn of targeted therapies, the issue of multifactorial resistance appears to be an 'elephant-in-the-room'-type problem. That is, a type of phenomenon that is impossible to ignore when scrutinized but preferable to not note due to convenience or comfort considerations.

36.6. Implications of multifactorial resistance

Why should we care? After all, the current target discovery pipelines provide pharma with candidates worth focusing on; pharmacological inhibition of at least some of these targets does translate into longer remissions. Obviously, there is time and place for reductionistic science as these efforts are balanced by commensurable efforts to integrate knowledge. The problem is that the dominance of the extreme reductionistic paradigm created a moat where investigators are incentivized to ignore the complexity and nuances. The ability to reduce complex phenomena to a clean, linear chain of interactions between genes, proteins, or metabolites became an implicit standard for judging the quality of basic research studies. Conversely, making a case for more nuanced and complex scenarios is a risky proposition. In a highly competitive funding environment, deviation from the expectations of neat, linear causation not only lacks a clear upside but also presents a tangible danger. Far too many investigators, proven right in retrospect, were forced to reconsider their research focus or even lost the ability to practice scientific enterprise. Thus, the dominance of reductionistic causality freezes the development of the field, limiting our understanding of cancer and therapy resistance.

In turn, this limitation impacts our ability to advance the ultimate purpose of cancer research, i.e. to minimize the toll of cancer-related morbidity and mortality. If resistance is not reducible to a single cause, at least at the level of individual subpopulations, how can we hope to achieve cures while staying within the current target discovery-drug development paradigms? Paraphrasing a famous saying of Vincent Feliti ('It's hard to get enough of something that almost works.'), the oncologists might be getting better at playing the whack-a-mole game with cancers, but this game is doomed to eventual failure unless we re-evaluate our basic assumptions and strategies?

36.7. Conclusions and future directions

What is the way out of the impasse? Firstly, the field needs to recognize the existence of the issue created by the dominance of the

single-cause fallacy. However, by itself, acknowledging the complexity and potential existence of an 'elephant in the room' of multifactorial causation will not solve the problem. We need sensible strategies to decipher and conceptualize the interaction of multiple causes in a way that enables us to predict the outcomes of these interactions. The challenge is exacerbated by the fact that cancers are complex dynamic systems that change in response to therapeutic perturbations, and we need to account for these changes. While reductionistic studies are indispensable in identifying genes, metabolites, and proteins that influence phenotypes of interest, integrating these 'building blocks' requires systems biology and mathematical modelling concepts and approaches.

Considering all of the complexity of signalling, gene expression, and metabolic networks across different scales (cells, cell populations, and tissues) is an impossible task, especially if we aim above a mere description. However, perfect understanding through simultaneous consideration of everything is not necessary for useful predictability, which can be achieved at a level of approximation and abstraction. For example, meteorology can make reasonably accurate near-future weather forecasts based on a limited set of measurements and parameters without the need to account for every factor that has been linked with influencing weather at progressively increased spatiotemporal resolution. Instead, predictability is achieved by developing adequate conceptual and mathematical models that capture the interplay of several germane factors that influence weather changes. These abstract models guide the selection of the parameters that need to be measured, as well as the spatial and temporal resolution of measurements, sufficient to achieve a desired level of accuracy. The resulting predictions are not perfect, but they are clearly superior to both random guesses or trying to make sense and predict weather changes by focusing on a single 'key driver' (such as temperature) in isolation.

Analogously, achieving more than incremental progress in dealing with therapy resistance would require stepping beyond cataloguing individual resistance mechanisms and matching the list to a clinical scenario to identify a single factor, continuing the treatment until it becomes obvious that it no longer works, and then repeating the cycle. Instead, we must develop adequate conceptual models of tumour therapy responses using systems biology and mathematical modelling tools. The predictive power of the models should be harnessed to develop a flexible therapeutic strategy that considers the use of all relevant tools within an oncologist's arsenal within clinically feasible constraints (such as frequency of administration, toxicities, etc.). Tumour responses need to be monitored with the parameters (such as tumour burden and phenotypic and genetic markers) measured at the level of resolution guided by the interplay of feasibility and utility. If warranted by the model's predictions, therapy is then adjusted to maximize the likelihood of optimal long-term outcomes (tumour eradication or maximal prolongation of remission while managing toxicity).

Obviously, the models and strategies cannot be produced out of thin air. Developing an adequate level of understanding and predictive capabilities requires dedicated and focused effort, resources, time, and patience within an interdisciplinary collaboration between system biologists, mathematical modellers, clinicians, and experimentalists. Much like the re-assembly of a complex mechanism is a much more challenging task than its disassembly, integrating multiple causes responsible for a complex phenotype is a much more daunting task than identifying individual molecular mechanisms. Thus, it calls for at least matching the scale of the investment and support received by the reductionist molecular biology/drug development enterprise in a futile search for the elusive silver bullet.

REFERENCES

1. Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell. 2010;**141**(1):69–80.
2. França GS, Baron M, Pour M, King BR, Rao A, Misirlioglu S, et al. Drug-induced adaptation along a resistance continuum in cancer cells. bioRxiv. **2022**:2022.06.21.496830.
3. Vander Velde R, Shaffer S, Marusyk A. Integrating mutational and nonmutational mechanisms of acquired therapy resistance within the Darwinian paradigm. Trends Cancer. 2022;**8**(6):456–466.
4. Wu YL, Soo RA, Locatelli G, Stammberger U, Scagliotti G, Park K. Does c-Met remain a rational target for therapy in patients with EGFR TKI-resistant non-small cell lung cancer? Cancer Treat Rev. 2017;**61**:70–81.
5. Sieber OM, Tomlinson SR, Tomlinson IP. Tissue, cell and stage specificity of (epi)mutations in cancers. Nat Rev Cancer. 2005;**5**(8):649–55.
6. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? Nat Rev Cancer. 2012;**12**(5):323–34.
7. Marusyk A, Janiszewska M, Polyak K. Intratumor heterogeneity: the Rosetta stone of therapy resistance. Cancer Cell. 2020;**37**(4):471–84.
8. Welch DR. Tumor heterogeneity—a 'contemporary concept' founded on historical insights and predictions. Cancer Res. 2016;**76**(1):4–6.
9. Cahill DP, Kinzler KW, Vogelstein B, Lengauer C. Genetic instability and darwinian selection in tumours. Trends Cell Biol. 1999;**9**(12):M57–60.
10. Wodarz D, Komarova NL. Emergence and prevention of resistance against small molecule inhibitors. Semin Cancer Biol. 2005;**15**(6):506–14.
11. Bozic I, Reiter JG, Allen B, Antal T, Chatterjee K, Shah P, et al. Evolutionary dynamics of cancer in response to targeted combination therapy. Elife. 2013;**2**:e00747.
12. Cojoc M, Mabert K, Muders MH, Dubrovska A. A role for cancer stem cells in therapy resistance: cellular and molecular mechanisms. Semin Cancer Biol. 2015;**31**:16–27.
13. Prasad S, Ramachandran S, Gupta N, Kaushik I, Srivastava SK. Cancer cells stemness: a doorstep to targeted therapy. Biochim Biophys Acta Mol Basis Dis. 2020;**1866**(4):165424.
14. Smith BN, Bhowmick NA. Role of EMT in metastasis and therapy resistance. J Clin Med. 2016;**5**(2):17.
15. Menzies P. Against causal reductionism. Mind. 1988;**97**(388):551–74.
16. Grasso M, Albantakis L, Lang JP, Tononi G. Causal reductionism and causal structures. Nat Neurosci. 2021;**24**(10):1348–55.
17. Warmuth M, Kim S, Gu XJ, Xia G, Adrian F. Ba/F3 cells and their use in kinase drug discovery. Curr Opin Oncol. 2007;**19**(1):55–60.
18. Piotrowska Z, Niederst MJ, Karlovich CA, Wakelee HA, Neal JW, Mino-Kenudson M, et al. Heterogeneity underlies the emergence of EGFRT790 wild-type clones following treatment of

T790M-positive cancers with a third-generation EGFR inhibitor. Cancer Discov. 2015;**5**(7):713–22.
19. Chabon JJ, Simmons AD, Lovejoy AF, Esfahani MS, Newman AM, Haringsma HJ, et al. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. Nat Commun. 2016;**7**:11815.
20. Suda K, Murakami I, Sakai K, Tomizawa K, Mizuuchi H, Sato K, et al. Heterogeneity in resistance mechanisms causes shorter duration of epidermal growth factor receptor kinase inhibitor treatment in lung cancer. Lung Cancer. 2016;**91**:36–40.
21. Lovly CM, McDonald NT, Chen HD, Ortiz-Cuaran S, Heukamp LC, Yan YJ, et al. Rationale for co-targeting IGF-1R and ALK in ALK fusion-positive lung cancer. Nat Med. 2014;**20**(9):1027–34.
22. Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature. 2017;**546**(7658):431–5.
23. Vander Velde R, Yoon N, Marusyk V, Durmaz A, Dhawan A, Miroshnychenko D, et al. Resistance to targeted therapies as a multifactorial, gradual adaptation to inhibitor specific selective pressures. Nat Commun. 2020;**11**(1):2393.

37

Group behaviour and drug resistance in cancer

Supriyo Bhattacharya, Atish Mohanty, Govindan Rangarajan, and Ravi Salgia

37.1. Introduction

Group behaviour refers to the collective property of a group or population that emerges due to interactions among individual members of the group [1], often benefiting the group as a whole. Cancer cells reside in a communal environment, in close proximity to one another as well as other cell types that constitute the tumour microenvironment (TME). These cells communicate and influence each other's behaviour through exchange of chemical messengers [2,3]. Such mutual interactions, coupled with the heterogeneity of the TME, give rise to complex ecological dependencies (e.g. competition and cooperation) within the tumour. Therefore, group behaviour is a major mechanism that is relevant to every aspect of cancer evolution, from initial progression to metastasis and therapy resistance.

One critical yet overlooked property of cancer cells is phenotypic plasticity that allows rapid adaptation to environmental changes without undergoing genetic mutations [4]. Mechanisms such as phenotypic plasticity and group behaviour facilitate the survival of cancer cells, especially during stressful events such as therapeutic intervention. These mechanisms are of considerable relevance to the emergence of drug resistance, alongside the more familiar mechanism through genetic mutations. Therefore, understanding the role of group behaviour, and the underlying non-genetic mechanisms, can lead to more efficacious treatment designs and minimize or delay the emergence of resistance.

For well over the past 150 years, cancer has been thought to be a predominantly genetic disease, where individual clones acquire driver mutations of increasing fitness through natural selection [5–11]. Furthermore, this thinking has also helped ingrain the idea that drug resistance in cancer, whether innate or acquired, is primarily driven by genetic mutations [12,13]. However, there is a growing appreciation that genetic evolution is unlikely to represent the only mechanism for acquiring drug resistance. Emerging evidence indicates that non-genetic mechanisms such as epigenetic modifications and rewiring of protein interaction networks also contribute to various aspects of cancer, including its origin, progression, and emergence of drug resistance [14–16].

A hallmark of cancer cells is their phenotypic plasticity, that is the ability to exhibit different phenotypes when exposed to variable environmental conditions without undergoing any genotypic changes [17]. The underlying mechanisms contributing to the development of phenotypic plasticity are non-genetic. Cancer cells can switch phenotypes reversibly which allows them to evade the toxic effects of a drug without acquiring any mutation or genetic alteration(s) while contributing to intra-tumour heterogeneity. Indeed, such heterogeneity induced through non-genetic mechanisms serves as an effective bet-hedging strategy that can help overcome the varying selection pressures faced by cancer cells [18,19]. Therefore, it is important to recognize the pervasive contribution of phenotypic plasticity and to develop strategies to effectively counteract this feature of cancer cells, in addition to the genomic-guided approach frequently used with targeted therapies. Of note, while genetic and non-genetic mechanisms of drug resistance are often recognized as separate entities to illustrate the concepts associated with them, most cancers appear to leverage both processes for therapeutic evasion that are not mutually exclusive evolutionary trajectories [14,20].

There is evidence that drug-resistant clones pre-exist within tumours prior to drug treatment, whereas the emergence of a drug-tolerant (i.e. weakly or moderately resistant) state is stochastic which could be exhibited by any cell in the tumour [21–23]. The cells exhibiting tolerance phenotypes are called persisters that are not very well characterized and usually present as a minor fraction of drug-sensitive cells [24]. Different processes, such as pathway rebound through the release of negative feedback loops, transcriptional rewiring mediated by chromatin remodelling, and autocrine/paracrine communication among tumour cells and between tumour cells and other cell types in the TME, are thought to contribute to the emergence of these cells [25]. Nonetheless, this begs the question how drug-sensitive and tolerant or resistant cells in a tumour influence each other's fitness (growth), and whether cooperation and competition (group behaviour) between the sensitive and tolerant cells influence their response to drug treatment. Several studies in the literature have demonstrated the power of game-theory-based approaches to understand group behaviour and its contribution to

drug resistance in different cancers. Furthermore, with advances in technology enabling live cell imaging and in the power of computing with big datasets, discerning group behaviour by monitoring interactions between drug-tolerant and drug-sensitive cells in real time in the absence or presence of the drug has emerged as a powerful tool to elucidate the role of group behaviour [26–34].

In this chapter, firstly, we briefly review game-theory-based models developed for understanding drug resistance, especially in non-small-cell lung cancer (NSCLC) [31,32,34], and draw attention to some of the challenges associated with applying classical game theory to cancer. We then discuss Phenotype Switch Model with Stress Response (PSMSR), a new mathematical approach with game theoretical underpinnings that we developed to model real-time growth data of NSCLC cells to discern patterns in response to treatment with cisplatin [1,35]. We show that the cisplatin-sensitive and cisplatin-tolerant NSCLC cells, when co-cultured in the absence or presence of the drug, display dynamic group behaviour strategies. Tolerant cells exhibit a 'persister-like' behaviour and are attenuated by sensitive cells; they also appear to 'educate' sensitive cells to evade chemotherapy. Further, tolerant cells can switch phenotypes to become sensitive, especially at low cisplatin concentrations. Finally, switching treatment from continuous to an intermittent regimen can attenuate the emergence of tolerant cells, suggesting that intermittent chemotherapy may improve outcomes in NSCLC. We conclude by summarizing the enormous potential of mathematical modelling and quantitative cancer biology.

37.2. Introduction to game theory

Game theory is the mathematical framework for studying the strategic interactions between competing players in a communal environment. A strategy refers to a set of choices or actions adopted by a player at a certain time. The type of strategy that a player chooses will depend on their goals, the goals of the other players, and the rules of the game. Mathematically, each strategy is associated with a payoff matrix that lists the possible outcomes of the strategy and the costs and benefits incurred by different players under alternative scenarios. Classical game theory was developed to analyse the behaviours and strategies followed by human players or organizations, whose decisions are often expected to be rational and geared towards maximizing their payoffs [36]. This often leads to zero-sum games (one player's win is counterbalanced by the opponent's loss), equilibrium situations (e.g. Nash equilibrium, where a single-player strategy change does not lead to any gain, unless others change their strategies as well), or various forms of cooperations and collaborations [37].

The branch of game theory that studies evolutionary processes involving biological species is called evolutionary game theory. Unlike human players, the strategies adopted by biological entities (e.g. animal or plant species, microorganisms, and cancer cells) are not rational rather inherited through generations and evolved for the survival benefit of individuals or communities under a given environment. The dynamics of evolutionary games are driven by competition, cooperation, or other more complex strategies (e.g. bet-hedging and defection) among groups of individuals, where the outcomes (payoffs) of the strategies depend on the opponent strategies as well as relative group populations. In recent years, researchers have shown that phenotypic plasticity exhibited by certain microorganisms and cancer cells can lead to complex game landscapes [35,38], where strategies need not be fixed through inheritance rather switch depending on the environment or, in some cases, learned *de novo* from other players, as seen in the case of drug-sensitive lung cancer cells in the presence of drug-tolerant cells [35].

37.3. Game theory and drug resistance in cancer

Evolutionary game theory has been a valuable conceptual tool to understand the behaviour of cancer cells, the role of tumour heterogeneity, interaction with the microenvironment and immune system, and forecast disease prognosis and design effective therapy [2,27,28,30,31,33–35, 39–44]. The tumour ecosystem is comprised of multiple cell types, such as proliferative cancer cells, supportive stromal cells, immune cells, and fibroblasts (**Figure 37.1**A), each of which can be treated as players in an evolutionary game. Interaction among these cell types and with the microenvironment shapes the cellular phenotypes within the tumour. The group behaviours of individual clones and subclones that result from such interactions can be considered as heritable game strategies. For example, the cooperative subclones in the cancer milieu benefit each other by secreting diffusible factors [45]. However, non-cooperative subclones (cheaters) can compete with the cooperative cells to free ride on the diffusible factors for their own benefit.

These game strategies are subject to selection pressure from the microenvironment, and their payoffs (survival benefit) depend on the group populations. The more successful strategies that increase survival of the tumour as a whole become dominant over time. The strategies dominant within a tumour determines the type of evolutionary game being played. Several studies have elegantly demonstrated that games, such as Prisoner's dilemma, Hawk-Dove, stag hunt, snowdrift, rock–paper–scissors, and Leader and Deadlock, are excellent models to explain many of the observations in clonal dynamics [31,39,41,42].

Therapeutic intervention represents a significant change in the tumour environment that reshapes the clonal composition by altering the fitness of existing clones and promoting newer ones. Drug treatment also leads to the emergence of new game strategies tailored towards evading drug toxicity. Conceptually, treatment and the emergence of drug resistance may be considered as an evolutionary predator–prey game between the cancer (prey) and the oncologist (predator). By analysing the nature of the game being played by the tumour (and carefully anticipating its future moves), the oncologist can design treatment strategies to effectively defeat the tumour [45]. Such strategies typically leverage the evolutionary costs associated with synthesis, maintenance, and operation of the molecular mechanisms necessary for evading and surviving drug treatment. Here, the benefit of resistance outweighs the costs. However, in the absence of treatment, particularly in the TME where resources are limited, the cost renders resistant cells less fit than their drug-sensitive counterparts. Thus, treatment withdrawal at regular intervals interspersed between treatment cycles (intermittent therapy) can encourage residual populations of drug-sensitive cells to exploit their fitness advantage at the expense of the less-fit resistant phenotypes. While withholding

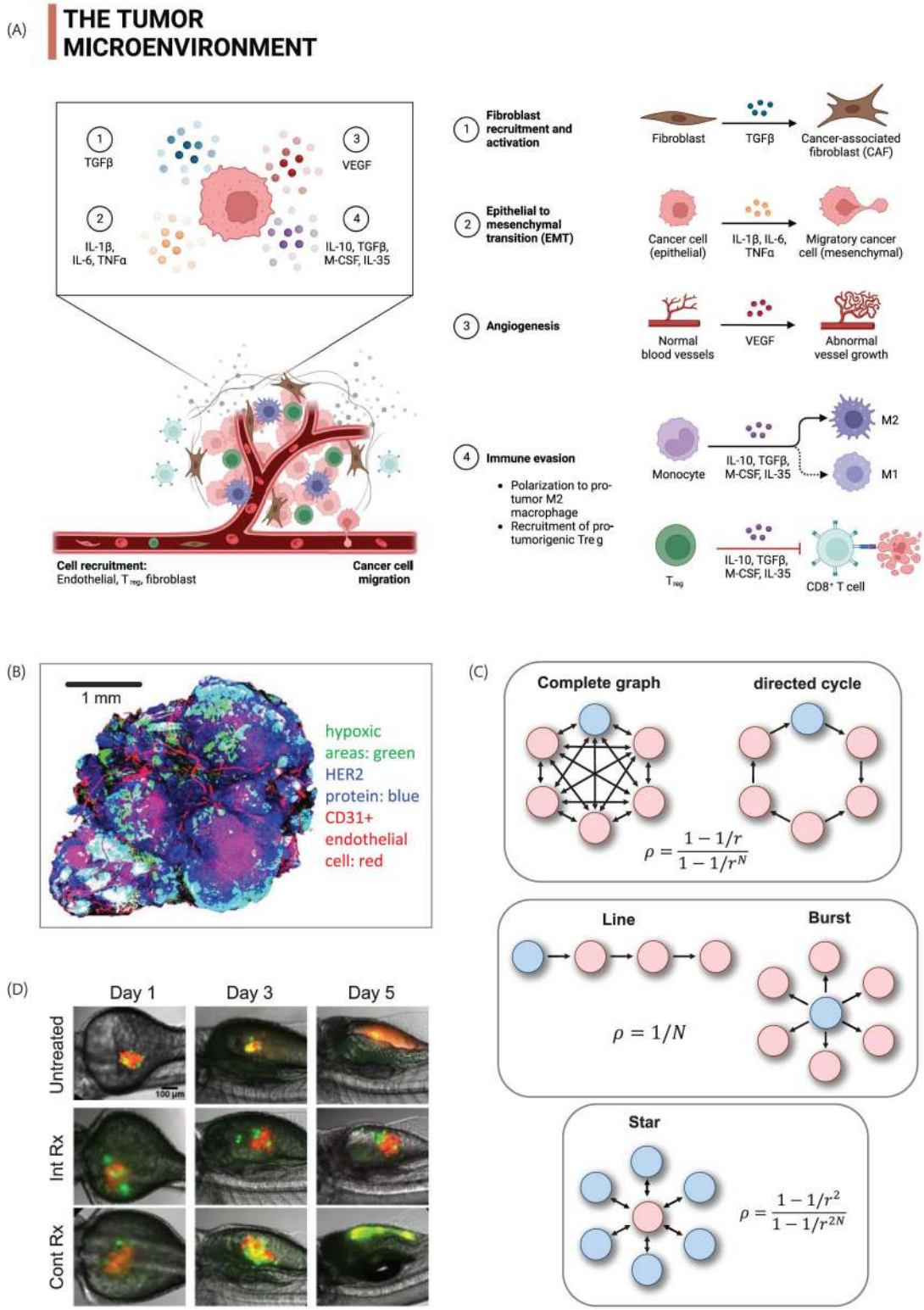


Figure 37.1. (A) Overview of the tumour microenvironment. Few of the common diffusible factors released by proliferating cancer cells are shown in the box. Role of each diffusible factor is depicted on the right. Cancer cells utilize the functional effects of such diffusible factors as part of their game strategies to survive in the host environment and compete/cooperate with other cells. *Source:* Created using Biorender (https//www.biorender.com). (B) High-resolution image of a growing tumour obtained using transparent tumour tomography, showing spatially heterogeneous regions expressing different biomarkers. *Source:* Obtained from https://www.flickr.com/photos/nihgov/27390448613 under license CC BY 2.0. Credit: Steve Seung-Young Lee, University of Chicago Comprehensive Cancer Center, National Cancer Institute, National Institutes of Health. (C) Description of evolutionary graph theory that studies evolutionary games in a spatially restricted environment. Parents and offsprings are organized as nodes in a graph. Mutants (blue nodes) can replace parental population only along the edges of the graph, in the direction specified by the arrows. Under such scenarios, fixation probability (taking over the entire population through successive generations) of a mutant, ρ, depends on the graph topology. Several topologies are shown along with their respective fixation probabilities. Here, r denotes the relative fitness of a mutant (compared to the parental population) and N is the population size. Topology names are according to Nowak et al. (D) Comparison of tumour growth in zebrafish over time, seeded with a mixture of fluorescence-tagged drug-sensitive (red) and drug-tolerant (green) cells under untreated, continuous, and intermittent therapy conditions. *Source:*

treatment allows tumour regrowth, the resistant subpopulation remains small so that retreatment with the same drug(s) remains effective [44]. Thus, it follows that game theory-based studies have provided a novel framework for evolutionarily informed therapies wherein the physician aims to guide the eco-evolutionary dynamics of cancer towards better outcomes or outright cure [32,46,47]. Taken together, it is obvious that mathematical models provide valuable tools for formulating hypotheses and evaluating different scenarios pertaining to the interactions between cancer cell types and therapy [44].

Despite the apparent success, the complexity of cellular behaviour, such as soluble factors with non-linear effects on different cells as well as phenotypic plasticity, remains considerable challenges. For example, as discussed below, unlike in classical ecology models where players do not switch identities, cancer cells, drug-resistant, and drug-sensitive cells can switch their phenotypes because of their innate plasticity. Furthermore, since the behaviour of cancer cells is highly dependent on their unique microenvironment, it is often challenging to translate the findings from in vitro studies to *in vivo*. Secondly, the stochasticity involved at different levels of cellular behaviour, from intracellular biochemical networks to interactions among groups of cells, can result in multistability, and therefore, challenge claims of causal connection between quantitative phenotypic markers, such as the expression of cell surface receptors and their behavioural effects [48].

37.4. Tumour heterogeneity and significance of the spatial dimension

Cellular heterogeneity is a hallmark of cancer progression and reflected in the coexistence of multiple clones within the tumour and the diversity of the microenvironment. Much work has been carried out to explore the mutational landscape of developing tumours and its effect on treatment and resistance [49,50]. For instance, the positive correlation between tumour heterogeneity and worse clinical prognosis has been reported in multiple cancer types [51–54]. Recently, it was shown that linear (same clone successively acquiring multiple mutations) vs. branched (driver mutations distributed among multiple clones) evolution, as well as the sequence of acquiring mutations, makes a difference in clinical outcome in acute myeloid leukaemia [55]. Moreover, the frequency of certain mutations among subclones determines drug sensitivity. Such studies suggest that tumour heterogeneity evolves in response to selection pressure rather than as a by-product of cancer proliferation. In addition to mutations, non-genetic mechanisms play a major but underappreciated role in conferring phenotypic heterogeneity to the tumour ecosystem, which has been extensively discussed in several recent reviews by us and others [20,56,57]. Difficulty of measurement and the mutation-centric view of cancer are some of the reasons why non-genetic mechanisms are less appreciated. Besides clonal diversity, another major source of tumour heterogeneity is the TME (**Figure 37.1**A and B). As the disease progresses, crosstalk between the tumour and the TME shapes each other's heterogeneity and spatial organization, and this interaction is likely to be critical for the long-term survival of the disease. The benefits of the TME include providing proliferative and metabolic factors to the tumour and maintaining an immunosuppressive environment for the tumour to thrive. The ecological forces that shape tumour composition are challenging to study using current experimental techniques, leading to the development of theoretical and simulation frameworks as discussed next.

Solid tumours proliferate within a dense environment of host cells and extra-cellular matrix, where each cell mainly interacts with its nearest neighbours. This is in sharp contrast to scenarios, such as leukaemia, where cells are fully mobile and free to interact with any other cell in the environment. For example, in a spatially restricted environment, beneficial diffusible elements such as growth factors will primarily affect the near neighbourhood of the source cells since their levels will fall off with distance from the origin. Therefore, it has been proposed that spatial organization plays a vital role in tumour evolution, heterogeneity, and development of game strategies. To understand these effects from a theory standpoint, Nowak developed the evolutionary graph theory, where cells are organized as nodes in a graph [58]. Competition and cooperation are only allowed between neighbouring cells that share common edges in the graph network (**Figure 37.1**C). Nowak and co-workers have showed that evolutionary dynamics follows different trajectories in a spatially constrained environment [59], depending on the graph layout (i.e. fully connected, scale free or circular, etc.). Key properties, such as fixation probability (certain graph layouts can amplify or suppress natural selection) and payoffs, vary depending on the graph layouts, thus highlighting the importance of spatial dimension in tumour evolution. One interesting outcome of such theoretical analysis is the emergence of cooperation. In simulations of cooperator–defector dynamics on a spatial grid, it was shown that defectors invade the cooperator cells and outcompete them when the benefit to defection was above a certain threshold. However, the cooperators were never completely wiped out, and they survived by organizing themselves into tight clusters.

The above narrative carries a strong parallel to our observation of the interplay between drug-sensitive and drug-tolerant NSCLC cells in *in vivo* zebrafish model [35], where the drug-tolerant cells (cooperators) formed tight clusters, surrounded by the drug-sensitive cells (defectors) (**Figure 37.1**D). NSCLC can be classified as adenocarcinoma or squamous cell carcinoma or large cell carcinoma histologically. There are a number of oncogenes that can be abnormal in NSCLC, such as EGFR mutation, ALK fusions, ROS1 fusions, MET exon 14 splice variants, and others. There is large heterogeneity of lung cancer; however, the majority of NSCLC respond to platinum-based therapy. Independent *in vitro* studies, coupled with mathematical modelling, confirmed the cooperative trait of the drug-tolerant cells in the form of diffusible factors. Works like the above that explore cancer progression from the ecological perspective are rare due to the challenges faced in monitoring the phenotypic behaviour of tumour components at the cellular level. One exciting development in this field is the microfluidic death galaxy developed by Austin and co-workers who can monitor the growth and spatial organization of multiple cell types under different ecological conditions, such as drug concentrations [43]. Combining experimental observations with game theory models, such constructs, can estimate hard to obtain parameters such as payoffs under varying selection pressures, which can be used for future prediction of prognosis under therapeutic intervention.

37.5. Group behaviour via non-genetic mechanisms facilitates therapy resistance

Historically, cancer research has focused on genetic alterations (e.g. mutations, copy number variations, and chromosomal instability) as the primary drivers of the disease. Emergence of resistance-conferring mutations in response to therapy is typically ascribed to the cause of tumour survival and disease relapse. Recently, however, the importance of non-genetic resistance mechanisms has come to light. Due to their genetic and epigenomic alterations, cancer cells are adept in rewiring their signalling networks to bypass the effect of anti-cancer drugs, as shown in the case of melanoma [60,61], or more recently through our own works on NSCLC [35,62]. These network rewiring acts are typically carried out at the transcriptional (e.g. regulation of transcription factors, DNA looping, and chromatin accessibility), post-transcriptional (e.g. alternative splicing and selective RNA degradation), translational (e.g. ribosomal regulation), or post-translational (e.g. protein phosphorylation and ubiquitination) levels and through various autocrine and paracrine processes. By network rewiring, cancer cells can temporarily switch to a different phenotype that enables them to survive under a rapidly changing environment, such as ambient drug concentrations. This behaviour is called phenotypic switching and has a strong presence in the fungal and bacterial world, where such mechanisms are used to combat sudden toxicity or invasion by competing microorganisms [63,64]. In the cancer world, one prime example of phenotypic switching is melanoma cells transitioning from a proliferative (drug sensitive) to an invasive phenotype (drug resistant) in response to mitogen-activated protein kinase (MAPK) pathway inhibitors, without undergoing genetic mutations [60].

An important indicator of non-genetic resistance mechanisms is drug-tolerant persister (DTP), a small subpopulation of drug-tolerant cancer cells that naturally exist within an otherwise sensitive population and survive drug treatment without undergoing genetic alterations. DTPs have been studied and referred to multiple times in the literature over the past decade, after they were first reported in 2010 [25,65]. Upon extended drug exposure, the DTPs were found to proliferate and reestablish *in vitro* colony of drug-tolerant cells, and these cells reverted to drug sensitivity upon drug withdrawal within a few cell divisions (~30). These reversible phenotypic transitions within a few generations indicate that heritable epigenetic modifications can stabilize drug-tolerant phenotypes and are supported by the involvement of histone demethylation, as in the case of melanoma cells.

Available evidence therefore indicates that in some cases non-genetic mechanisms, such as phenotypic switching and heritable epigenetic modifications, may be preferred over genetic mutations in developing therapy resistance. The question is why? From a game theory perspective where cells or groups of cells can be considered players, both non-genetic and genetic mechanisms are survival strategies that evolve under environmental pressures. If we analyse the benefits and costs associated with each type of strategy, it may be understandable why one would be preferable over the other. Non-genetic mechanisms are rapid and reversible and are therefore capable of addressing sudden changes in the environment. Genetic mechanisms lead to the fixation of heritable alterations in the genome that once acquired are not easily lost. However, evolution of such strategies through fitness conferring mutations require cellular proliferation and selection over multiple generations (through stochastic trial-and-error attempts) and are typically slower in response to environmental changes. Moreover, genetic changes are permanent since reversing a mutation is not a spontaneous process. In comparison, non-genetic mechanisms can be reversible. While switching from a proliferative to an invasive phenotype, melanoma cells still retained their proliferative potential so that when the drug pressure was lifted, they were able to switch to the proliferative phenotype and reestablish their colony [60]. However, it is important to note that non-genetic mechanisms often come with a cost. For instance, they may require increased energy expenditure due to elevated transcription, protein synthesis, and kinase recruitment for phenotypic switching mechanisms.

The choice between non-genetic mechanisms and permanent genetic alterations as survival strategies depends on a careful assessment of the associated benefits and costs. In situations where the advantages of rapid response outweigh the costs, non-genetic mechanisms become the preferred option. It is hypothesized that the administration of anti-cancer drugs can induce an environmental shift that favours non-genetic mechanisms during the initial phase of adaptability as opposed to genetic alterations. Currently, the proposed hypothesis by our research team and other experts in the field suggests that non-genetic mechanisms can serve as a survival strategy for tumours facing environmental stress, such as exposure to cytotoxic drugs. These mechanisms provide a temporary solution until more permanent modifications, such as genetic mutations or epigenetic changes, evolve [1,20,66]. It could also mean that therapists can manipulate the environmental conditions that promote the dominance of non-genetic mechanisms within the tumour and delay the emergence of permanently resistant clones [35,66]. We will hold that thought and revisit it in a subsequent section (intermittent therapy).

37.6. Phenotypic switching, stress response, and intra-tumour cooperation

In the previous section, we have argued how drug-sensitive cancer cells can avert the effect of environmental stress by temporarily switching to a drug-tolerant phenotype. We recently addressed the question whether drug-tolerant phenotypes can also cooperate with drug-sensitive phenotypes and assist their survival under stress [35]. Such cooperation (and altruism) will benefit the tumour as a whole, being embodied in the theoretical framework of the Price equation, and was explored in the context of drug resistance evolution in bacteria [67]. To this end, we monitored the *in vitro* growth of drug-sensitive and tolerant NSCLC cells (shown to undergo phenotypic switching [62]) both in monotypic cultures and mixed at various ratios over a period of several weeks, in the presence or absence of the chemotherapeutic drug cisplatin. The growth rates of these cells when cultured together showed significant differences compared to their respective monotypic cultures. Additionally, the growth rates varied according to the proportion of sensitive to tolerant cells. Clearly, these cells altered their behaviour by sensing each other's presence in a frequency-dependent manner, underscoring the importance of group behaviour in

tumour growth. We quickly realized that the complex growth dynamics of these cells could not be explained by invoking simple competition or cooperation, thus necessitating more intricate models. We also observed two interesting characteristics of these cells: (1) the sensitive cells secreted a diffusible factor that negatively affected cellular growth, including that of tolerant cells; (2) the tolerant cells could be reverted to cisplatin sensitivity by using a histone deacetylase inhibitor, indicating an epigenetic basis for their drug tolerance. These observations motivated us to develop a mathematical theory of drug resistance that incorporates the role of phenotypic switching, stress response, and cooperation of cancer cells in a community environment.

37.7. Phenotypic switching enables cancer cells to adapt to rapid environmental changes

Detailed formulation of PSMSR is already published elsewhere [35], so we briefly discuss it here. The chief hypothesis behind PSMSR is that cancer cell phenotype is not rigid but stochastically switch between drug-sensitive and drug-tolerant states. Further, this phenotypic switching can be influenced by environmental factors that adversely affect cellular growth, such as stress elements, lack of oxygen, or diffusible factors (collectively referred to as stress in this model) (**Figure 37.2**A). We also explored competition and cooperation through dissemination of 'public goods' and neutralization of stress by the drug-tolerant phenotype. As a result, evolutionary strategies dynamically altered with the level of stress in the environment (**Figure 37.2**B). The different parameters of PSMSR (e.g. cellular growth, phenotypic switching, and stress generation/neutralization rates) were derived by fitting real-time growth data of sensitive and tolerant cells in mixed cultures at different ratios (**Figure 37.2**C) [35]. In comparison to several other cellular growth models that account for group interactions, PSMSR fits the experimental data the best.

PSMSR predicted that in addition to the cellular frequencies the level of growth-retarding diffusible factors released by the sensitive cells (termed stress in the model) determined the evolutionary strategy adopted by the tolerant cells. At low stress, passive competition existed between the two phenotypes, which quickly changed into cooperation by the tolerant cells (due to increased stress neutralization, thus benefiting the whole community) as stress built up (**Figure 37.2**D). The optimum fraction of tolerant phenotypes in the ecosystem (to ensure positive payoff for the whole community) was determined by the phenotypic switching rates, which dynamically altered in response to stress and cellular frequencies. By combining PSMSR with competitive Lotka–Volterra model, we determined the magnitude of cooperative interaction between the sensitive and tolerant phenotypes as a function of time and seeding ratio, thus generating evolutionary strategy landscapes for both cell types (**Figure 37.2**D and E). We found that compared to the sensitive phenotype, the tolerant phenotype was more flexible in altering their game strategies depending on the environment. This suggests that the tolerant phenotype is more adaptive to increased stress, such as therapeutic pressure, and their cooperation with the sensitive cells could ensure the survival of the tumour as a whole. One caveat is that these inferences are drawn from *in vitro* studies and therefore cannot account for the effects of a real TME, which can alter the strategy landscape of the tumour and direct resistance evolution significantly.

37.8. Non-genetic resistance mechanism underscores the benefit of adaptive/intermittent therapy in delaying resistance

Traditionally, the first-line therapy in cancer involves continuous administration of targeted drugs or chemotherapeutic agents at the maximum tolerated dose. This inevitably creates a condition that favours the emergence of resistant disease. Recent awareness of the role of ecology in resistance development has led to innovations in therapy design, such as adaptive therapy, that aims to delay the onset of resistance [4,20,30,35,46]. The goal of adaptive therapy is to suppress the emergence of the drug-resistant phenotype and maintain drug sensitivity of the tumour by adjusting the dosage amounts and intervals. Adaptive therapy designs are usually based on assumptions, such as (1) pre-existence of drug-resistant clones in the original tumour and (2) low fitness of resistant cells in the absence of therapeutic pressure [40]. In a 2018 review article from our group, we argued that adaptive therapy through intermittent dosage can also be beneficial in cases where the resistance mechanism is non-genetic (as opposed to therapy-induced selection of pre-existing resistant clones) [20]. As was shown in multiple cases (and discussed in previous sections of this chapter), drug-tolerant phenotypes can emerge from drug-sensitive cellular population through non-genetic mechanisms, such as phenotypic switching and epigenetic alterations. Drug resistance in such cases do not need the pre-existence of resistant clones in the tumour. However, adaptive therapies can still be beneficial by suppressing the rate of transition to the tolerant phenotype or altering the group dynamics among the sensitive and tolerant phenotypes as well as the microenvironment.

We have shown that intermittent rather than continuous cisplatin treatment can suppress the growth of tolerant NSCLC cells and retain drug sensitivity in both *in vitro* and zebrafish models [35]. Starting with a mixture of fluorescence-tagged sensitive and tolerant cells at different seeding ratios, we cultured them for two weeks under continuous exposure to 1 μM cisplatin, as well as an initial exposure of three days, followed by growth in cisplatin-free media (intermittent dosage). The cellular growth was quantitatively measured in real time using Incucyte live cell analyser. As expected, we observed a massive expansion in tolerant cell population (60–80 times sensitive cell population) under continuous cisplatin treatment with the increasing trend continuing at the end of two weeks (**Figure 37.2**F). Conversely, under intermittent treatment, the tolerant-to-sensitive ratio was moderate [4–6] and was stabilized by day 10. Intermittent cisplatin treatment also suppressed the proliferation of tolerant cells *in vivo* in zebrafish, although the experiments had to be concluded within five days due to regulatory reasons (**Figure 37.1**D).

To see the long-term effect of intermittent cisplatin treatment, we continued the *in vitro* experiments for an extended period and noted that the tolerant-to-sensitive cell ratio was still low after 43 days (**Figure 37.2**G). Interestingly, a second cisplatin dose added on day 14 (intermittent: two cycles) after the first cisplatin treatment triggered the proliferation of tolerant cells (Figure 37.2G) suggesting that the exposure of the sensitive cell population to the first cisplatin

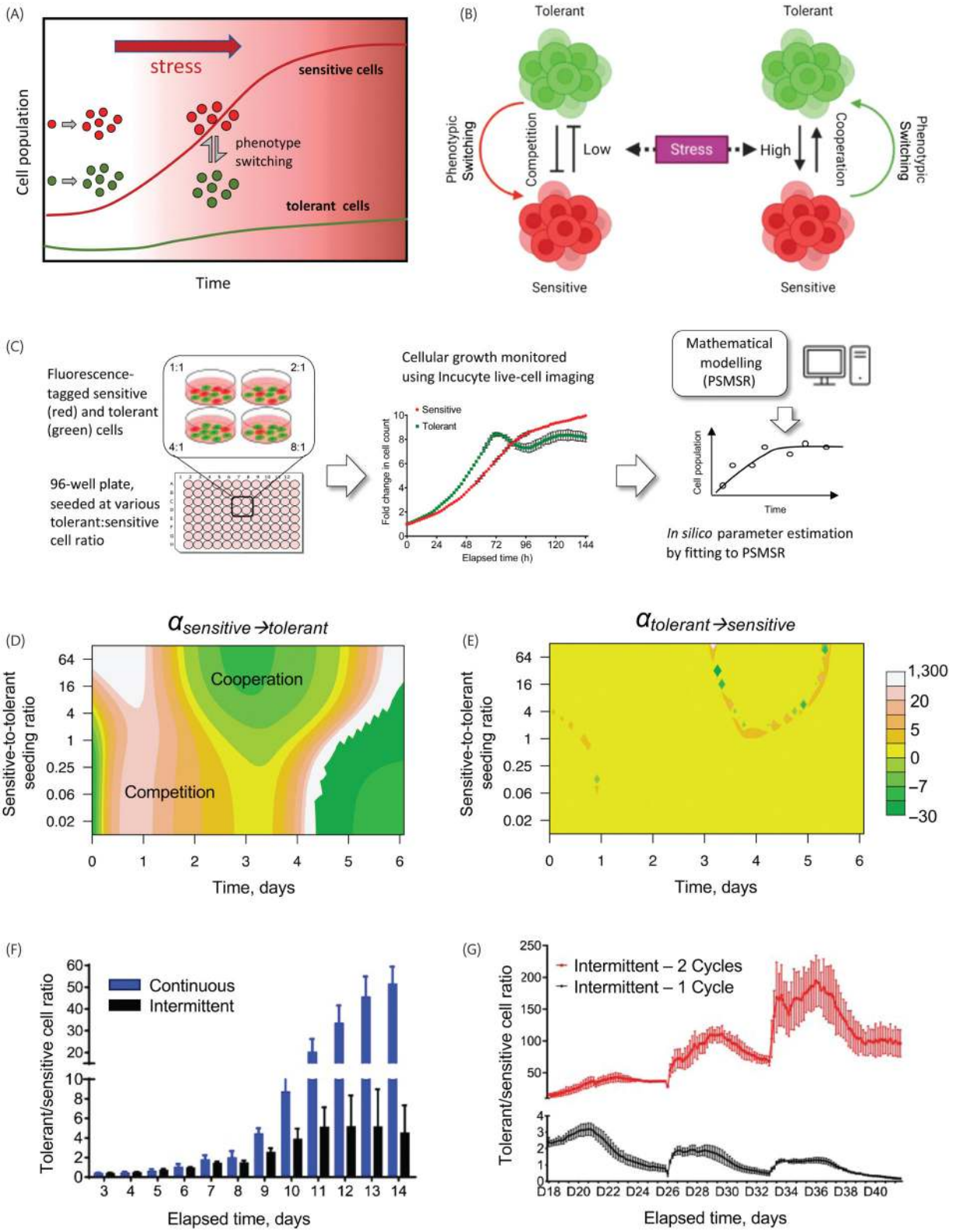


Figure 37.2. (A) Schematic describing the physical process (cellular growth and stress generation over time) that forms the basis of PSMSR. (B) Depiction of phenotypic switching between drug-sensitive and drug-tolerant lung cancer cells in response to environmental stress. (C) Schematic of cell growth monitoring process to determine group behaviour by fitting to PSMSR. (D, E) Dynamic game strategic landscapes of drug-sensitive and drug-tolerant phenotypes as a function of time, under various seeding conditions. Starting from initial populations of sensitive and tolerant cells mixed at different proportions (seeding ratios), cellular growth was simulated using PSMSR. Competition/cooperation was estimated by piecewise fitting of the competitive Lotka–Volterra equation to the PSMSR growth trends over a window of three days. Contour plots show the *α* parameters of the LV equation for different time and seeding ratios. Contours are coloured according to quantiles. Positive and negative values are indicative of competition and cooperation, respectively. For ease of comparison, the same scale is used in both (D) and (E). For more information, see Nam et al. Biomolecules, 2022. (F) Tolerant-to-sensitive cell ratio as a function of time under continuous and intermittent cisplatin therapy, monitored using the set-up depicted in (C). Sensitive and tolerant cells were cultured at an initial seeding ratio of 4:1. (G) Comparison of tolerant cell expansion under two different intermittent therapy regimens (see the text for more details). *Source:* Panels (F) and (G) are reproduced under Creative Common CC BY license. For more details, see Nam et al. 2021 [35].

dose may have compromised their proliferative potential, thereby allowing tolerant cells to proliferate. We also used PSMSR to model the cellular growth under cisplatin treatment and showed that PSMSR captured the difference in growth dynamics between the continuous and intermittent conditions [35]. Further experiments are needed, potentially using mouse models of NSCLC to further establish the effectiveness of such therapeutic strategies under *in vivo* conditions. Mathematical and agent-based models that account for the group behaviour of mixed populations of sensitive and tolerant cells can be valuable tools to determine the optimal drug doses and intervals under intermittent treatment.

37.9. Conclusions and future directions

It is well established that, even within a given cancer type, there exist multiple mechanisms that regulate phenotypic switching and drug resistance. Furthermore, as discussed here, although intermittent therapy appears promising in some cases, several challenges still remain. Nonetheless, from the foregoing, it is obvious that insights from cancer systems biology and mathematical modelling can help identify new treatment strategies based on the principles of ecology and evolution. By incorporating these new concepts in clinical protocols, we could enhance the precision in which we deliver personalized medicine to all our patients, regardless of their economic status or their ability to access advanced medical centres. Furthermore, lowering the dose of the drug and its frequency as a result of intermittent rather than continuous therapy could not only lower the toxicity and undesirable side effects of the drugs but may also positively impact the financial burden carried by the patient and insurance providers [68].

REFERENCES

1. Bhattacharya S, Mohanty A, Achuthan S, Kotnala S, Jolly MK, Kulkarni P, et al. Group behavior and emergence of cancer drug resistance. Trends Cancer. 2021;**7**(4):323–34.
2. Dominiak A, Chelstowska B, Olejarz W, Nowicka G. Communication in the cancer microenvironment as a target for therapeutic interventions. Cancers (Basel). 2020;**12**(5).
3. Brucher BL, Jamall IS. Cell-cell communication in the tumor microenvironment, carcinogenesis, and anticancer treatment. Cell Physiol Biochem. 2014;**34**(2):213–43.
4. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic plasticity: driver of cancer initiation, progression, and therapy resistance. Cell Stem Cell. 2019;**24**(1):65–78.
5. Hansemann DV. Ueber asymmetrische Zelltheilung in Epithelkrebsen und deren biologische Bedeutung. Arch Pathol Anat Physiol Klin Med. 1890;**119**:299–326.
6. Boveri T. Zur Frage der Entstehung maligner Tumoren. Jena: Gustav Fischer; 1914.
7. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976;**194**(4260):23–8.
8. Knudson AG. Two genetic hits (more or less) to cancer. Nat Rev Cancer. 2001;**1**(2):157–62.
9. Loeb LA, Harris CC. Advances in chemical carcinogenesis: a historical review and prospective. Cancer Res. 2008;**68**(17):6863–72.
10. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;**100**(1):57–70.
11. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science. 2015;**349**(6255):1483–9.
12. Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. Nat Rev Cancer. 2013;**13**(10):714–26.
13. Nussinov R, Tsai CJ, Jang H. Anticancer drug resistance: an update and perspective. Drug Resist Updat. 2021;**59**:100796.
14. Marine JC, Dawson SJ, Dawson MA. Non-genetic mechanisms of therapeutic resistance in cancer. Nat Rev Cancer. 2020;**20**(12):743–56.
15. Bell CC, Gilan O. Principles and mechanisms of non-genetic resistance in cancer. Br J Cancer. 2020;**122**(4):465–72.
16. Huang S. Reconciling non-genetic plasticity with somatic evolution in cancer. Trends Cancer. 2021;**7**(4):309–22.
17. Pigliucci M, Murren CJ, Schlichting CD. Phenotypic plasticity and evolution by genetic assimilation. J Exp Biol. 2006;**209**(Pt 12):2362–7.
18. Mathis RA, Sokol ES, Gupta PB. Cancer cells exhibit clonal diversity in phenotypic plasticity. Open Biol. 2017;**7**(2).
19. Jolly MK, Kulkarni P, Weninger K, Orban J, Levine H. Phenotypic plasticity, bet-hedging, and androgen independence in prostate cancer: role of non-genetic heterogeneity. Front Oncol. 2018;**8**:50.
20. Salgia R, Kulkarni P. The genetic/non-genetic duality of drug 'resistance' in cancer. Trends Cancer. 2018;**4**(2):110–8.
21. Hata AN, Niederst MJ, Archibald HL, Gomez-Caraballo M, Siddiqui FM, Mulvey HE, et al. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. Nat Med. 2016;**22**(3):262–9.
22. Wang X, Zhang H, Chen X. Drug resistance and combating drug resistance in cancer. Cancer Drug Resist. 2019;**2**(2):141–60.
23. Mullard A. Stemming the tide of drug resistance in cancer. Nat Rev Drug Discov. 2020;**19**(4):221–3.
24. Venkatesan S, Swanton C, Taylor BS, Costello JF. Treatment-induced mutagenesis and selective pressures sculpt cancer evolution. Cold Spring Harb Perspect Med. 2017;**7**(8).
25. Swayden M, Chhouri H, Anouar Y, Grumolato L. Tolerant/persister cancer cells and the path to resistance to targeted therapy. Cells. 2020;**9**(12).
26. Tomlinson IP. Game-theory models of interactions between tumour cells. Eur J Cancer. 1997;**33**(9):1495–500.
27. Gatenby RA, Vincent TL. An evolutionary model of carcinogenesis. Cancer Res. 2003;**63**(19):6212–20.
28. Archetti M. Evolutionary game theory of growth factor production: implications for tumour heterogeneity and resistance to therapies. Br J Cancer. 2013;**109**(4):1056–62.
29. Hiltunen T, Virta M, Laine AL. Antibiotic resistance in the wild: an eco-evolutionary perspective. Philos Trans R Soc Lond B Biol Sci. 2017;**372**(1712).
30. Stanková K, Brown JS, Dalton WS, Gatenby RA. Optimizing cancer treatment using game theory: a review. JAMA Oncol. 2019;**5**(1):96–103.
31. Kaznatcheev A, Peacock J, Basanta D, Marusyk A, Scott JG. Fibroblasts and alectinib switch the evolutionary games played by non-small cell lung cancer. Nat Ecol Evol. 2019;**3**(3):450–6.
32. Staňková K. Resistance games. Nat Ecol Evol. 2019;**3**(3):336–7.
33. Grolmusz VK, Chen J, Emond R, Cosgrove PA, Pflieger L, Nath A, et al. Exploiting collateral sensitivity controls growth of mixed culture of sensitive and resistant cells and decreases selection for resistant cells in a cell line model. Cancer Cell Int. 2020;**20**:253.

34. Archetti M. Collapse of intra-tumor cooperation induced by engineered defector cells. Cancers (Basel). 2021;**13**(15).
35. Nam A, Mohanty A, Bhattacharya S, Kotnala S, Achuthan S, Hari K, et al. Dynamic phenotypic switching and group behavior help non-small cell lung cancer cells evade chemotherapy. Biomolecules. 2021;**12**(1).
36. von Neumann J, Morgenstern O, Rubinstein A. Theory of games and economic behavior (60th Anniversary Commemorative Edition). Princeton University Press; 1944.
37. Osborne MJ, Rubinstein A. A course in game theory. Cambridge, MA: MIT Press; 1994.
38. Borenstein E, Meilijson I, Ruppin E. The effect of phenotypic plasticity on evolution in multipeaked fitness landscapes. J Evol Biol. 2006;**19**(5):1555–70.
39. Bayer P, Gatenby RA, McDonald PH, Duckett DR, Staňková K, Brown JS. Coordination games in cancer. PLoS ONE. 2022;**17**(1):e0261578.
40. Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive therapy. Cancer Res. 2009;**69**(11):4894–903.
41. McEvoy JW. Evolutionary game theory: lessons and limitations, a cancer perspective. Br J Cancer. 2009;**101**(12):2060–1; author reply 2–3.
42. Pacheco JM, Santos FC, Dingli D. The ecology of cancer from an evolutionary game theory perspective. Interface Focus. 2014;**4**(4):20140019.
43. Wu A, Liao D, Tlsty TD, Sturm JC, Austin RH. Game theory in the death galaxy: interaction of cancer and stromal cells in tumour microenvironment. Interface Focus. 2014;**4**(4):20140028.
44. Zhang J, Cunningham JJ, Brown JS, Gatenby RA. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. Nat Commun. 2017;**8**(1):1816.
45. Archetti M, Pienta KJ. Cooperation among cancer cells: applying game theory to cancer. Nat Rev Cancer. 2019;**19**(2):110–7.
46. Gatenby RA. A change of strategy in the war on cancer. Nature. 2009;**459**(7246):508–9.
47. Ibrahim-Hashim A, Robertson-Tessi M, Enriquez-Navas PM, Damaghi M, Balagurunathan Y, Wojtkowiak JW, et al. Defining cancer subpopulations by adaptive strategies rather than molecular properties provides novel insights into intratumoral evolution. Cancer Res. 2017;**77**(9):2242–54.
48. Pisco AO, Brock A, Zhou J, Moor A, Mojtahedi M, Jackson D, et al. Non-Darwinian dynamics in therapy-induced cancer drug resistance. Nat Commun. 2013;**4**:2467.
49. Ramón YCS, Sesé M, Capdevila C, Aasen T, De Mattos-Arruda L, Diaz-Cano SJ, et al. Clinical implications of intratumor heterogeneity: challenges and opportunities. J Mol Med (Berl). 2020;**98**(2):161–77.
50. Diaz-Cano SJ. Tumor heterogeneity: mechanisms and bases for a reliable application of molecular marker design. Int J Mol Sci. 2012;**13**(2):1951–2011.
51. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic classification and prognosis in acute myeloid leukemia. New Engl J Med. 2016;**374**(23):2209–21.
52. Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. Oral Oncol. 2013;**49**(3):211–5.
53. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science. 2014;**346**(6206):256–9.
54. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014;**344**(6190):1396–401.
55. Benard BA, Leak LB, Azizi A, Thomas D, Gentles AJ, Majeti R. Clonal architecture predicts clinical outcomes and drug sensitivity in acute myeloid leukemia. Nat Commun. 2021;**12**(1):7244.
56. Sharma A, Merritt E, Hu X, Cruz A, Jiang C, Sarkodie H, et al. Non-genetic intra-tumor heterogeneity is a major predictor of phenotypic heterogeneity and ongoing evolutionary dynamics in lung tumors. Cell Rep. 2019;**29**(8):2164–74.e5.
57. Black JRM, McGranahan N. Genetic and non-genetic clonal diversity in cancer evolution. Nat Rev Cancer. 2021;**21**(6):379–92.
58. Nowak MA. Evolutionary dynamics exploring the equations of life. Cambridge, MA: Harvard University Press; 2006.
59. Noble R, Burri D, Le Sueur C, Lemant J, Viossat Y, Kather JN, et al. Spatial structure governs the mode of tumour evolution. Nat Ecol Evol. 2022;**6**(2):207–17.
60. Kemper K, de Goeje PL, Peeper DS, van Amerongen R. Phenotype switching: tumor cell plasticity as a resistance mechanism and target for therapy. Cancer Res. 2014;**74**(21):5937–41.
61. Zipser MC, Eichhoff OM, Widmer DS, Schlegel NC, Schoenewolf NL, Stuart D, et al. A proliferative melanoma cell phenotype is responsive to RAF/MEK inhibition independent of BRAF mutation status. Pigment Cell Melanoma Res. 2011;**24**(2):326–33.
62. Mohanty A, Nam A, Pozhitkov A, Yang L, Srivastava S, Nathan A, et al. A non-genetic mechanism involving the integrin beta4/Paxillin axis contributes to chemoresistance in lung cancer. iScience. 2020;**23**(9):101496.
63. Farquhar KS, Charlebois DA, Szenk M, Cohen J, Nevozhay D, Balázsi G. Role of network-mediated stochasticity in mammalian drug resistance. Nat Commun. 2019;**10**(1):2766.
64. Jain N, Hasan F, Fries BC. Phenotypic switching in fungi. Curr Fungal Infect Rep. 2008;**2**(3):180–8.
65. Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, et al. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell. 2010;**141**(1):69–80.
66. Gunnarsson EB, De S, Leder K, Foo J. Understanding the role of phenotypic switching in cancer drug resistance. J Theor Biol. 2020;**490**:110162.
67. Gjini E, Wood KB. Price equation captures the role of drug interactions and collateral effects in the evolution of multidrug resistance. eLife. 2021;**10**:e64851.
68. Mason NT, Burkett JM, Nelson RS, Pow-Sang JM, Gatenby RA, Kubal T, et al. Budget impact of adaptive abiraterone therapy for castration-resistant prostate cancer. Am Health Drug Benefits. 2021;**14**(1):15–20.

38

The fundamentals of evolutionary therapy in cancer

Jeffrey West, Jill Gallaher, Maximilian A.R. Strobl, Mark Robertson-Tessi, and Alexander R.A. Anderson

38.1. Introduction

38.1.1. The mathematics of treatment scheduling in oncology

At its core, a mathematical model contains a set of assumptions from which a set of conclusions are deduced. As Gunawardena has pointed out: if the model is correct and if you accept its assumptions, you must as a principle of logic also accept the model's conclusions [1]. The exercise of writing down a set of equations is precisely the exercise of carefully stating those assumptions and seeking their logical conclusions.

Thus, it comes as no surprise that mathematical modelling has found a long tradition within clinical decision-making, especially in scheduling treatment with anti-cancer drugs. We begin by tracing the history of assumptions made in the mathematical modelling of cancer treatment response made since the 1960s. We will find that mathematical modelling was necessarily limited by the current biological understanding of the day, but it still proved useful to periodically plant a flag in the ground and identify testable hypotheses or derive quantitative predictions.

Mathematics formalizes and crystalizes assumptions, enabling clear and concise hypotheses. Viewing the history of treatment scheduling as an ever-evolving and refined list of biological assumptions cast into the formalized logic language of mathematical modelling helps us understand why models throughout the decades provide different and even sometimes contradictory results. Each model refines a new set of assumptions that represents the scientific community's maturing view of cancer's underlying complexity.

For example, early attempts at mathematical models encompass only a limited, homogeneous view of tumours (in Skipper's law [2], every cancer cell is an identical copy), while subsequent models aim for a more faithful representation of tumour heterogeneity. In this way, the history of mathematics in cancer is nearly inseparable from the history of cancer evolution. In this chapter, we explain how the nature of cancer as an evolutionary and ecological disease [3] renders traditional treatment scheduling paradigms (e.g. maximum tolerable dosing) less effective than evolutionary therapy. Mathematics has played a key role in the design and implementation of these novel paradigms. Firstly, we review the important advances in treatment scheduling, aided by mathematics, across each decade from the 1960s until present, before turning our attention to evolutionary therapy (**Figure 38.1**).

38.1.2. Accounting for cellular and microenvironmental heterogeneity

The oft-cited first example of the translational utility of mathematics in cancer is the seminal work by Skipper, Schabel, and Wilcox introducing the log-kill dosing law in the 1960s. They showed that more durable responses could be achieved if treatment is given at the highest dose and frequency which toxicity permits [2,4]. Each dose kills a constant proportion (e.g. 90%) of tumour cells; therefore, to maximize the likelihood of cure, doses should be given early and at a high level. This mathematical thinking revolutionized treatment of paediatric leukaemia, wherein the VAMP regimen combined four chemotherapies at maximum dose and frequency [5].

The log-kill paradigm proved its utility in leukaemia, but translation to solid tumours observed the creation of a new modelling paradigm introduced by Norton and Simon in the 1970s [6]. Motivated by the observation that the proliferating fraction of cells is a function of tumour size, these authors revised the log-kill model. Maximizing cure within the new mathematical model indicates that residual disease should be treated at maximum dose and frequency. However, to mitigate toxicity trade-offs, the authors propose a back-loaded approach of gradual dose escalation so that residual disease coincides with maximal doses while minimizing toxicity breaks.

While Norton and Simon did indeed account for heterogeneity in proliferating fraction of cells at each point in time, they did not explicitly account for drug resistance [6]. Goldie and Coldman popularized a model in the 1980s that considered drug-resistance

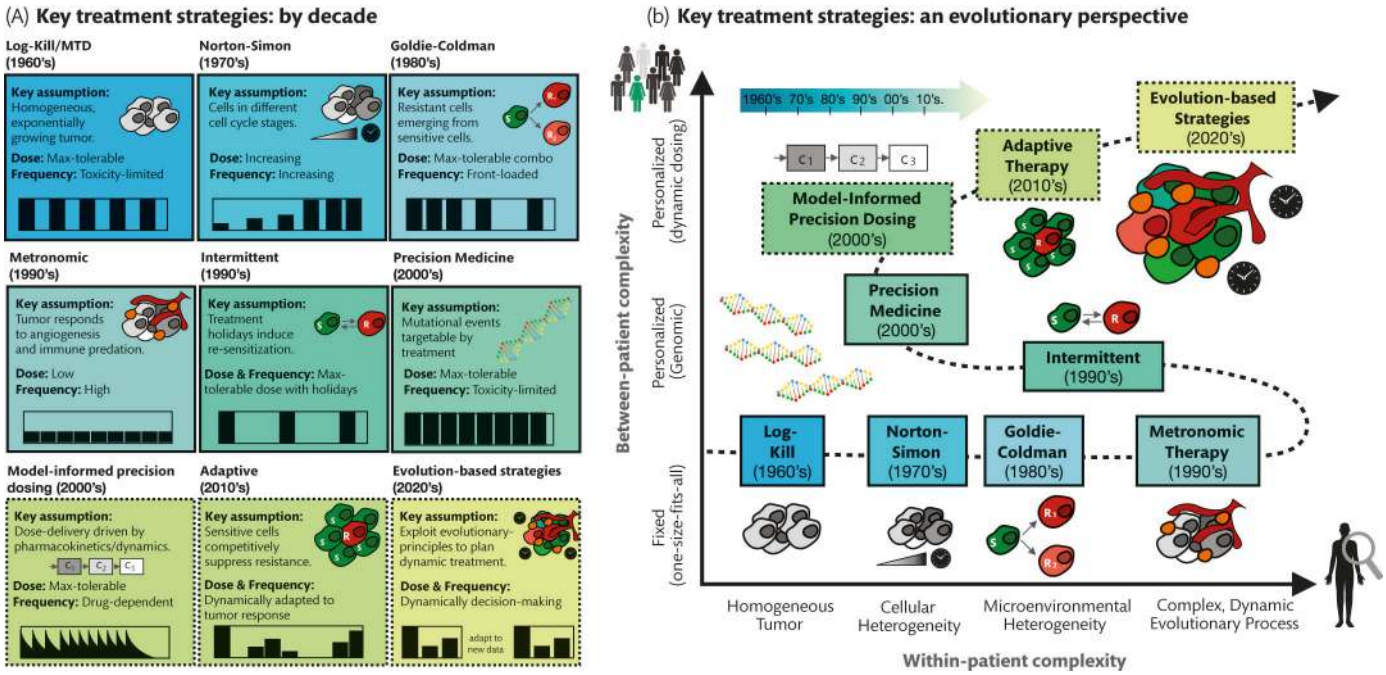


Figure 38.1. (A) Summaries of key treatment strategies by decade, illustrated schematically by black bars that represent treatment dosing. Boxes with solid outlines are static schedules in which the schedule is defined at the start of treatment, whereas boxes with dashed outlines represent dynamic schedules that are adjusted according to patient response. (B) The key treatment strategies can be categorized onto two axes: within-patient complexity (*x*-axis) and between-patient complexity (*y*-axis). As scientific understanding of tumour heterogeneity and evolution was refined, new approaches have increasingly sought to incorporate a more holistic picture of the tumour in its microenvironmental context (left-to-right trajectory). Other strategies have focused on patient-specific improvements to tailor schedules that account for individual patient differences (bottom-to-top trajectory).

Figure modified from Strobl et al [73] under a Creative Commons CC-BY licence.

heterogeneity. Their mathematical formulation indicated that avoiding acquired resistance is most likely when treating early and rapidly alternating multiple drugs.

The next leap forward in treatment scheduling introduced consideration of microenvironmental conditions. Folkman, Hannahan, Kerbael, and others observed that prolonged rest periods during maximum tolerable dosing may allow tumour vasculature to recover, driving subsequent recurrence. Instead, they proposed a low-dose metronomic paradigm that avoids such treatment breaks while increasing the anti-angiogenic effect of chemotherapies. In contrast, in a somewhat controverting fashion, research in different treatment and disease settings (e.g. hormone therapy or targeted drugs) suggested that resistance acquisition could be slowed or reversed during treatment breaks. This led to the development of intermittent treatment strategies, of which promising preclinical results [7–10] have not yet proved convincing in the clinic [11–15].

38.1.3. Accounting for inter-patient complexity

The history we have traced so far illustrates a maturing set of assumptions that increasingly focuses on the effects of cellular and microenvironmental heterogeneity. The ensuing decades shift focus to inter-patient variability. The fields of pharmacometrics and quantitative systems pharmacology have successfully used mathematical modelling to predict drug exposure, response, and toxicity, and are routinely used in modern industrial drug development [16,17]. One key challenge is addressing differences in physical, genetic, metabolic, or environmental factors (body weight, diet, concurrent medications, etc.) that drive significant pharmacokinetic or pharmacodynamic differences between patients. More recently, the fields of therapeutic drug monitoring or model-informed precision dosing have attempted dynamic adjustment of treatment dose for efficacy, safety, and toxicity [18–20]. These approaches have heavily relied upon mathematics, combining patient-specific data with drug pharmacokinetics to develop a more holistic and personalized medicine approach for designing treatment schedules.

38.1.4. Tumours as a complex, dynamic evolutionary process

In reviewing the historical timeline above, treatment scheduling protocols have undergone steady refinement decade by decade to account for either heterogeneity at the tumour scale or variability on the patient scale. Yet, up until the later part of the first decade of the millennium, missing from the personalized medicine revolution was the combination of both: evolution-based treatment strategies to directly address tumour heterogeneity and resistance but which also address heterogeneity at the patient population level.

To rectify this, Gatenby et al. introduced adaptive therapy, a method of maintaining a sizeable tumour population of treatment-sensitive cells for a prolonged period of time in order to suppress any resistant populations that exist in small numbers [21]. No different than many of the other advancements in treatment scheduling protocols, adaptive therapy protocols were designed in coordination with mathematical modelling, which encapsulated principles of cancer evolution to derive new adaptive algorithms and test hypotheses from clinical data [22–24].

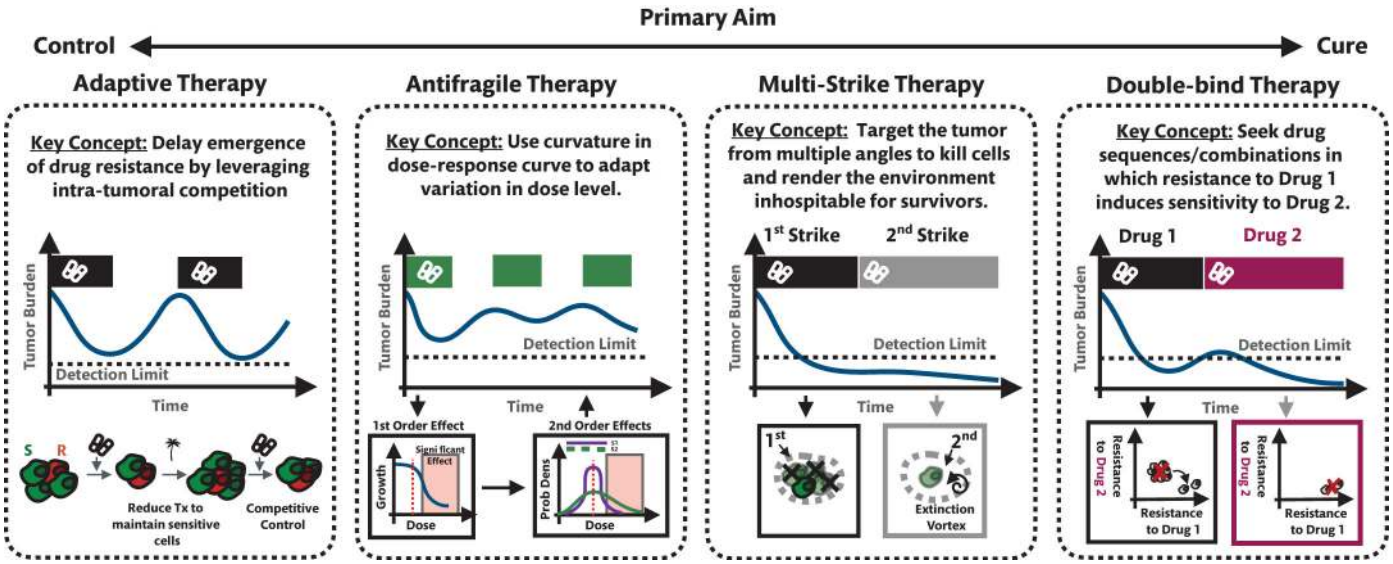


Figure 38.2. Evolutionary therapy strategies. The primary goal of treatment shifts from tumour control to disease cure from left to right.

In the next section, we describe the fundamentals of designing an evolutionary therapy, providing several canonical examples that broadly classify along a spectrum of two goals: cure vs. control (Figure 38.2). Some of these therapeutic paradigms, such as adaptive therapy or multi-strike therapy, have been successfully implemented in clinical trials. Others, such as antifragile therapy or double-bind therapy, are in the preclinical development stage, aided through the construction of mathematical models to list key assumptions and define hypotheses to be tested and validated in the near future.

38.2. The fundamentals of evolutionary therapy for control

38.2.1. Adaptive therapy

Adaptive therapy is a dynamic cancer treatment protocol that modifies treatment decisions based on a tumour's growth and treatment response dynamics. The goal is to maintain a stable tumour burden by retaining treatment-sensitive cells to compete with and suppress the outgrowth of treatment-resistant cells. On treatment, the sensitive cells die off and allow the resistant population to expand, while off treatment the sensitive cells rebound faster, often due to a growth advantage of sensitive cells. So adaptive therapies, in general, work by changing the dosing to control the larger sensitive population and thus prevent selection for resistant populations that are ultimately not controllable.

With continuous treatment, there may be an initial response, but any pre-existing resistant cells will eventually be selected for or may be acquired (Figure 38.3A). Fortunately, a strategy was devised countering similar behaviour in the agriculture industry with the failure to eradicate destructive pests [25]. The widespread use of synthetic pesticides in the 1940s in farming progressed like an evolutionary arms race as more resistant species to survive these pesticides evolved in parallel with the development of newer and better compounds to kill them [26,27]. The solution to this problem was integrated pest management [28,29], in which the goal was modified from complete eradication of pests to instead achieve control below damaging levels to the crops. By acknowledging that insecticides may promote the selection for more resistant bugs, and that more sensitive species are an important source of control by competition, pesticides are used more sparingly and the field is kept in a more manageable condition. Similar dynamics may occur with the large, proliferative, and invasive populations within heterogeneous tumours that originally respond but eventually become resistant to current therapies [25].

The concept for adaptive therapy applied to cancer was originally presented as an information theory formalism to account for selection for subpopulations and tested in ovarian cancer cell lines with the chemotherapy drug, carboplatin, in mice [30]. The theory was developed subsequently alongside preclinical experiments [31,32] that led to the first clinical trials in metastatic castrate-resistant prostate cancer (NCT02415621) [33,34]. In the years following, this evolution-based treatment approach has been implemented into clinical trials for several different cancers, including BRAF-mutant melanoma (NCT03543969), metastatic castrate sensitive prostate cancer (NCT03511196) [35], thyroid cancer (NCT03630120), ovarian cancer (NCT05080556), and basal cell carcinoma (NCT05651828). Open questions remain in designing useful adaptive therapies [36], but practically, for successful adaptive therapy outcomes in the clinic, it must have a useful tumour burden biomarker to facilitate decision-making. Serologic biomarkers are relatively cheap and can be noninvasively collected over time, like prostate-specific antigen (PSA) for prostate cancer [33–35] or lactate dehydrogenase for advanced melanoma [37]. In other cancers, imaging or circulating DNA markers may be used for burden quantification, but not all cancers have easily accessible ways to track tumour burden dynamics.

Competition and interactions between species are important factors in the success of adaptive therapy. Competitive release refers to the expansion of a species that is usually restricted due to competition with another species when that species is removed. Long cyclers in metastatic prostate cancer were found to have a large and

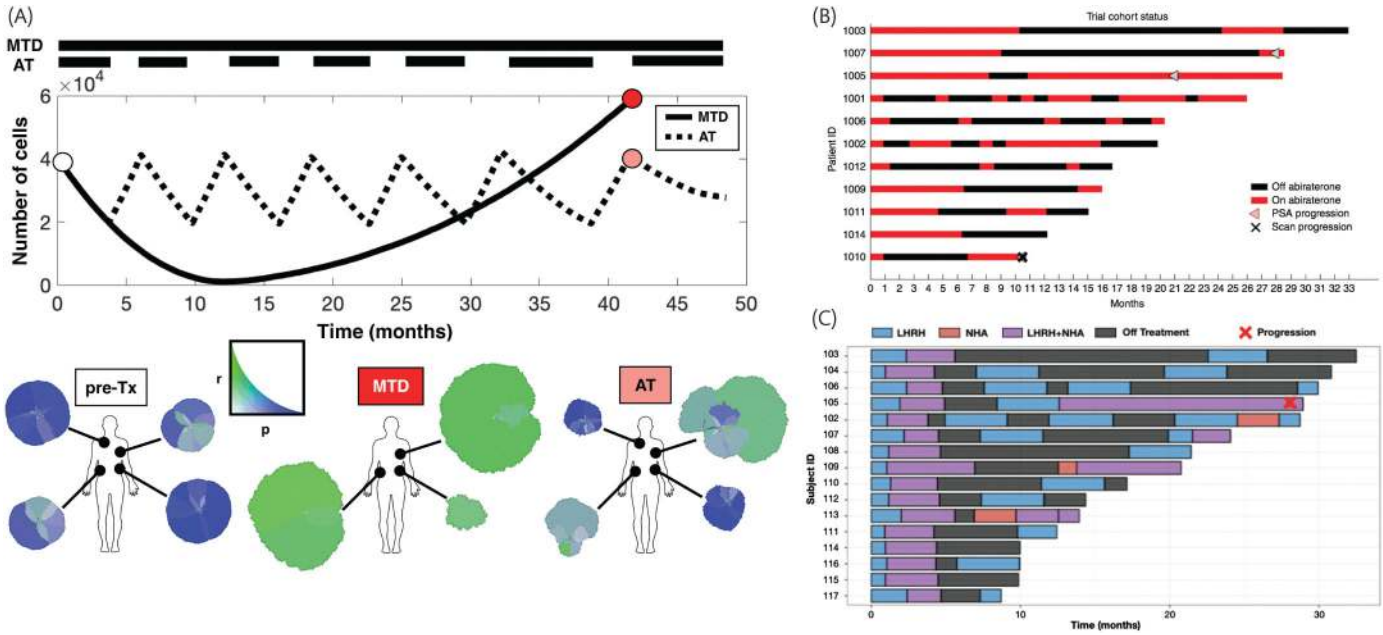


Figure 38.3. Adaptive therapy. (A) Heterogeneous tumours treated with continuous treatment eventually select for resistant cells, while adaptive therapy can delay the outgrowth of resistant cells by keeping more sensitive cells around for competition. (B) Treatment-resistant prostate cancer trial and (C) treatment-sensitive prostate cancer trial swimmer plots showing patient-specific schedules with treatment vacations (black bars) and treatment application periods (non-black bars).

asymmetrical competitive effect of sensitive cells on resistant cells [34]. Trade-offs between therapy effectiveness and competitiveness between cell subtypes in the absence of therapy favour adaptive therapies over treat to cure [36]. Many of these models also assume that there is a trade-off between proliferation and resistance such that some modes of resistance require sufficient energy that slows proliferation. This may account for the often larger numbers of sensitive cells. However, trade-offs like this are not always present [38], nor necessary for the success of adaptive therapy as long as there is competition and a limited carrying capacity [39].

Other factors that influence the efficacy of adaptive therapy include tumour burden, space, and plasticity. Adaptive therapy maintains high levels of tumour burden to benefit from the competitive suppression of treatment-sensitive subpopulations on treatment-resistant subpopulations [36]. This large burden may lead to pain and other symptoms, or this large sensitive pool may lead to more possibilities of cells becoming resistant or metastatic. Hansen and Read suggest that the threshold burden may need to be adjusted for optimal control [40]. Further, space may play an important role in competition as a limited resource. If there is a trade-off, more proliferative cells may take over most of the space and leave the rarer, more resistant subtypes to be contained within the interior [41,42]. Yet, resistant phenotypes can also compete with each other [43], and different anatomical architectures present different selective pressures on cells [44]. Spatial context may have an effect on cell–cell interactions. The effect of competition and trade-offs may lessen if cells are not in direct contact, which could occur from separation into clusters within the tumour [43], be due to dispersion from migration/invasiveness [40] or distributed among individual metastases [45]. Plasticity or switching between sensitive and resistant states has been shown to benefit adaptive therapy control if the sensitive cell growth rate is slow [37], but more phenotypic drift when there is a trade-off between proliferation and resistance can reduce the effectiveness of adaptive therapy [41].

Many possible dynamic treatment protocols can be termed adaptive therapy. However, adaptive therapy differs from intermittent therapy, which tends to have either regular treatment intervals and vacations or switches at fixed burden values, while adaptive therapy cycles are determined by the treatment response and regrowth dynamics, and switch points are generally patient specific. Maintenance therapies can also be given for long-term, low-dose, low-toxicity control, but they also tend to be fixed and patient generic [46]. The most common adaptive protocols for single drugs that have been tested in clinical, experimental, and theoretical settings are treatment vacations and dose modulation. For treatment vacations, treatment is applied until the tumour burden decreases to a certain threshold and then treatment is withdrawn until the burden reaches its initial value. This kind of protocol was used in metastatic castrate-resistant prostate cancer (**Figure 38.3**B) [33] as it was simpler and more practical to take or not take a daily pill. Dose-modulation strategies instead adjust the dose to decrease when the tumour burden decreases and increase when the burden increases to maintain a steady burden. These two strategies were compared using chemotherapy for breast cancer in mice [31]. They found that both adaptive therapies had less toxicity but still maintained control but the dose-modulation better preserved the vasculature, which aids drug delivery, similar to the low-dose metronomic schedule. These schedules with an anti-proliferative agent were compared in a spatial agent-based model with focus on tumour heterogeneity and not the microenvironment [41]. They found that the dose skipping allowed better control for subsequent cycles, while dose modulations selected for more and more resistant cells over time. However, models and experiments designed by Strobl et al. to treat ovarian cancers with poly (ADP-ribose) polymerase (PARP) inhibitors adaptively showed that treatment vacations gave up too much control during regrowth, so a small maintenance dose was needed for control [47]. In that case, a stepping algorithm was designed to switch between high and low

similar to dose-modulation switching in [31] but more simplified and therefore probably more clinically tractable. Other modifications include adjusting the 50% switch point [37,40,48] and cycling treatment within patient-specific bounds [49].

With just a single drug, the goal is often to shift between tumour control with toxicity to restoring sensitivity. However, with more drugs, the number of possible treatment combinations, doses, and timing increases dramatically. West et al. consider how fixed treatment plans can be selected to drive tumours into repeatable evolutionary cycles by carefully choosing the ordering and timing of treatment combinations based on the dynamics of the subpopulations and their interactions [50,51]. In an adaptive therapy clinical trial for metastatic castrate-sensitive prostate cancer, androgen deprivation treatment is used alongside another novel hormonal agent to prevent castrate resistance. The treatment combination for each time point in that case is determined by both the dynamics of the tumour burden (PSA) and the testosterone level [35] (Figure 38.3C). Other strategies, for multidrug adaptive therapy, that have been considered by Thomas et al. include dose-modulation combinations, dose-modulation alternating drugs, and switching drugs on progression [52].

38.2.2. Antifragile therapy

This field of antifragile research describes and quantifies the fragility of systems in response to external volatility or perturbations. Many such systems (financial systems, manufacturing processes, organization structures, etc.) are fragile to external stressors, i.e. external stressors can negatively impact the system. The key insight of this field and its application to evolutionary therapy in cancer, however, is to recognize that some systems respond to external stressors in a positively beneficial manner. For example, evolution by natural selection can lead to a greater population-level fitness in the presence of environmental stressors [53]. Applying an antifragile strategy to cancer treatment requires an understanding of how the dosing is personalized with the optimal dose rate and timing.

These features can be understood within the context of adaptive therapy. A keystone of the first prostate cancer adaptive trial is the degree of personalization. The 50% threshold for stopping therapy led to differences in the timing of drug holidays between patients. The algorithmic design of the trial allows the schedule to be responsive to inter-patient differences in tumour dynamics. Personalization (indirectly) leads to a second keystone feature of the trial: a reduction in the cumulative dose delivered within the patient's treatment course. This second feature has a distinct difference: it is an emergent property of the algorithm. The cumulative dose is not prescribed *a priori*, but it is an emergent metric that is the result of how the 50% threshold algorithm interacts with the patient's tumour response over time. Figure 38.4 illustrates the emergent dosing metrics in the metastatic castration-resistant

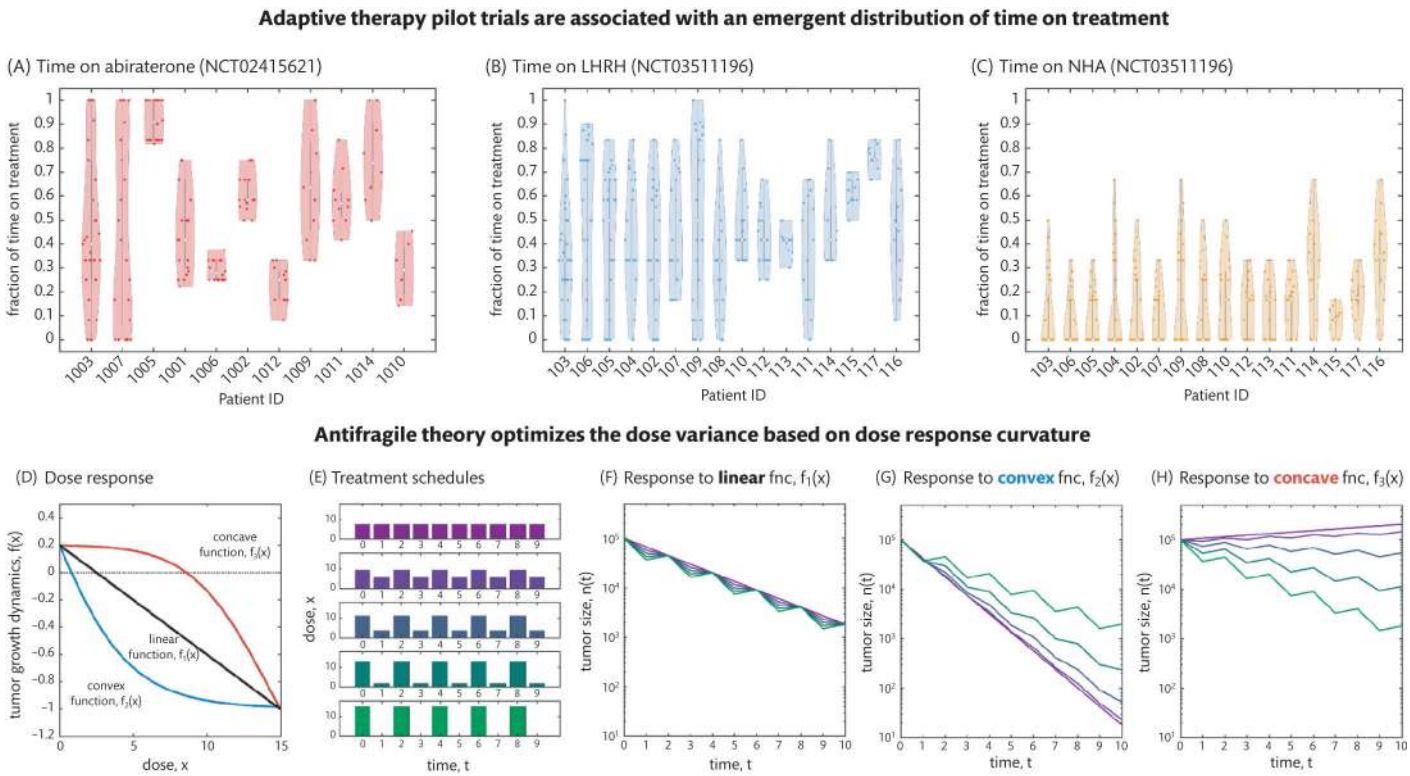


Figure 38.4. Panels (A)–(C) illustrate dosing data from the two adaptive pilot trials in Figure 38.3B (panel (A)) and Figure 38.3C (panels (B) and (C)). Shown is the distribution of time on treatment (the 12-month rolling average of time on treatment, per patient). Each individual patient has an *emergent* mean and variance associated with the adaptive treatment algorithm. White dot shows median and grey bars show 25th and 75th percentiles. Panels (D)–(H) illustrate the utility of antifragile theory, where optimal treatment schedule variance depends on the shape of the dose response curve. (D) Example dose–response curves: convex (blue), concave (red), and linear (black). (E) Example treatment schedules ranging from low variance (even dosing; purple) to high variance (uneven dosing; green). (F) With a linear response curve (panel (D), black), all schedules result in equal control. (G) With a convex curve (panel (D), blue), lower dose variance results in better control, and (H) with a concave curve (panel (D), red), higher dose variance results in the best tumour control.

(Figure 38.4A) and castration-sensitive (Figure 38.4B and C) trials that are shown in Figure 38.3B and C, respectively.

Arguably, it is difficult to determine which of these keystone features is most important: (1) the degree of dose personalization or (2) the reduction in cumulative dose. We have taken steps to address this question by disentangling these first- and second-order effects in treatment scheduling. First-order effects are those effects that are driven by the average/mean dose delivered (or the cumulative dose). Second-order effects depend on the variance of dose delivered over time (the timing of treatment). As stated above, adaptive therapy alters both first-and second-order characteristics of a schedule (cumulative dose and dose timing). First- and second-order effects can be quantified by viewing the dose–response function through the lens of antifragile theory, which we have previously applied to cancer by measuring dose–response convexity to predict optimal treatment scheduling [54,55].

38.2.2.1. First-order effects

Firstly, let's assume that the dose–response function, $f(x)$, is a *decreasing* function of dose (some example decreasing functions are shown in Figure 38.4D). Each function describes the per capita growth rate of a tumour as a function of the dose concentration, x:

$$\frac{\dot{n}}{n} = f(x)$$

Decreasing functions, $f(x)$, of dose concentration indicate that increasing the mean dose delivered always leads to *greater tumour reduction*. Stated in an alternative way, the first-order effects are positive (beneficial) with respect to tumour reduction. First-order effects are determined by the first derivative of the dose–response function, $f'(x)$. If this slope is decreasing, increasing the mean dose is beneficial (and vice versa). Thus, first-order effects determine the optimal mean dose delivered based on whether the derivative of the response function is increasing or decreasing.

Positive first-order effects are commonly seen in mathematical models of treatment response. The aforementioned Norton–Simon mathematical model developed in the 1970s assumes that the effect of treatment is *linearly* proportional to the instantaneous growth rate of a tumour. It is a well-known result that tumour size at time t is proportional to the cumulative dose delivered [56].

38.2.2.2. Second-order effects

We have now described the first-order effect as a function of the slope of the dose response. Next, we illustrate a second-order effect as a function of the curvature (the second derivative) of the dose response. Three example functions are shown in Figure 38.4D: linear (black), convex (blue), and concave (red) dose–response functions.

As the first derivative of the response function determines the optimal mean dose delivered, the curvature (linear, concave, and convex) of the dose response determines the optimal dose variance. Example treatment schedules are shown in Figure 38.4E, ranging from low variance, continuous therapy (purple) to a high variance, intermittent therapy (green). To control for confounding first-order effects, here we consider only treatment schedules with an identical mean dose delivered.

If the dose response is linear (Figure 38.4D, black), then all these treatment schedules lead to the same outcomes (Figure 38.4F)—indicating the absence of any second-order effects. Again, referring back to the Norton–Simon model, we see that linear dose–response functions are common in literature. In these models, rearranging the pattern of dosing has no effect: front-loaded dosing is as effective as back-loaded dosing (and every other arrangement) as long as the cumulative dose is identical[1]. There is a mathematical rule that can be employed to arrive at this result simply: because tumour growth is a linear function of dose, x, so second-order effects become null and all schedules with equivalent cumulative dose lead to equivalent outcomes (e.g. see Jensen's inequality).

In contrast, second-order effects are negative for convex dose–response functions (Figure 38.4D, blue). The result in Figure 38.4G illustrates that it is most beneficial to reduce dose volatility by employing the purple, continuous treatment strategies. The opposite is also true: in the event of a concave dose–response function (Figure 38.4D, red), it is most beneficial to maximize dose volatility by employing the green, intermittent treatment strategy (Figure 38.4H).

38.2.2.3. Second-order effects in evolutionary therapy

Thus far, we have illustrated a series of simple and contrived fixed treatment schedules. However, every treatment schedule has an associated mean dose delivered and dose variance. The adaptive algorithm has an emergent effect on altering each individual patient's mean and variance. More research is required to determine how to optimize mean and variance for adaptive algorithms, and antifragility theory may provide a path forward.

Adaptive therapy was designed with the purpose of leveraging cell–cell interactions for patient benefit: treatment-sensitive cells suppress treatment-resistant cells in the absence of treatment. Using a theoretical modelling approach, we have recently shown that cell–cell interactions also influence second-order effects in cancer treatment [57]. Mathematical modelling has shown that pharmacokinetic clearance can expand the range of doses over which second-order effects are beneficial for tumour reduction and mitigating treatment-induced cachexia [58].

38.3. The fundamentals of evolutionary therapy for cure

38.3.1. Multi-strike therapy

Multi-strike therapy is an evolutionary therapy approach that emulates the dynamics of Anthropocene extinction events [59]. What is often seen in such events is that a number of factors contribute to the extinction by occurring concurrently or in close sequence. Furthermore, none of these factors would be able to cause the extinction individually. In the case of the passenger pigeon (E. migratorius), which went extinct in 1914, the coincidence of multiple factors led to its demise. The pigeons numbered in the billions, but hunting by humans reduced their numbers significantly over the course of a century. The decline was further exacerbated by the deforestation of land that fragmented the habitat; the species' dependence on

1. Note: We do not wish to imply that the back-loading strategy proposed via the original Norton–Simon model was incorrect: the dose backloading was proposed as a strategy to achieve the minimum tumor size *at any point in time*, not just the final tumor size. Both are mathematically valid points.

large social flocks for breeding was highly sensitive to this ecological change. These factors, coupled with a natural tendency for the species to experience large fluctuations in population numbers even under favourable conditions, combined to eliminate all members of what was North America's most abundant bird. Interestingly, it is thought that none of these factors alone would have necessarily led to extinction; it was the combination that led to the outcome.

Today, most conservation efforts seek to avoid extinctions of species by preserving habitats, limiting hunting, and similar measures. In oncology, however, the idea that a cancer is a 'species' that can be made extinct within the habitat of the patient is a potential way to design therapies aiming for cure. By using lessons learned from Anthropocene extinctions, we can combine and sequence available agents in ways that affect both the cancer cells and the tumour microenvironment similar to the illustration above. Tumour-cell killing needs not be the only consideration when treating a patient. Indeed, one can see analogies between the pigeons and tumours: hunting and cytotoxic drugs like chemotherapies; deforestation and agents like anti-angiogenics; and population-level breeding dependencies and tumour growth within the microenvironment.

Unlike adaptive therapy, where the guiding philosophy is preventing the development of drug resistance for as long as possible, in multi-strike therapy the aim is maximal tumour burden reduction with available agents. Ideally, the interventions are sequenced for maximal synergy, and unlike most standard-of-care approaches, switching should occur before observation of any progression (i.e. progressive disease via RECIST or biomarker increase).

Multi-strike therapy is typically applied when the following conditions exist: (1) for cancers where standard of care does not lead to cure in the great majority of cases; (2) there are multiple available treatment modalities and agents for the disease, with varied modes of action; (3) these agents can have a significant effect on reducing tumour burden and/or affecting the tumour microenvironment. These three conditions can vary during the course of a given disease. For example, early stage breast cancer is highly curable with current practice; point 1 would not be satisfied. Mid- to late-stage breast cancers often have lower rates of cure, have numerous available agents, and in many cases these agents have significant observable efficacy. This phase of disease would be a prime target for implementing multi-strike therapies. Very late-stage breast cancer, while mostly incurable (point 1) and potentially having multiple agents to select from (point 2), often displays very little response to therapy, and therefore point 3 would limit the effectiveness of multi-strike therapy.

As with any therapeutic approach, there are both opportunities and challenges. The most significant advantage is the potential to induce cures in patients who would likely see disease progression under standard of care. A sequence of closely scheduled strikes has the potential to reduce the tumour volume below a minimum viable disease (MVD) threshold, whereas the same sequence delivered sequentially only upon evidence of progression for each strike might not (see **Figure 38.5**). Conceptually, MVD is the threshold below which the tumour will go extinct. At minimum, MVD is one cell, representing 100% killing of all tumour cells. However, MVD can be greater than 1 cell when considering other mechanisms affecting tumour dynamics. For example, the immune system is known to fight tumours and generally would have more leverage over smaller tumours. In principle, if a series of strikes reduces the tumour burden

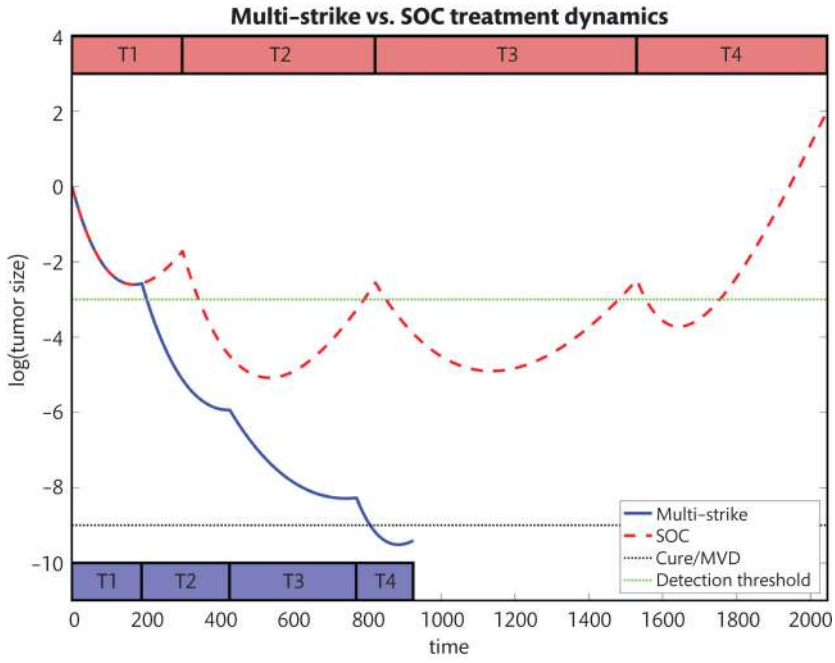


Figure 38.5. Illustration of multi-strike therapy versus standard of care for the same sequence of four treatment agents (T1–T4). The red boxes and line plot show the schedule and tumour size respectively for standard of care, where treatment is switched when either (A) the tumour diameter grows by greater than 20% from nadir (seen at the end of T1), or (B) when a tumour driven to NED becomes detectable again (shown at the end of T2 and T3). The blue boxes and line plot show a multi-strike approach, where the treatments are switched near the nadir of tumour volume. The treatments have the same efficacy in both approaches, but the early switching leads to a deeper tumour nadir, and in this case that nadir drives the tumour below the level of minimum viable disease.

below a certain point, the immune system may be able to kill the remaining cells after the efficacy of the therapy has concluded. Similarly, spatial fragmentation of the tumour, changes to the blood supply, and other microenvironmental shifts may lead to MVD sizes greater than 1 cell.

Another advantage of multi-strike therapy is that by cycling agents more quickly and for shorter periods of time, the potential to reuse agents at a later date becomes more likely. This is because selection pressures for resistance mechanisms are generally higher the longer a treatment continues.

Multi-strike therapy has some disadvantages as well. Designing sequences and combinations that will maximize the overall regimen efficacy is challenging given the lack of clinical data on using multiple agents in this fashion. Another issue arises due to the limits of tumour detection via current imaging and biomarkers. If the patient's tumour burden falls below the threshold of detection, the timing of subsequent strikes will not be informed by the dynamics of tumour burden, and therefore the schedule will have to be applied in a less patient-specific way. One may miss the optimal switching windows. A third challenge is that of toxicity: sequencing therapies closer together has the potential to increase side effects and adverse events due to dose densities.

Our first approach to delivering the needed decision support for multi-strike regimens has been implemented via an 'Evolutionary Tumour Board'. In this approach [60], we work with individual incurable patients to identify multi-strike therapy opportunities. We developed mathematical models for several diseases of interest; collected detailed patient data regarding the dynamics of their disease; studied retrospective patients within the same disease cohort to understand tumour growth and response to treatment; and used clinical trial data to constrain our predictions based on population-level outcomes for each agent available to the patient.

Growing interest in the eco-evolutionary dynamics of cancer, coupled with continuing improvements in both clinical care and diagnostic technology, hold great promise for the future of implementing multi-strike therapy. Preclinical work is underway to explore these approaches for different agents, shedding light on mechanism and drug interactions. The advent of blood-based biomarkers, such as circulating tumour cells and circulating DNA, may lower our thresholds of detection for tracking tumour burden as well as give insight into tumour and stromal subpopulations and their composition. Furthermore, being less invasive than biopsies and often less costly than imaging modalities, the potential for high-frequency and high-resolution data collection is a significant opportunity that multi-strike therapy is poised to leverage.

38.3.2. Double-bind therapy

As stated previously, adaptive therapy was designed and implemented based on the premise of a cost to developing resistance mechanisms. Resistant subpopulations divert resources to resistance mechanisms that would often otherwise be devoted to proliferation [22,61,62]. This allows exploitation of this cost of resistance through smart allocation of treatment to maintain treatment-sensitive competitors that capitalize on the cost.

By drawing lessons from applied ecology, Gatenby, Brown, and Vincent proposed an alternative paradigm that incorporates the cost of resistance into therapeutic strategies: an evolutionary double bind [63]. In ecology, species that develop compensatory behaviours in avoidance of natural predators often come at a trade-off with fecundity. For example, in response to a predatory hawk, mice may limit foraging behaviour patterns, leading to less food intake and subsequently fewer offspring. Predator avoidance causes an unavoidable cost in reproductive fitness, which may have a net effect as great as the lethality of predation [64]. This trade-off between adaptation and fecundity has been termed an evolutionary double bind.

A single biological agent may be insufficient to produce a true evolutionary double bind, owing (for example) to a heterogeneous response to the agent, or insufficient loss of fitness. This is evident in the field of pest management, where a number of predators, pathogens, or treatments may still be insufficient for controlling pests [28,29]. Thus, it has been suggested to rely on 'predator facilitation' as an effective method for producing an evolutionary double bind. A classic example in ecology is desert rodents that naturally hide within the safety of shrubs in response to owl predation. However, predatory snakes lie in waiting underneath these shrubs to ambush the rodents, completing the double bind. Similarly, in ovarian cancer care, it may be possible to induce sensitivity to chemotherapy or radiation therapy through the use of PARP inhibitors [65].

This concept has other exciting parallels in cancer treatment: designing a treatment protocol where a first-line therapy induces an evolutionary double-bind response within the tumour that can be exploited using a second-line therapy. First-line therapy resistance mechanisms can produce an increased sensitivity to a second treatment or produce specific vulnerabilities that are targetable. Another possible example is that first-line epidermal growth factor receptor (EGFR) inhibitors (e.g. gefitinib) induce resistance mediated by T790M mutations that can be targeted by third-line inhibitors (e.g. osimertinib) [66]. Our group has shown the potential synergistic double bind between radiation therapy [72]. Radiation therapy induces double-strand DNA breaks, and thus resistance is mediated by cells that up-regulate DNA damage response pathways. This up-regulation results in an increased natural killer cell ligand expression on tumour cells, which we hypothesize would induce increased sensitivity to NK cell-based adoptive therapy.

A promising strategy for identifying double bind in cancer is confront cell lines with evolved resistance to front lines with a suite of possible candidate second-line drugs, to identify collaterally-sensitive options (**Figure 38.6**A). Collateral sensitivity has been used to target multi-drug resistance [67,68] and has demonstrated promising results in targeting antibiotic resistance [69,70]. However, it is important to note that this second therapy may only be effective in the context of the double bind but much less so when given as an initial front-line treatment. Basanta et al., and evolutionary game theoretic model of the synergistic effect between a p53 cancer vaccine and chemotherapy, showed that chemotherapy first led to optimal outcomes but not the reverse [71]. As shown in Figure 38.6B, properly leveraging a double bind requires inducing the desired resistance (to drug A) mechanism before applying the synergistic drug B. Reversing the order negates any advantage of the double bind (Figure 38.6C).

38.4. Discussion

The history of mathematical modelling in cancer treatment can be traced back to the 1960s and, just as today, these models were driven

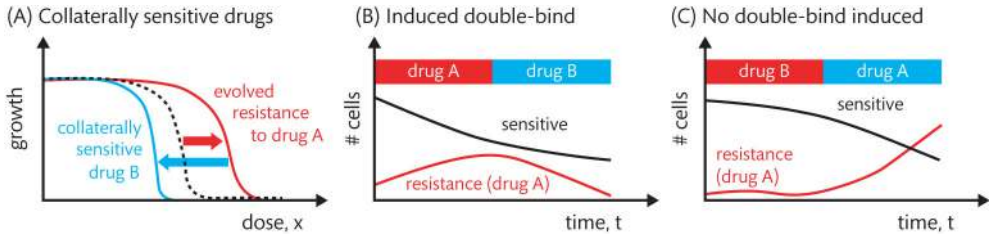


Figure 38.6. A promising strategy for identifying potential evolutionary double-bind drugs. (A) Screen-evolved resistance cell lines (red) across a range of drugs to identify collaterally sensitive options (blue). Treatment naive cell line is shown in black. (B) An evolutionary double bind is induced when drug B is applied after resistance to drug A occurs. (C) Reversing the order of drugs negates any advantage of the double bind.

by the biological understanding of cancer at the time they were developed. At the turn of the millennium, the role of tumour heterogeneity emerged as part of the molecular revolution and along with it came a greater understanding of the impact of drug resistance as well as the role of the microenvironment. Surprisingly, it's only been in the past decade that the evolutionary nature of cancer has begun to impact clinical decisions. Importantly, this has forced us into making a critical but realistic decision concerning patient treatment: cure or control.

While historically the goal of all cancer treatment has been to aim for cure, evolutionary therapies also embrace control that may be a more realistic goal when metastatic disease is being treated. Adaptive therapy is a control strategy and the metastatic prostate cancer trial has, so far, been the only evolutionary clinical trial (NCT02415621) to be completed and showed significant benefits over the standard of care [33]. There are multiple adaptive trials actively underway in a range of cancers, including prostate (NCT03511196 and NCT05393791), ovarian (NCT05080556), melanoma (NCT03543969), basal cell carcinoma (NCT05651828), and the Evolutionary Tumor Board (NCT03543969). The latter three trials are all actively incorporating mathematical modelling as a decision support tool rather than using a simple rule of thumb.

When we consider evolutionary trials that are more focused on cure or extinction (NCT04388839 and NCT05189457), the role of mathematical modelling in terms of treatment decisions becomes even more important. How drugs should be sequenced (or combined), at what dose and duration, and when should switch points be made are all important questions in developing a treatment strategy for cure that cannot easily be intuited. Mathematical modelling is emerging as a key player in helping to forecast the impact of specific treatment decisions, and getting these predictions into the hands of clinicians in easy to understand and practical ways is still very much a work in progress. We have articulated some of this process through our recent paper on the first year of the Evolutionary Tumor Board [60], but many questions remain. For example, (i) how best to combine patient data from different spatial and temporal scales; (ii) how often should data be collected on a patient when there are financial and practical trade-offs; (iii) how frequently can we ask a patient to practically alter treatment; (iv) how statistically confident do we need to be in a predicted decision to share it with an oncologist; (v) how best to balance model complexity with patient data to calibrate it; (vi) since any model is an abstraction of reality should we generate multiple models and make ensemble predictions.

What has become clear is that, while evolution-based therapeutic strategies are only just emerging in the clinic through mostly pilot clinical trials, there is so much we do not fully understand, illustrating the need to further investigate these ideas *in silico*, *in vitro*, *in vivo*, and in the clinic.

Acknowledgements

The authors gratefully acknowledge funding by the National Cancer Institute via the Cancer Systems Biology Consortium (CSBC) U01CA232382, the Physical Sciences Oncology Network (PSON) U54CA193489, and support from the Moffitt Center of Excellence for Evolutionary Therapy.

REFERENCES

1. Gunawardena J. Models in biology: accurate descriptions of our pathetic thinking. BMC Biol. 2014;**12**(1):29. doi: 10.1186/1741-7007-12-29.
2. Skipper HE, Schabel FM. Experimental evaluation of potential anticancer agents. VII. Cross resistance of alkylating agent-resistant neoplasms. Cancer Chemoth Rep. 1962;**22**:1–22.
3. Merlo LMF, Pepper JW, Reid BJ, Maley CC. Cancer as an evolutionary and ecological process. Nat Rev Cancer. 2006;**6**(12):924–35. doi: 10.1038/nrc2013.
4. Skipper HE. The effects of chemotherapy on the kinetics of leukemic cell behavior. Cancer Res. 1965;**25**(9):1544–50.
5. Lichtman MA, Spivak JL, Boxer LA, Shattil SJ, Henderson ES. Hematology. Section 3: Advances in treatment of hematologic disease. 2000. p. 655–57. doi: 10.1016/b978-012448510-5.50152-7.
6. Norton L, Simon R, Tumor size, sensitivity to therapy, and design of treatment schedules. Cancer Treat Rep. 1977;**61**(7):1307–17.
7. Akakura K, Bruchovsky N, Goldenberg SL, Rennie PS, Buckley AR, Sullivan LD, Effects of intermittent androgen suppression on androgen-dependent tumors. Apoptosis and serum prostate-specific antigen. Cancer. 1993;**71**(9):2782–90. doi: 10.1002/1097-0142(19930501)71:9<2782::aid-cncr2820710916>3.0.co;2-z.
8. Bruchovsky N, Rennie PS, Coldman AJ, Goldenberg SL, To M, Lawson D. Effects of androgen withdrawal on the stem cell composition of the Shionogi carcinoma. Cancer Res. 1990;**50**(8):2275–82.
9. Thakur MD, et al. Modelling vemurafenib resistance in melanoma reveals a strategy to forestall drug resistance. Nature. 2013;**494**(7436):251–55. doi: 10.1038/nature11814.
10. Smalley I, et al. Leveraging transcriptional dynamics to improve BRAF inhibitor responses in melanoma. Ebiomedicine. 2019;**48**:178–90. doi: 10.1016/j.ebiom.2019.09.023.

11. Bruchovsky N, et al. Final results of the Canadian prospective phase II trial of intermittent androgen suppression for men in biochemical recurrence after radiotherapy for locally advanced prostate cancer: clinical parameters. Cancer. 2006;**107**(2):389–95. doi: 10.1002/cncr.21989.
12. Maha H, et al. Intermittent versus continuous androgen deprivation in prostate cancer. New Engl J Med. 2013;**368**(14):1314–25. doi: 10.1056/nejmoa1212299.
13. Hussain M, Tangen C, Higano C, Vogelzang N, Thompson I. Evaluating intermittent androgen-deprivation therapy phase III clinical trials: the devil is in the details. J Clin Oncol. 2015;**34**(3):280–85. doi: 10.1200/jco.2015.62.8065.
14. Lee JH, Carlino MS, Rizos H. Dosing of BRAK and MEK inhibitors in melanoma: no point in taking a break. Cancer Cell. 2020;**38**(6):779–81. doi: 10.1016/j.ccell.2020.11.010.
15. Kuczynski EA, Sargent DJ, Grothey A, Kerbel RS. Drug rechallenge and treatment beyond progression—implications for drug resistance. Nat Rev Clin Oncol. 2013;**10**(10):571–87. doi: 10.1038/nrclinonc.2013.158.
16. Benzekry S. Artificial intelligence and mechanistic modeling for clinical decision making in oncology. Clin Pharmacol Ther, 2020;**108**(3):471–86. doi: 10.1002/cpt.1951.
17. Bruno R, et al. Progress and opportunities to advance clinical cancer therapeutics using tumor dynamic models. Clin Cancer Res. 2020;**26**(8):1787–95. doi: 10.1158/1078-0432.ccr-19-0287.
18. Beumer JH, Without therapeutic drug monitoring, there is no personalized cancer care. Clin Pharmacol Ther. 2013;**93**(3):228–30. doi: 10.1038/clpt.2012.243.
19. Darwich AS, et al. Why has model-informed precision dosing not yet become common clinical reality? Lessons from the past and a roadmap for the future. Clin Pharmacol Ther. 2017;**101**(5):646–56. doi: 10.1002/cpt.659.
20. Ferrer F, Fanciullino R, Milano G, Ciccolini J. Towards rational cancer therapeutics: optimizing dosing, delivery, scheduling, and combinations. Clin Pharmacol Ther. 2020; **108**(3):458–70. doi: 10.1002/cpt.1954.
21. Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive therapy. Cancer Res. 2009;**69**(11): 4894–903. doi: 10.1158/0008-5472.can-08-3658.
22. Gatenby RA. A change of strategy in the war on cancer. Nature. 2009;**459**(7246):508–509. doi: 10.1038/459508a.
23. Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive therapy. Cancer Res. 2009;**69**(11):4894–903.
24. Enriquez-Navas PM, Wojtkowiak JW, Gatenby RA. Application of evolutionary principles to cancer therapy. Cancer Res. 2015;**75**(22):4675–80.
25. Gatenby RA. A change of strategy in the war on cancer. Nature. 2009;**459**(7246):508–9. doi: 10.1038/459508a.
26. Denholm I, Rowland MW. Tactics for managing pesticide resistance in arthropods: theory and practice. Annu Rev Entomol. 1992;**37**(1):91–112. doi: 10.1146/annurev.en.37.010192.000515.
27. Talekar NS, Shelton AM. Biology, ecology, and management of the diamondback moth. Annu Rev Entomol. 1993;**38**(1):275–301. doi: 10.1146/annurev.en.38.010193.001423.
28. Cunningham JJ. A call for integrated metastatic management. Nat Ecol Evol. 2019;**3**(7): 996–8. doi: 10.1038/s41559-019-0927-x.
29. Whelan CJ, Cunningham JJ. Resistance is not the end: lessons from pest management. Cancer Control J Moffitt Cancer Cent. 2020;**27**(1):1073274820922543. doi: 10.1177/1073274820922543.
30. Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive therapy. Cancer Res. 2009;**69**(11): 4894–903. doi: 10.1158/0008-5472.can-08-3658.
31. Enriquez-Navas PM, et al. Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. Sci Transl Med. 2016;**8**(327):327ra24–327ra24. doi: 10.1126/scitranslmed.aad7842.
32. Silva AS, Kam Y, Khin ZP, Minton SE, Gillies RJ, Gatenby RA. Evolutionary approaches to prolong progression-free survival in breast cancer. Cancer Res. 2012;**72**(24):6362–70. doi: 10.1158/0008-5472.can-12-2235.
33. Zhang J, Cunningham JJ, Brown JS, Gatenby RA. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. Nat Commun. 2017;**8**(1):1816. doi: 10.1038/s41467-017-01968-5.
34. Zhang J, Cunningham J, Brown J, Gatenby R. Evolution-based mathematical models significantly prolong response to abiraterone in metastatic castrate-resistant prostate cancer and identify strategies to further improve outcomes. Elife. 2022;**11**:e76284. doi: 10.7554/elife.76284.
35. Zhang J, et al. A phase 1b adaptive androgen deprivation therapy trial in metastatic castration sensitive prostate cancer. Cancers. 2022;**14**(21):5225. doi: 10.3390/cancers14215225.
36. West J, et al. A survey of open questions in adaptive therapy: Bridging mathematics and clinical translation. Elife. 2023;**12**:e84263. doi: 10.7554/elife.84263.
37. Kim E, Brown JS, Eroglu Z, Anderson ARA. Adaptive therapy for metastatic melanoma: predictions from patient calibrated mathematical models. Cancers. 2021;**13**(4):823. doi: 10.3390/cancers13040823.
38. Gallaher JA, Brown JS, Anderson ARA. The impact of proliferation-migration tradeoffs on phenotypic evolution in cancer. Sci Rep-UK. 2019;**9**(1):2425. doi: 10.1038/s41598-019-39636-x.
39. Strobl MAR, et al. Turnover modulates the need for a cost of resistance in adaptive therapy. Cancer Res. 2021;**81**(4):1135–47. doi: 10.1158/0008-5472.can-20-0806.
40. Hansen E, Read AF. Modifying adaptive therapy to enhance competitive suppression. Cancers. 2020;**12**(12):3556. doi: 10.3390/cancers12123556.
41. Gallaher J, Enriquez-Navas PM, Luddy KA, Gatenby RA, Anderson AR. Spatial heterogeneity and evolutionary dynamics modulate time to recurrence in continuous and adaptive cancer therapies. Cancer Res. 2018;**78**(8):canres.2649.2017. doi: 10.1158/0008-5472.can-17-2649.
42. Bacevic K, et al. Spatial competition constrains resistance to targeted cancer therapy. Nat Commun. 2017;**8**(1):1995. doi: 10.1038/s41467-017-01516-1.
43. Strobl MAR, Gallaher J, West J, Robertson-Tessi M, Maini PK, Anderson ARA. Spatial structure impacts adaptive therapy by shaping intra-tumoral competition. Commun Med. 2022;**2**(1):46. doi: 10.1038/s43856-022-00110-x.
44. Noble R, et al. Spatial structure governs the mode of tumour evolution. Nat Ecol Evol. 2022;**6**(2):207–17. doi: 10.1038/s41559-021-01615-9.
45. Gallaher J, et al. The sum and the parts: dynamics of multiple and individual metastases during adaptive therapy. Biorxiv. 2022;2022.08.04.502852. doi: 10.1101/2022.08.04.502852.
46. Benzekry S, et al. Metronomic reloaded: Theoretical models bringing chemotherapy into the era of precision medicine. Semin Cancer Biol. 2015;**35**: 53–61. doi: 10.1016/j.semcancer.2015.09.002.
47. Strobl M, et al. Adaptive therapy for ovarian cancer: an integrated approach to PARP inhibitor scheduling. Biorxiv Prepr Serv Biol. 2023. doi: 10.1101/2023.03.22.533721.

48. Viossat Y, Noble R. A theoretical analysis of tumour containment. Nat Ecol Evol. 2021;**5**(6):826–35. doi: 10.1038/s41559-021-01428-w.
49. Brady-Nicholls R, Enderling H. Range-bounded adaptive therapy in metastatic prostate cancer. Cancers. 2022;**14**(21):5319. doi: 10.3390/cancers14215319.
50. West J, et al. Towards multidrug adaptive therapy. Cancer Res. 2020;**80**(7):1578–89. doi: 10.1158/0008-5472.can-19-2669.
51. West JB, Dinh MN, Brown JS, Zhang J, Anderson AR, Gatenby RA. Multidrug cancer therapy in metastatic castrate-resistant prostate cancer: an evolution-based strategy. Clin Cancer Res. 2019;**25**(14):4413–21. doi: 10.1158/1078-0432.ccr-19-0006.
52. Thomas DS, Cisneros LH, Anderson ARA, Maley CC. In silico investigations of multi-drug adaptive therapy protocols. Cancers. 2022;**14**(11):2699. doi: 10.3390/cancers14112699.
53. Danchin A, Binder PM, Noria S. Antifragility and tinkering in biology (and in business) flexibility provides an efficient epigenetic way to manage risk. Genes. 2011;**2**(4):998–1016.
54. Taleb NN, West J. Working with convex responses: antifragility from finance to oncology. Arxiv. 2022. doi: 10.48550/arxiv.2209.14631.
55. West J, Desai B, Strobl M, Pierik L, Velde, RV, Armagost C, et al. Antifragile therapy. Biorxiv. 2020;2020.10.08.331678. doi: 10.1101/2020.10.08.331678.
56. Viossat Y, Noble R. A theoretical analysis of tumour containment. Nat Ecol Evol. 2021;**5**(6):826–35.
57. Bayer P, West J. Games and the treatment convexity of cancer. Dyn Games Appl. 2023;**13**(4):1088–105. doi: 10.1101/2023.02.27.530257.
58. Pierik L, McDonald P, Anderson ARA, West J. Second-order effects of chemotherapy pharmacodynamics and pharmacokinetics on tumor regression and cachexia. Bull Math Biol. 2024;**86**(5):47 .
59. Gatenby RA, Zhang J, Brown JS. First strike–second strike strategies in metastatic cancer: lessons from the evolutionary dynamics of extinction. Cancer Res. 2019;**79**(13):3174–77. doi: 10.1158/0008-5472.can-19-0807.
60. Robertson-Tessi M, et al. Feasibility of an evolutionary tumor board for generating novel personalized therapeutic strategies. Medrxiv. 2023;2023.01.18.23284628. doi: 10.1101/2023.01.18.23284628.
61. West J, Ma Y, Newton PK. Capitalizing on competition: an evolutionary model of competitive release in metastatic castration resistant prostate cancer treatment. J Theor Biol. 2018;**455**:249–60. doi: 10.1016/j.jtbi.2018.07.028.
62. Enriquez-Navas PM, et al. Exploiting evolutionary principles to prolong tumor control in preclinical models of breast cancer. Sci Transl Med. 2016;**8**(327):327ra24. doi: 10.1126/scitranslmed.aad7842.
63. Gatenby RA, Brown J, Vincent T. Lessons from applied ecology: cancer control using an evolutionary double bind. Cancer Res. 2009;**69**(19):7499–502. doi: 10.1158/0008-5472.can-09-1354.
64. Schmitz OJ, Beckerman AP, O'Brien KM. Behaviorally mediated trophic cascades: effects of predation risk on food web interactions. Ecology. 1997;**78**(5):1388–99. doi: 10.1890/0012-9658(1997)078[1388:bmtceo]2.0.co;2.
65. Calabrese CR, et al., Anticancer chemosensitization and radiosensitization by the novel poly(ADP-ribose) polymerase-1 inhibitor AG14361. JNCI J Natl Cancer Inst. 2004;**96**(1): 56–67. doi: 10.1093/jnci/djh005.
66. Poels KE, et al. Identification of optimal dosing schedules of dacomitinib and osimertinib for a phase I/II trial in advanced EGFR-mutant non-small cell lung cancer. Nat Commun. 2021;**12**(1):3697. doi: 10.1038/s41467-021-23912-4.
67. Pluchino KM, Hall MD, Goldsborough AS, Callaghan R, Gottesman MM. Collateral sensitivity as a strategy against cancer multidrug resistance. Drug Resist Update. 2012; **15**(1–2):98–105. doi: 10.1016/j.drup.2012.03.002.
68. Hall MD, Handley MD, Gottesman MM. Is resistance useless? Multidrug resistance and collateral sensitivity. Trends Pharmacol Sci. 2009;**30**(10):546–56. doi: 10.1016/j.tips.2009.07.003.
69. Aulin LBS, Liakopoulos A, van der Graaf PH, Rozen DE, van Hasselt JGC. Design principles of collateral sensitivity-based dosing strategies. Nat Commun. 2021;**12**(1):5691. doi: 10.1038/s41467-021-25927-3.
70. Nichol D, et al. Antibiotic collateral sensitivity is contingent on the repeatability of evolution. Nat Commun. 2019;**10**(1):334. doi: 10.1038/s41467-018-08098-6.
71. Basanta D, Gatenby R, Anderson A. Exploiting evolution to treat drug resistance: combination therapy and the double bind. Nat Pré cé d. 2011; **9**(4):914–21. doi: 10.1038/npre.2011.6380.1.
72. Luddy KA, West J, Robertson-Tessi M, Desai B, Bursell TM, Barrett S, et al. Evolutionary double-bind treatment using radiotherapy and NK cell-based immunotherapy in prostate cancer. bioRxiv. 2024;2024.03.11.584452.
73. Strobl, M. A. R., Gallaher, J., Robertson-Tessi, M., West, J. & Anderson, A. R. A. Treatment of evolving cancers will require dynamic decision support. *Ann. Oncol.* **34**, 867–884 (2023).

SECTION 9

Critical transitions and chaos in cancer

39

Methods for identifying critical transitions during cancer progression

Smita Deb, Subhendu Bhandary, Mohit Kumar Jolly, and Partha Sharathi Dutta

39.1. Introduction

Casualty due to cancer is increasing with the growing and ageing of the world population [1]. The challenges in the treatment of cancer can be attributed to factors such as difficulty in discerning normal cells from cancer cells and thereby ending up in the survival of cancer cells or killing normal cells [2]. Other factors, such as the rapid evolution of cancer cells [3,4], allow them to get over suppressing mechanisms and developing drug resistance and recurrence post-therapeutic intervention [5,6]. Albeit shreds of evidence advocate a sudden and drastic change in cancer progression of a range of organs, detecting the pre-disease stage can save human life while utilizing fewer medical resources. In general, cancer and a wide range of other disease progressions exhibit alternate stable states; the normal state involves acute inflammation duration followed by the pre-disease and the disease states (see **Figure 39.1**A). A pre-disease state can be characterized by a tipping point where a normal or healthy state may transition to a disease state. Notably, at the pre-disease state, diagnosis and suitable treatment can reverse the state back to the normal stable state. However, beyond the pre-disease state, the system often undergoes an irreversible transition to the alternate (disease) stable state [7–10]. This motivates the need for detecting the pre-disease state from the normal state rather than distinguishing the normal and disease states. Identifying the pre-disease state is pivotal to prevent a critical transition to the disease state, thereby saving human lives at the hands of numerous deathly diseases. Concerning the *a priori* detection of diseases, researchers have developed biochemical indicators termed as biomarkers [11,12], which mark the structural and functional changes in the cells and tissues. Primarily, biomarkers have been employed in differentiating the disease stage from the normal state and measuring the efficiency of the drugs and therapeutic interventions.

Traditional search for biomarkers developed molecular biomarkers from cell concentrations and network biomarkers using information from groups of molecules. Molecular biomarkers include genes, RNAs, proteins, and metabolites or other biological molecules that are key units for proper functioning in a cell. With the availability of high-resolution throughput molecular level data, researchers have been able to evaluate biological homeostasis, distinguishing the disease state from the normal state in cancer of various organs as well as other diseases [13,14]. With a large number of methods developed to detect molecular biomarkers and efforts to visualize the severity of the disease from expressions of molecular biomarkers, they still bear limitations.

Though molecules are the basic and fundamental units of biological systems, practically the genesis of complex diseases can be attributed to groups of genes or molecules rather than individual ones [15]. The causes of diseases are diverse, correlations or clustering of molecules, genes, and proteins play their part. On the contrary, network biomarkers [16,17] that are composed of interacting molecules can provide with more reliable methods for detecting the disease state. Network biomarkers have gained excellence in analysing the occurrence and progression of the disease over their molecular biomarker counterpart [18]. Nonetheless, both these approaches are static in nature and can only distinguish between normal states and disease states (see Figures 39.1B–F). Also, changes at the molecular level are only observed once the system has reached the disease state. But evidence claims that the system transitions gradually from the normal state to the pre-disease state beyond which there is a rapid shift to the disease state [7,8]. As mentioned previously, the latter is an irreversible transition. To prevent transition to the disease state, identifying the critical (pre-disease) state is essential. For achieving the same, researchers have developed dynamic network biomarkers (DNBs) deploying information contained in the network and integrating non-linear dynamics theory that successfully distinguishes a pre-disease state from the normal and disease states [8,19,20]. DNB captures correlations between groups of molecules and their fluctuations, and the characteristic changes are observed in the DNB genes or proteins uniquely prior to a tipping point (see Figures 39.1G and H). DNBs incorporating the dynamics of various biological processes within the system are found to detect the pre-disease state in cancer and other rare diseases [21] with only a small amount of high-throughput data.

In this chapter, we briefly discuss the methods developed till now for detecting cancer progression and their applicability in the diagnosis and treatment of cancer. We present mathematical and theoretical

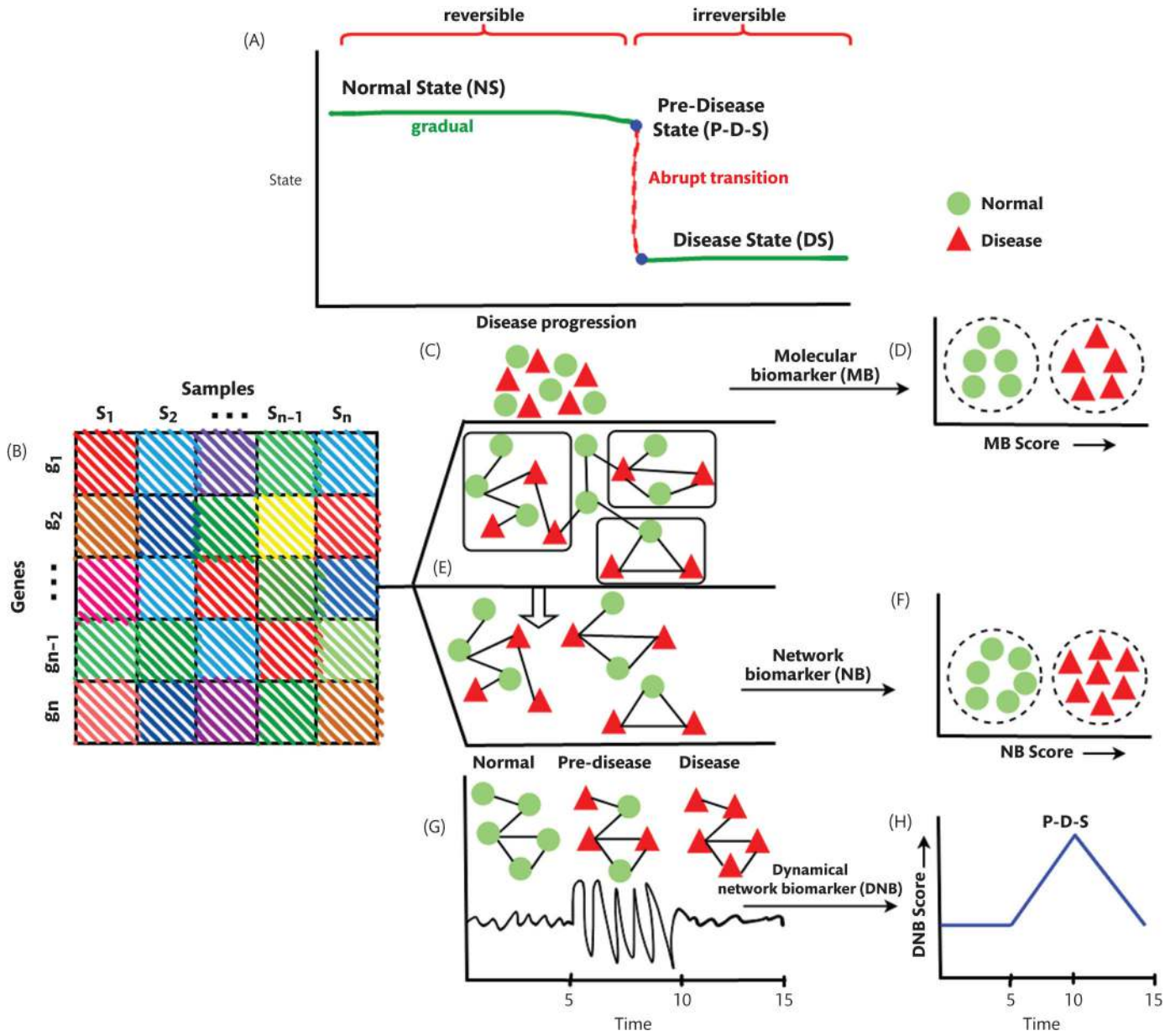


Figure 39.1. A schematic representation of disease states in the course of disease progression and biomarkers. (A) A system state undergoes gradual change until it crosses the pre-disease state and experiences an abrupt shift to the alternate (disease) state. (B) Gene expressions from samples. (C, D) Classification of normal and disease states by molecular biomarkers. (E, F) Subnetworks or network biomarkers identified which distinguish disease and normal phenotypes and detect disease. (G, H) Dynamic network biomarkers (DNBs) identified and a corresponding peak observed in the DNB score for these genes as the system approaches the pre-disease tipping point.

frameworks that guide the working of conventional biomarkers (molecular and network biomarkers) and DNBs. We discuss the limitations of each of these methods and outline other available methods for detecting cancer progression. Additionally, we discuss some mechanisms in cancer cells—cooperation and competition that promote the hallmarks of cancer. Cooperation in cancer cells evolves over time when cells interact strategically and is better modelled by incorporating evolutionary game theory. Such cooperation among cells can influence tumour proliferation, stability, and resilience of tumours, and thereby trigger tipping points. Altogether, these have implications for the development of therapeutic interventions at an early stage. We point this out motivating more in-depth study of tumours and associated cancer mechanisms exploiting game theory. This will bridge the gap between the evolution of cancer cells and the development of DNB and anti-cancer drugs *a priori*.

39.2. Conventional biomarkers

39.2.1. Molecular biomarkers: methods, applications, and limitations

Molecular biomarkers for cancer constitute genes, RNAs, proteins, and other metabolites that are identified using different methods

exploiting large amounts of accumulated omics data [18]. Some methods for identifying molecular biomarkers are as follows.

39.2.1.1. Multivariate (logistic regression) analysis

Recognizing a molecular biomarker, one can determine the number of molecules, the expression of which can distinguish normal and disease states in patients. One of the simplest and classical approaches for such a classification is logistic regression (LR). Multivariate LR [22] is a method used to determine relations between dependent and multiple independent variables. It calculates the probability of an event occurring depending on multiple variables and takes the form $\pi(X) = \frac{e^{\beta_0+\beta_1X_1+\cdots+\beta_pX_p}}{1+e^{\beta_0+\beta_1X_1+\cdots+\beta_pX_p}}$, where $X_1, X_2, \ldots, X_p$ are p independent variables.

The corresponding Logit function takes the form

$$\text{logit}\left[\pi(X)\right] = ln\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \beta_1X_1 + \cdots + \beta_pX_p,$$

where X represents the whole set of covariates $X_1, X_2, \ldots, X_p$. Here, π is the probability that an observation is in a specified category of the response variable Y. The distribution associated with the LR of p variables is binomial and the likelihood function takes the form

$$\sum_{i=1}^{p}\Pi\left(x_i\right)^{y_i}\left(1-\Pi\left(x_i\right)\right)^{\left(1-y_i\right)}.$$

where y_i are the components of the response variable. The coefficients β_i's can be estimated by maximizing the likelihood function. The multivariate LR analysis draws a boundary clearly separating the normal and disease samples, identifying molecular biomarkers for disease. Previously, it has been employed in the calculation of prostate-specific antigen (PSA) tests to detect the onset of prostate cancer. These methods have also been used to identify biomarkers for breast cancer [23,24], cardiac complications [25], as well as other rare diseases [26]. However, researchers have switched to improved methods to overcome associated shortcomings such as poor accuracy and specificity when distributions deviate from normal.

39.2.1.2. Classification and regression tree

It is a non-parametric machine-learning (ML) approach also used for detecting molecular biomarkers apart from its wide utility in interdisciplinary fields. This is a data-driven method that does not rely on any predefined decision boundary. The classification and regression tree (CART) [27] is a decision tree and consists of 3 three main parts: root nodes, internal nodes, and leaf nodes. CART is a predictive model that follows a pattern similar to other decision trees in inferring in steps from observations about a sample represented by branches and obtaining the final output at the target value of the sample denoted by leaves. A CART composed of 5 five protein peaks had successfully differentiated ovarian cancer samples from control samples with high accuracy [28]. Decision tree algorithms have also appeared fruitful in detecting pancreatic cancer [29] from serum samples of patients. Nonetheless, the utility of CART extends to detecting cancer risks, and rectal tumour response to preoperative radiotherapy [30]. Despite its potential in detecting biomarkers, it suffers from model convergence problems for tasks involving a large number of nodes accounting for huge complexity of computation.

39.2.1.3. Voting panel approach

This is a straightforward method where it returns positive or negative results for normal and diseased samples, respectively. It returns a value based on the threshold for individual inputs. This is followed by logical AND, OR gates that operate on an individual molecule to obtain the final output. This finds application in detecting glioma samples, nasopharyngeal carcinoma [31], and differentiating malignant and benign pelvic mass [32]. It has also provided with biomarkers for ovarian cancer [33] and other diseases, such as Crohn's disease and ulcerative colitis [34]. Despite the simplicity of the approach, voting scheme lacks accuracy.

39.2.1.4. Artificial neural networks

Artificial neural networks (ANNs) [35] are algorithms that mimic the functioning of neurons in the human brain. It is a parameterized function mapping an n-dimensional input space R^N to an output in R. ANNs consist of 3 three layers—the input, hidden, and output layers. The hidden layers extract the most relevant information pertaining to the classification task. Neural network architectures are universal approximators mapping non-linear relationships between inputs and outputs. This is performed by training the ANN and modifying weights using a back-propagation algorithm, optimization of the loss function, and obtaining learned weights on which unseen samples with similar distributions are tested. Particularly, in the task of identifying molecular biomarkers and ranking the significance of molecules, ANNs are suited for robust predictions when trained on ample training samples. While determining the loss function and tuning the hyper-parameters may be a tedious task, the goodness of predictions trades off these limitations. In an experiment by Ball et al. [36], ANN with a back-propagation algorithm has successfully detected spectral peaks with intensity values similar to that of tumour grade when trained on data derived from matrix-assisted laser desorption/ionization (MALDI) mass spectroscopy. The performance of ANNs to identify molecular biomarkers is superior over other methods and have been applied for the diagnosis and anticipating treatment outcomes in prostate cancer patients and elsewhere [37,38].

39.2.1.5. Support vector machine

It is a classical supervised learning algorithm. The primary goal of a support vector machine (SVM) classifier is to separate an n-dimensional space by drawing decision boundaries. An SVM classifier can be linear or non-linear. Quite obviously, interactions in the human body that cause cancer or other diseases are highly complex. In context, a non-linear SVM classifier can be used for such purposes to decide the best decision boundary or the hyperplane. Data points that are closest to all the classes are called support vectors, and the distance between the hyperplane and support vectors is the margin. The objective of the SVM algorithm is to find a hyperplane that maximizes the margin and is known as the optimal hyperplane. This simple algorithm is limited by the ability to choose appropriate kernel functions for complex tasks. SVM has effectively provided with classification algorithms for cancer of the lungs, prostate, ovary, colorectum, kidney, liver, pancreas, bladder, and gastroesophagus [39]. They distinguish a set of genes from larger groups of genes that characterize the type of tumour. Researchers have also developed SVM with applications in the multi-category classification problems and diagnosis of multi-class cancer.

Other molecular biomarker identification methods include genetic algorithm [40], dimension reduction approach [41], risk stratification [42], and heterogeneous expression profile analysis [43]. In the presence of high-throughput data, DNA micro-array data at the gene level has provided biomarkers for cancer of the brain, breast, pancreatic, bladder, and several other diseases including diabetes mellitus.

39.2.2. Network biomarkers: methods, applications, and limitations

These are modular biomarkers composed of various interacting molecules similar in character and functioning [15]. Measuring the change in a network of interconnected genes or molecules rather than changes in a single gene gives a better understanding of the mechanisms within cells, tissues, or other larger units of biological systems often responsible for ailments. Network biomarkers may be constructed directly or built from molecular biomarkers. Some of the most commonly used computational methods for identifying network biomarkers are as follows:

39.2.2.1. Active sub-network identifying method

It is a set of genes that foster a connected component in an existing protein–protein interaction (PPI) network. For a sub-network say N, let the activity score of a tumour sample be a vector S whose j th component is computed after normalization of expression values g_{ij} and averaging over each gene in the sub-network, the corresponding class label be denoted by C, the discriminative score $S(N)$ that is the mutual information (MI) between S', a discretization of S and C is given by

$$S(N) = \mathrm{MI}(S', C) = \sum_{x \epsilon S'} \sum_{y \epsilon C} p(x, y) \log \frac{p(x, y)}{p(x)p(y)},$$

where $p(x, y)$ is the joint probability density function of S' and C, $p(x)$ and $p(y)$ are the marginal pdfs of S' and C. Computing the MI scores, the heterogeneity in expressions, and both the differences in the mean and variance is thereby suitable for classification purposes in cancer patients. Set of genes that maximize $S(N)$ is regarded as optimal and works as a biomarker. This set of genes or sub-network is identified by performing a greedy search algorithm within the PPI network. While this is done, one needs to measure the reliability of the results by calculating the false-positive rates. The method has improved accuracy in classifying metastatic and non-metastatic tumours associated with breast cancer [44]. This finds potential application in identifying disease-related genes that are not differentially expressed in systems [45].

39.2.2.2. Disease-specific module identification

In this approach, taking into account gene expressions and information from the PPI network, a network-based biomarker is built. After the initial steps of laying the gene expressions on their corresponding proteins, and choosing differentially expressed proteins (proteins with a significant difference in expression level in two or more different experimental scenarios) using statistical methods [46], highly connected proteins based on PPI information statistics are also put in the protein pool. Gathering PPI information and a protein pool, a PPI network is constructed. This is further modified by means of a regression model. For a protein i in the PPI network:

$$\gamma_i[n] = \sum_{k=1}^{N_i} \alpha_{ik} \gamma_{ik}[n] + \epsilon_i[n],$$

where $\gamma_i[n]$ is the gene expression level, N_i represents the number of interacting protein with the target protein i, α_{ik} is the association ability between the target protein i and the k th interacting protein, and ϵ_i represents stochasticity or model uncertainty.

After the construction of the PPI network, the parameters of the model are estimated using maximum likelihood estimation separately for cancerous and non-cancerous data ($\alpha_{ik,C}$ and $\alpha_{ik,N}$). Post this Akaike information criterion is used to quantify the significance of the association of proteins with either data. C and N are the matrices representing protein association with cancer and no cancer. The entries of the matrix correspond to the impact of i th and j th interacting proteins. Further, to measure the effect or correlation of protein with cancer, a carcinogenesis relevance value (CRV) is defined such that

$$\mathrm{CRV}_i = \sum_{j=1}^{k} d_{ij},$$

where $d_{ij} = \alpha_{ij,C} - \alpha_{ij,N}$. CRV values with $p \leq 0.5$ are considered to be significant proteins associated with cancer. This network-based biomarker has aided in distinguishing 40 different proteins associated with lung cancer [46]. Similar graph-based methods incorporating co-expression dynamics have been applied to different cancerous and non-cancerous data and appeared to better identify interrelations between interacting genes or proteins often revealing dysregulated pathways specific to a disease. Nonetheless, the method is computationally expensive requiring an exhaustive search process and may suffer from over-fitting in the regression model when only a small number of samples are available much lower than the number of parameters to be estimated.

39.2.2.3. Classification of differential interactions

In this method, available gene expressions from cancerous and non-cancerous tissues are divided into groups, and the correlation coefficient between interacting proteins within the PPI network is calculated. PPIs with correlation coefficients beyond a threshold value are considered. A new rewired PPI is then created, and normal and diseased samples are distinguished in phases. Genes or proteins that are common in each phase are considered diseased genes and form network-based biomarkers after testing their significance. This approach differs from other traditional methods on grounds that it investigates differential interactions between normal and diseased samples while the latter separate differential gene expressions. It is successful in finding dysfunctional modules in gastric cancer [7] and identifying nitro-proteins associated with pituitary adenoma apart from cell deaths and cancer. This claims its efficiency in rightfully distinguishing normal and disease states and thus should be validated on other disease datasets for evaluating its generality [47].

39.2.2.4. Information flow approach

Here, dysregulated pathways are identified from the pathway interaction network comparing activity scores for each pathway. Dysregulated pathways are features that can discern between normal and disease states. Pathways are identified in steps and added to the previous pathway biomarker and continued till no more such pathways are found. This comprises the pathway

interaction network where nodes are pathways and edges are present in between pathways if they share a common gene or they have genes that share connections on grounds of PPI. Identification of 5 five dysregulation pathways in pancreatic cancer not only provides with a network-based biomarker but also identifies biological processes that play a major role in the formation and maturation of tumour [48]. Other accomplishments of this method include the classification of states in lung, breast, prostate tumours, and congenial heart diseases [49,50].

39.2.2.5. Support vector machine

As mentioned in the previous section, SVM is an ML method that is used to separate data into classes grouping them based on common features. It employs a linear or non-linear kernel to perform the classification task based on the complexity of the task. In a similar approach to detecting molecular biomarkers, here it segregates groups of molecules or genes integrating information from interactions of the same. It incorporates various aspects, such as gene co-expression, regulatory networks, interrelations, and functional similarity in genes to provide improved performance in detecting disease and normal samples. SVM has provided convincing results in identification of network biomarkers for cancer of various organs [51].

39.2.2.6. Differential dependency network

This method classifies states depending on topological changes in the network for different conditions. Consider a set of random variables $X = \{X_1, X_2, X_3, ..., X_n\}$, a dependency network for X modelled by a set of local conditional probability distribution which follows

$$P(X_i \mid Z_i) = P(X_i \mid \mathbf{X}_{-i}),$$

where $\mathbf{X}_{-i} = \{X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots X_m\}$ and $Z_i \subseteq \mathbf{X}_{-i}$.

This can be generalized for M genes in the network where the dependencies of the i th gene on others are given by

$$P_i = \{P(X_i \mid Z_{i,s_1}), P(X_i \mid Z_{i,s_2}), \ldots, P(X_i \mid Z_{i,s_i})\}$$

Here $Z_{i,s_1}, \ldots, Z_{i,s_i}$ are all in $\mathbf{X}_{-i}$, and s_i for $i \in 1,2,\ldots,m$ denotes the conditional probability for random variable X_i. Further, the conditional probabilities obtained can be inferred by performing linear regression where Z_i predicts X_i as

$$X_i = \beta^T Z_i + \epsilon_i.$$

The error ϵ_i follows a normal distribution $N(0, \sigma_i^2)$ and local conditional probability $P(X_i \mid Z_i) = N(\beta^T Z_i, \sigma_i^2)$ for $i \in \{1,2,...,m\}$. The parameters β_i may be estimated using LASSO estimator. To evaluate the goodness of prediction of X_i by Z_i, one can calculate the coefficient of determination (COD) which takes the form

$$\mathrm{COD} = \frac{\mathrm{var}\{X_i\} - \mathrm{var}\{X_i - f_{X_i|Z_i}(Z_i)\}}{\mathrm{var}\{X_i\}}.$$

Here, f denotes the best function that reduces the residual variance. The differential dependency network (DDN) when applied to identify network topological changes for normal or disease states, the local conditional probabilities P_i^N and P_i^D are computed, and $P_i = P_i^N \cup P_i^D$. For sample corresponding to each condition, COD and test statistics $\hat{\theta} = |\mathrm{COD}^1 - \mathrm{COD}^2|$ are calculated. Multiple simulations are performed to test the significance of the differences observed. The method detected genes with network topological changes in oestrogen-dependent T-47D oestrogen receptor–positive breast cancer cell dataset [52] and human and mouse embryonic stem cell datasets [53]. Other methods of identifying network biomarkers include the incorporation of network information with sequence data. In particular, single nucleotide polymorphic data and the analysis of genome-wide association at SNP level captured aggregated effects of various processes and provided added information to the occurrence and progression of the disease [54].

39.3. Dynamic network biomarkers

As evident from the discussions in the previous sections on molecular biomarkers and network biomarkers, they only differentiate between normal and disease states. However, to prevent or cure a terminal disease, like cancer, it requires an early diagnosis of the pre-disease state prior to the critical transition to a disease state. While identifying the pre-disease state is a challenging task, there occur minute changes in the interval of normal to pre-disease state progression. To develop early warnings of a critical transition to a disease state, researchers have developed a model-free approach, DNB [8], that detects the pre-disease state with minimal but high-throughput data. As proved theoretically near a tipping point, there exist groups of genes or molecules that exhibit characteristic features that can be used to characterize a DNB. Characteristic properties that are common to DNB irrespective of the concerned system and are necessary for identifying a DNB for a disease are as follows:

1. Correlation between molecules comprising DNB increases sharply.
2. Correlation between molecules comprising DNB and that of other non-DNB elements weaken.
3. Standard deviation of molecules comprising DNB increases rapidly.

The composite index (CI) for calculating the signal for the DNB is given by,

$$\mathrm{CI} = \frac{\mathrm{PCC}_i \times \mathrm{SD}_i}{\mathrm{PCC}_o} \tag{39.1}$$

Here, PCC_i denotes the average of Pearson correlation coefficient (PCC) in the molecules of the DNB, PCC_o is the average of PCC of the molecules in DNB with other molecules, and SD_i is the average standard deviation of molecules in the DNB. A peak in the CI value with a rising trend marks the onset of a disease tipping point. DNBs have provided early warnings indicating a pre-disease state in cancer of various organs and other complex diseases [19]. In an instance, DNB was identified for GSE21510 and TCGA colon and rectal cancers [55]. At each stage $I - IV$, the DNB scores were computed. At stage III, the DNB score showed a sharp rise assuming a peak value at stage IV. Stage III is therefore marked as the pre-disease state corresponding to the progression of colon and rectal cancers. They

also identify genes that amplify and suppress tumour proliferation. Other instances where DNB has provided early warning signals are hepatocellular carcinoma [56], lung injury [57], breast cancer [58], type-1 and type-2 diabetes [20,59], etc. Though this traditional DNB detection method is applicable in cell-fate determination, cell differentiation, and determining the pre-disease stage, it requires multiple samples at each time point often not available in clinical and practical instances. Given the limitations of the number of samples in the critical state, researchers have developed methods that can anticipate the pre-disease state even with a single sample. A single sample dynamic biomarker network (sDNB) is based on the information of differential association allowing predictions of the occurrence of the disease *a priori*.

39.3.1. Working of sDNB

Identifying pre-disease state using DNB from a single sample requires the presence of a control sample, and the CI is computed using the same formula as that for traditional DNBs (see Equation (39.1)). In the process, the deviation of a gene expression in a single sample is calculated as the difference of gene expression of new sample (x) and the average expression of gene x in the reference sample ($\check{x}$), i.e.

$$\text{sED} = |x - \check{x}|.$$

Further, $\text{PCC}_n(x, y)$ and $\text{PCC}_{n+1}(x, y)$, the PCC between two genes of the reference sample of size n and that of the new sample, are calculated, respectively. The single sample PCC of the two genes (new and reference) is given by

$$\text{sPCC} = \text{PCC}_{n+1}(x, y) - \text{PCC}_n(x, y).$$

The significance of $\text{sPCC}(x, y)$ for two genes x and y can be evaluated by observing the p-value for the z-score where

$$z = \frac{\text{sPCC} \times (n-1)}{(1 - \text{PCC}_n^2)}.$$

Further, the CI of sDNB can be computed by using

$$\text{CI}_s = \frac{\text{sPCC}_{in} \times \text{sED}_{in}}{\text{sPCC}_{out}}, \tag{39.2}$$

where sED_{in} is the single sample expression deviation, sPCC_{in} denotes sPCC among inner genes of a module, and sED_{in} is the sPCC between inner and outer genes of a module. Further details of the working of sDNB are presented in [60]. This method not only anticipates the pre-disease state in cancer metastasis but also aids in deciphering processes that govern the dynamics of the complex disease at a single sample level. Other available single-sample DNB approaches employ the hidden Markov model [61] and the Kullback–Leiber divergence method [62] that manifest improved performance and are widely in practice. Moreover, sDNB built using local network entropy (LNE) has successfully classified biomarkers for the pre-disease state of 10 types of cancers [63], namely kidney renal clear cell carcinoma, lung squamous cell carcinoma, stomach adenocarcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, oesophageal carcinoma, colon adenocarcinoma, rectum adenocarcinoma, thyroid carcinoma, and kidney renal papillary cell carcinoma available in The Cancer Genomic Atlas (TCGA) (cancer.gov https://portal.gdc.cancer.gov). Further, Liu et al. [63] made significant contributions in finding two variants of prognostic biomarkers, optimistic LNE and pessimistic LNE, that identify pre-disease states and enable *a priori* predictions.

39.3.2. Working of landscape-DNB

Landscape-DNB (L-DNB) [64] is a single sample DNB method used to detect the pre-disease state from a single sample. It differs from other sDNB approaches on grounds that it computes the local DNB score (I_s) in the same manner using Equation (39.2) and then integrates the local DNBs into a landscape. Similar to other sDNBs, here a reference of n samples is considered as a control. A single sample network is constructed and the local DNB score I_s is calculated for the local module centred at a gene x. Further, the global DNB score I_g is computed from the landscape of the sample by selecting genes with the highest scores as components of the DNB. DNB genes are characterized by the highest k-local DNB scores, and the global DNB score takes the form $I_g = \sum_{x=1}^{k} \frac{I_s(x)}{k}$. The sample with the highest I_g is considered to be at or approaching a disease tipping. The advantage of the L-DNB is that it allows for predicting the disease tipping or identifying the pre-disease state efficiently and is computationally less expensive as it avoids lengthy and complicated algorithms. The L-DNB method has been applied to three tumour disease data from TCGA, and it successfully identifies the pre-disease states using a single DNB for each patient. Its application also extends to the identification of symptoms associated with influenza prior to its actual occurrence. The precedence of DNB over other detection methods is apparently evident as it detects tipping points prior to the occurrence of disease while using minimal samples and in some instances by only utilizing an individual sample.

Overall, DNBs greatly advance the study of complex diseases and aid in the diagnosis of various fatal diseases that would have been otherwise impossible to prevent. Nonetheless, the battle against cancer is ceaseless as real data are limiting; on the other hand, cancer cells and tumours are continuously evolving and interacting strategically which are often overlooked. This is better done by integrating cancer dynamics and evolutionary game theory and then predicting cell fate using the above methods for more accurate and precise drugs and therapeutic interventions.

39.4. Game theory and effective cancer therapies

Evolutionary game theory [65] explains reproductive selection and frequency-dependent growth in a population of evolving individuals. Games or strategic interactions between living organisms are observed in the smallest of cells to the largest of animals whenever an individual's payoff or reward depends not only on his action but also on other individuals in the population. While such interactions are more apparent in financial markets, political science, psychology, etc., concepts from evolutionary games also provide a better understanding of interactions within and across cancer cells [2] such as stromal cells that have frequency-dependent fitness. Frequency-dependent fitness may be defined as the selection where the fitness or growth of a phenotype also depends on that of other phenotypes present in the population. While

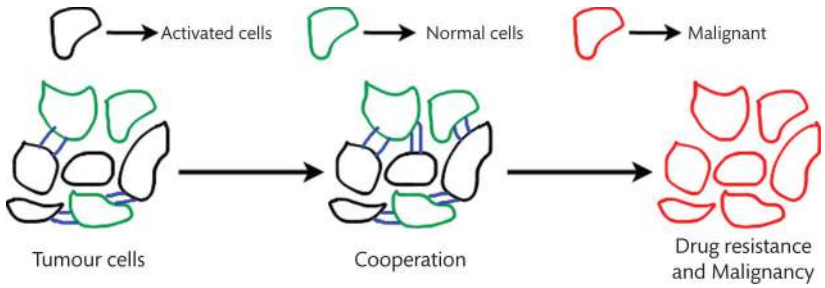


Figure 39.2. A schematic representation of simplified tumour–cell interactions: increase in cooperation between tumour cells leading to drug resistance and hence tumour progression, resulting in malignancy.

the simplest strategies in game theory are cooperation and defection, in the context of cancer cell metastasis it implies that by cooperating cells aids in the generation of diffusible factors that enhance proliferation (see Figure 39.2). Whether cells continue to cooperate with other cells or may change their strategy to defect has much to do with the development of more effective therapies [66]. Although the decision of cells to cooperate or defect depends on cost–benefit ratios, cooperation can lead to a stationary state with reduced fitness (as benefits are shared by all cells) depending upon various factors such as diffusion of the secreted molecule, concentration, etc. In an instance of pancreatic cancer, it is inferred from experiments that cooperation evolves over time at low-cost scenarios and may diminish beyond an upper threshold value. The experiment by [67] maintained a low cost–benefit ratio to enable cooperation to persist by titrating the extent of exogenous growth factor accessible to the cell. Another experiment by [68] exhibited mutualism between tumour sub-clones for the exchange of diffusible substances between cells. Thus, one can claim that pure defection and an interplay of cooperation and defection are stable strategies depending on the initial population densities.

These have implications for achieving effective and long-lasting therapies as a treatment to cancer experiences setbacks due to drug resistance and failure in responding to targeted drugs and chemotherapy. It has been observed that cancer cells are unstable in nature and mutants that do not succumb to drug/therapy emerge and evolve by natural selection and lead to patients not responding to treatments. Game theory could come to the rescue to design therapies or drugs whose therapeutic effects remain unchanged over time with more efficacy and not give in to evolving mutants. Experts have tried formulating therapy by changing drug dosage to bring in competition among cells. Instead of targeting a specific cell or tumour, reducing the dosage would slow down the rate at which resistance is developed. This has achieved some success in cancer treatment as well as for other diseases [69,70]. Using multiple therapies with collective impact on cancer cells that compete to survive toe-to-toe turns out to be harmful to each other leading to another form of adaptive therapy [71,72]. Thus, linking more real evidence to evolutionary game theory and finding evolutionary stable strategies will be a giant leap in achieving evolution-proof therapies.

39.5. Conclusions

Altogether, we have discussed various methods for the detection of pre-disease, normal, and disease states in cancer of various organs. Traditional methods, such as molecular and network biomarkers, have been widely used in clinical practices owing to their straightforward application. However, guided by theory accompanied by observations, detecting the pre-disease state is crucial to prevent the occurrence of diseases as in many cases once the system resides in the disease state the casualty is unavoidable. Although the disease state has not yet occurred, but at the limit of the normal (pre-disease) state, genes exhibit features that can be captured by the DNBs. The scores for the respective DNBs or components of the DNB constitute the best indicators of marking the onset of the disease. Needless to mention that DNBs requiring only 3 three prerequisite conditions to be satisfied are superior to other traditional methods as they classify a pre-disease sample from the normal sample, have a data-driven approach, do not rely on *a priori* information, and consequently reduce bias. Methods for classifying pre-disease state from normal samples are only recently developed and require validation on a larger class of cancer and disease samples. While DNBs are robust and generalizable to a wide range of diseases and show high accuracy, there are still limitations. Lack of data availability for cancer of different organs has prevented investigating the applicability of DNB pertaining to a disease. Moreover, while DNB efficiently classifies pre-disease state for long sequences, in real scenarios often only single or smaller number of samples are available. Improving data sampling technique, data quality, and incorporating time information, developing more improved methods with reduced human interventions can help enhance the practicality of DNBs.

Alternative suitable candidate methods for improving predictions of the pre-disease state are deep-learning tools that have shown success in anticipating critical transitions while discerning them from other continuous transitions [73]. Nonetheless, the application of deep learning in forewarning sudden transitions is in its infancy and requires more in-depth investigations. Further, the extent of perturbation in the presence of which DNB identification holds is usually small, and it requires the data not to be aperiodic. While such scenarios are omnipresent, the developments of DNBs for diseases with chaotic behaviour *a priori* or in course of disease progression are promising future directions towards diagnosing and curing cancer. The evolutionary nature of cancer cells [74] has been recognized for ages. As mutants become resistant to drugs, treatments fail to suppress tumour growth [6,75]. Yet, it is less explored in the context of developing effective treatments and target-specific drugs. Thus, integrating game theory and identifying strategic interactions among cells, mutation rates, and cell size will add to the development of adaptive therapies against cancer.

Acknowledgements

S.D. acknowledges the Ministry of Education (MoE), Government of India, for Prime Minister's Research Fellowship (PMRF). P.S.D. acknowledges the Science and Engineering Research Board (SERB), Government of India (Grant No. CRG/2019/002402).

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. CA: Cancer J Clin. 2022 Jan;**72**(1):7–33.
2. Korolev KS, Xavier JB, Gore J. Turning ecology and evolution against cancer. Nat Rev Cancer. 2014 May;**14**(5):371–80.
3. Wodarz, D. Dynamics of cancer: incidence, inheritance, and evolution. 2009;117–8.
4. Nowell PC. The clonal evolution of tumor cell populations: acquired genetic lability permits stepwise selection of variant sublines and underlies tumor progression. Science. 1976 Oct 1;**194**(4260):23–8.
5. Valkenburg KC, De Groot AE, Pienta KJ. Targeting the tumour stroma to improve cancer therapy. Nat Rev Clin Oncol. 2018 Jun;**15**(6):366–81.
6. Camidge DR, Pao W, Sequist LV. Acquired resistance to TKIs in solid tumours: learning from lung cancer. Nat Rev Clin Oncol. 2014 Aug;**11**(8):473–81.
7. Liu R, Li M, Liu ZP, Wu J, Chen L, Aihara K. Identifying critical transitions and their leading biomolecular networks in complex diseases. Sci Rep. 2012 Dec 10;**2**(1):813.
8. Chen L, Liu R, Liu ZP, Li M, Aihara K. Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers. Sci Rep. 2012 Mar 29;**2**(1):1–8.
9. Deb S, Bhandary S, Sinha SK, Jolly MK, Dutta PS. Identifying critical transitions in complex diseases. J Biosci. 2022 Apr 20;**47**(2):25.
10. Sarkar S, Sinha SK, Levine H, Jolly MK, Dutta PS. Anticipating critical transitions in epithelial–hybrid-mesenchymal cell-fate determination. Proc Natl Acad Sci U S A. 2019 Dec 16;**116**(52):26343–52.
11. Wang X, Ward PA. Opportunities and challenges of disease biomarkers: a new section in the journal of translational medicine. J Transl Med. 2012 Dec;**10**:1–4.
12. Ptolemy AS, Rifai N. What is a biomarker? Research investments and lack of clinical integration necessitate a review of biomarker terminology and validation schema. Scand J Clin Lab Invest. 2010 Jan 1;**70**(sup242):6–14.
13. Duffy MJ, Crown J. A personalized approach to cancer treatment: how biomarkers can help. Clin Chem. 2008 Nov 1;**54**(11):1770–9.
14. Poste G, Carbone DP, Parkinson DR, Verweij J, Hewitt SM, Jessup JM. Leveling the playing field: bringing development of biomarkers and molecular diagnostics up to the standards for drug development. Clin Cancer Res. 2012 Mar 15;**18**(6):1515–23.
15. Schadt EE. Molecular networks as sensors and drivers of common human diseases. Nature. 2009 Sep 10;**461**(7261):218–23.
16. Barabási AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nat Rev Genet. 2011 Jan;**12**(1):56–68.
17. Chen L, Wang RS, Zhang XS. Biomolecular networks: methods and applications in systems biology. USA: John Wiley & Sons; 2009 Jun 29.
18. Liu R, Wang X, Aihara K, Chen L. Early diagnosis of complex diseases by molecular biomarkers, network biomarkers, and dynamical network biomarkers. Med Res Rev. 2014 May;**34**(3):455–78.
19. Liu R, Aihara K, Chen L. Dynamical network biomarkers for identifying critical transitions and their driving networks of biologic processes. Quant Biol. 2013 Jun;**1**(2):105–14.
20. Li M, Zeng T, Liu R, Chen L. Detecting tissue-specific early warning signals for complex diseases based on dynamical network biomarkers: study of type 2 diabetes by cross-tissue analysis. Brief Bioinform. 2014 Mar 1;**15**(2):229–43.
21. Tang S, Yuan K, Chen L. Molecular biomarkers, network biomarkers, and dynamic network biomarkers for diagnosis and prediction of rare diseases. Fundam Res. 2022 Aug 9;2(6):894–902.
22. Hosner DW, Lemeshow S. Applied logistic regression. New York: John Wiley & Sons; 1989. p. 581.
23. Rouzier R, Perou CM, Symmans WF, Ibrahim N, Cristofanilli M, Anderson K, et al. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res. 2005 Aug 15;**11**(16):5678–85.
24. Bhatavdekar JM, Patel DD, Shah NG, Vora HH, Suthar TP, Chikhlikar PR, et al. Prognostic significance of immunohistochemically localized biomarkers in stage II and stage III breast cancer: a multivariate analysis. Ann Surg Oncol. 2000 May;**7**:305–11.
25. Lette J, Colletti BW, Cerino M, Mcnamara D, Eybalin MC, Levasseur A, et al. Artificial intelligence versus logistic regression statistical modelling to predict cardiac complications after noncardiac surgery. Clin Cardiol. 1994 Nov;**17**(11):609–14.
26. Sell S. In: Diamandis EP, Fritche HA, Lilja H, Chan DW, Schwartz MK, editors. Tumor markers: physiology, pathobiology, technology and clinical applications. Washington, DC: AACC Press; 2002. 513 pp. Clin Chem. 2003 Feb 1;**49**(2):342.
27. Muller R, Möckel M. Logistic regression and CART in the analysis of multimarker studies. Clin Chim Acta. 2008 Aug 1;**394**(1–2):1–6.
28. Vlahou A, Schorge JO, Gregory BW, Coleman RL. Diagnosis of ovarian cancer using decision tree classification of mass spectral data. J Biomed Biotechnol. 2003 Dec 4;**2003**(5):308–14.
29. Yu Y, Chen S, Wang LS, Chen WL, Guo WJ, Yan H, et al. Prediction of pancreatic cancer by serum biomarkers using surface-enhanced laser desorption/ionization-based decision tree classification. Oncology. 2005;**68**(1):79–86.
30. Zlobec I, Steele R, Nigam N, Compton CC. A predictive model of rectal tumor response to preoperative radiotherapy using classification and regression tree methods. Clin Cancer Res. 2005 Aug 1;**11**(15):5440–3.
31. Oluwadara O, Chiappelli F. Biomarkers for early detection of high risk cancers: from gliomas to nasopharyngeal carcinoma. Bioinformation. 2009;**3**(8):332.
32. Woolas RP, Conaway MR, Xu F, Jacobs IJ, Yu Y, Daly L, et al. Combinations of multiple serum markers are superior to individual assays for discriminating malignant from benign pelvic masses. Gynecol Oncol. 1995 Oct 1;**59**(1):111–6.
33. Woolas RP, Xu FJ, Jacabs IJ, Yu YH, Daly L, Brechuck A, et al. Elevation of multiple serum markers in patients with stage I ovarian cancer. JNCI: J Natl Cancer Inst. 1993 Nov 3;**85**(21):1748–51.
34. Burczynski ME, Peterson RL, Twine NC, Zuberek KA, Brodeur BJ, Casciotti L, et al. Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in

peripheral blood mononuclear cells. J Mol Diagn. 2006 Feb 1;**8**(1):51–61.
35. Rodvold DM, McLeod DG, Brandt JM, Snow PB, Murphy GP. Introduction to artificial neural networks for physicians: taking the lid off the black box. The Prostate. 2001 Jan;**46**(1):39–44.
36. Ball G, Mian S, Holding F, Allibone RO, Lowe J, Ali S, et al. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. Bioinformatics. 2002 Mar 1;**18**(3):395–404.
37. Babaian RJ, Zhang Z. Computer-assisted diagnostics: application to prostate cancer. Mol Urol. 2001 Dec 1;**5**(4):175–80.
38. Wei JT, Zhang Z, Barnhill SD, Madyastha KR, Zhang H, Oesterling JE. Understanding artificial neural networks and exploring their potential applications for the practicing urologist. Urology. 1998 Aug 1;**52**(2):161–72.
39. Lee Y, Lee CK. Classification of multiple cancer types by multicategory support vector machines using gene expression data. Bioinformatics. 2003 Jun 12;**19**(9):1132–9.
40. Liu JJ, Cutler G, Li W, Pan Z, Peng S, Hoey T, et al. Multiclass cancer classification and biomarker discovery using GA-based algorithms. Bioinformatics. 2005 Jun 1;**21**(11):2691–7.
41. Hilario M, Kalousis A. Approaches to dimensionality reduction in proteomic biomarker studies. Brief Bioinform. 2008 Mar 1;**9**(2):102–18.
42. Apple FS, Wu AH, Mair J, Ravkilde J, Panteghini M, Tate J, et al. Future biomarkers for detection of ischemia and risk stratification in acute coronary syndrome. Clin Chem. 2005 May 1;**51**(5):810–24.
43. Xu QW, Zhao W, Wang Y, Sartor MA, Han DM, Deng J, et al. An integrated genome-wide approach to discover tumor-specific antigens as potential immunologic and clinical targets in cancergenome-wide discovery of TSAs as clinical cancer targets. Cancer Res. 2012 Dec 15;**72**(24):6351–61.
44. Chuang HY, Lee E, Liu YT, Lee D, Ideker T. Network-based classification of breast cancer metastasis. Mol Syst Biol. 2007;**3**(1):140.
45. Lee E, Chuang HY, Kim JW, Ideker T, Lee D. Inferring pathway activity toward precise disease classification. PLoS Comput Biol. 2008 Nov 7;**4**(11):e1000217.
46. Wang YC, Chen BS. A network-based biomarker approach for molecular investigation and diagnosis of lung cancer. BMC Med Genomics. 2011 Dec;**4**(1):1–5.
47. Zhan X, Desiderio DM. Signaling pathway networks mined from human pituitary adenoma proteomics data. BMC Med Genomics. 2010 Dec;**3**:1–26.
48. Liu KQ, Liu ZP, Hao JK, Chen L, Zhao XM. Identifying dysregulated pathways in cancers from pathway interaction networks. BMC Bioinform. 2012 Dec;**13**(1):1–11.
49. Kreeger PK, Lauffenburger DA. Cancer systems biology: a network modeling perspective. Carcinogenesis. 2010 Jan 1;**31**(1):2–8.
50. He D, Liu ZP, Chen L. Identification of dysfunctional modules and disease genes in congenital heart disease by a network-based approach. BMC Genomics. 2011 Dec;**12**:1–6.
51. Zhu Y, Shen X, Pan W. Network-based support vector machine for classification of microarray samples. BMC Bioinform. 2009 Jan;**10**(1):1–1.
52. Zhang B, Li H, Riggins RB, Zhan M, Xuan J, Zhang Z, et al. Differential dependency network analysis to identify condition-specific topological changes in biological networks. Bioinformatics. 2009 Feb 15;**25**(4):526–32.
53. Liu Y, Shin S, Zeng X, Zhan M, Gonzalez R, Mueller FJ, et al. Genome wide profiling of human embryonic stem cells (hESCs), their derivatives and embryonal carcinoma cells to develop base profiles of US Federal government approved hESC lines. BMC Dev Biol. 2006 Dec;**6**:1–6.
54. Li M, Wang K, Grant SF, Hakonarson H, Li C. ATOM: a powerful gene-based association test by combining optimally weighted markers. Bioinformatics. 2009 Feb 15;**25**(4):497–503.
55. Tong Y, Song Y, Xia C, Deng S. Theoretical and in silico analyses reveal MYC as a dynamic network biomarker in colon and rectal cancer. Front Genet. 2020 Oct 20;**11**:555540.
56. Yang B, Li M, Tang W, Liu W, Zhang S, Chen L, et al. Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. Nat Commun. 2018 Feb 14;**9**(1):678.
57. Sciuto AM, Phillips CS, Orzolek LD, Hege AI, Moran TS, Dillman JF. Genomic analysis of murine pulmonary tissue following carbonyl chloride inhalation. Chem Res Toxicol. 2005 Nov 21;**18**(11):1654–60.
58. Liu R, Wang J, Ukai M, Sewon K, Chen P, Suzuki Y, et al. Hunt for the tipping point during endocrine resistance process in breast cancer by dynamic network biomarkers. J Mol Cell Biol. 2019 Aug;**11**(8):649–64.
59. Liu X, Liu R, Zhao XM, Chen L. Detecting early-warning signals of type 1 diabetes and its leading biomolecular networks by dynamical network biomarkers. BMC Med Genomics. 2013 May;**6**(2):1–10.
60. Liu X, Chang X, Liu R, Yu X, Chen L, Aihara K. Quantifying critical states of complex diseases using single-sample dynamic network biomarkers. PLoS Comput Biol. 2017 Jul 5;**13**(7):e1005633.
61. Liu R, Zhong J, Yu X, Li Y, Chen P. Identifying critical state of complex diseases by single-sample-based hidden Markov model. Front Genet. 2019 Apr 4;**10**:285.
62. Zhong J, Liu R, Chen P. Identifying critical state of complex diseases by single-sample Kullback–Leibler divergence. BMC Genomics. 2020 Dec;**21**:1–5.
63. Liu J, Ding D, Zhong J, Liu R. Identifying the critical states and dynamic network biomarkers of cancers based on network entropy. J Transl Med. 2022 Dec;**20**(1):1–3.
64. Liu X, Chang X, Leng S, Tang H, Aihara K, Chen L. Detection for disease tipping points by landscape dynamic network biomarkers. Natl Sci Rev. 2019 Jul 1;**6**(4):775–85.
65. Weibull JW. Evolutionary game theory.Cambridge: MIT Press; 1997.
66. Archetti M, Pienta KJ. Cooperation among cancer cells: applying game theory to cancer. Nat Rev Cancer. 2019 Feb;**19**(2):110–7.
67. Archetti M, Ferraro DA, Christofori G. Heterogeneity for IGF-II production maintained by public goods dynamics in neuroendocrine pancreatic cancer. Proc Natl Acad Sci U S A. 2015 Feb 10;**112**(6):1833–8.
68. Cleary AS, Leonard TL, Gestl SA, Gunther EJ. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. Nature. 2014 Apr 3;**508**(7494):113–7.
69. Gatenby RA, Silva AS, Gillies RJ, Frieden BR. Adaptive therapy. Cancer Res. 2009 Jun 1;**69**(11):4894–903.
70. Gatenby RA. A change of strategy in the war on cancer. Nature. 2009 May 28;**459**(7246):508–9.

71. Gatenby RA, Gillies RJ, Brown JS. The evolutionary dynamics of cancer prevention. Nat Rev Cancer. 2010 Aug;**10**(8):526–7.
72. Basanta D, Gatenby RA, Anderson AR. Exploiting evolution to treat drug resistance: combination therapy and the double bind. Mol Pharm. 2012 Apr 2;**9**(4):914–21.
73. Deb S, Sidheekh S, Clements CF, Krishnan NC, Dutta PS. Machine learning methods trained on simple models can predict critical transitions in complex natural systems. R Soc Open Sci. 2022 Feb 16;**9**(2):211475.
74. Wilkins A, Corbett R, Eeles R. Age distribution and a multi-stage theory of carcinogenesis: 70 years on. Br J Cancer. 2022 Oct 28;1–3.
75. Obenauf AC, Zou Y, Ji AL, Vanharanta S, Shu W, Shi H, et al. Therapy-induced tumour secretomes promote resistance and tumour progression. Nature. 2015 Apr 16;**520**(7547):368–72.

40

Chaos and complexity: Hallmarks of cancer progression

Abicumaran Uthamacumaran

40.1. Introduction: complexity science in precision oncology

In the era of Big Data, fast-paced advancements within data science, bioinformatics, and evidence-based medicine are evolving towards integrating the vast computational power of machine intelligence and physics-driven quantitative models in precision healthcare. At the heart of this culmination lies the *complex systems paradigm*, a transdisciplinary science and art devoted to the study of otherwise incomputable problems, or intractable diseases like *cancers*, as *complex adaptive systems*. Cancers can be studied within the complexity framework, via bridging computational medicine, mathematical oncology, and systems biology, to better identify *targeted therapies*, improve cancer screening (i.e. diagnostics and prognostics), accelerate patient-specific biomarkers discovery, and optimize clinical decision-making. Forecasting the complex dynamics of cancers fosters best practices within the *art of healing*, by allowing clinicians to better adapt drug delivery mechanisms or control therapeutic dosage, to prevent adaptive behaviours within tumour ecosystems, such as therapy resistance [1–3]. Single-cell multi-omics are emerging as powerful tools at the frontier of precision and personalized cancer medicine [4]. Single-cell multi-omics through the lens of complexity science holds great promise in further advancing precision patient care by helping us identify causal patterns driving tumour behaviours. As such, healthcare practitioners and interdisciplinary scientists must be educated on the toolkits and concepts pertaining to complexity science.

With such pedagogical intentions, this chapter is devoted to a bridge between precision oncology and *complex systems theory*, in understanding cancers as ecosystems, or simply, *complex systems* [5,6]. To achieve this, the multiscale physics and machine-learning algorithms for characterizing the multilevel patterns of behaviours within single-cell cancer *multi-omics* will be investigated herein. This marriage will help elucidate the genetic and non-genetic instabilities (e.g. epigenetic modifications) underlying collective behaviours such as *phenotypic plasticity* and its inter-related intra-tumoral heterogeneity, serving as fundamental barriers in treating cancers [7]. Machine-learning algorithms are state-of-the-art tools in systems oncology with counterfactual purposes. For instance, they can be employed for capturing drug response prediction and cancer classification using multi-omics data [8]. Further, they are being used for drug repurposing and discovering targeted therapies in precision oncology [9,10]. Systems oncology views cancer as a dynamic, multiscale process—a form of collective intelligence emerging from complex biopsychosocial interactions. Let us begin by defining *complexity science*. In short, *complex systems theory* or *complexity theory* is the transdisciplinary study of patterns (structures) and behaviours (processes) within *complex systems*. It is a recent paradigm in physics intersecting many branches of knowledge systems, including the sciences, and arts/social humanities, such as the *study of wholeness* (i.e. process ontology), dynamical systems theory, statistical mechanics (information theory), cybernetics (the study of regulation, navigation, and feedback loops in control systems), and the computational sciences [11–14]. Mathematicians and physicists have been using these tools to study dynamical systems, such as weather patterns (fluid turbulence), the fractality of clouds, the flocking or swarming of ecological systems (i.e. many-body systems behaving as a coordinated whole), and chemical oscillations in morphogenesis, to name a few. However, the study of cancers as *complex adaptive systems* remains a relatively new stream of thought. A *complex system* is typically a many-body dynamical system, with many interconnected variables or parts exhibiting adaptive behaviours and phenomena across different scales [14]. Thus, a *complex system* is composed of many non-linearly interacting constituents or processes that cannot be reduced from the whole [11]. Living systems, such as ecosystems, and the dynamics of single cells, proteins, and biomolecules, are such complex systems. The irreducibility of these dynamic interactions forms feedback loops across many scales, as a result of which the system exhibits adaptive or *emergent behaviours*. Emergence denotes the concept of 'more is different', a state of synergy or stigmergy, wherein the non-linear dynamics between the parts gives rise to new causal patterns and behaviours [15].

Most scientists equate emergence or irreducibility to complexity, while others make it synonymous with non-linear dynamics. However, complexity entails the meta-study of all these

characteristics. The emergent properties allow the complex system to learn and adapt from and within its environmental context(s). In contrast, to the many-body physics view, even very simple physical and computational or mathematical systems (e.g. cellular automata or a difference equation) might give rise to unanticipated complex behaviours and patterns. The recursive iterations of the Mandelbrot set, a fractal structure, is a great example. Toolkits, such as chaos theory, criticality, fractal geometry, network theory, and statistical physics, can well quantify, describe, or predict these complex behaviours and patterns across a wide range of *complex systems*, and hence, the complex systems *paradigm* is an umbrella framework applicable across many realms of study. For instance, tools such as network science derived from the social humanities (e.g. social networks theory) or ecological sciences (i.e. ecosystems dynamics) can be universally applied across multiscale biological processes, from molecular networks to morphological relationships [14]. Moreover, within contemporary art practice, and the social humanities, complexity science shares syntax with post-modernism, *new materialism* (post-humanism), Jung's analytical psychology, Gestalt psychology, and *process theory/ontology*, to name a few schools of thought.

In complexity theory, we are interested in quantitative approaches for causal discovery, to reveal the *hidden patterns* of organization or processes/mechanisms under which a complex system operates, steers through its environment, and makes decisions. Network dynamics and attractor reconstruction are two core approaches to causality inference in cancer cell-fate dynamics. Key signatures of complex systems include wholeness, non-linear dynamics, non-Gaussian/nonequilibrium statistics, long-term unpredictability (epistemological uncertainty), interconnectedness, multi-nested causal processes such as feedback loops (i.e. cybernetics), sudden abrupt changes in behaviour (i.e. *criticality* or phase transitions), spontaneous pattern formation, purposive or goal-oriented behaviours, evolvability (adaptation), self-organization (globally organized behaviours on a small timescale), collective intelligence (e.g. spatial navigation for resources or metastatic invasion in cancer), and emergence (collective behaviours) [11–14]. Many of these complexity features are signatures of chaotic dynamics, i.e. non-linear systems with sensitive dependence on initial conditions and fluctuations/perturbations. Chaotic systems are a subset of complex systems, generating counterfactual possibilities, wherein a small error on the initial parameters can grow extremely fast, making long-term predictions difficult. In systems biology, self-organization can also denote *autopoiesis*, the ability of living systems to self-replicate, self-regulate, and organize themselves [16]. Some cognitive neuroscientists would argue that complexity is defined by second-order cybernetics and self-meta/reorganization (i.e. the self-ability to remodify and adapt the organization process), as advocated by the theory of *embodied cognition* [17]. Similarly, cancers can be seen as a *collective intelligence* emerging from the coupling between the tumour and its environment. Most often, emergent behaviours and intricate complex patterns form simply by fluctuations and updates in local interactions. However, the interconnectedness of complex systems results in foundational problems, bridging quantum mechanics with complexity theory, such as the issue of non-locality vs. local information dynamics, and the *observer effect* (e.g. participatory or passive role of an agent influencing the behaviours) [11,18].

In summary, the key defining feature of complex systems is that new properties will emerge at every scale of interactions. Perhaps the most evident hallmark of complexity in cancer dynamics is its *phenotypic plasticity* [19]. As argued herewith, phenotypic plasticity serves as an interface to study complex dynamics, such as chaos, fractals, and criticality within cancer systems, and will be a central theme of this chapter. For instance, high-grade gliomas (HGGs) such as glioblastomas, the deadliest of brain tumours, are hierarchically organized tumour ecosystems driven by glioma stem cells (GSCs) that retain partial differentiation potential. GSCs are a subpopulation of therapy-resistant tumour cells, integrating the complex microenvironmental cues spanning across the physiological spaces of the patient/host system, including neurons, immune cells, stromal cells, microglia, astrocytes, oligodendrocytes, the blood–brain barrier, neuroendocrine secretory cells, and extra-cellular matrices. We refer to this niche of interactions as the tumour microenvironment (TME). The integration allows the emergence of adaptive behaviours, such as metastatic invasion and *phenotypic plasticity* dynamics (i.e. cell state transitions) [19]. For instance, the GSC–TME crosstalk allows cancer cells to aggressively proliferate, evolve, invade the circulatory systems, and communicate with immune cells to hijack or suppress local tumour immunity [20,21]. These multiscale processes span across multidimensional physiological spaces, such as the brain–microbiome–immune axis and neuroendocrine axes, promoting the inflammatory TME, tumour progression, and therapy efficiency (or resistance) [22–24].

Therapy-induced stressors can also promote the TME to induce adaptive traits like cancer plasticity [19]. Increasing evidence amounts to the critical role of neural–tumour interactions in driving gliomas. Epigenetic and transcriptomic signatures distinguishing high-neural glioblastomas (hypomethylated CpG sites and up-regulation of genes involved in synaptic integration) from low profiles [25] show a gradient of plasticity emerging from therapy-induced stressors. The high-neural glioblastoma show higher levels of stem-cell-like malignant cells (GSCs), which differentiate towards neural progenitor-like cells that can lineage bifurcate towards oligodendrocyte-like and neural precursor-like cells, suggesting that the neural–glioma synaptic connections drive glioma (cancer) plasticity dynamics [19,25,26]. The GSCs can also differentiate towards glial progenitor-like cells, giving rise to astrocyte-like and mesenchymal-like glioma cells, via the reactivation of wound healing pathways, hijacked immune signalling, and the embryonic neurodevelopmental landscape [19]. However, the locality or non-locality of the TME's governance of phenotypic plasticity is largely questioned. That is, whether the plasticity dynamics depends on the local context/stem cell niches or is a non-local behaviour wherein the plasticity program is internally embedded within every hybrid or partially 'differentiated' tumour cell remains a debated problem in systems medicine. Tumour heterogeneity mapping via *single-cell multi-omics* sequencing is hinting towards the latter wherein the entire tumour system (population) is believed to exhibit a gradient of plasticity or stemness potential [4,19]. This spectrum of plastic, phenotypic states might also arise from cells "stuck" or trapped along their developmental trajectories away from terminal cell fates. The plasticity dynamics also depends on long-range cell–cell communication systems such as exosome-mediated interactions from distant tissue microenvironments, promoting cancer migration and metastatic invasion [27,28].

While the origins and mechanisms of therapy-induced plasticity are still being unravelled, advanced computational and mathematical

methods provide critical insights into quantifying and modelling phenotypic transitions and cell-fate trajectories in cancer. Such methods suggest that chaos, turbulence, and complex dynamics may help elucidate the emergence of stem-like or plastic traits in cancers. Thus, to germinate complexity and chaos, as hallmarks of cancer aggressivity, the chapter is divided into four major sections. The first is devoted to some insights on state-of-the-art algorithms and computational tools in overlaying multi-omics information to elucidate phenotypic plasticity, i.e. cell-fate transition dynamics. The second section focuses on network science, a powerful lens capable of identifying critical drivers (network links and nodes) regulating adaptive behaviours like phenotypic transitions. The last two sections question the limitations of normative science. They are devoted to the *physics of pattern formation*, suggesting a more non-conventional, explorative research program of complexity and chaos in tumour self-organization, interweaving concepts, such as chaos, fractals, *strange attractors* (the *hidden geometric patterns* of chaos), chemical turbulence, and active fluids. Albeit seeming unorthodox and counterfactual, the *chaos theory* syntax is universally inherent to the temporal patterns and behavioural sequences of complex natural systems, thereby, warranting its exploration in cancer dynamics.

40.2. Cancer dynamics: data science approaches in forecasting cell-fate trajectories

Our discussion in this section is centred on reconstructing behavioural patterns, such as cell-fate transitions from gene expression dynamics (transcriptomics), and their epigenetic regulation/control. However, they can be extended to proteogenomics, single-cell proteomics (e.g. mass spectrometry (MS)/CyTOF) and metabolomics [29]. While epigenetic control mechanisms guide normal cell-fate decisions, dysregulation of these epigenetic processes is a hallmark of many cancers, including gliomas, where altered epigenetic landscapes shape tumour–microenvironment interactions, immune evasion, and phenotypic plasticity [30]. Epigenetic modifications include histone modifications, DNA methylation, post-translational modifications, and 3D-chromatin reorganization. For instance, in pediatric HGGs, oncohistones, such as mutations of the histone 3 genes (H3C1 and H3F3A), and of driver genes, such as isocitrate dehydrogenase 1/2 (IDH1/2), α-thalassemia/mental retardation, and X-linked (ATRX), steer epigenetic dysregulation, shaping TME interactions, and thereby, regulate disease or therapeutic progression and patient survival [30]. Furthermore, in paediatric HGGs and leukaemia, the oncohistone variant K27M amino acid substitution inhibits the polycomb repressive methyltransferase complex 2, resulting in decreased trimethylation of H3 proteins, and hence, alters gene expression patterns to confer cancer stem-cell-like malignant states [31].

The terms *phenotypic switching*, *plasticity dynamics*, and cell *state transitions* are synonymously used throughout to denote cell-fate dynamics during stalled differentiation, de-differentiation, and trans-differentiation to molecular/environmental perturbations in cancer ecosystems (e.g. trajectories during metastatic invasion or decision-making of cancer stem cells/malignant cells under therapeutic resistance). Cell-fate dynamics, or simply, *cancer dynamics*, are quantitatively studied via a diverse set of systems-theoretic (mathematical) tools and data-driven models. These multi-physics predictive models are generally found within two major classes, namely, discrete and continuous models, borrowed from dynamical systems theory and statistical mechanics/information theory. Discrete models involve approaches, such as difference equations or computational methods like cellular automata. In contrast, continuous models involve differential equations and are solved analytically or via numerical methods. Continuous models, as focused herein, are further subdivided into deterministic and stochastic models. One of our main arguments herewith is that the lack of deterministic models due to the normative preference/dogma of stochastic approaches largely reduces the complexity of processes underlying cancer dynamics, such as the emergence of chaos and strange attractors, within the state-space reconstruction of the cellular trajectories. This is explained in part by the challenge of distinguishing stochasticity from deterministic chaos. More recently, a third approach is emerging: artificial intelligence (AI)-driven modelling. AI-driven models, although intersecting with the other two dynamical systems approaches, is an umbrella term transcending their limits, including computational techniques, such as algorithmic complexity approaches, model-driven statistical algorithms, deep-learning methods such as generative AI, multifeatured-selection-based pattern recognition, quantum machine intelligence, etc.

Although our discussions focuses on single-cell multi-omics, bulk-cell analysis reveals complementary insights that should be contrasted with the single-cell resolution behaviours. Computational pipelines, such as Noiseq or DESEQ2 [32], are available for bulk RNA-seq analysis, wherein multiple sample measurements can be integrated, but the diverse phenotypes will be averaged to a single point in the state space. The general approach is to first filter out lowly expressed genes and then identify differentially expressed genes/markers (DEGs) beyond some cut-off threshold of log-fold change in expression (log2FC). The filtering step itself could be problematic as they may be bursting or express complex/chaotic dynamics in time-series analysis. DEG analysis is performed such that their statistical significance is assessed by the Wald test following correction using the Benjamini and Hochberg method or other statistical approaches [33]. Whether we include these corrections or not can drastically change the number of DEGs. Dimensionality reduction, such as principal component analysis (PCA), is also performed to restrict the DEGs/markers within the top PCA components as most variable features across the tumour samples/clusters. The Jaccard similarity index is often used to identify phenotypic cluster modules (i.e. hierarchical clustering) via the similarly expressed genes. Lastly, signalling pathways and transcription modules pertinent to the DEGs are identified via databases, such as KEGG, GSEA, gprofiler, and other gene set enrichment analysis tools.

In contrast, single-cell analyses differ in the sense that we can map the heterogeneity in differential expression of biomarkers at a single-cell resolution, and hence, trace population dynamics via the respective markers (e.g. DEGs) in time-course progression. PCA could help visualize attractor manifolds in such high-dimensional complex systems. We can also perform non-linear dimensionality reduction techniques to visualize the DEG expression across distinct phenotypes and how they vary across communities of cell-fate clusters (subpopulations). Multiple single-cell platforms can be overlayed simultaneously to reconstruct the phase-space trajectories of cancer dynamics more accurately. To illustrate, a study by

Granja et al. [34] co-analysed scRNA-seq, scATAC-seq (transposase-based chromatin accessibility profiling), and scCITE-seq (cellular indexing of transcriptomes and epitopes) to sequence the transcriptome, epigenome, and surface protein markers, respectively, in comparing healthy and leukaemic blood cells. Using these approaches, they identified leukaemia-specific regulatory genes, such as *RUNX1*, driving blood differentiation dynamics.

Given the paramount importance of single-cell multi-omics in reconstructing cancer dynamics, algorithmic approaches capturing the complex attractor dynamics underlying these cell-fate decisions are of utmost interest to precision medicine and systems biology. For instance, aggressive cancers such as paediatric HGGs are currently believed to be stuck in terminal differentiation (i.e. unstable attractors), unable to commit to a glial phenotype (stable attractor), and they exhibit a spectrum of phenotypic variability (i.e. multistability) [35,36], ranging from mesenchymal-like states to neuro-glial phenotypes. These paediatric brain tumours also show a lower mutational burden than their adult counterparts, with epigenetic drivers shaping the tumour–immune landscapes [37,38]. Aberrant methylation profiles, at the level of chromatin, and reprogrammed patterns of histone modifications, i.e. oncohistones, are signature hallmarks of many paediatric tumours [38]. However, even in adult cancers, such as glioblastoma, epigenetic markers are becoming increasingly apparent and being used to predict longer survival and therapy response [7].

The increasing relevance of epigenetic drivers in cancer progression, therapy response, and plasticity dynamics, shifts our dogmatic focus from the genetic instabilities of the somatic mutation landscape (view of Darwinian selection) towards the integration of non-genetic instabilities in cell-fate behaviours, and their corresponding *attractor dynamics* [36]. Therefore, model-driven algorithms rooted in *dynamical systems theory*, a pillar of complexity theory, are ideal candidates to map complex cellular processes such as cell-fate decisions emerging from complex disease–environment interactions. As such cell-fate dynamics and differentiation processes can be metaphorically captured as a *Waddington landscape* or energy landscape [39,40]. However, different formulations/theories can replace the potential energy states with an information-theoretic, 'entropic' landscape, or an optimal transport theory-derived 'optimization' landscape to characterize the cell-fate trajectories (attractor reconstruction). The central differences across these three landscape reconstruction methods lie within whether we employ an energy function (e.g. Lyapunov energy) for energy minimization, an entropy function for global entropy maximization, or a cost function for cost minimization, in defining the attractor states, respectively. The latter approach can visualize the cell-fate dynamics as a complex problem-solving task, such as navigating a transcriptomic phase space or optimizing resource gradients within a game-theoretic, predator–prey-like model of adaptive dynamics. [Note: While Shannon entropy is usually used to identify stochasticity or measure uncertainty, the Kolmogorov–Sinai entropy or metric entropy analogue can be used for quantifying complex attractor dynamics (chaos)].

Let us consider a few such Waddington landscape algorithms to gain some intuition into cell-fate trajectory (attractor) reconstruction. Waddington-OT is an attractor reconstruction algorithm inferring cellular dynamics via time-series mapping, using optimal transport theory [41]. Waddington-OT infers the temporal couplings of state transitions, from samples collected at various time points, using probabilistic distance metric methods [41]. The coupling refers to the Wasserstein distance between two probability distributions characterizing cell states with similar transcriptomic profiles in gene expression state space, which minimizes the transportation cost (i.e. for the two distributions to merge into each other) [41,42]. The method allows the identification of transient dynamics in the over-expression of a set of transcription factors (TFs), as gene regulatory programs for the transition-state dynamics [41]. However, the downfall of this approach, alike most bioinformatic pipelines in single-cell analyses, is that the conventionally filtered *lowly expressed genes* can also exhibit complex dynamics, such as irregularity/non-periodicity, bursting, or intermittency, i.e. signatures of chaotic/complex dynamics. Furthermore, these algorithms *a priori* assume *stochastic dynamics*, such as Brownian motion, to characterize the cell-fate trajectories. Complex dynamics, such as critical dynamics (intermittency), collective cell migration (e.g. flocking or wave-like effects), and (chaotic) turbulence, may exist within tumour populations, which are ignored by these assumptions. Similarly, another optimal transport algorithm called Lineage-OT quantifies levels of entropy between the temporal couplings to infer a lineage tree constructed from a heuristic called neighbour joining to infer cell-fate trajectories [43]. The entropic optimal transport can be understood as the maximum-likelihood coupling between the cell population, the entropy parameter, and the assumed Brownian motion of cell-fate trajectories [43].

Multi-model approaches combining differential equations, network science, and dynamical systems theory (Waddington-like landscape reconstruction) have been used to model phenotypic switching in cancer epithelial–mesenchymal transition (EMT) dynamics. For instance, Jolly et al. demonstrated that EMT plasticity results in the emergence of a spectrum (heterogeneity) of cellular states, with distinct regulatory network modules characterizing the phenotypic communities. Their analysis revealed that network topology can characterize the phenotypic landscape (distributions), and hence, these landscapes and their underlying network dynamics are powerful predictive tools in cancer cell-fate control [44]. Similarly, Pillai et al. [45] used landscape reconstruction models and *in silico* perturbation analysis to show plasticity-induced state transitions as an emergent property of cancer systems. They demonstrated that cusp-like catastrophic transitions or criticality, in cancer cell-fate dynamics, can give rise to distinctive attractor states (phenotypes). For example, Saez et al. [46] combined Bayesian inference and catastrophe theory to reconstruct cell-fate transitions from single-cell data onto a Waddington landscape. Bayesian inference is a commonly employed causality inference tool in statistical physics, wherein the integration of prior knowledge with multivariate or multimodal posterior distributions optimizes our decision-making [46,47]. For instance, Sanity, a Bayesian inference algorithm and BASiCS (Bayesian Analysis of Single-Cell Sequencing data) have been proposed as predictive machines to better optimize nearest-neighbours cell clustering into phenotypic subtypes in single-cell analysis, outperforming current normalization techniques [48,49].

Another multi-model attractor landscape reconstruction algorithm is MuTrans [50]. MuTrans infers cell-state transition dynamics from single-cell gene expression using *stochastic differential equations* (SDE) and coarse-grained transition path theory. Downstream analysis can be used to identify gene targets underlying the transient

states. The method is consistent with the well-established Langevin equation and transition rate theory. The dynamics of cell fates can be described by the SDE:

$$dX_t = f(X_t)dt + \sigma(X_t)dW_t$$

where X_t denotes the cell's gene expression count at a time t, $f(x)$ is the non-linear gene regulations, $\sigma(x)$ is the noise strength (internal and environmental fluctuations), and W_t is the Brownian motion noise [50]. Coarse-grained state-space dynamics can be deduced with time-series data. MuTrans takes as input the pre-processed single-cell gene expression matrix and learns the cellular random walk transition probability matrix using a Gaussian-like kernel because of which the cell dynamics fit to an over-damped Langevin equation. Coarse-graining methods identify optimized cell clusters, which are fit onto a dynamical manifold using a Gaussian mixture distribution model connecting the attractor basins to stable phenotypes, identified via steady-state distributions of the Fokker–Planck equation. Transition/hybrid cell states are mapped along transition paths connecting the attractor basins. The general message established by these examples is that the choice of models, their hyperparameters tuning, and the mathematics/physics underlying the algorithms (i.e. assumptions, constraints, model parameters, and conditions) can drastically alter or frame the interpretations of complex cellular dynamics. Thereby, to assess the robustness of identified attractors or biomarkers, we should overlay multiple algorithmic approaches and use multivariate statistical metrics, such as significance testing and cross-validation.

Meanwhile, AI applications are witnessing an expansion of quantitative applications in systems medicine. The discovery of targeted therapies via drug repurposing and the profiling of patient-derived liquid biopsies are some of its many promising applications in precision oncology [51–53]. For instance, AI platforms, such as deep-learning network (DLN) algorithms like AlphaFold 2, have solved one of the most difficult problems in systems biology: the protein folding prediction problem [54]. Given cancer biomarkers typically consist of malformed or misfolded (onco-mutant) proteins, or malfunctioning protein expression networks, AlphaFold-like algorithms can be used for AI-driven drug design and drug repurposing. Machine-learning algorithms can also be coupled with network science methods for anti-cancer drug discovery [55]. Similarly, these AI systems can be used to quantify complex cell-fate dynamics. For instance, the Hopfield network is a type of artificial neural network that intakes a gene expression matrix and performs an energy minimization of the Lyapunov function (of gene expression state space), recursively trying to find the local energy minima (i.e. cell state attractors) and infer a differentiation energy landscape by finding the minimal distance/path connecting the local attractors [56]. The recursive search for solutions (the energy minima or attractor basins for cell states) can be found in many other AI systems like *evolutionary algorithms*, warranting further research in cancer dynamics modelling.

Generative adversarial networks (GANs) and autoencoders, a subset of DLNs, can learn complex data distributions allowing the interface between artistic creativity and science. Generative AI systems of pre-trained DLN models, such as OpenAI's DALLE-2 and Midjourney (art generation AI systems) (see **Figure 40.1**), and ChatGPT-4, have revolutionized the fields of computer vision, art making, and natural language generation (NLP), allowing complex tasks such as code interpretation and data analysis with context reasoning. These systems are capable, to a decent degree, of interpreting complex datasets (with a certain window size) within patient-specific physiological contexts, allowing predictive health modelling in precision healthcare. Recently, such DLN tools have been used in reconstructing high-dimensional phenotypes using expression quantitative loci [57]. As mentioned, we typically assume a linear relation between the observed gene expression variables and hence use dimensionality-reduced latent spaces like PCA to reduce the gene expression state space. However, given the multidimensional complexity of tumour ecosystems, non-linear latent spaces can be better captured by generative AI systems, such as variational autoencoders, deep Boltzmann machines, or GANs, with imputation and generative techniques, to infer the plasticity/lineage trajectories [58]. Stable diffusion is a popular generative AI technique used for text-to-image generation. Similar approaches with data diffusion have been used in leveraging scRNA-seq pattern spaces, into manifold reconstruction, such as within the MAGIC imputation algorithm [59]. Diffusion models combined with generative AI are now being used for time-series scRNA-seq manifold (attractor) reconstruction to predict cell differentiation trajectories [60]. More recently, scGPT, a generative pre-trained transformer like the working principles of ChatGPT, was shown within the developmental pipelines, as capable of decoding complex cell-fate dynamics and downstream analyses in single-cell data, including cell-type classification, multi-batch or multi-omics integration, gene network reconstruction, and molecular/genetic perturbation predictions [61]. Hence, generative AI and deep-learning algorithms provide a powerful causal inference tool for reconstructing multidimensional attractors from single-cell gene expression spaces [62].

Figure 40.1. Chaos and turbulence as hallmarks of cancer progression. Cancer dynamics as imagined by the Midjourney algorithm, a text-to-image generative AI system, using the following keywords: cancer dynamics, turbulence, and chaos.

Lastly, although in its infancy, we are currently witnessing the second revolution of quantum mechanics, with the rise of quantum information sciences, and quantum machine intelligence, which holds great promises for studying cancer nanosystems [63–65]. Quantum machine learning may help significantly increase the predictive power of AI by increasing the vastness and complexity of the search space of solutions in systems oncology problems [66]. For instance, Kao et al. [67] demonstrated a hybrid quantum GAN system capable of innovating drug design. Genome (and multi-omics) analysis and molecular diagnostics using quantum computing are expected to vastly accelerate personalized cancer care [64,68]. The future also holds great promise for quantum AI applications in forecasting cancer dynamics and attractor reconstruction. Still, we must keep in mind the fundamental limitations posed by mathematical theorems underlying the *computability of algorithms*. These include Godel's incompleteness theorem (i.e. there are mathematical truths built on unprovable or irreducible axioms, and thus, the set of these truths is much larger than those provable by algorithms), as well as the inter-related Turing's proof and Church's theorem on the *Entscheidungsproblem* (Hilbert's decision problem), such as the Halting problem. This is to say that there are mathematical equations or problems that are unprovable or (currently) without solutions (e.g. the P vs. NP problem), highlighting intractability or lack of computability, as fundamental properties of complex computational systems. For instance, finding the network modules or shortest paths connecting the central nodes of a *complex network*, such as a transcriptional cancer network, is an NP-hard problem [69]. As discussed, cancer data science uses dimensionality reduction and searching/optimization heuristics to overcome these incomputability challenges. Whether the parallel worlds accessed by quantum computing can help resolve or update these theorems in computational complexity is largely debated. The perspective undertaken herein is that quantum computing's accessibility of the vast Hilbert space will be advantageous to clinical cancer, especially in predictive and preventive care, such as in subtyping cancer phenotypes, forecasting longitudinal patterns of behaviours, and simulating or identifying personalized therapies [68].

40.3. Network medicine: mapping interactions within the cancer multiome

The pioneering research of Phil Gold in clinical immuno-oncology paved *systems oncology*, bridging the gap between cancer physiology, biomarker discovery, and precision medicine [70]. Systems oncology advocates the view of tumour ecosystems, as being coordinated by multiscale *complex networks* [22,71–73]. Network science, or network medicine, is a branch of complexity theory using graph-theoretic methods to quantify and study the interconnectedness of such complex networks. Network medicine is being used to identify therapeutic targets and central drivers of phenotypic transitions [74,75]. Further, network medicine and AI can be combined to repurpose or reposition drugs in personalized medicine [76] and innovate therapeutic discovery, such as by identifying candidate immunotherapies using protein–protein interaction (PPI) regulatory networks [77]. Therefore, *complex networks* are both a hallmark property and a mathematical tool to study, quantify, and predict the behavioural patterns of cancer dynamics [74].

In graph theory and network science, the topology (i.e. connectivity pattern or spatial organization) of the network controls the dynamics and thereby evolves towards *attractors* within the network state-space dynamics [74,78]. With single-cell analyses, we see that the collective cell dynamics (of populations) can form a global attractor in the multi-omics state space, while the individual cell states (phenotypes) can also form local attractors, i.e. the local vs. nonlocality (global) debate [40]. Network theory is a robust tool to investigate how local dynamics interconnect with global behaviours. There are unique spatial features and temporal characteristics specific to complex biological networks. Biological networks, such as PPI networks, gene regulatory networks (GRN), and metabolic networks, often show *fractal structures*, i.e. a *scale-free topology* [79,80]. A scale-free topology means that the probability $P(k)$ of the degree distribution, or some other network property, forms a power-law behaviour such that an arbitrary element of the network is connected to exactly k other elements and has the form $P(k) = Ck^{-\gamma}$, where γ is usually called the scale-free exponent. The fractal scaling implies that biological networks in healthy contexts, and healthy physiological computations, exhibit self-similarity across scales and operate based on *criticality* (or critical dynamics) [79,80]. For instance, the *critical brain hypothesis* suggests that the healthy brain's neural computations are poised at a critical system [81].

Criticality is an essential property of complex systems, a state of *slippage* or juxtaposition of being poised in between 'order/regularity and chaos', which optimizes between robustness and adaptability [14]. Perturbations of the internal or external environment can cause the complex system to adapt (change dynamics), resulting in *critical transitions* above certain tipping points (thresholds). Thus, criticality allows the complex network to generate new structures when encountering dynamics or fluctuations above these thresholds, while maintaining the previous structures, below the threshold [82]. In statistical physics, the order can be seen as low entropy (low uncertainty or information content) resulting in stability and predictable dynamics, whereas chaos denotes a high-entropy system, with constant change and information flow, high adaptability, and long-term unpredictability. Critical systems are described by power-law distributions, or scaling behaviours, such as those seen in fractals and complex biological networks. As discussed in the next section, *fractals are the geometry of chaos* [13]. Hence, critical networks allow biosystems to optimize cellular decisions (computations), but under certain perturbations they can transition to *chaotic dynamics*. In proposition, this transition to chaos may be a hallmark of cancer dynamics. A good intuition of critical dynamics is obtained by the self-organized criticality (SOC) in Bak's sandpile model, a paradigm model for natural complex systems, exhibiting power-law behaviours and self-similarity or fractality in the avalanche sizes of sandpiles, simulated by cellular automatons [83]. However, SOC is only one of many mechanisms, perhaps the simplest one, for the cybernetics of self-organization and critical dynamics. Critical dynamics are also referred to as *intermittency*, in complex systems such as turbulence or chaotic systems. Thus, non-Gaussian dynamics and intermittency are hallmarks of complex network dynamics.

Another property of complex biological networks is *multi-nestedness* or *recursivity* [74], a topological feature associated with the fractality of cancer networks, and their non-linear feedback loops with the environment. In network science, we refer to this embodiment or embeddedness as *modularity* [74]. The information

flow from local contexts to non-local (global) behaviours results in the compartmentalization of networks into modules or clusters with distinct functional properties (e.g. clusters of phenotypes in cancer systems). This modularity suggests we may need multimodal treatments targeting the distinct modules (clusters) to devise effective therapeutic strategies against complex systems like cancers. In cancer network science, we use Louvain community detection and similar modularity optimization algorithms to identify such network modules [74]. Modularity is often a property of critical dynamics, allowing the robustness of individual clusters (modules or phenotypic communities), and the adaptability or evolvability from one to another (e.g. state transitions).

Now, let us explore various approaches to infer cancer network dynamics, the simplest of which are Boolean networks. In Boolean networks, also known as NK models, the existence of a phase transition or critical transition is governed by the value of the scale-free exponent of the network [78]. These Boolean network models show that many biological networks (systems) exist at the *edge of chaos*, as coined by Kauffman [78], allowing the systems to coexist between robustness and flexibility, i.e. allowing the switching between multiple phenotypes (network states) [82]. For example, Chu et al. [84] used a stochastic Boolean network model to infer attractor states, wherein tumorigenesis can be hypothesized as a critical transition between proliferative and apoptotic attractors on the state space of the molecular networks. Boolean models show that skewed topologies of networks, as denoted by higher correlations in gene expression or PPIs, may correspond to fractal (strange) attractors. In principle, a higher entropy and higher mutual information can be used to identify these fractal attractors in such networks [78]. However, within highly complex systems from multiscale experimental data, better-optimized information-theoretics or multivariate statistical tools are needed, as discussed below. Network science metrics, such as modularity optimization (community detection algorithms), shortest path algorithms between the nodes of the network and network centrality measures, can be used to identify key driver genes/markers within complex networks [55]. Further, some scholars suggest that *nonlinear network dynamics* may help the epigenetic (in)stability and controllability in adaptive cancer dynamics [85].

At the experimental frontier, many network inference tools exist for reconstructing cancer networks from multi-omics data. For instance, we can trace the TFs (activators and repressors) that control phenotypic plasticity and heterogeneity, thereby identifying the network clusters (modules) that allow the reprogramming of cell states [86]. To illustrate, DriverNet, is an R-software that reconstructs genomic–transcriptomic cancer networks, by mapping driver mutations onto the mRNA expression networks from cancer patients' data [69]. It builds an influence graph to reconstruct the effects of a mutation in a gene on the change of expression of another gene using (1) a binary matrix for the mutation data (0 for the absence and 1 for the presence of a mutation) inferred from PPI (PPIs) or copy numbers, and possibly, epigenomic data; and (2) the other gene matrix being the real-valued gene expression or transcriptomic matrix (RNA-seq expression values), i.e. a bipartite graph. The minimum set cover problem created by this complex network is NP-hard (i.e. recall the computability limits of algorithms). Thus, a heuristic like the greedy approximation algorithm is used to solve this optimization problem, and statistical significance tests of the driver genes are assessed using network randomization [69]. While the statistical tools (e.g. correlation metrics or probability distributions for genes) and clustering approaches may vary, the general network inference approach outlined herein remains the baseline algorithm for most cancer networks inference tools from real-patient (experimental) multi-omics cancer data. That is, there will always be the use of some heuristics to solve otherwise intractable (NP-hard) optimization problems in complex networks.

Many other network science approaches have been used to identify regulatory gene expression modules of cancer networks. Some examples include gene co-expression analysis networks in predicting clusters of breast cancer subtypes with poorer survival (prognostic) outcomes [87]. Another example is the transcriptional regulatory networks of tumour-infiltrating tumour-associated macrophages (TAMs) driving malignancy in glioblastoma. Sa et al. [88] identified that TME-dependent loss of NF1 and PI3K/mTOR signalling in glioblastoma cells was critical for driving TAMs towards the mesenchymal phenotypes. For instance, a network inference method called InPheRNo can identify the transcriptional networks controlling phenotypes, and distinguish healthy and disease states, based on a probabilistic graphical model that predicts the synergistic effects of multiple TFs in cancer networks [89]. Similarly, OncoPPI is a network inference method for identifying dysregulated PPI network targets or hubs (interactomes) serving as candidate biomarkers for effective anti-cancer therapeutics [90]. PPI network-based precision therapies and drug discovery platforms are emerging via identifying hijacked or reprogrammed (dysregulated) network interactions from MS-based patient proteomics data [91]. The recent advances in single-cell methods such as CyTOF can further help elucidate the complex physiological networks, such as the immune–inflammasome networks, from whole-blood analyses and liquid biopsies of cancer patients [92].

An example of a multilevel network dynamics inference model is the SCENIC+ algorithm. SCENIC+ can infer gene regulatory networks from the enriched regulons identified from the single-cell gene expression matrix, as well as genomic patterns, such as the enhancer regulatory networks (i.e. chromatin accessibility) by integrating scATAC-seq and correlating them with the molecular phenotypes, and their differentiation dynamics [93,94]. In cancer systems, this combined multi-omics analysis may be useful for identifying gene expression programs underlying cell state transitions during de-differentiation or trans-differentiation (i.e. plasticity dynamics), with simultaneous detection of regions with chromatin activity [94]. At the core of SCENIC+ is a package called CisTopic, a probabilistic framework used to simultaneously discover co-accessible enhancers and stable cell states from sparse single-cell epigenomics data. The regulatory networks can be visualized using the iRegulon package and Scope visualization tools employing the SCENIC software. However, it remains a challenge how to exploit single-cell epigenomic data, such as Hi-C sequencing (to infer 3D-chromatin remodelling) for resolving spatiotemporal enhancer activity and GRN dynamics, both experimentally and computationally. RNA-labelling methods, CITE-seq, and RNA-velocity algorithms can also be coupled with SCENIC+ to capture time-resolved differentiation dynamics from cancer systems [95–97]. Such vector analysis of the multi-omics state space allows attractor reconstruction and its underlying network inference in pan-cancer analyses [97].

Lastly, our recent results have shown network science-driven cancer biomarker discovery. Using network science measures, we

identified gene candidates regulating cell-fate transitions in glioblastoma and GSCs using single-cell transcriptomics [98]. The roles of the transition genes we identified are being verified as central regulators of phenotypic dynamics. For instance, Robertson et al. [99] recently demonstrated that elevated levels of FOXG1, one of the network markers we identified, drive the exit of GSCs from quiescence to promote tumour growth. Another study revealed that the transcription factor YY1, another network marker we identified, promotes hepatocellular carcinoma migration and invasion, via forming local hubs of phase-separated liquid condensates at super-enhancers [100]. Furthermore, our identified network markers play central roles in the *epigenetic regulation* of glioblastoma, such as involvement in histone demethylase complexes, suggesting an underlying *epigenetic/histone code* [98].

40.4. Complex dynamics: chaos, fractals, and strange attractors

We have now arrived at a *critical point*, a gap between current practices in cancer data science and mathematical oncology. The experimental data in cancer single-cell multi-omics shows complex spatial patterns and emergent behaviours such as plasticity. However, the syntax within complexity, namely, chaos, fractals, self-organization, and non-linear dynamics, the toolkit for complex attractor reconstruction, remains vastly unexplored in experimental cancer dynamics. As such, we provide the theoretical frameworks and concepts for chaos, fractals, and complexity, along with mathematical (theoretical) models supporting the emergence of chaos in cancer dynamics such that they can be extended to experimental systems oncology.

Chaos arises from non-linear interactions between components (like genes, proteins, cells, physiological systems, and environment) in complex systems (like cancer). Insights from mathematical oncology indicate that the phase transition from criticality (in healthy systems) to *chaos* may be a hallmark of cancer dynamics. Prigogine describes this complex instability as a causal pattern. Unlike disorder or randomness (uncertainty), as measured by entropy, chaos is a type of *unstable order*, with complex temporal sequences of motion or patterns of behaviour [12]. These unstable orders take geometric/topological forms known as attractors. Attractors are mathematical patterns, i.e. the *implicate order*(s), characterizing the evolution/*behaviour* of the complex system in phase space (state space). It is a state, or set of states, towards which a complex system tends to evolve or unfold [11,13]. While networks represent connectivity patterns—structure-function relationships defined by interactions within the complex system—the temporal trajectories that shape the system's adaptive behaviors are governed by attractors within the network's state space. As the critical parameter changes, forks or *bifurcations* emerge in possible system states, cascades of which are a hallmark of chaotic transitions. Each bifurcation expands possible behaviours into new complexity. Tracking shifts through bifurcations elucidates how complex dynamics emerge from simpler origins via feedback systems. In higher dimensions, Hopf and other bifurcations bring about several types of attractors, including limit cycles, tori, and strange attractors [13].

There are diverse types of transitions/bifurcation routes to chaos, including the Ruelle–Takens route (i.e. the emergence of a strange attractor), the Li–Yorke route (period 3 means chaos), *period-doubling bifurcations*, the *intermittency* seen in turbulent fluid dynamics, and the Smale horseshoe map, to name a few. Mathematical oncology reveals that our discussions on cancer dynamics in multi-dimensional state spaces are most fruitful with the Ruelle–Takens route. In the 1970s, Ruelle proposed the concept of *strange attractors* to describe the behaviour of chaotic systems such as turbulence, while scientists like Lorenz discovered these fractal structures in weather forecasting patterns [13,101]. Strange attractors exhibit sensitivity to initial conditions—minor changes (fluctuations) or perturbations within the initial conditions can lead to widely divergent outcomes/trajectories over time (i.e. errors can exponentially grow over time, making long-term forecasting unpredictable) [13,102]. Paradoxically, while the trajectory of the chaotic system is unpredictable over time, its global dynamics or evolution is bound to predictable states, i.e. the fractal attractors, within phase space [13,103]. The chaotic dynamics can be visualized by the *stretching and folding* of phase space into pastries or taffies, the cross-section of which resembles a Cantor's set (a fractal) [13,104] (see **Figure 40.2**).

Quantitative studies in mathematical oncology have repeatedly shown the importance of strange attractors in cancer dynamics prediction, such as within chaotic models of cancer growth dynamics and tumour progression. These models indicate strange attractors may be a biomarker, i.e. a pathological hallmark, of cancer states, in time-series or longitudinal cancer datasets. This may help explain clinically relevant barriers to treatment efficiency, such as cancer heterogeneity, metastatic emergence, and therapy resistance. This also means two genetically identical tumours may progress very differently over time. Mathematical models suggest strange attractors are signatures of cancer progression, and prognostic indicators of tumour recurrence, long-term tumour relapse, aggressivity, and

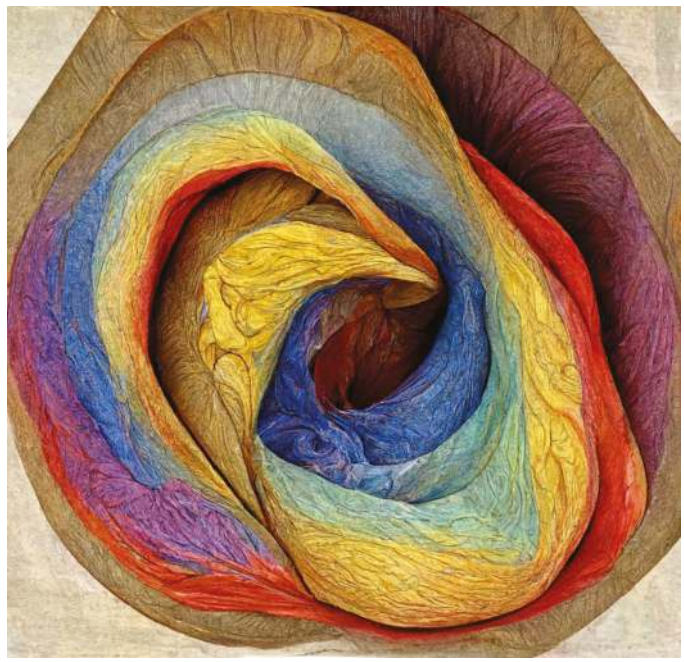


Figure 40.2. Strange attractor. An artwork of a strange attractor made by generative AI (Midjourney) in the style of Hilma af Klint. The stretching–folding dynamics of the chaotic attractor is well visualized by the AI system.

therapy evasion [102,105–107]. Therefore, let us briefly gain some insights into *strange attractors* in the context of some mathematical/computational models of cancer dynamics that can be overlaid on experimental cancer data.

The most common continuum approach for growth models is *reaction–diffusion systems* characterized by partial differential equations describing the pattern formation system of interest, such as the extra-cellular matrix remodelling in cancer dynamics [105,106]. For instance, Itik and Banks [102] used reaction–diffusion equations to compute the (positive) Lyapunov exponents and fractal dimensions of the tumour growth models. They explicitly demonstrated a Lorenz-like chaotic attractor in their 3D tumour–host–immune model, with stretching and folding trajectories. Ruelle [101] mentions Smale's work on proving the existence of strange attractors via horseshoe-type mappings. Similarly, Itik and Banks [102] use Shilnikov's theorem, which builds on Smale's horseshoe idea, to demonstrate chaos. Further, Khajanchi et al. [107] used time-delay differential equations to study chaotic dynamics in a similar simplified model of tumour ecology. Time lags are the standard approach for embedding algorithms and attractor reconstruction. More recently, Debbouche et al. [108] studied chaotic behaviours in mathematical cancer models of cancer–immune–host cell dynamics. They used bifurcation diagrams, Lyapunov exponents, and phase plots to confirm chaotic dynamics. Further, the fractional-order differential equations used by Debbouche et al. [108] are capable of merging fuzzy logic systems with chaotic dynamics, allowing fuzzy AI systems as a candidate machine-learning model for cancer ecosystem dynamics modelling.

Most of these models were reduced to Poincare's famous *three-body problem* from which chaos theory took birth [109]. That is, they involve interactions between the tumour cells, host/healthy tissue cells, and immune system cells. The vast complexity of their phenotypes has been limited to a homogeneous three-body system, and yet even such simple models exhibit chaotic dynamics. More complex cancer models suggest we need a multidimensional attractor state space in four dimensions or more, requiring *hyperchaos* and multifractal attractors (i.e. strange attractors with a fractal dimension > 3) to define cancer behavioural dynamics [110]. For instance, Ivankevic et al. [105] proposed a 4D spatiotemporal model coupling reaction–diffusion equations to non-linear interactions between tumour cells, matrix metalloproteinases, matrix-degradative enzymes, and oxygen.

Various other interdisciplinary approaches are using complex systems tools available in the study of tumour behaviours as chaotic attractor dynamics. For example, a chaotic model of tumour growth and decay [111] was demonstrated using control theory principles. Izquierdo-Kulich et al. [112] used the entropy production rate of tumours to show the fractal growth dynamics of avascular tumours. Ecological–evolutionary models of cancer treatment regimens and tumour responses have also emerged within *game theory* approaches to devise adaptive treatment strategies [113]. Game-theoretic approaches with predator–prey-like dynamics of the tumour ecosystem is an alternative to reaction–diffusion-based patterning/population dynamics. The fractal dimension measure has also been used to spatially profile the contours and surface complexity of tumour structures. For instance, fractal dimension analysis of lung cancers and CT scan analyses of tumours suggest fractals as signatures of tumour patterning [5,114,115]. The fractal dimension has also been proposed as a prognostic indicator for assessing the irregularities or complexity of tumour-associated chromatin structures [116].

In principle, the above-discussed chaos detection methods, from the mathematical models such as fractal dimension analysis, bifurcation analysis, and time-delay embedding algorithms such as Takens's theorem or recurrence quantification analysis, can be used for attractor reconstruction. However, their applications may be limited for complex data spaces such as those of single-cell multi-omics largely due to its size and multi-dimensionality. As such, AI provides a robust tool for chaos detection and strange attractor reconstruction. One such AI algorithm is reservoir computing (RC). RC is a predictive machine, generated from several prior recurrent neural network (RNN) models, such as the echo-state networks and liquid-state machines, allowing model-free prediction of non-linear dynamics. RC has been used for complex attractor reconstruction within multidimensional spatiotemporally chaotic systems [117–119]. RC provides an efficient approach for dimensionality expansion and attractor reconstruction from sequential gene expression data [120]. To demonstrate, Sayari et al. [121] used an echo-state network, a type of RNN at the basis of RC, with mathematical cancer models of host–immune–tumour cell dynamics. They performed a time-series machine-learning analysis and predicted the parameters at which complex oscillatory dynamics emerge in the phase space of these cancer systems. A more recent approach denotes *liquid neural networks*, a class of continuous-time neural networks, and an adaptive AI that learns beyond the training phase—may be most apt for modelling complex decision-making problems, involving continuous differential equations, such as those relevant to cell-fate transition dynamics, and reaction–diffusion-based pattern formation systems [122]. Time-series predictions can also be extended to AI architectures such as transformers and graph neural networks. Therefore, future research in systems oncology should exploit these mathematical models of chaos and AI systems to infer strange attractor dynamics from experimental cancer datasets.

40.5. Chemical turbulence: soft matter physics, fluid dynamics, and chaotic oscillations

As discussed, by the works of Ruelle [101] and Lorenz [123], in many ways, cancer dynamics prediction is analogous to weather forecasting. Like the discussed cancer models, weather patterns and their underlying fluid turbulence exhibit intermittency and fractal patterns across many scales of interactions. Turbulence, a complex system with infinite complexity, is a novelty generator, creating counterfactual flow patterns in its temporal progression. Yet, the infinite complexity is paradoxically bound to the (multi)fractal patterns and strange attractors [101,124]. The intersection of strange attractor dynamics and tumour pattern formation (morphogenesis) suggests that *chemical turbulence* may be a hallmark of cancer dynamics, explaining its heterogeneity, aggressivity, and tumour progression/invasion. Turbulence may occur in various forms of chemical systems, from protein-mediated heterogeneous pattern formation systems to the collective cellular migration patterns seen in metastatic invasion. Turing, the father of modern computing, first devised a simplistic mathematical model of morphogenesis (1952) [125]. Turing patterns are spatial patterns that emerge spontaneously from the

interaction of diffusing chemical morphogens through *a reaction–diffusion system*. They represent a paradigm for how complex order and patterns can arise in biology through self-organization and have been hypothesized to underlie tissue differentiation and developmental processes. Reaction–diffusion systems may shed light on the mechanisms of cancer initiation and progression, such as tumour heterogeneity of phenotypes, hypoxia, and angiogenesis. While these systems can describe 'solid tumours', when considering more fluid-like cancer models, i.e. liquid malignancies, such as leukaemia and lymphoma, hydrodynamic models or clonal evolution models may be needed [126,127]. Regardless of the spectrum of fluidity, all tumours fall within the study of biological soft matter systems and warrant fluid dynamics description [128]. This fluid, adaptive nature may also underlie cancer's process of evolvability, i.e., capacity for novelty generation—framing cancer intelligence as a form of biological creativity.

In extension to reaction–diffusion systems, works by Frey and Halatek [129] have demonstrated both via theoretical simulations and experimental *in vitro* reconstitution of the bacterial Min protein systems so that intracellular protein flows can create chaotic attractors. The self-organized protein patterns of the oscillation in between the membrane-bound conformation (on lipid bilayers) and the free cytosolic protein conformation, for even simple isolated chemical systems like the Min proteins, were shown to exhibit *chaotic oscillations* [128,130]. The wave-like turbulence can move with or against the flow direction, depending on the protein concentration ratios [130]. Hence, imagine the complexity emerging from the orchestra and non-linear cascades of thousands of protein interactions, in a disease context, like cancer dynamics (Frey and Brauns, 2022). The mammalian equivalent of the Min system, the Par protein system, plays a key role in cancer decision-making processes, such as cell polarity in morphogenesis, cell division, and 3D-chromatin organization/modelling during cell replication [131] or invasion. As such, these multiscale models of chaotic dynamics warrant extension towards cancer protein-mediated patterning systems, such as the Par system [129].

Chaotic oscillations in many-body fluid systems are referred to as 'chemical turbulence', as was first discussed by Rössler and Kuramoto in their study of reaction–diffusion patterning systems. Kuramoto and Rössler argue that chaotic dynamics like strange attractors could arise spontaneously in chemical systems *far from equilibrium*, analogous to hydrodynamic turbulence of fluids [132–134]. Similar theories were proposed by Prigogine for the emergence of self-organized patterns in chemical kinetic reactions in nonequilibrium systems [12]. The above-discussed mathematical cancer growth models exhibited strange attractors emerging from non-linear interactions between various cell populations and protein molecules like matrix remodelling enzymes [105,110]. This spontaneous pattern formation in population dynamics can also be referred to as *chemical turbulence*. Intermittency is a common property of such turbulent systems giving rise to these fractal attractor structures. Rössler is also credited for deriving the simplest of strange attractors known to mathematics, the *Rössler attractor* [132,133].

Chemical turbulence can occur at the scale of intracellular protein dynamics. In principle, any oscillating chemical reaction can give rise to spatiotemporal patterns of chaos in a complex environment, such as the TME. For instance, enzymatic reaction–diffusion models have exhibited hyperchaotic dynamics, referred to as *chemical turbulence* [135]. Chaotic and turbulent states were shown to emerge from the interaction of a Hopf and Turing instability within activator–inhibitor models characterizing these enzymatic processes [135]. Heltberg and Jensen studied simple mathematical models of cellular protein reaction networks and demonstrated they could exhibit chaotic oscillations. They demonstrated how chaotic oscillations can emerge in metabolic processes, such as glycolytic oscillations, and within the transcription factor NF-κB, to affect downstream protein production, modulate gene expression, and up-regulate certain families of low-affinity gene networks [136,137]. Hence, they proposed that chaotic oscillations of protein networks create a *heterogeneous* population of cell states, as seen in cancer ecosystems. More recently, they demonstrated that chaotic oscillations may occur at the level of transcriptional protein systems that have shown liquid droplet formations (i.e. liquid–liquid phase transitions), such as in the cancer p53 DNA-repair system's oscillatory dynamics [138]. The liquid–liquid phase separations have significant roles in cancer processes ranging from chromatin remodelling to pattern formation/morphogenesis. The intrinsically disordered proteins, at the basis of these emergent behaviours, are defined as *edge of chaos* (critical) systems [128].

Turbulence dynamics as described by Kolmogorov's fluid turbulence theory have also been applied to protein folding dynamics, like that of the SH3 domain found in focal adhesion complexes regulating cancer metastasis and ECM remodelling. The simulations of the protein folding resembled turbulent fluid flows, with cascading vortex structures and fractal eddies containing strange attractors [139,140]. Hence, perhaps the protein folding of mutant oncoproteins exhibits a variation of such turbulence dynamics, suggesting that this avenue of research in combination with AI systems like AlphaFold 2 is a prospective direction for proteomics in disease contexts and cancer interactomes [54].

Mesoscale turbulence can also occur in tumour patterning via an emerging system in (condensed) soft matter physics known as active fluids. Active fluids, or active matter systems, are complex systems wherein energy-driven, self-propelled particles, or particles, such as cytoskeletal suspensions of microtubules equipped with energy-driven motors like ATP, or ecological dynamics such as the flocking of birds, swarming of bacteria, and crowding of insects, exhibit fluid turbulence-like behaviours [141,142]. Active turbulence requires continuum-level descriptions by hydrodynamic equations, such as variants of the *Navier–Stokes equations*. Similarly, cancer cells may also be exhibiting swarming or flocking-like collective cell behaviours during migration, invasion, and pattern formation, and in theory, within their nanoscale dynamics such as extra-cellular vesicle-mediated communications. Extra-cellular vesicles are nanoscopic systems released by cancer cells to hijack host cells outside of the local tumour context, reprogramming them towards tumorigenic or tumour-supporting cells (e.g. support tumour-immune evasion) and as pre-metastatic niches [27,143]. Modelling the complex hydrodynamics of these nanosystems may be critical for quantitative precision medicine and personalized healthcare, namely in liquid biopsies screening, and targeted anti-metastatic drug development. It is proposed herein that we can better tailor precision therapies such as engineer nanomedicine (e.g. nanovesicle or nanoparticle drug delivery, oncolytic viral therapy, and other immunotherapies) via studying the collective behaviours of such nanosystems as complex fluid dynamics [144].

The counter-intuitive idea in this field of research is that active matter systems, such as cytoskeletal protein fluids or cellular *flocking*, exhibit turbulence at low Reynolds numbers [145–147]. To illustrate a cancer-pertinent example, Tan et al. [148] showed that defect-mediated turbulence underlies complex wave propagation patterns of Rho-GTP signalling proteins on starfish egg cell membranes. These Rho-GTP protein networks play central roles in cancer cell motility, migration, and invasion [149]. Active turbulence models may help us understand collective migrations and metastatic invasion in cancer progression [150]. For instance, a study by Lin et al. [151] demonstrated that the energy spectra of a 2D monolayer cell migration model exhibited a power-decaying law at large wavenumbers, characteristic of turbulent/complex dynamics. In collective cell migration, the mechanical waves of the collective cell movements may also give rise to a phenomenon known as viscoelastic turbulence, experimentally observed in low Reynolds numbers [152]. Thus, cell flocking, or *active turbulence*, as seen in developmental epithelial systems with correlation lengths spanning several to dozens of cells [153], can help understand collective cell dynamics in cancer systems [150].

Another system of key interest in cancer dynamics exhibiting chemical turbulence is microtubules, involved in cancer processes, such as proliferation, scaffolding for intracellular macromolecular transport, and cytoskeletal/morphology rearrangements in cancer plasticity dynamics [154,155]. Microtubules have been shown to exhibit *collective emergent behaviours*, such as active matter turbulence, and may have an underlying *strange attractor* structure. Microtubule dynamics/remodelling is also the basis of *chromatin segregation* or aneuploidy in (cancer) cell division [155]. As such, it is proposed herein that investigating attractor reconstruction in cytoskeletal dynamics may be insightful to controlling cell-fate decisions, such as plasticity/state transitions, and understanding cancer dynamics through the lens of chaos and complexity.

It must also be emphasized that as we go down to the nanoscale and below, the regimes and laws of physics change, i.e. the *quantum* domain of biophysics dominates. There are many nanoscale molecular systems, such as the microtubule dynamics, and energy transport machineries of cells, which overlap within the emerging field of *quantum biology*, which proposes that the macroscopic quantum effects/behaviours of biomolecules may have functional roles in living systems. Some examples include quantum tunnelling in DNA as a basis for some tautomeric mutations, enzymatic activities, and extended quantum coherence in protein condensates, i.e. collective electronic excitations such as dipole–dipole interactions in protein molecules, such as microtubules [156–158]. If further empirically validated, these quantum processes at the level of DNA mutations and microtubule dynamics may be relevant to carcinogenesis and cancer cell division and progression [157,159,160]. These quantum biology processes operate as another layer of complexity, for the future of systems oncology. Quantum nanophotonic/optics, single-molecule quantum biosensors, and tools, such as optical multidimensional spectroscopy, may help probe these complex nanoscale quantum systems and ultrafast processes in cancer systems.

Lastly, there may be bioelectric analogues to explain pattern formation dynamics, describing cancer cell-fate decision-making as a goal-directed computational system driven by bioelectric fields [161]. As discussed, neural–glial interactions and a hijacked or reprogrammed neurodevelopmental (embryonic) landscape promote phenotypic plasticity in HGGs [19]. The neural interactions underlying plasticity suggest that perhaps there may be an underlying *bioelectric code*, i.e. bioelectric signalling as a driver of the morphogenetic patterns coordinating cancer development. In support of this counterfactual idea, Levin et al. propose that voltage membrane potentials and bioelectric fields generated by the energy metabolome provide a basis for the collective intelligence of cancer cells, instructing patterning cues that guide cell behaviours, such as spatial navigation and large-scale anatomy during development [160,162]. Disruption of these endogenous bioelectric signals can induce changes in the collective intelligence of cancer cells leading to aggressive, and metastatic, cancerous phenotypes [161].

40.6. Conclusions

In summary, this chapter discusses a culmination of techniques, quantitative approaches, and causality inference tools to investigate *complex dynamics*, such as chemical turbulence, and strange attractors, as biomarkers of cancer progression. *Strange attractors* have been repeatedly confirmed in cancer growth and progression models in mathematical oncology, wherein the number of cellular components or number of population subtypes used in the model is an indicator of the dimensionality of the chaotic attractor. Chaotic dynamics may confer adaptability and aggressiveness to cancer cells, highlighting the utility of complex systems methods for their modelling, prediction, and control of cancer dynamics. As evidenced by the discussed mathematical models, it is proposed herein that TME dynamics and therapy-induced stressors can increase tumour plasticity, an adaptive (emergent) behaviour, towards complex chaotic attractors, as hallmarks of cancer progression. Strange attractors may be signatures of therapy evasion, tumour relapse, and recurrence [19,98,107]. Overall, our discussions highlight how mathematical modelling and analysis of non-linear tumour–immune dynamics can reveal complex behaviours like chaos that may impact cancer progression and clinical outcomes.

Complex and chaotic chemical oscillations, such as the oscillations of protein networks or gene products, can create chaotic dynamics. Experimental approaches such as cytoplasmic live-cell biopsies for temporal profiling of single-cell dynamics or cell-labelled reporter systems, such as RNA-labelling (gene expression dynamics), CITE-seq, and fluorescent reporter protein imaging, can be used to capture time-series dynamics followed by chaos detection tools. Traditional methods like time-delay embedding, Lyapunov exponents, and fractal analysis can help identify strange attractors in cancer dynamics. However, these approaches have dimensionality limits within large datasets posing an attractor reconstruction problem in cancer data science. Hence, combining multiscale physics models with computational algorithms, including machine-learning algorithms, such as RC networks, liquid neural networks, and generative AI, can reconstruct high-dimensional attractors and predict chaotic dynamics from the multidimensional pattern spaces of cancer multi-omics. In prospect, the hierarchical 3D organization of chromatin, as mapped by Hi-C sequencing, is critical for cancer development and differentiation [96]. However, we lack a full understanding of the dynamic interplay between 3D-chromatin remodelling in nuclear space and the gene expression patterns that drive cancer progression. Thus, complex systems research should shed light on this

realm of complexity [96]. Further, photonic quantum technologies, and quantum information sciences, such as quantum machine intelligence, hold great promise in the study of cancer dynamics.

On a final note, top-down causation (e.g. social risk factors, cultural contexts, psychosocial profiles, etc.) warrant equal importance in the time-series/longitudinal multi-omics sequencing analysis, as bottom-up approaches (genes, proteins, molecules, cells, etc.), to fully integrate cancer dynamics in patient-centred medicine. The top-down factors may be critical especially to grasp the epigenetic control of cancer dynamics and pave trauma-informed precision care [163,164]. For instance, environmental stressors, such as those mediated by a lack of support from family and social/community structures, developmental histories, and personalized cultural contexts, can largely affect therapeutic progression, and hence, cancer dynamics. The role of these psychosocial determinants in cancer development is emerging in the field of *psychosocial epigenetics* as well [163,164]. While therapy-induced stressors are known to induce cancer phenotypic plasticity, research into such psychosocial trauma and stress-induced cancer plasticity remains at its infancy and can largely affect personalized medicine. It remains a fundamental challenge how to integrate these psychosocial–cultural axes of health dynamics into biopsy-derived cellular systems and their data science. The awareness and thought of such complexity are planted herein. To conclude, the complex systems toolkit and longitudinal (time-series) cancer data need to address how the outlined intersectionality framework of these psychosocial determinants affects gene expression dynamics or multi-omics profiles, in cancer patients, as the future of *human-centred* systems medicine. As such, systems oncology and computational medicine should serve as a calling for human-centred best practices in science and medicine, fostering love for humanity, and compassionate care within the art of healing *humans*, the 'patients' devoid of the patient label and stigma [165].

Acknowledgements: My humble gratitude goes to the children in B7 at the Montreal Children's Hospital for all they have taught me.

REFERENCES

1. Sanga S, Sinek JP, Frieboes HB, Ferrari M, Fruehauf JP, Cristini V. Mathematical modeling of cancer progression and response to chemotherapy. Exp Rev Anticancer Ther. 2006;**6**(10):1361–76. https://doi.org/10.1586/14737140.6.10.1361
2. Frost JJ. What cancer Is. In: Balaz I, Adamatzky A, editors. Cancer, complexity, computation. emergence, complexity and computation. Vol **46**. Cham: Springer; 2022. https://doi.org/10.1007/978-3-031-04379-6_1
3. Casadei B, Giacosa M, Maula A, Plos S, Zappulla L, Viotto C, et al. Complexities of drug resistance in cancer: an overview of strategies and mathematical models. In: Balaz I, Adamatzky A, editors. Cancer, complexity, computation. Emergence, complexity and computation. Vol. **46**. Cham: Springer; 2022. https://doi.org/10.1007/978-3-031-04379-6_14
4. Sun G, Li Z, Rong D, Zhang H, Shi X, Yang W, et al. Single-cell RNA sequencing in cancer: applications, advances, and emerging challenges. Mol Ther Oncolytics. 2021;**21**: 183–206. https://doi.org/10.1016/j.omto.2021.04.001
5. Coffey DS. Self-organization, complexity and chaos: the new biology for medicine. Nat Med. 1998;**4**(8):882–5. https://doi.org/10.1038/nm0898-882
6. Mallick P. Complexity and information: cancer as a multi-scale complex adaptive system. In: Janmey P, et al., editors. Physical sciences and engineering advances in life sciences and oncology. Science Policy Reports. Cham: Springer; 2016. https://doi.org/10.1007/978-3-319-17930-8_2
7. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;**12**(1):31–46. https://doi.org/10.1158/2159-8290.CD-21-1059
8. Cai Z, Poulos RC, Liu J, Zhong Q. Machine learning for multi-omics data integration in cancer. iScience. 2022;**25**(2):103798. https://doi.org/10.1016/j.isci.2022.103798
9. de Anda-Jáuregui G, Hernández-Lemus, E. Computational oncology in the multi-omics era: state of the art. Front Oncol. 2020;**10**:423. https://doi.org/10.3389/fonc.2020.00423
10. Heo YJ, Hwa C, Lee GH, Park JM, An JY. Integrative multi-omics approaches in cancer research: from biological networks to clinical subtypes. Mol Cells. 2021 Jul 31;**44**(7):433–43. doi: 10.14348/molcells.2021.0042. PMID: 34238766; PMCID: PMC8334347.
11. Bohm, D. Wholeness and the implicate order. New York: Routledge; 1980.
12. Prigogine I. From being to becoming. Time and complexity in the physical sciences. San Francisco: W.H. Freeman and Company; 1980.
13. Gleick J. Chaos: making a new science. New York, NY: Penguin; 1988.
14. Bossomaier TRJ, Green DG. Complex systems. Cambridge, UK: Cambridge University Press; 2000.
15. Anderson PW. More is different. Science. 1972;**177**:393–6. http://dx.doi.org/10.1126/science.177.4047.393
16. Varela FG, Maturana HR, Uribe R. Autopoiesis: the organization of living systems, its characterization and a model. Curr Mod Biol. 1974;**5**(4):187–96. https://doi.org/10.1016/0303-2647(74)90031-8
17. Noe A, Thompson E, editors. Vision and mind: selected readings in the philosophy of perception. Cambridge, MA: MIT Press; 2002.
18. Bohm D, Hiley, BJ. In Hiley BJ, editor. The undivided universe: an ontological interpretation of quantum theory. New York: Routledge; 1993.
19. De Silva MI, Stringer BW, Bardy C. Neuronal and tumourigenic boundaries of glioblastoma plasticity. Trends Cancer. 2023;**9**(3):223–36. https://doi.org/10.1016/j.trecan.2022.10.010
20. Liu J, Qu S, Zhang T, Gao Y, Shi H, Song K, et al. Applications of single-cell omics in tumor immunology. Front Immunol. 2021;**12**:697412. https://doi.org/10.3389/fimmu.2021.697412
21. Ma A, Xin G. Ma Q. The use of single-cell multi-omics in immuno-oncology. Nat Commun. 2022;**13**:2728. https://doi.org/10.1038/s41467-022-30549-4
22. Swartz MA, Iida N, Roberts EW, Sangaletti S, Wong MH, Yull FE, et al. Tumor microenvironment complexity: emerging roles in cancer therapy. Cancer Res. 2012; **72**(10):2473–80. https://doi.org/10.1158/0008-5472.CAN-12-0122
23. Buchta Rosean C, Feng TY, Azar FN, Rutkowski MR. Impact of the microbiome on cancer progression and response to anti-cancer therapies. Adv Cancer Res. 2019;**143**: 255–94. https://doi.org/10.1016/bs.acr.2019.03.005
24. Zong Y, Zhou Y, Liao B, Liao M, Shi Y, Wei Y, et al. The interaction between the microbiome and tumors. Front Cell Infect Microbiol. 2021;**11**:673724. https://doi.org/10.3389/fcimb.2021.673724

25. Drexler R, Khatri R, Sauvigny T, Mohme M, Maire CL, Ryba A, et al. Epigenetic neural glioblastoma enhances synaptic integration and predicts therapeutic vulnerability. bioRxiv 2023.08.04.552017. 2023. doi: https://doi.org/10.1101/2023.08.04.552017
26. Venkatesh HS, Morishita W, Geraghty AC, Silverbush D, Gillespie SM, Arzt M, et al. Electrical and synaptic integration of glioma into neural circuits. Nature. 2019; **573**(7775):539–45. https://doi.org/10.1038/s41586-019-1563-y
27. Dong Q, Liu X, Cheng K, Sheng J, Kong J, Liu T. Pre-metastatic niche formation in different organs induced by tumor extracellular vesicles. Front Cell Dev Biol. 2021;**9**: 733627. https://doi.org/10.3389/fcell.2021.733627
28. Ghanam J, Chetty VK, Barthel L, Reinhardt D, Hoyer P-F, Thakur BK, et al. DNA in extracellular vesicles: from evolution to its current application in health and disease. Cell Biosci. 2022;**12**:37. https://doi.org/10.1186/s13578-022-00771-0
29. Ahmad R, Budnik B. A review of the current state of single-cell proteomics and future perspective. Anal Bioanal Chem. 2023;**415**:6889–99. https://doi.org/10.1007/s00216-023-04759-8
30. McClellan BL, Haase S, Nunez FJ, Alghamri MS, Dabaja AA, Lowenstein PR, et al. Impact of epigenetic reprogramming on antitumor immune responses in glioma. J Clin Invest. 2023;**133**(2):e163450. https://doi.org/10.1172/JCI163450
31. Harutyunyan AS, Krug B, Chen H, Papillon-Cavanagh S, Zeinieh M, De Jay N, et al. H3K27M induces defective chromatin spread of PRC2-mediated repressive H3K27me2/me3 and is essential for glioma tumorigenesis. Nat Commun. 2019;**10**(1):1262.
32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;**15**:550. doi: 10.1186/s13059-014-0550-8.
33. Li Y, Ge X, Peng F, Li W, Li JJ. Exaggerated false positives by popular differential expression methods when analyzing human population samples. Genome Biol. 2022;**23**:79. https://doi.org/10.1186/s13059-022-02648-4
34. Granja JM, Klemm S, McGinnis LM, Kathiria AS, Mezger A, Corces MR, et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. Nat Biotechnol. 2019;**37**(12):1458–65. https://doi.org/10.1038/s41587-019-0332-7
35. Jessa S, Blanchet-Cohen A, Krug B, Vladoiu M, Coutelier M, Faury D, et al. Stalled developmental programs at the root of pediatric brain tumors. Nat Genet. 2019;**51**(12):1702–13. https://doi.org/10.1038/s41588-019-0531-7
36. Huang S. Reconciling non-genetic plasticity with somatic evolution in cancer. Trends Cancer. 2021;**7**(4):309–22. https://doi.org/10.1016/j.trecan.2020.12.007
37. McEachron TA, Helman LJ. Recent advances in pediatric cancer research. Cancer Res. 2021;**81**(23):5783–99. https://doi.org/10.1158/0008-5472.CAN-21-1191
38. Deshmukh S, Ptack A, Krug B, Jabado N. Oncohistones: a roadmap to stalled development. FEBS J. 2022;**289**(5):1315–28. https://doi.org/10.1111/febs.15963
39. Huang S, Ernberg I, Kauffman S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. Semin Cell Dev Biol. 2009;**20**(7):869–76. https://doi.org/10.1016/j.semcdb.2009.07.003
40. Li Q, Wennborg A, Aurell E, Dekel E, Zou JZ, Xu Y, et al. Dynamics inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape. Proc Natl Acad Sci U S A. 2016;**113**(10):2672–77. https://doi.org/10.1073/pnas.1519210113
41. Schiebinger G, Shu J, Tabaka M, Cleary B, Subramanian V, Solomon A, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. Cell. 2019;**176**(4):928–43.e22. https://doi.org/10.1016/j.cell.2019.01.006
42. Huizing GJ, Peyré G, Cantini L. Optimal transport improves cell-cell similarity inference in single-cell omics data. Bioinformatics (Oxford, England). 2022;**38**(8):2169–177. https://doi.org/10.1093/bioinformatics/btac084
43. Forrow A, Schiebinger G. LineageOT is a unified framework for lineage tracing and trajectory inference. Nat Commun. 2021;**12**:4940. https://doi.org/10.1038/s41467-021-25133-1
44. Hari K, Ullanat V, Balasubramanian A, Gopalan A, Jolly MK. Landscape of epithelial-mesenchymal plasticity as an emergent property of coordinated teams in regulatory networks. eLife. 2022;**11**:e76535. https://doi.org/10.7554/eLife.76535
45. Pillai M, Chen Z, Jolly MK, Li C. Quantitative landscapes reveal trajectories of cell-state transitions associated with drug resistance in melanoma. iScience. 2022;**25**(12):105499. https://doi.org/10.1016/j.isci.2022.105499
46. Sáez M, Blassberg R, Camacho-Aguilar E, Siggia ED, Rand DA, Briscoe J. Statistically derived geometrical landscapes capture principles of decision-making dynamics during cell fate transitions. Cell Syst. 2022;**13**(1):12–28.e3. https://doi.org/10.1016/j.cels.2021.08.013
47. Shi J, Teschendorff AE, Chen W, Chen L, Li T. Quantifying Waddington's epigenetic landscape: a comparison of single-cell potency measures. Brief Bioinform. 2020;**21**(1): 248–61. https://doi.org/10.1093/bib/bby093
48. Vallejos CA, Marioni JC, Richardson S. BASiCS: Bayesian analysis of single-cell sequencing data. PLoS Comput Biol. 2015;**11**(6):e1004333. https://doi.org/10.1371/journal.pcbi.1004333
49. Breda J, Zavolan M, van Nimwegen E. Bayesian inference of gene expression states from single-cell RNA-seq data. Nat Biotechnol. 2021;**39**(8):1008–16. https://doi.org/10.1038/s41587-021-00875-x
50. Zhou P, Wang S, Li T, Nie Q. Dissecting transition cells from single-cell transcriptome data through multiscale stochastic dynamics. Nat Commun. 2021;**12**:5609. https://doi.org/10.1038/s41467-021-25548-w
51. Shin H, Choi BH, Shim O, Kim J, Park Y, Cho SK, et al. Single test-based diagnosis of multiple cancer types using Exosome-SERS-AI for early stage cancers. Nat Commun. 2023;**14**:1644. https://doi.org/10.1038/s41467-023-37403-1
52. Stackpole ML, Zeng W, Li S, Liu CC, Zhou Y, He S, et al. Cost-effective methylome sequencing of cell-free DNA for accurately detecting and locating cancer. Nat Commun. 2022;**13**(1):5566. https://doi.org/10.1038/s41467-022-32995-6
53. Tanoli Z, Vähä-Koskela M, Aittokallio T. Artificial intelligence, machine learning, and drug repurposing in cancer. Exp Opin Drug Discov. 2021;**16**(9):977–89. https://doi.org/10.1080/17460441.2021.1883585
54. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;**596**:583–9. https://doi.org/10.1038/s41586-021-03819-2
55. You Y, Lai X, Pan Y, Zheng H, Vera J, Liu S, et al. Artificial intelligence in cancer target identification and drug discovery. Sig Transduct Target Ther. 2022;**7**:156. https://doi.org/10.1038/s41392-022-00994-0
56. Guo J, Zheng J. HopLand: Single-cell pseudotime recovery using continuous Hopfield network-based modeling of Waddington's

epigenetic landscape. Bioinformatics. 2017;**33**(14):i102–9. doi: 10.1093/bioinformatics/btx232
57. Liu Y, Baggerly KA, Orouji E, Manyam G, Chen H, Lam M, et al. Methylation-eQTL analysis in cancer research. Bioinformatics (Oxford, England), 2021;**37**(22):4014–22. https://doi.org/10.1093/bioinformatics/btab443
58. Treppner M, Binder H, Hess M. Interpretable generative deep learning: an illustration with single cell gene expression data. Hum Genet. 2022;**141**:1481–98. https://doi.org/10.1007/s00439-021-02417-6
59. van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ, et al. Recovering gene interactions from single-cell data using data diffusion. Cell. 2018;**174**(3):716–29.e27. https://doi.org/10.1016/j.cell.2018.05.061
60. Yeo GHT, Saksena SD. Gifford DK. Generative modeling of single-cell time series with PRESCIENT enables prediction of cell trajectories with interventions. Nat Commun. 2021;**12**:3222. https://doi.org/10.1038/s41467-021-23518-w
61. Cui H, Wang C, Maan H, Pang K, Luo F, Wang B. scGPT: towards building a foundation model for single-cell multi-omics using generative AI. bioRXiv [online preprint]. 2023.
62. Guo Z, Liu J, Wang Y, Chen M, Wang D, Xu D, et al. Diffusion models in bioinformatics: a new wave of deep learning revolution in action. arXiv, 13 Feb. 2023. arXiv.org, https://doi.org/10.48550/arXiv.2302.10907.
63. Altug H, Oh SH, Maier SA, Homola J. Advances and applications of nanophotonic biosensors. Nat Nanotechnol. 2022;**17**:5–16. https://doi.org/10.1038/s41565-021-01045-5
64. Taniguchi M, Ohshiro T, Tada T. Single-molecule identification of nucleotides using a quantum computer. J Phys Chem B. 2023. **127**(30):6636–42. https://doi.org/10.1021/acs.jpcb.3c02918
65. Yu CJ, von Kugelgen S, Laorenza DW, Freedman DE. A molecular approach to quantum sensing. ACS Cent Sci. 2021;**7**(5):712–23. https://doi.org/10.1021/acscentsci.0c00737
66. Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S. Quantum machine learning. Nature. 2017;**549**: 195–202. https://doi.org/10.1038/nature23474
67. Kao PY, Yang YC, Chiang WY, Hsiao JY, Cao Y, Aliper A, et al. Exploring the advantages of quantum generative adversarial networks in generative chemistry. J Chem Inform Model. 2023;**63**(11):3307–18. https://doi.org/10.1021/acs.jcim.3c00562
68. Cordier BA, Sawaya NPD, Guerreschi GG, McWeeney SK. Biology and medicine in the landscape of quantum advantages. J R Soc Interface. 2022;**19**(196):20220541. https://doi.org/10.1098/rsif.2022.0541
69. Bashashati A, Haffari G, Ding J, Ha G, Lui K, Rosner J, et al. DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. Genome Biol. 2012;**13**:R124. https://doi.org/10.1186/gb-2012-13-12-r124
70. Gold P, Freedman SO. Specific carcinoembryonic antigens of the human digestive system. J Exp Med. 1965;**122**(3):467–81. https://doi.org/10.1084/jem.122.3.467
71. Yu J, Peng J, Chi H. Systems immunology: Integrating multi-omics data to infer regulatory networks and hidden drivers of immunity. Curr Opin Syst Biol. 2019;**15**:19–29. https://doi.org/10.1016/j.coisb.2019.03.003
72. Hao S, Yan KK, Ding L, Qian C, Chi H, Yu J. Network approaches for dissecting the immune system. iScience. 2020;**23**(8):101354. https://doi.org/10.1016/j.isci.2020.101354
73. Karimi MR, Karimi AH, Abolmaali S, Sadeghi M, Schmitz U. Prospects and challenges of cancer systems medicine: from genes to disease networks. Brief Bioinform. 2022;**23**(1):bbab343. https://doi.org/10.1093/bib/bbab343
74. Barabasi A. Network science. Cambridge, UK: Cambridge University Press; 2016.
75. Vulliard L, Menche J. Complex networks in health and disease. Syst Med. 2021;**3**:26–33. https://doi.org/10.1016/B978-0-12-801238-3.11640-X
76. Zhang L, Fan S, Vera J, Lai X. A network medicine approach for identifying diagnostic and prognostic biomarkers and exploring drug repurposing in human cancer. Comput Struct Biotechnol J. 2022;**21**:34–45. https://doi.org/10.1016/j.csbj.2022.11.037
77. Horaira MA, Islam MA, Kibria MK, Alam MJ, Kabir SR, Mollah MNH. Bioinformatics screening of colorectal-cancer causing molecular signatures through gene expression profiles to discover therapeutic targets and candidate agents. BMC Med Genomics. 2023;**16**:64. https://doi.org/10.1186/s12920-023-01488-w
78. Aldana M. Boolean dynamics of networks with scale-free topology. Phys D: Nonlin Phenom. 2003;**185**(1):45–66.
79. Albert R. Scale-free networks in cell biology. J Cell Sci. 2005;**118**(Pt 21):4947–57. https://doi.org/10.1242/jcs.02714
80. Newman MEJ. Power laws, Pareto distributions and Zipf's law. Contemp Phys. 2005;**46**: 323–51.
81. Zimmern V. Why brain criticality is clinically relevant: a scoping review. Front Neural Circuits. 2020;**14**:54. https://doi.org/10.3389/fncir.2020.00054
82. Torres-Sosa C, Huang S, Aldana M. Criticality is an emergent property of genetic networks that exhibit evolvability. PLoS Comput Biol. 2012;**8**(9):e1002669. https://doi.org/10.1371/journal.pcbi.1002669
83. Bak P. How nature works: the science of self-organized criticality/Per Bak. New York, NY: Copernicus; 1996.
84. Chu H, Lee D, Cho KH. Precritical state transition dynamics in the attractor landscape of a molecular interaction network underlying colorectal tumorigenesis. PLoS ONE. 2015; **10**(10):e0140172. https://doi.org/10.1371/journal.pone.0140172
85. Manicka S, Johnson K, Levin M, Murrugarra D. The nonlinearity of regulation in biological networks. NPJ Syst Biol Appl. 2023;**9**:10. https://doi.org/10.1038/s41540-023-00273-w
86. Manzar N, Ganguly P, Khan UK, Ateeq B. Transcription networks rewire gene repertoire to coordinate cellular reprograming in prostate cancer. Semin Cancer Biol. 2023;**89**:76–91. https://doi.org/10.1016/j.semcancer.2023.01.004
87. Clarke C, Madden SF, Doolan P, Aherne ST, Joyce H, O'Driscoll L, et al. Correlating transcriptional networks to breast cancer survival: a large-scale coexpression analysis. Carcinogenesis. 2013;**34**(10):2300–8. doi: 10.1093/carcin/bgt208
88. Sa JK, Chang N, Lee HW, Cho HJ, Ceccarelli M, Cerulo L, et al. Transcriptional regulatory networks of tumor-associated macrophages that drive malignancy in mesenchymal glioblastoma. Genome Biol. 2020;**21**:216. https://doi.org/10.1186/s13059-020-02140-x
89. Emad A, Sinha S. Inference of phenotype-relevant transcriptional regulatory networks elucidates cancer type-specific regulatory mechanisms in a pan-cancer study. NPJ Syst Biol Appl. 2021;**7**:9. https://doi.org/10.1038/s41540-021-00169-7
90. Li Z, Ivanov A, Su R, Gonzalez-Pecchi V, Qi Q, Liu S, et al. The OncoPPi network of cancer-focused protein–protein interactions to inform biological insights and therapeutic strategies. Nat Commun. 2017;**8**:14356. https://doi.org/10.1038/ncomms14356

91. Sharifi Tabar M, Francis H, Yeo D, Bailey CG, Rasko JEJ. Mapping oncogenic protein interactions for precision medicine. Int J Cancer. 2022;**151**(1):7–19. https://doi.org/10.1002/ijc.33954
92. Maby P, Corneau A, Galon J. Phenotyping of tumor infiltrating immune cells using mass-cytometry (CyTOF). Methods Enzymol. 2020;**632**:339–68. https://doi.org/10.1016/bs.mie.2019.07.025
93. Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc. 2020;**15**(7):2247–76. https://doi.org/10.1038/s41596-020-0336-2
94. Bravo González-Blas C, De Winter S, Hulselmans G, Hecker N, Matetovici I, Christiaens V, et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. Nat Methods. 2023. **20**:1355–67. https://doi.org/10.1038/s41592-023-01938-4
95. La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. 2018;**560**:494–8. https://doi.org/10.1038/s41586-018-0414-6
96. Lebeau B, Jangal M, Zhao T, Wong CK, Wong N, Cañedo EC, et al. 3D chromatin remodeling potentiates transcriptional programs driving cell invasion. Proc Natl Acad Sci. 2022;**119**(36):e2203452119. https://doi.org/10.1073/pnas.2203452119
97. Qiu X, Zhang Y, Martin-Rufino JD, Weng C, Hosseinzadeh S, Yang D, et al. Mapping transcriptomic vector fields of single cells. Cell. 2022;**185**(4):690–711.e45. https://doi.org/10.1016/j.cell.2021.12.045
98. Uthamacumaran A, Craig M. Algorithmic reconstruction of glioblastoma network complexity. iScience. 2022;**25**(5):104179. https://doi.org/10.1016/j.isci.2022.104179
99. Robertson FL, O'Duibhir E, Gangoso E, Bressan RB, Bulstrode H, Marqués-Torrejón MÁ, et al. Elevated FOXG1 in glioblastoma stem cells cooperates with Wnt/β-catenin to induce exit from quiescence. Cell Rep. 2023;**42**(6):112561. Advance online publication. https://doi.org/10.1016/j.celrep.2023.112561
100. Meng J, Han J, Wang X, Wu T, Zhang H, An H, et al. Twist1-YY1-p300 complex promotes the malignant progression of HCC through activation of miR-9 by forming phase-separated condensates at super-enhancers and relieved by metformin. Pharmacol Res. 2023;**188**:106661. https://doi.org/10.1016/j.phrs.2023.106661
101. Ruelle, D. Strange attractors. Math Intell. 1980;**2**:126–37 https://doi.org/10.1007/BF03023053
102. Itik M, Banks SP. Chaos in a three-dimensional cancer model. Int J Bifurcat Chaos. 2010;**20**(01):71–9.
103. Abraham RH, Shaw CD. Dynamics—the geometry of behavior, Books 0–4. Aerial Press, Inc. 1984.
104. Letellier C, Abraham R, Shepelyansky DL, Rössler OE, Holmes P, Lozi R, et al. Some elements for a history of the dynamical systems theory. Chaos (Woodbury, NY). 2021; **31**(5):053110. https://doi.org/10.1063/5.0047851
105. Ivancevic T, Bottema Murk J, Jain Lakhmi C. A mathematical model of chaotic attractor in tumor growth and decay. 2008. https://doi.org/10.48550/arXiv.0810.4580
106. Letellier C, Denis F, Aguirre LA. What can be learned from a chaotic cancer model?. J Theor Biol. 2013;**322**:7–16. https://doi.org/10.1016/j.jtbi.2013.01.003
107. Khajanchi S, Perc M, Ghosh D. The influence of time delay in a chaotic cancer model. Chaos (Woodbury, NY). 2018;**28**(10):103101. https://doi.org/10.1063/1.5052496
108. Debbouche N, Ouannas A, Grassi G, Al-Hussein AA, Tahir FR, Saad KM, et al. Chaos in cancer tumor growth model with commensurate and incommensurate fractional-order derivatives. Comput Math Methods Med. 2022;**2022**:5227503. https://doi.org/10.1155/2022/5227503
109. Oestreicher C. A history of chaos theory. Dialogues Clin Neurosci. 2007;**9**(3):279–89. https://doi.org/10.31887/DCNS.2007.9.3/coestreicher
110. Singh P, Krishna Roy B. Chaos and multistability behaviors in 4D dissipative cancer growth/decay model with unstable line of equilibria Chaos. Soliton Fract. 2022;**161**:112312. https://doi.org/10.1016/j.chaos.2022.112312
111. Sabarathinam S, Thamilmaran K. Controlling of chaos in a tumour growth cancer model: an experimental study. Electron Lett. 2018;**54**:1160–62. https://doi.org/10.1049/el.2018.5126
112. Izquierdo-Kulich E, Alonso-Becerra E, Nieto-Villar J. Entropy production rate for avascular tumor growth. J Mod Phys. 2011;**2**(6A):615–20. doi: 10.4236/jmp.2011.226071.
113. Stanková K, Brown JS, Dalton WS, Gatenby RA. Optimizing cancer treatment using game theory: a review. JAMA Oncol. 2019;**5**(1):96–103. https://doi.org/10.1001/jamaoncol.2018.3395
114. Hayano K, Yoshida H, Zhu AX, Sahani DV. Fractal analysis of contrast-enhanced CT images to predict survival of patients with hepatocellular carcinoma treated with sunitinib. Digest Dis Sci. 2014;**59**(8):1996–2003. https://doi.org/10.1007/s10620-014-3064-z
115. Lennon FE, Cianci GC, Cipriani NA, Hensing TA, Zhang HJ, Chen CT, et al. Lung cancer-a fractal viewpoint. Nat Rev Clin Oncol. 2015;**12**(11):664–75. https://doi.org/10.1038/nrclinonc.2015.108
116. Metze K. Fractal dimension of chromatin: potential molecular diagnostic applications for cancer prognosis. Exp Rev Mol Diagn. 2013;**13**(7):719–35. https://doi.org/10.1586/14737159.2013.828889
117. Lu Z, Hunt BR, Ott E. Attractor reconstruction by machine learning. Chaos (Woodbury, NY), 2018;**28**(6):061104. https://doi.org/10.1063/1.5039508
118. Pathak J, Hunt B, Girvan M, Lu Z, Ott E. Model-free prediction of large spatiotemporally chaotic systems from data: a reservoir computing approach. Phys Rev Lett. 2018;**120**(2): 024102. https://doi.org/10.1103/PhysRevLett.120.024102
119. Vlachas PR, Pathak J, Hunt BR, Sapsis TP, Girvan M, Ott E, et al. Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics. Neural Netw: Off J Int Neural Netw Soc. 2020;**126**: 191–217. https://doi.org/10.1016/j.neunet.2020.02.016
120. Gauthier DJ, Bollt E, Griffith A, Barbosa WAS. Next generation reservoir computing. Nat Commun. 2021;**12**:5564. https://doi.org/10.1038/s41467-021-25801-2
121. Sayari E, da Silva ST, Iarosz KC, Viana RL, Szezech JD, Batista AM. Prediction of fluctuations in a chaotic cancer model using machine learning. Chaos Soliton Fract. Elsevier, 2022;**164**(C). doi: 10.1016/j.chaos.2022.112616
122. Hasani R, Lechner M, Amini A, Liebenwein L, Ray A, Tschaikowski M, et al. Closed-form continuous-time neural networks. Nat Mach Intell. 2022;**4**:992–1003. https://doi.org/10.1038/s42256-022-00556-7
123. Lorenz EN. Deterministic nonperiodic flow. J Atmos Sci. 1963;**20**: 130–41. http://dx.doi.org/10.1175/1520-0469

124. Rössler OE, Letellier C. Chaos and turbulence. In: Chaos. Understanding complex systems. Cham: Springer; 2020.
125. Turing, A. M. The chemical basis of morphogenesis. Phil. Trans. R. Soc. Lond. B **237**, 37–72 (1952).
126. Ding, L, Ley, T. J, Larson, D. E, Miller, C. A, Koboldt, D. C, Welch, J. S, Ritchey, J. K, Young, M. A, Lamprecht, T, McLellan, M. D, McMichael, J. F., Wallis, J. W., Lu, C., Shen, D., Harris, C. C., Dooling, D. J., Fulton, R. S., Fulton, L. L., Chen, K., Schmidt, H., … DiPersio, J. F. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature, (2012). **481**(7382), 506–510. https://doi.org/10.1038/nature10738
127. Kwok, M, Wu, C. J. Clonal evolution of high-risk chronic lymphocytic leukemia: a contemporary perspective. Frontiers in Oncology, (2021). **11**, 790004. https://doi.org/10.3389/fonc.2021.790004
128. Turoverov, K. K, Kuznetsova, I. M, Fonin, A. V, Darling, A. L, Zaslavsky, B. Y, Uversky, V. N. Stochasticity of biological soft matter: emerging concepts in intrinsically disordered proteins and biological phase separation. Trends in Biochemical Sciences, (2019). **44**(8), 716–728. https://doi.org/10.1016/j.tibs.2019.03.005
129. Halatek, J, Frey, E. Rethinking pattern formation in reaction–diffusion systems. Nature Phys **14**, 507–514 (2018). https://doi.org/10.1038/s41567-017-0040-5
130. Würthner, L, Brauns, F, Pawlik, G, Halatek, J, Kerssemakers, J, Dekker, C, Frey, E. Bridging scales in a multiscale pattern-forming system. Proceedings of the National Academy of Sciences of the United States of America, (2022). **119**(33), e2206888119. https://doi.org/10.1073/pnas.2206888119
131. Merino-Salomón, A, Babl, L, Schwille, P. Self-organized protein patterns: the MinCDE and ParABS systems. Current Opinion in Cell Biology, (2021). **72**, 106–115. https://doi.org/10.1016/j.ceb.2021.07.001
132. Rössler, O. (1976). Chemical turbulence: chaos in a simple reaction-diffusion system. Zeitschrift für Naturforschung A, **31**(10), 1168–1172. https://doi.org/10.1515/zna-1976-1006
133. Rössler, O.E. (1977). Chemical turbulence a synopsis. In: Haken, H. (eds) Synergetics. Springer Series in Synergetics, Vol **2**. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-66784-8_16
134. Kuramoto, Y. Chemical turbulence. In: Chemical oscillations, waves, and turbulence. springer series in synergetics, Vol **19**. Springer, Berlin, Heidelberg.(1984). https://doi.org/10.1007/978-3-642-69689-3_7
135. Strasser P, Rössler OE, Baier G. Hyperchaos and chemical turbulence in enzymatic reaction-diffusion systems. J Chem Phys. 1996;**104**(24):9974–82.
136. Heltberg ML, Krishna S, Jensen MH. On chaotic dynamics in transcription factors and the associated effects in differential gene regulation. Nat Commun. 2019;**10**:71. https://doi.org/10.1038/s41467-018-07932-1
137. Heltberg ML, Krishna S, Kadanoff LP, Jensen MH. A tale of two rhythms: locked clocks and chaos in biology. Cell Syst. 2021;**12**(4):291–303. https://doi.org/10.1016/j.cels.2021.03.003
138. Heltberg MS, Lucchetti A, Hsieh FS, Minh Nguyen DP, Chen SH, Jensen MH. Enhanced DNA repair through droplet formation and p53 oscillations. Cell. 2022;**185**(23):4394–408.e10. https://doi.org/10.1016/j.cell.2022.10.004
139. Kalgin IV, Karplus M, Chekmarev SF. Folding of a SH3 domain: standard and 'hydrodynamic' analyses. J Phys Chem B. 2009;**113**(38):12759–72. https://doi.org/10.1021/jp903325z
140. Kalgin IV, Chekmarev SF. Turbulent phenomena in protein folding. Phys Rev E, Stat Nonlin Soft Matter Phys. 2011;**83**(1 Pt 1):011920. https://doi.org/10.1103/PhysRevE.83.011920
141. Wensink HH, Dunkel J, Heidenreich S, Drescher K, Goldstein RE, Löwen H, et al. Meso-scale turbulence in living fluids. Proc Natl Acad Sci U S A. 2012 Sep 4;**109**(36):14308–13. doi: 10.1073/pnas.1202032109. Epub 2012 Aug 20. PMID: 22908244; PMCID: PMC3437854.
142. Alert R, Casademunt J, Joanny J-F. Active turbulence. Annu Rev Condens Matter Phys. 2022;**13**(1):143–70.
143. Guo Y, Ji X, Liu J, Fan D, Zhou Q, Chen C, et al. Effects of exosomes on pre-metastatic niche formation in tumors. Mol Cancer. 2019;**18**:39. https://doi.org/10.1186/s12943-019-0995-1
144. Swana M, Blee J, Stillman N, Ives J, Hauert S. Swarms: the next frontier for cancer nanomedicine. In: Balaz I, Adamatzky A, editors. Cancer, complexity, computation. emergence, complexity and computation. Vol **46**. Cham: Springer; 2022. https://doi.org/10.1007/978-3-031-04379-6_12
145. Doostmohammadi A, Shendruk T, Thijssen K, Yeomans, JM. Onset of meso-scale turbulence in active nematics. Nat Commun. 2017;**8**:15326. https://doi.org/10.1038/ncomms15326
146. Doostmohammadi A, Ignés-Mullol J, Yeomans JM, Sagues F. Active nematics. Nat Commun. 2018;**9**:3246. https://doi.org/10.1038/s41467-018-05666-8
147. Wagner CG, Norton MM, Park JS, Grover P. Exact coherent structures and phase space geometry of preturbulent 2D active nematic channel flow. Phys Rev Lett. 2022;**128**(2): 028003. https://doi.org/10.1103/PhysRevLett.128.028003
148. Tan TH, Liu J, Miller PW, Tekant M, Dunkel J, Fakhri N. Topological turbulence in the membrane of a living cell. Nat. Phys. 2020;**16**:657–62. https://doi.org/10.1038/s41567-020-0841-9
149. Haga RB, Ridley AJ. Rho GTPases: regulation and roles in cancer cell biology. Small GTPases. 2016;**7**(4):207–21. https://doi.org/10.1080/21541248.2016.1232583
150. Henshaw RJ, Martin OG, Guasto JS. Dynamic mode structure of active turbulence. Phys Rev Fluids. 2023;**8**:023101.
151. Lin SZ, Zhang WY, Bi D, Li B, Feng X-Q. Energetics of mesoscale cell turbulence in two-dimensional monolayers. Commun Phys. 2021;**4**:21. https://doi.org/10.1038/s42005-021-00530-6
152. Pajic-Lijakovic I, Milivojevic M. Mechanical waves caused by collective cell migration: generation. Eur Biophys J. 2022 Jan;**51**(1):1–13. doi: 10.1007/s00249-021-01581-x. Epub 2022 Jan 24. PMID: 35072747.
153. Lin SZ, Ye S, Xu GK, Li B, Feng XQ. Dynamic migration modes of collective cells. Biophys J. 2018 Nov 6;**115**(9):1826–35. doi: 10.1016/j.bpj.2018.09.010. Epub 2018 Sep 20. PMID: 30297134; PMCID: PMC6224637.
154. Scholz C, Schröder-Turk GE, Mecke K. Pattern-fluid interpretation of chemical turbulence. Phys Rev E Stat Nonlin Soft Matter Phys. 2015;**91**(4): 042907. https://doi.org/10.1103/PhysRevE.91.042907
155. Kalra AP, Benny A, Travis SM, Zizzi EA, Morales-Sanchez A, Oblinsky DG, et al. Electronic energy migration in microtubules. ACS Cent Sci. 2023;**9**(3):352–61. https://doi.org/10.1021/acscentsci.2c01114
156. Davydov AS. Biology and quantum mechanics. Oxford: Pergamon Press; 1982.
157. Marais A, Adams B, Ringsmuth AK, Ferretti M, Gruber JM, Hendrikx R, et al. The future of quantum biology. J R Soc Interface. 2018 Nov 14;**15**(148):20180640. doi: 10.1098/rsif.2018.0640. PMID: 30429265; PMCID: PMC6283985.

158. Slocombe L, Sacchi M, Al-Khalili J. An open quantum systems approach to proton tunnelling in DNA. Commun Phys. 2022;**5**:109. https://doi.org/10.1038/s42005-022-00881-8
159. Löwdin PO. Proton tunneling in DNA and its biological implications. Rev Mod Phys. 1963;**35**:724–32. 10.1103/RevModPhys.35.724
160. Craddock TJ, Friesen D, Mane J, Hameroff S, Tuszynski JA. The feasibility of coherent energy transfer in microtubules. J R Soc Interface. 2014;**11**(100):20140677. https://doi.org/10.1098/rsif.2014.0677
161. Levin M. Bioelectric signaling: reprogrammable circuits underlying embryogenesis, regeneration, and cancer. Cell. 2021;**184**(8):1971–89. https://doi.org/10.1016/j.cell.2021.02.034
162. Chernet B, Levin M. Endogenous voltage potentials and the microenvironment: bioelectric signals that reveal, induce and normalize cancer. J Clin Exp Oncol. 2013;Suppl **1**:S1–002. https://doi.org/10.4172/2324-9110.S1-002
163. Litzelman K, Verma M. Epigenetic regulation in biopsychosocial pathways. Methods Mol Biol. (Clifton, NJ). 2015;**1238**:549–67. https://doi.org/10.1007/978-1-4939-1804-1_29
164. Chen J, Long MD, Sribenja S, Ma SJ, Yan L, Hu Q, et al. An epigenome-wide analysis of socioeconomic position and tumor DNA methylation in breast cancer patients. Clin Epigenet. 2023;**15**(68):1–10. https://doi.org/10.1186/s13148-023-01470-4
165. Sontag S. Illness as metaphor; and aids and its metaphors (1st Anchor books). Doubleday; 1990.

41

Cancer formation as creation and penetration of unknown life spaces

Andrzej Kasperski and Henry H. Heng

41.1. Introduction

More than eight million people die from cancer each year, which means that more than 20 thousand people die from cancer every day. It also means that over 400 people will die of cancer while reading this chapter (i.e. for about 30 min). The high rate of cancer mortality forcefully questions the current gene-mutation-theory-based cancer concepts and strategies to fight cancer. In recent years, various -omic approaches, especially the cancer genome sequence project, have illustrated that chromosome instability (CIN), rather than a few dominant gene mutations, functions as a common driver of cancer [1,2]. Equally important, cancer evolution can be described by the two-phased evolution (punctuated macroevolution followed by stepwise gradual microevolution) with genome and gene-based mechanisms, respectively. Furthermore, the ultimate significance of the CIN is explained using karyotype coding, as the system coding organizes the gene interaction, which provides the cellular structural basis of genomic networks, a core concept of applying systems biology to cancer research [3]. With the new knowledge in hand, the time is ripe to integrate the concept of cancer attractor with two-phased cancer evolution and the information management between gene and genome, individual cell to population, and normal tissue to cancerous tissue. Establishing a universal model of cancer transformation and development may provide an answer to which direction our efforts should go in finding a way to reduce the impact of cancer on human health.

41.2. Review of cancer attractor ideas

In 1858, Rudolf Virchow formulated the idea that cancer cells are the organism's own cells [4]. Since then, many hypotheses have been proposed explaining the origin of cancer cells and how such heterogeneous morphology, increased proliferation, metastatic capacity, and invasive behaviour develop [5,6]. For example, before gene mutation-centric cancer research becomes dominant, investigators studied cancer through the lens of pathology and developmental biology [7–9]. Since the 1970s, the promises from cancer genes have transformed cancer research into hunting key oncogene and tumour suppressor genes. With the increased number of cancer genes being identified far more than predicted, however, the systems biology approaches gained attention, as the gene interaction network is more important than individual genes. Equally important, the cancer genome can be considered a complex network of mutually regulating genes [10]. This network can become unstable, among others, due to genome instability (GIN) (including chromosomal instability (CIN)) [1,11]. The proposed by Stuart Kauffman cancer attractor idea shows that this unstable network has the potential to create hundreds of stable equilibrium states called attractors [12–14]. These stable states depend on the gene expression profiles associated with each cell type [13,15,16]. As a result, a variety of stable, discreetly distinct cell phenotypes can be created [16,17]. The idea of cancer attractors has been finding experimental support through genomic technologies [12,16]. This cancer attractor model is influenced by the developmental landscape model as well as Neo–Darwinian stepwise evolutionary concept where the accumulation of microevolution over time leads to macroevolution [18]. However, since cancer is not just an issue of cell status reflected, by genes and/or epigenetic landscapes, but also the new system emergence reflected by new genome formation via genome reorganization [3], the gene and genome difference and their corresponding evolutionary pattern should be integrated into the cancer attractor model. The multiple-level landscape model of cancer, therefore, was suggested to separate 'local microevolutionary potential' (adaptative potential provided primarily by gene-level or non-genetic changes) and 'global macroevolutionary potential' [6,19]. Interestingly, the gene–genome relationship has been integrated into the cancer attractor model [20]. Unfortunately, most attractor models have continuously ignored the important distinction between genes and genomes, micro and macroevolution, and developmental landscape and the macroevolutionary landscape.

41.3. The new attractor model of cancer formation

Despite that cancer has been extensively linked to a large number of molecular mechanisms, cancer fundamentally represents a

macroevolutionary process where new systems emerge from hosts' tissue by breaking various constraints [3,6,21]. This relationship between cancer evolutionary dynamics and vast individual molecular mechanisms has been referred to as the evolutionary mechanism of cancer. It is proposed that many molecular triggering factors can contribute to cancer via GIN-mediated evolution, but the clinical predictability is low if only based on individual mechanisms when there are so many triggers during cancer evolution [22]. For example, in the light of the unified cell bioenergetics, the cause of cancer transformation can also be linked to overenergization of organism normal cells [23–26]. Cancer transformation, a two-phased evolutionary event, causes switch of cell fate of overenergized normal cell to cancerous/atavistic cell fate and initiates, among others, the cloning process as an associated phenomenon that occurs after transformation [23,27,28]. After cancer transformation, additional microevolution can be involved. The microevolution is related to development of clone phenotypes. Microevolution also contains gene mutations as a factor that can influence the development of the phenotype. Microevolution is obligatory and occurs through, among others, changes of cell fates of cancerous cells. Macroevolution may occur or not (i.e. it is not obligatory) during cancer development depending on if further phase transition is needed and is related to changes of genome. Macroevolution occurs through genome destabilizations, followed by genome chaos, genome rearrangements, and obtaining genome stability in new genome attractors [27,28]. That means that macroevolution can be considered as a process of changing genome attractors [27,28]. During the process of changing genome attractors, the whole genome undergoes reorganization and remodelling [27,28]. In the light of this concept, macroevolution is based on the whole genome changes. In accordance with information presented in [6,21], cancer macroevolution is based on the change of whole genome information package rather than on specific genes. After attaining genome attractor cells undergo microevolution, i.e. when cells are trapped in the genome attractor, changes of cell-fate attractors occur resulting in changes of phenotypes of cancer clones [27,28]. Changes of cell fates can occur as a result of destabilizations of current cell fates. Destabilizations of cell fates can occur as a result of cell bioenergetic problems. These destabilizations can be additionally stimulated by mutations. In the light of this concept, two types of attractors can be distinguished during cancer development that are important to this process, i.e. cancer genome attractors and cancer cell-fate attractors [27,28]. No matter which types of attractors form which kinds of genomic and non-genomic landscapes, the common consequence is the new information-mediated cellular changes that create and penetrate unknown life spaces (for normal cells), albeit involving different organizational scales [21,38]. Interestingly, this conclusion also applies to the organismal evolution beyond cancer [29].

According to the recently proposed idea, cancer formation can be considered as a phenomenon that is associated with the development of clouds of cancerous cell fates [27]. Cancerous cell fates are characterized by cancerous/atavistic phenotypes. Cancerous clouds of cell fates are 'generated' from genome attractors, which means that cancer macroevolution gives the possibility of generation of new and new cancerous clouds [27,28]. Cancerous phenotypes of cells that belong to different cancerous clouds differ significantly from each other because they are generated from different genome attractors (this is a consequence and a hallmark of cancer macroevolution). Cancerous phenotypes of cells in cancerous cloud are similar because all cell fates are generated from one genome attractor (this is a consequence and a hallmark of cancer microevolution).

Even though the above genome attractor model was illustrated through the analysis of unified cell bioenergetics and cell fate [27], it can be applied to explain the common relationship between different triggering factors of cancer, the two-phased cancer evolution, and the cancer-related genome/gene attractors. As illustrated in **Figure 41.1**, the upper portion (yellow colour) belongs to macroevolution to generate different genome attractors while the lower portion belongs to microevolution to generate different statuses of cell populations. Both the horizontal and vertical information passages can be found in macroevolutionary and microevolutionary phases, representing different meanings of generating, preserving, modifying, and using the system or parts information [3,18], which is essential for the success of cancer evolution by searching for and even creating the new life space. It is important to separate genome attractors (involving reconstruction of the boundary of a given genetic network) from traditional attractors (involving network rewiring without altering the genome system). The latter category includes an array of genomic and non-genomic cellular mechanisms that regulate cellular states, such as epigenetic function, usage of intrinsically disordered proteins (IDPs) [30], and developmental process [3,6,21].

41.4. The multiple levels of biological systems and their corresponding information management

One of the key rationales for promoting the genome attractor model is to popularize the concept of distinguishing multiple levels of genomic systems with the evolutionary and informational context. Recent studies suggest that separating gene-coded and genome-coded information is of ultimate importance. Furthermore, understanding information management is crucial for proposing a cancer genome attractor model [3,18].

41.4.1. Activation/deactivation of functionalities and karyotype coding

Chromosomes contain sets of functions that are coded by genes. Requested functionalities can be activated by activation of a set of needed functions. For example, activation of the Crabtree effect needs activation of set of genes expression, i.e. activation of set of functions to realize needed functionality (in this case to realize the Crabtree effect functionality). In the light of this concept, functionality is something more than a function of coding proteins. Appropriate and stable functionality can be established by chromatin remodelling. Post-translational modifications of chromatin (acetylation, methylation, and phosphorylation of the histones and DNA methylation) play a major role in the activation or repression of gene transcription [31]. Some of these chromatin modifications are involved in the maintenance of gene expression stable patterns (usually referred to as epigenetic regulation) [31]. Chromatin remodelling causes activation (and then gradual stimulation of expressions of genes) and gradual inhibition of expressions of other genes (including deactivation). So, chromatin remodelling acts on sets of genes, causing remodelling of expressions of groups of genes in order to obtain appropriate and stable functionalities.

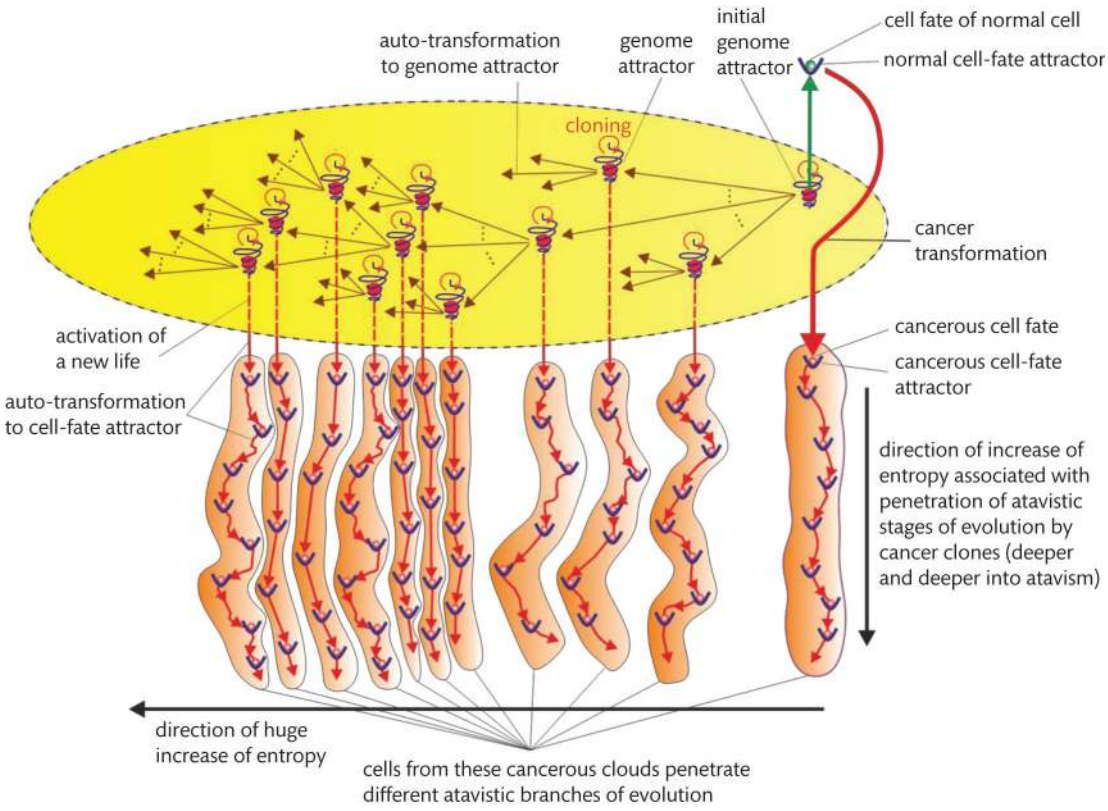


Figure 41.1. Cancer formation as creation and simultaneous penetration of different atavistic stages (including unknown stages/space) of evolution by cancer clones. CIN causes genome chaos and formation of unstable genomes. Auto-transformation to genome attractor can cause aneuploidy, rearrangements, and other ordered genome changes, the aim of which is to achieve new genome. Auto-transformation to genome attractor occurs as a result of genome destabilization and causes stabilization of unstable genome in a new genome attractor. Auto-transformation to cell-fate attractor occurs after cell-fate destabilization and causes stabilization of cell fate in a new cell-fate attractor [27,28].

Traditionally, systems biologists have focused on genes and the epigenetic aspect of chromatin remodelling, as the genome-level alterations have been long ignored in the gene-centric era. The two-phased cancer evolutionary model has highlighted the importance of the karyotype that organizes the genetic networks, including the epigenetic landscape. This organizational function defines system inheritance via a new genomic coding mechanism called karyotype coding [32]. According to karyotype coding, genes are organized along chromosomes with a fixed physical order, which preserves species genome information and provides a platform for other genetic and non-genetic information to develop and accumulate [6,21,32]. Because cancer evolution involves both macroevolution and microevolution, the multiple levels of genomic and non-genomic changes are often involved, including karyotype alterations, gene mutations, and epigenetic variations. Thus, the cell-fate changes are multi-scales and multi-faced, not just the same genome with altered epigenetic landscapes. This is in accordance with Theodor Boveri, who has stated that cancer is due to a certain permanent change in the chromosome complex [33]. Indeed, the two-phased model of cancer evolution has nicely reconciled the contribution from gene mutations, epigenetic changes, and karyotype reorganization.

41.4.2. Role of polyploidy during cancer development

In recent years, chaotic genomes including various subtypes, such as chromothripsis and giant polyploidy giant cancer cells (PGCCs), have become a hot topic [1,8,34–36]. Traditionally, researchers have focused on specific structural or numerical aberrations without the information context and evolutionary process. For example, the significance of chromosomal translocation is on the interrupted genes, and as it is presented in [37], polyploidization increases the probability of survival of cancer cells. But a polyploid cell is not clonal, and this proposed advantage cannot last. Yes, depolyploidization allows the Hayflick's limit to be restored, and hence the activation of the cell cycle and mitotic cell cycle [37]. After depolyploidization, cancer cells again become clonal. As a result of this mechanism (i.e. circular polyploidization and depolyploidization), cancer cells are more resistant to changes in environmental conditions (that occur as a result of, for example, chemotherapy) maintaining cloning potential. All these changes can be explained by the cycles of two-phased cancer evolution, where all these changes are phenotypes of a newly emergent genome system. No matter what type of genome chaos (massive translocations, or circular polyploidization and depolyploidization), the ultimate importance is to reorganize the old genomes to form new genomes, followed by the microevolution to grow the cellular populations. This is the process of creating, preserving, and amplifying new system information [3,6,21].

41.4.3. Unify multiple levels of variations via the information management

Even though the relationship between different genomic and non-genomic variations is highly dynamic, it can be understood by the

mechanism of information management during different phases of evolution [3,32]. In particular, separating genome alteration-mediated system information and gene alteration-mediated parts information can promote the effort to unify multiple levels of variations and their corresponding attractors using a multiple-level landscape model [38].

41.5. Implications of the new attractor model

Establishing a genome-based cancer attractor model is important. To date, most attractor models have failed to include karyotype dynamics that define the physical basis of genomic networks. The genomic topology is an essential component of the genomic information, and the topological relationship among genes and regulating elements is needed to explain the attractor model. With the new frameworks, many observations can be reconciled. The followings are some case studies.

41.5.1. Overenergized mitochondria as a main driver of cancer evolution

Cancer transformation, which mainly results from macroevolutionary genome restructuring but can also occur without mutations under specific conditions, leads to the activation of the Warburg effect (aerobic glycolysis) as part of the transformation process. By stimulation of aerobic glycolysis, cells can try to prevent their mitochondria against overenergization and against too high level of reactive oxygen species (ROS) [23,28]. But obtaining energy from aerobic glycolysis is much less efficient way as compared to oxidative phosphorylation (OXPHOS). Cells compensate for this drop in efficiency by processing more glucose. Even in the presence of oxygen, rapidly proliferating tumour cells have typically a glycolysis rates of up to 200 times higher compared to cells of normal tissue of origin [39,40]. The increase in glucose uptake is a major feature distinguishing cancer cells from normal cells [41]. Possible explanations for the increase in the rate of glycolysis of cancer cells is the overexpression of glucose transporters and all enzymes of the glycolytic pathway as a result of oncogene activation [42]. Accordingly, high levels of glucose uptake in cancer cells are associated with increased expression of glucose transporter proteins (GLUT1, GLUT3, and/or GLUT12) [42]. In addition, many cancers overexpress and/or overactivate the major enzymes that control the glycolytic pathway (i.e. hexokinase, phospho-fructokinase, and pyruvate kinase) [42,43].

A large amount of obtained energy in the glycolysis–fermentation pathway (aerobic fermentation) prevents the discharge of mitochondria from high-energy molecules, thus causing cancer mitochondria to remain overenergized [23]. In this way, the cancer cells get a lot of the energy they need and their mitochondria remain overenergized, but at the same time prevented against too high overenergization that can lead to apoptosis (because of too high level of ROS). Overenergized mitochondria are the main driver of macroevolution leading to permanent changes of genome attractors. Overenergized mitochondria are also driver of cancer microevolution, stimulating cancer phenotype evolution by stimulating permanent changes of cell fates. It should be noted that there is a great deal of evidence that challenges the paradigm of pure 'glycolytic' cancer cells [43,44]. Some glioma, hepatoma, and breast cancer cell lines have been shown to possess functional mitochondria, obtaining ATP primarily from OXPHOS [44]. The phenomenon of obtaining ATP mainly from OXPHOS additionally stimulates macroevolution leading to more intensive ROS generation and, as a result, permanent losses of genome stability followed by changes of genome attractors. More intensive ROS generation also stimulates microevolution (by increasing the probability of random DNA mutations), resulting in stimulation of cancer phenotype evolution by changing cell fates. Here, it should be noted that cancer phenotype evolution allows the cancer clones to increase phenotypic diversity and also allows clones to flexibly adapt to changing environment. Cancer phenotype evolution can be stimulated by random DNA mutations by elevated ROS, but this evolution can also be a response to the cell bioenergetic problems and can occur without mutations [27,45]. It should be pointed out that despite all these complicated descriptions, this chain of events belongs to the microevolutionary phase followed by macroevolution. There are many different microevolutionary stories reported in the literature, for this reason, it is challenging to cure cancer just based on individual case studies [18,46].

41.5.2. Cancerous tissue

Solid tumours are not only clones of cancer cells but also abnormal organs composed of many types of cells and the extracellular matrix [47]. Some aspects of cancer development resemble those seen in organ development, while others are more like tissue remodelling [47]. Cancer cells and stromal components are organized into tissues, which in turn are organized into cancerous organs that interact with the entire organism [47].

As it was presented in [27], cancer transformation can occur as a result of huge bioenergetic disturbances in multicellular-level functionalities which lead to a loss of control over atavistic/unicellular functionalities. Maintaining the ability to create tissue indicates that (although disturbed multicellular layer functionalities do not fully control atavistic functionalities after cancer transformation) part of multicellular-level functionalities responsible for creation of multicellular organism remains active. Activity of atavistic functionalities and disturbed activity of multicellular-level functionalities cause that development of cancer shows two faces, i.e. development of cancer occurs as a development of population of individual cells (due to activity of atavistic/unicellular functionalities) along with maintaining the ability to create cancerous tissue (due to activity, albeit disturbed, of multicellular-level functionalities). In accordance with, e.g. [48–50], the genes of unicellular origin are overexpressed in cancerous tissues, while the genes appearing in the multicellular evolutionary stages are down-regulated in cancerous tissues, which can support considerations presented in this section. It is interesting to note that cancerous tissue consists of genetically heterogeneous cells that may be very different, because of cancer macroevolution, from each other. This is in contrast to organism normal tissue that is genetically homogeneous. Additionally, microevolution causes that clones that form cancerous tissue are phenotypically diverse, even if generated from the same genome attractor.

Cancerous tissue undergoes constant changes related to macroevolution and microevolution. Macroevolution and microevolution cause asynchronous point changes of cancerous tissue through genome rearrangements and changes of cell fates of the tissue clones. Microevolution causes asynchronous point adaptation of cancerous tissue to environment. Point changes and adaptation mean that

macroevolution and microevolution are related to individual clones that form cancerous tissue. Asynchronous means that macroevolution and microevolution of individual clones (i.e. macroevolution and microevolution in different points of cancerous tissue) occur at different times (i.e. asynchronously). Cancer clones constantly adapt to the environment through microevolution, and through macroevolution, altered clones can penetrate genome space in distant places. In addition, created new clones can be attached to the existing cancerous tissue. In the light of this concept, cancer can be considered as a tissue-based disease. It should be noted that macroevolution is optional, i.e. it may occur or not (there are known cancers that can develop without genome rearrangements (i.e. without changes of genome attractors)). Cancer microevolution always occurs (regardless of cancer type) because cancer cells always have potential to change their cell fates to adapt to environment. Cancer phenotype evolution by changes of cell fates can occur with mutations as an associated phenomenon or without mutations (there are known cancers that can develop without mutations).

41.5.2.1. Cancer as a tissue-based disease

As discussed in the previous section, cancer can be considered as tissue-based disease. The tissue organization field theory (TOFT) also posits that cancer is a tissue-based disease and that proliferation is the default state of all cells [51]. In accordance with TOFT, cancer is a disease of tissue organization akin to development gone awry [52]. TOFT states that cancer is a disease that occurs at the level of biological organization of tissues, resulting from disruption of the morphogenetic field that organizes histogenesis and organogenesis from fertilization to senescence [53].

Some important issues need to be addressed here, however. Clearly, cancer is not just a proliferation problem of the same system but new systems. Any tissue-based theory should explain somatic evolution as well. Interestingly, the genome attractor model that is based on two-phased cancer evolution can also be applied to the tissue level. For example, giant cancer cells with chaotic genomes can rapidly produce a large number of cancer cells, the basis of cancerous tissue. Importantly, those newly emergent cells often display altered karyotype coding, suggesting that they represent new systems rather than modified old systems [54]. Dr Jinsong Liu's group has directly illustrated the rapid cancerous tissue formation via PGCCs (personal communications).

41.5.2.2. Cancer metastasis as a tissue-related phenomenon

Metastasis is a term related to the development of secondary tumours in a part of the body that is far from the original primary cancer [55]. Being responsible for about 90% of cancer deaths, cancer metastasis is the leading cause of cancer morbidity and mortality [56]. During cancer metastasis process, metastatic cells go through four essential, metastatic steps, i.e. detachment, migration, invasion, and adhesion [56]. The detachment of cancer cell from the primary tumour is a necessary first condition and an initial early stage of metastasis [57,58]. The detachment of a cancer cell has to be followed by tissue invasion, entry into the bloodstream, colonization of a distant site, and proliferation [57]. The probability of metastasis occurring is very small, i.e. only one in the many millions of cells that have gone through all of this can turn out to regenerate a cancer in a distant site [57].

In the light of the presented model, macroevolution involves genome rearrangements (including changes of karyotypes). It is also likely that without macroevolution, most cells will not survive during those four stages. These genome rearrangements can cause deactivation (or lost) of functionalities of multicellular level. This indicates that, during cancer development related to macroevolution, the ability to create tissue is decreasing and simultaneously a loss of control over atavistic/unicellular functionalities is increasing. Moreover, deactivating functionalities of multicellular level causes an increase in the probability of cancerous cells detaching from existing cancerous tissue. This also indicates that macroevolution stimulates an increase in aggressiveness of cancer simultaneously leading to an increased likelihood of metastasis.

Again, metastasis is achieved by another cycle of two-phased evolution where CIN is the common driver [21,38,59,60], which can be explained by the new genome attractor model.

41.5.3. Cancer development is not the only reversion to pure atavistic stages

During the development of cancer, a return to an atavistic life can be observed based on certain features [61]. Due to the atavistic theory of oncogenesis, the biogenetic law is currently gaining particular importance [61–64,50]. In accordance with the Serial Atavism Model, multi-stage cancer progression reverses the multi-stage evolutionary chronology [61]. In the light of this concept, during cancer development, several deep evolutionary transitions can occur towards more and more primitive prokaryotes [61]. This point of view is correct, as long as the cancer cells do not change their genome attractors. During development of some cancers that is characterized by changes of genome attractors, the original organic code is also gradually changed (i.e. new information is created and old information is modified). These significant changes in the organic code, which are associated with changes of genome attractors, cause cancer clones to simultaneously penetrate series of unknown, atavistic branches that may substantially deviate from original established during evolution pathways. In the light of this concept, cancer development is not the only reversion to pure atavistic stages. This is especially related to aggressive cancers, which are characterized by higher level of GIN and resulted in frequent changes of genome attractors.

Cancer cells have been proved to have a higher entropy generation than healthy cells, where entropy is expected surprise or uncertainty [65–67]. The rate of entropy generation shows the intensity of proliferation, invasion, and robustness in cancer cells. For these reasons, modelling of dynamic entropy can be an effective combined tool in cancer research [66]. Penetration of a series of unknown, atavistic branches that may deviate substantially from original evolutionary pathways lead to a huge increase of surprise (i.e. increase of entropy). Simultaneous penetration of these unknown area additionally multiply this surprise. This penetration of unknown atavistic life spaces is associated with a gradual significant loss (mainly because of macroevolution) of original, stored in genome information that defines the life of healthy cells, and consequently with a gradual build-up of strange, very difficult to understand and predict behaviours resulting in an enormous increase of surprise and leading to a huge increase of entropy (**Figure 41.1**). In the light of the presented considerations, cancerous tissue consists of different sets (due to macroevolution) of clones. The clones from these sets simultaneously penetrate different branches of atavistic life spaces (**Figure 41.1**). Moreover, microevolution causes that clones from different sets adapt to changes in the environment, establishing different

phenotypes by changing cell fates. In this way, cancerous tissues undergo flexible adaptation to the environment separately, simultaneously, and independently in each part of the cancerous tissue.

41.5.4. Fighting against cancer: concentration on causes not effects

In the article [27], it was proposed that cancer transformation, i.e. a change in normal cell fate to cancerous/atavistic cell fate, occurs as a result of loss of control over functionalities of the unicellular layer, resulting in a loss of control over atavistic functionalities. The Warburg effect, i.e. an increase in aerobic glycolysis, occurs as a result of cancer transformation. Reducing the intensity of aerobic glycolysis is one of the ways to fight cancer, but this approach is fighting the effects not the cause of cancer.

Surgery is often considered as an effective method of fighting cancer, because by surgically removing the cancer cells from the organism, it is possible to get rid of the problem. Other methods of fighting cancer should be effective as long as they focus on the causes not effects of cancer transformation and development. In the light of the idea presented in [27], the causes of cancer are disturbances and/or destruction of multicellular-level functionalities. Cancer cells contact with normal cells, which causes the occurrence of interaction between them (i.e. between cancer cells and normal cells), but probably only the establishment of contact with normal cells of eggs or embryos can cause restoration and activation of multicellular-level functionalities that can lead cancer cells to revert to normal cells. The success of such reversion, however, is often constrained by the karyotype-code-defined genome attractor [3,21,32].

41.6. Conclusions and future directions

The war on cancer is far from over, but every day is very important during this war. This chapter aims to unify some diverse phenomena that occur during cancer transformation and development. Striving to this evolutionary unification can lead to better approaches of understanding and fighting cancer.

(a) The concept of the attractor is an important one in systems biology. Currently, most cancer attractor models confuse the developmental landscape and evolutionary landscape, especially the macroevolutionary model is incorrectly assumed as a stepwise microevolutionary model plus time [29,68]. The novel aspect of the genome attractor model is separating macroevolution from microevolution, which effectively integrates the attractor concept with the evolutionary context.

(b) The concept of karyotype coding provides the genomic topological properties for understanding how gene regulatory network works. The traditional systems biology lacks the correct understanding of the relationship between genes/epigenetics and genomes, and ignores that cancer cells display different karyotypes compared with normal cells. All functional activities of genes and epigenetics are within the genome attractor's internal boundaries, and the genome attractor provides the most important system constraints. Macroevolution can create specific genome attractors, while microevolution can change the cellular states within the genome attractors and increase the cellular population.

(c) The genome attractor model can be used to explain the dynamic relationship between highly diverse molecular mechanisms of cancer and the two-phased cancer evolutionary model. Even though only selective issues are briefly discussed in this chapter (the overenergized mitochondria, atavistic phenotype, giant polyploidy cancer cells, cancer as a tissue-based disease, and metastasis), a similar discussion can be expanded into many different molecular understandings of cancer [2]. For example, macroevolution causes the cancer clones change genome attractors. Microevolution causes adaptive growth of cancers cells that are trapped in the genome attractors. As a result of microevolution, cancer clones change cell-fate attractors, flexibly adapting to the environment. During microevolution accumulation of mutations, affecting changes of cell fates can also occur and thus potentially leading to macroevolution.

(d) Even though multiple runs of two-phased cancer evolution can explain the vast majority of cancer cases (e.g. for those undifferentiated tumours, karyotype change is the common feature and CIN is the common driver), there might be some exceptions. Some differentiated tumours (arrested at earlier developmental stages) can be genomically stable, which can be no mutations, or with simple mutations, or MSI, or simple translocations.

(e) Further research needs to consider the information management in the attractor models. Macroevolution generates genome attractor that creates system information, while microevolution influences epigenetic regulation and population growth via information modification and usage. In a sense, new information package creation and preservation as well as usage represent the fundamental driver of cancer evolution. Interestingly, while increased entropy may indicate increased opportunities for cancer cells to searching/creating the new life space, too much entropy may also be associated with a reduced capability for information preservation, leading to the self-elimination of newly formed cells. Clearly, a certain degree of genome stability is also important for maintain genome attractors [70].

(f) Both the genome attractor model and the two-phased cancer evolutionary framework can have clinical implications. As the harsh treatment can lead to genome chaos that alters the genome attractor via genome reorganization, an alternative strategy is to build the constraints of attractors at a higher system level while slowing down the population growth, to reduce the treatment-induced rapid cancer macroevolution. It is thus necessary to identify and monitor the two-phased cancer evolution, and not to push new runs of macroevolution, which is highly unpredictable but moderately controls the cancer growth during the microevolutionary phase [3,6,21,36,69,70,71].

REFERENCES

1. Heng HH, Bremer SW, Stevens JB, Horne SD, Liu G, Abdallah BY, et al. Chromosomal instability (CIN): what it is and why it is crucial to cancer evolution. Cancer Metastasis Rev. 2013 Dec;**32**(3–4):325–40. doi: 10.1007/s10555-013-9427-7
2. Ye CJ, Sharpe Z, Heng HH. Origins and consequences of chromosomal instability: from cellular adaptation to genome chaos-mediated system survival. Genes (Basel). 2020 Sep 30;**11**(10):1162. doi: 10.3390/genes11101162

3. Heng J, Heng HH. Genome chaos: creating new genomic information essential for cancer macroevolution. Semin Cancer Biol. 2022 Jun;**81**:160–75. doi: 10.1016/j.semcancer.2020.11.003
4. Virchow R. Die cellularpathologie in ihrer begründung auf physiologische und pathologische gewebelehre. Zwanzig vorlesungen gehalten während der monate februar, märz und aprilim Pathologischen institute zu Berlin. Berlin, A. Hirschwald, 1858. Pdf. https://www.loc.gov/item/06041231/
5. Hanselmann RG, Welter C. Origin of cancer: an information, energy, and matter disease. Front Cell Dev Biol. 2016;**4**:121. doi: 10.3389/fcell.2016.00121
6. Heng HH. Debating cancer: the paradox in cancer research. Singapore: World Scientific Publish Co.; 2015.
7. Markert CL. Neoplasia: a disease of cell differentiation. Cancer Res. 1968;**28**(9):1908–14.
8. Liu J. The dualistic origin of human tumors. Semin Cancer Biol. 2018 Dec;**53**:1–16. doi: 10.1016/j.semcancer.2018.07.004
9. Zhu S, Wang J, Zellmer L, Xu N, Liu M, Hu Y, et al. Mutation or not, what directly establishes a neoplastic state, namely cellular immortality and autonomy, still remains unknown and should be prioritized in our research. J Cancer. 2022 Jul 4;**13**(9):2810–43. doi: 10.7150/jca.72628
10. Greaves M, Maley C. Clonal evolution in cancer. Nature. 2012;**481**:306–13. doi: doi.org/10.1038/nature10762
11. Stevens JB, Liu G, Abdallah BY, Horne SD, Ye KJ, Bremer SW, et al. Unstable genomes elevate transcriptome dynamics. Int J Cancer. 2014 May 1;**134**(9):2074–87. doi: 10.1002/ijc.28531
12. Kauffman S. Differentiation of malignant to benign cells. J Theor Biol. 1971; **31**(3):429–51. doi: 10.1016/0022-5193(71)90020-8
13. Kauffman S. Homeostasis and differentiation in random genetic control networks. Nature. 1969 Oct 11;**224**(5215):177–8. doi: 10.1038/224177a0
14. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol. 1969 Mar;**22**(3):437–67. doi: 10.1016/0022-5193(69)90015-0
15. Kauffman SA. The origins of order. New York: Oxford University Press; 1993.
16. Huang S, Ernberg I, Kauffman S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. Semin Cell Dev Biol. 2009 Sep;**20**(7):869–76. doi: 10.1016/j.semcdb.2009.07.003
17. Huang S, Kauffman SA. Complex gene regulatory networks—from structure to biological observables: cell fate determination. In: Meyers R, editor, Computational complexity. New York, NY: Springer; 2012. doi: 10.1007/978-1-4614-1800-9_35
18. Heng J, Heng HH. Two-phased evolution: genome chaos-mediated information creation and maintenance. Prog Biophys Mol Biol. 2021 Oct;**165**:29–42. doi: 10.1016/j.pbiomolbio.2021.04.003
19. Heng HH, Stevens JB, Bremer SW, Liu G, Abdallah BY, Ye CJ. Evolutionary mechanisms and diversity in cancer. Adv Cancer Res. 2011;**112**:217–53. doi: 10.1016/B978-0-12-387688-1.00008-9
20. Huang S. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. Cancer Metastasis Rev. 2013 Dec;**32**(3–4):423–48. doi: 10.1007/s10555-013-9435-7
21. Heng HH. Genome chaos: rethinking genetics, evolution, and molecular medicine. San Diego, CA, USA: Academic Press; 2019.
22. Ye CJ, Stevens JB, Liu G, Bremer SW, Jaiswal AS, Ye KJ, et al. Genome based cell population heterogeneity promotes tumorigenicity: the evolutionary mechanism of cancer. J Cell Physiol. 2009 May;**219**(2):288–300. doi: 10.1002/jcp.21663
23. Kasperski A, Kasperska R. Bioenergetics of life, disease and death phenomena. Theory Biosci. 2018;**137**(2):155–68. doi: 10.1007/s12064-018-0266-5
24. Kasperski A. Modelling of cells bioenergetics. Acta Biotheor. 2008 Sep;**56**(3):233–47. doi: 10.1007/s10441-008-9050-0
25. Kasperski A, Kasperska R. Selected disease fundamentals based on the unified cell bioenergetics. J Invest Biochem. 2013;**2**(2):93–100. doi: 10.5455/jib.20130227041230
26. Kasperski, A. Genome attractors as places of evolution and oases of life. Processes. 2021;**9**(9):1646. doi: 10.3390/pr9091646
27. Kasperski A. Life entrapped in a network of atavistic attractors: how to find a rescue. Int J Mol Sci. 2022;**23**(7):4017. doi: doi.org/10.3390/ijms23074017
28. Kasperski A, Kasperska R. Study on attractors during organism evolution. Sci Rep. 2021;**11**:9637. doi: 10.1038/s41598-021-89001-0
29. Crkvenjakov R, Heng HH. Further illusions: on key evolutionary mechanisms that could never fit with Modern Synthesis. Prog Biophys Mol Biol. 2022 Mar–May;169–170:3–11. doi: 10.1016/j.pbiomolbio.2021.10.002
30. Kulkarni P, Shiraishi T, Kulkarni RV. Cancer: tilting at windmills? Mol Cancer. 2013 Sep 24;12(1):108. doi: 10.1186/1476-4598-12-10
31. van der Knaap JA, Verrijzer CP. Undercover: gene control by metabolites and metabolic enzymes. Genes Dev. 2016;**30**(21):2345–69. doi: 10.1101/gad.289140.116
32. Heng J, Heng HH. Karyotype coding: the creation and maintenance of system information for complexity and biodiversity. Biosystems. 2021;**208**:104476. doi: 10.1016/j.biosystems.2021.104476
33. Boveri T. The origin of malignant tumors. Baltimore, MD: Williams & Wilkins; 1929. p. 62–3.
34. Liu J. Giant cells: linking McClintock's heredity to early embryogenesis and tumor origin throughout millennia of evolution on Earth. Semin Cancer Biol. 2022 Jun; 81:176–92. doi: 10.1016/j.semcancer.2021.06.007
35. Liu G, Stevens JB, Horne SD, Abdallah BY, Ye KJ, Bremer SW, et al. Genome chaos: survival strategy during crisis. Cell Cycle. 2014;**13**(4):528–37. doi: 10.4161/cc.27378
36. Ye JC, Horne S, Zhang JZ, Jackson L, Heng HH. Therapy induced genome chaos: a novel mechanism of rapid cancer drug resistance. Front Cell Dev Biol. 2021 Jun 10; **9**:676344. doi: 10.3389/fcell.2021.676344
37. Vainshelbaum NM, Salmina K, Gerashchenko BI, Lazovska M, Zayakin P, Cragg MS, et al. Role of the circadian clock 'death-loop' in the DNA damage response underpinning cancer treatment resistance. Cells. 2022. **11**(5):880. doi: 10.3390/cells11050880
38. Heng HH. Chapter 5, The genomic landscape of cancers. In: Ujvari B, Roche B, Thomas F, editors. Ecology and evolution of cancer. London: Elsevier, Academic Press; 2017. p. 69–86.
39. Warburg O. On the origin of cancer cells. Science. 1956;**123**(3191):309–14. doi: 10.1126/science.123.3191.309
40. Pecqueur C, Oliver L, Oizel K, Lalier L, Vallette FM. Targeting metabolism to induce cell death in cancer cells and cancer stem cells. Int J Cell Biol. 2013;**2013**:805975. doi: 10.1155/2013/805975
41. Phan LM, Yeung SC, Lee MH. Cancer metabolic reprogramming: importance, main features, and potentials for precise targeted anti-cancer therapies. Cancer Biol Med. 2014;**11**(1):1–19. doi: 10.7497/j.issn.2095-3941.2014.01.001
42. de Souza AC, Justo GZ, de Araújo DR, Cavagis AD. Defining the molecular basis of tumor metabolism: a continuing

challenge since Warburg's discovery. Cell Physiol Biochem. 2011;**28**(5):771–92. doi: 10.1159/000335792
43. Moreno-Sánchez R, Rodríguez-Enríquez S, Marín-Hernández A, Saavedra E. Energy metabolism in tumor cells. FEBS J. 2007 Mar;**274**(6):1393–418. doi: 10.1111/j.1742-4658.2007.05686.x
44. Diaz-Ruiz R, Rigoulet M, Devin A. The Warburg and Crabtree effects: On the origin of cancer cell energy metabolism and of yeast glucose repression. Biochim Biophys Acta. 2011 Jun;**1807**(6):568–76. doi: 10.1016/j.bbabio.2010.08.010
45. Zimatore G, Tsuchiya M, Hashimoto M, Kasperski A, Giuliani A. Self-organization of whole-gene expression through coordinated chromatin structural transition. Biophys Rev. 2021;**2**:031303. doi: 10.1063/5.0058511
46. Ye CJ, Sharpe Z, Alemara S, Mackenzie S, Liu G, Abdallah B, et al. Micronuclei and genome chaos: changing the system inheritance. Genes (Basel). 2019 May 13;**10**(5):366. doi: 10.3390/genes10050366
47. Egeblad M, Nakasone ES, Werb Z. Tumors as organs: complex tissues that interface with the entire organism. Dev Cell. 2010;**18**(6):884–901. doi: 10.1016/j.devcel.2010.05.012
48. Trigos AS, Pearson RB, Papenfuss AT, Goode DL. Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. Proc Natl Acad Sci U S A. 2017 Jun 13;**114**(24):6406–11. doi: 10.1073/pnas.1617743114
49. Bussey KJ, Cisneros LH, Lineweaver CH, Davies PCW. Ancestral gene regulatory networks drive cancer. Proc Natl Acad Sci U S A. 2017 Jun 13;**114**(24):6160–62. doi: 10.1073/pnas.1706990114
50. Vinogradov AE, Anatskaya OV. Cellular biogenetic law and its distortion by protein interactions: a possible unified framework for cancer biology and regenerative medicine. Int J Mol Sci. 2022;**23**(19):11486. https://doi.org/10.3390/ijms231911486
51. Soto AM, Sonnenschein C. The tissue organization field theory of cancer: a testable replacement for the somatic mutation theory. Bioessays. 2011;**33**(5):332–40. doi: 10.1002/bies.201100025
52. Soto AM, Sonnenschein C. The cancer puzzle: welcome to organicism. Prog Biophys Mol Biol. 2021;**165**:114–9. doi: 10.1016/j.pbiomolbio.2021.07.001
53. Bizzarri M, Cucina A. SMT and TOFT: why and how they are opposite and incompatible paradigms. Acta Biotheor. 2016;(64):221–39. doi:10.1007/s10441-016-9281-4
54. Heng HH, Regan SM, Liu G, Ye CJ. Why it is crucial to analyze non clonal chromosome aberrations or NCCAs? Mol Cytogenet. 2016 Feb 13;9:15. doi: 10.1186/s13039-016-0223-2
55. Fares J, Fares MY, Khachfe HH, Salhab HA, Fares Y. Molecular principles of metastasis: a hallmark of cancer revisited. Sig Transduct Target Ther. 2020;**5**:28. doi: doi.org/10.1038/s41392-020-0134-x
56. Guan X. Cancer metastases: challenges and opportunities. Acta Pharm Sin B. 2015 Sep;**5**(5):402–18. doi: 10.1016/j.apsb.2015.07.005
57. Beans C. Targeting metastasis to halt cancer's spread. Proc Natl Acad Sci U S A. 2018;**115**(50): 12539–43. doi: 10.1073/pnas.1818892115
58. Weiss L, Ward PM. Cell detachment and metastasis. Cancer Metastasis Rev. 1983;**2**(2):111–27. doi: 10.1007/BF00048965. PMID: 6352010
59. Gao C, Su Y, Koeman J, Haak E, Dykema K, Essenberg C, et al. Chromosome instability drives phenotypic switching to metastasis. Proc Natl Acad Sci U S A. 2016 Dec 20;**113**(51):14793–8. doi: 10.1073/pnas.1618215113
60. Bakhoum SF, Ngo B, Laughney AM, Cavallo JA, Murphy CJ, Ly P, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. Nature. 2018 Jan 25;**553**(7689):467–72. doi: 10.1038/nature25432
61. Lineweaver CH, Bussey KJ, Blackburn AC, Davies PCW. Cancer progression as a sequence of atavistic reversions. Bioessays. 2021;**43**(7):e2000305. doi: 10.1002/bies.202000305
62. Davies PC, Lineweaver CH. Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors. Phys Biol. 2011 Feb;**8**(1):015001. doi: 10.1088/1478-3975/8/1/015001
63. Vincent M. Cancer: a de-repression of a default survival program common to all cells?: a life-history perspective on the nature of cancer. Bioessays. 2012 Jan;**34**(1):72–82. doi: 10.1002/bies.201100049
64. Bussey KJ, Davies PCW. Reverting to single-cell biology: the predictions of the atavism theory of cancer. Prog Biophys Mol Biol. 2021 Oct;**165**:49–55. doi: 10.1016/j.pbiomolbio.2021.08.002
65. Ciompi L, Tschacher W. Affect-logic, embodiment, synergetics, and the free energy principle: new approaches to the understanding and treatment of schizophrenia. Entropy (Basel). 2021 Dec 1;**23**(12):1619. doi: 10.3390/e23121619
66. Movahed TM, Bidgoly HJ, Manesh MHK, Mirzaei HR. Predicting cancer cells progression via entropy generation based on AR and ARMA models. Int Commun Heat Mass Transf. 2021;127:105565. doi: 10.1016/j.icheatmasstransfer.2021.105565
67. Pitt MA. Increased temperature and entropy production in cancer: the role of anti-inflammatory drugs. Inflammopharmacology. 2015 Feb;**23**(1):17–20. doi: 10.1007/s10787-014-0224-x
68. Heng HH. The genome-centric concept: resynthesis of evolutionary theory. Bioessays. 2009 May;**31**(5):512–25. doi: 10.1002/bies.200800182
69. Somarelli JA, DeGregori J, Gerlinger M, Heng HH, Marusyk A, Welch DR, et al. Questions to guide cancer evolution as a framework for furthering progress in cancer research and sustainable patient outcomes. Med Oncol. 2022;**39**(9):137. doi: 10.1007/s12032-022-01721-z
70. Heng HH. Genome chaos: rethinking genetics, evolution, and molecular medicine. 2nd ed. Cambridge, MA: Academic Press, Elsevier; 2025 (in press).
71. Kasperski A, Heng HH. The spiral model of evolution: stable life forms of organisms and unstable life forms of cancers. Int J Mol Sci. 2024;**25**(17):9163. doi: 10.3390/ijms25179163

Index

For the benefit of digital users, indexed terms that span two pages (e.g., 52–53) may, on occasion, appear on only one of those pages.