

Methods in  
Molecular Biology 2577

Springer Protocols

Izuho Hatada · Takuro Horii *Editors*

# Epigenomics

Methods and Protocols

 Humana Press

# METHODS IN MOLECULAR BIOLOGY

*Series Editor*

**John M. Walker**

**School of Life and Medical Sciences**

**University of Hertfordshire**

**Hatfield, Hertfordshire, UK**

For further volumes:

<http://www.springer.com/series/7651>

For over 35 years, biological scientists have come to rely on the research protocols and methodologies in the critically acclaimed *Methods in Molecular Biology* series. The series was the first to introduce the step-by-step protocols approach that has become the standard in all biomedical protocol publishing. Each protocol is provided in readily-reproducible step-by-step fashion, opening with an introductory overview, a list of the materials and reagents needed to complete the experiment, and followed by a detailed procedure that is supported with a helpful notes section offering tips and tricks of the trade as well as troubleshooting advice. These hallmark features were introduced by series editor Dr. John Walker and constitute the key ingredient in each and every volume of the *Methods in Molecular Biology* series. Tested and trusted, comprehensive and reliable, all protocols from the series are indexed in PubMed.


# **Epigenomics**

## **Methods and Protocols**

Edited by

**Izuho Hatada and Takuro Horii**

*Institute for Molecular and Cellular Regulation, Gunma University, Maebashi, Gunma, Japan*

 **Humana Press**



*Editors*

Izuho Hatada  
Institute for Molecular and Cellular  
Regulation  
Gunma University  
Maebashi, Gunma, Japan

Takuro Horii  
Institute for Molecular and Cellular Regulation  
Gunma University  
Maebashi, Gunma, Japan

ISSN 1064-3745

ISSN 1940-6029 (electronic)

Methods in Molecular Biology

ISBN 978-1-0716-2723-5

ISBN 978-1-0716-2724-2 (eBook)

<https://doi.org/10.1007/978-1-0716-2724-2>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Science+Business Media, LLC, part of Springer Nature 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Humana imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.

The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

---

## Preface

Epigenetic modifications turn genes on and off and differ between individuals, tissues, and even cells, contributing to their unique characteristics. Diseases can also stem from epigenetic modifications that cause abnormal activation or silencing of genes. The advent of next-generation sequencing (NGS) technology at the turn of this century has made it possible to analyze epigenetic modifications across the entire genome, inaugurating the dawn of the epigenomics era. Ongoing research using NGS technology continues to show that epigenetic changes are at the basis of many human diseases. However, it was not until the early 2010s that a versatile method emerged to identify the epigenetic changes that cause a specific disease. Methods that enable the switching on and off of specific genes artificially also hold great promise for determining which gene(s) is responsible for a certain disease and could also have a wide range of medical and biotechnological applications.

The clustered regularly interspaced short palindromic repeats (CRISPR)/ Cas9 system from bacteria has been successfully adapted for RNA-guided genome editing in a wide range of organisms. Cas9 cleaves DNA upon recognition of a target sequence by the gRNA. Conversely, dCas9, a nuclease-defective Cas9, binds specifically to the target sequence without cleaving it. Therefore, dCas9 can recruit epigenetic enzymes to a specific gene and introduce changes in a single epigenetic mark. This new technology provides a platform for the specific manipulation of the epigenome, a capability that could only have been dreamed of a few years ago.

This book provides a comprehensive collection of protocols for epigenomic research. It consists of two main parts. The first part describes the methods used to analyze epigenomic modifications, such as DNA methylation (Chapters 1, 2, and 3), histone modifications, and chromatin structure (Chapters 4, 5, 6, 7, 8, 9, 10, and 11). The second describes methods for manipulating the epigenome, such as platforms for epigenome editing (Chapters 12, 13, 14, and 15), inducible systems for epigenome editing (Chapter 16), and epigenetically modified animals (Chapters 17, 18, 19, and 20).

We thank all of the authors for their outstanding contributions to this book. On behalf of all of us, we hope that you will find the information in this book useful and that it will further expand the utility and scope of epigenomics research.

*Maebashi, Gunma, Japan*

*Izaho Hatada  
Takuro Horii*

---

# Contents

<i>Preface</i> .....	<i>v</i>
<i>Contributors</i> .....	<i>ix</i>
PART I ANALYSIS OF THE EPIGENOME: DNA METHYLATION	
1 DNA Methylation Analysis Using Bisulfite Pyrosequencing .....	3
<i>Ken Higashimoto, Satoshi Hara, and Hidenobu Soejima</i>	
2 Post-bisulfite Adaptor Tagging Based on an ssDNA Ligation Technique (tPBAT) .....	21
<i>Fumihito Miura, Yukiko Shibata, Miki Miura, and Takashi Ito</i>	
3 Reduced Representation Bisulfite Sequencing (RRBS) .....	39
<i>Kazuhiko Nakabayashi, Michihiro Yamamura, Keita Hasegawa, and Kenichiro Hata</i>	
PART II ANALYSIS OF THE EPIGENOME: HISTONE MODIFICATION AND CHROMATIN STRUCTURE	
4 Chromatin Immunoprecipitation Sequencing (ChIP-seq) for Detecting Histone Modifications and Modifiers .....	55
<i>Shinjiro Hino, Tetsuya Sato, and Mitsuyoshi Nakao</i>	
5 ATAC-Seq Analysis of Accessible Chromatin: From Experimental Steps to Data Analysis .....	65
<i>Mayu Tataru, Taiyo Ikeda, Satoshi H. Namekawa, and So Maezawa</i>	
6 Low-Input CUT&RUN for Mouse Oocytes and Preimplantation Embryos .....	83
<i>Ryoya Hayashi and Azusa Inoue</i>	
7 Imaging Chromatin Accessibility by Assay of Transposase-Accessible Chromatin with Visualization .....	93
<i>Yusuke Miyanari</i>	
8 STREAMING-Tag System: Technology to Enable Visualization of Transcriptional Activity and Subnuclear Localization of Specific Endogenous Genes .....	103
<i>Hiroaki Ohishi and Hiroshi Ochiai</i>	
9 Bioinformatics Pipelines for Identification of Super-Enhancers and 3D Chromatin Contacts .....	123
<i>Akihiko Sakashita, Chikara Takeuchi, So Maezawa, and Satoshi H. Namekawa</i>	
10 A Method for Detection of Somatic LINE-1 Insertions at the Single-Cell Level from Postmortem Human Brain .....	147
<i>Miki Bundo and Kazuya Iwamoto</i>	

11 Solubilization of Mouse Sperm Chromatin for Sequencing Analyses Using a Chaperon Protein ..... 161  
*Yuko Fukuda, Keishi Shintomi, Kosuke Yamaguchi, Yasuhiro Fujiwara, and Yuki Okada*

PART III MANIPULATION OF THE EPIGENOME: PLATFORMS FOR EPIGENOME EDITING

12 Efficient Targeted DNA Methylation with dCas9-Coupled DNMT3A-DNMT3L Methyltransferase ..... 177  
*Pavel Bashtrykov, Nivethika Rajaram, and Albert Jeltsch*

13 Regulation of Gene Expression Using dCas9-SunTag Platforms ..... 189  
*Sumiyo Morita, Takuro Horii, and Izubo Hatada*

14 Concatenated Coiled-Coil Tag for Highly Efficient, Small Molecule-Inducible Upregulation of Endogenous Mammalian Genes ..... 197  
*Tina Lebar and Roman Jerala*

15 Design, Construction, and Validation of Targeted Gene Activation with TREE System in Human Cells ..... 211  
*Atsushi Kunii, Takashi Yamamoto, and Tetsushi Sakuma*

PART IV MANIPULATION OF THE EPIGENOME: INDUCIBLE SYSTEM FOR EPIGENOME EDITING

16 A Split CRISPR-CpfI Platform for Inducible Gene Activation ..... 229  
*Takahiro Otabe, Yuta Nihongaki, and Moritoshi Sato*

PART V MANIPULATION OF THE EPIGENOME: EPIGENETICALLY MODIFIED ANIMALS

17 Targeted DNA Methylation in Mouse Early Embryos ..... 243  
*Taiga Yamazaki, Yu Hatano, Noritada Kobayashi, and Kazuo Yamagata*

18 Generation of Epigenetic Disease Model Mice by Targeted Demethylation of the Epigenome ..... 255  
*Takuro Horii, Sumiyo Morita, and Izubo Hatada*

19 In Vivo Tissue-Specific DNA Demethylation in Mouse Liver Through a Hydrodynamic Tail Vein Injection ..... 269  
*Koshi Hashimoto and Nozomi Hanzawa*

20 Targeted Manipulation of Histone Modification in Medaka Embryos ..... 279  
*Hiroto S. Fukushima, Hiroyuki Takeda, and Ryobei Nakamura*

*Index* ..... 295

---

## Contributors

- PAVEL BASHTRYKOV • *Institute of Biochemistry and Technical Biochemistry, Department of Biochemistry, University of Stuttgart, Stuttgart, Germany*
- MIKI BUNDO • *Department of Molecular Brain Science, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan*
- YASUHIRO FUJIWARA • *Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan*
- YUKO FUKUDA • *Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan*
- HIROTO S. FUKUSHIMA • *Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan*
- NOZOMI HANZAWA • *Department of Diabetes and Endocrinology, National Disaster Medical Center, Tachikawa, Tokyo, Japan*
- SATOSHI HARA • *Division of Molecular Genetics and Epigenetics, Department of Biomolecular Sciences, Faculty of Medicine, Saga University, Saga, Japan*
- KEITA HASEGAWA • *Department of Obstetrics and Gynecology, Keio University School of Medicine, Shinjuku, Tokyo, Japan*
- KOSHI HASHIMOTO • *Department of Diabetes, Endocrinology and Hematology, Dokkyo Medical University Saitama Medical Center, Koshigaya, Saitama, Japan*
- KENICHIRO HATA • *Department of Maternal-Fetal Biology, Research Institute, National Center for Child Health and Development, Setagaya, Tokyo, Japan*
- IZUHO HATADA • *Laboratory of Genome Science, Biosignal Genome Resource Center, Institute for Molecular and Cellular Regulation, Gunma University, Maebashi, Gunma, Japan*
- YU HATANO • *Faculty of Biology-Oriented Science and Technology, Kindai University, Kinokawa, Wakayama, Japan*
- RYOYA HAYASHI • *Laboratory for Epigenome Inheritance, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan; Tokyo Metropolitan University, Hachioji, Tokyo, Japan*
- KEN HIGASHIMOTO • *Division of Molecular Genetics and Epigenetics, Department of Biomolecular Sciences, Faculty of Medicine, Saga University, Saga, Japan*
- SHINJIRO HINO • *Department of Medical Cell Biology, Institute of Molecular Embryology and Genetics, Kumamoto University, Kumamoto, Japan*
- TAKURO HORII • *Laboratory of Genome Science, Biosignal Genome Resource Center, Institute for Molecular and Cellular Regulation, Gunma University, Maebashi, Gunma, Japan*
- TAIYO IKEDA • *Department of Applied Biological Science, Tokyo University of Science, Chiba, Japan*
- AZUSA INOUE • *Laboratory for Epigenome Inheritance, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan; Tokyo Metropolitan University, Hachioji, Tokyo, Japan*
- TAKASHI ITO • *Department of Biochemistry, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan*
- KAZUYA IWAMOTO • *Department of Molecular Brain Science, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan*

- ALBERT JELTSCH • *Institute of Biochemistry and Technical Biochemistry, Department of Biochemistry, University of Stuttgart, Stuttgart, Germany*
- ROMAN JERALA • *Department of Synthetic Biology and Immunology, National Institute of Chemistry, Ljubljana, Slovenia*
- NORITADA KOBAYASHI • *Division of Biomedical Research, Kitasato University Medical Center, Kitasato University, Kitamoto, Saitama, Japan*
- ATSUSHI KUNII • *Division of Integrated Sciences for Life, Graduate School of Integrated Sciences for Life, Hiroshima University, Hiroshima, Japan*
- TINA LEBAR • *Department of Synthetic Biology and Immunology, National Institute of Chemistry, Ljubljana, Slovenia*
- SO MAEZAWA • *Department of Applied Biological Science, Tokyo University of Science, Chiba, Japan*
- FUMIHITO MIURA • *Department of Biochemistry, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan*
- MIKI MIURA • *Department of Biochemistry, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan*
- YUSUKE MIYANARI • *NanoLSI, Cancer Research Institute, Kanazawa University, Kanazawa, Japan*
- SUMIYO MORITA • *Laboratory of Genome Science, Biosignal Genome Resource Center, Institute for Molecular and Cellular Regulation, Gunma University, Maebashi, Gunma, Japan*
- KAZUHIKO NAKABAYASHI • *Department of Maternal-Fetal Biology, Research Institute, National Center for Child Health and Development, Setagaya, Tokyo, Japan*
- RYOHEI NAKAMURA • *Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan*
- MITSUYOSHI NAKAO • *Department of Medical Cell Biology, Institute of Molecular Embryology and Genetics, Kumamoto University, Kumamoto, Japan*
- SATOSHI H. NAMEKAWA • *Department of Microbiology and Molecular Genetics, University of California Davis, Davis, CA, USA*
- YUTA NIHONGAKI • *Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan; Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, MD, USA*
- HIROSHI OCHIAI • *Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, Japan*
- HIROAKI OHISHI • *Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, Japan*
- YUKI OKADA • *Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan*
- TAKAHIRO OTABE • *Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan; Kanagawa Institute of Industrial Science and Technology, Kanagawa, Japan*
- NIVETHIKA RAJARAM • *Institute of Biochemistry and Technical Biochemistry, Department of Biochemistry, University of Stuttgart, Stuttgart, Germany*
- AKIHIKO SAKASHITA • *Department of Molecular Biology, Keio University School of Medicine, Tokyo, Japan*
- TETSUSHI SAKUMA • *Division of Integrated Sciences for Life, Graduate School of Integrated Sciences for Life, Hiroshima University, Hiroshima, Japan*
- MORITOSHI SATO • *Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan; Kanagawa Institute of Industrial Science and Technology, Kanagawa, Japan*
- TETSUYA SATO • *Research Laboratory, H.U. Group Research Institute G.K., Tokyo, Japan*

- YUKIKO SHIBATA • *Department of Biochemistry, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan*
- KEISHI SHINTOMI • *Chromosome Dynamics Laboratory, RIKEN, Saitama, Japan*
- HIDENOBU SOEJIMA • *Division of Molecular Genetics and Epigenetics, Department of Biomolecular Sciences, Faculty of Medicine, Saga University, Saga, Japan*
- HIROYUKI TAKEDA • *Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan*
- CHIKARA TAKEUCHI • *Department of Molecular Biology, Keio University School of Medicine, Tokyo, Japan*
- MAYU TATARA • *Department of Applied Biological Science, Tokyo University of Science, Chiba, Japan*
- KAZUO YAMAGATA • *Faculty of Biology-Oriented Science and Technology, Kindai University, Kinokawa, Wakayama, Japan*
- KOSUKE YAMAGUCHI • *Institute for Quantitative Biosciences, The University of Tokyo, Tokyo, Japan*
- TAKASHI YAMAMOTO • *Division of Integrated Sciences for Life, Graduate School of Integrated Sciences for Life, Hiroshima University, Hiroshima, Japan*
- MICHIHIRO YAMAMURA • *Department of Obstetrics and Gynecology, The Jikei University School of Medicine, Minato, Tokyo, Japan*
- TAIGA YAMAZAKI • *Division of Biomedical Research, Kitasato University Medical Center, Kitasato University, Kitamoto, Saitama, Japan*

# Part I

## Analysis of the Epigenome: DNA Methylation





# Chapter 1

## DNA Methylation Analysis Using Bisulfite Pyrosequencing

Ken Higashimoto, Satoshi Hara, and Hidenobu Soejima

### Abstract

Pyrosequencing is a DNA sequencing-by-synthesis technique that can quantitatively detect single-nucleotide polymorphisms (SNPs). With pyrosequencing, the level of DNA methylation can be calculated according to the ratio of artificial cytosine/thymine SNPs produced by bisulfite conversion at each CpG site. This analysis method provides a reproducible and accurate measurement of methylation levels at CpG sites near sequencing primers with high quantitative resolution. DNA methylation plays an important role in mammalian development and cellular physiology; alterations in DNA methylation patterns have been implicated in several common diseases as well as cancers and imprinting disorders. Evaluating DNA methylation levels via pyrosequencing is useful for identifying biomarkers that could help with the diagnosis, prognosis, treatment selection, and onset risk assessment for several diseases. We describe the principles of pyrosequencing and detail a bisulfite pyrosequencing protocol based on our experience and the PyroMark Q24 User Manual.

**Key words** Pyrosequencing technology, SNPs, DNA methylation, CpG site, Bisulfite conversion, Biomarker detection, Primer design

---

## 1 Introduction

Epigenetics is the study of mitotically heritable alterations in gene expression without changing the DNA sequence [1]. Epigenetic mechanisms regulate various physiological processes, including development, genomic imprinting, X inactivation, and chromosomal stability [2–4]. Epigenetic processes, such as DNA methylation, histone modification, and noncoding RNA, influence relevant gene expression by altering the chromatin structure.

DNA methylation was the first epigenetic modification discovered in humans [5]. In mammals, methylation is a post-replication modification that predominantly affects cytosines (Cs) of CpG dinucleotide sequences. The modification is performed by DNA methyltransferases (DNMTs), which transfer a methyl group from S-adenosylmethionine (SAM) to the fifth carbon of C to form 5-methylcytosine (5mC). The role of DNA methylation in humans

has been investigated in the context of genomic imprinting. In imprinting disorders such as Beckwith–Wiedemann syndrome, aberrations have been found in differentially methylated regions (DMRs), which control the monoallelic expression of imprinted genes. Such aberrations have also been found in many cancer types, such as colon, breast, liver, bladder, Wilms, ovarian, and esophageal cancers [6]. Genome-wide DNA hypomethylation, which leads to chromosome instability and reactivation of transposable elements, and hypermethylation of tumor suppressor gene promoters, which results in gene silencing, have also been identified in cancers [7].

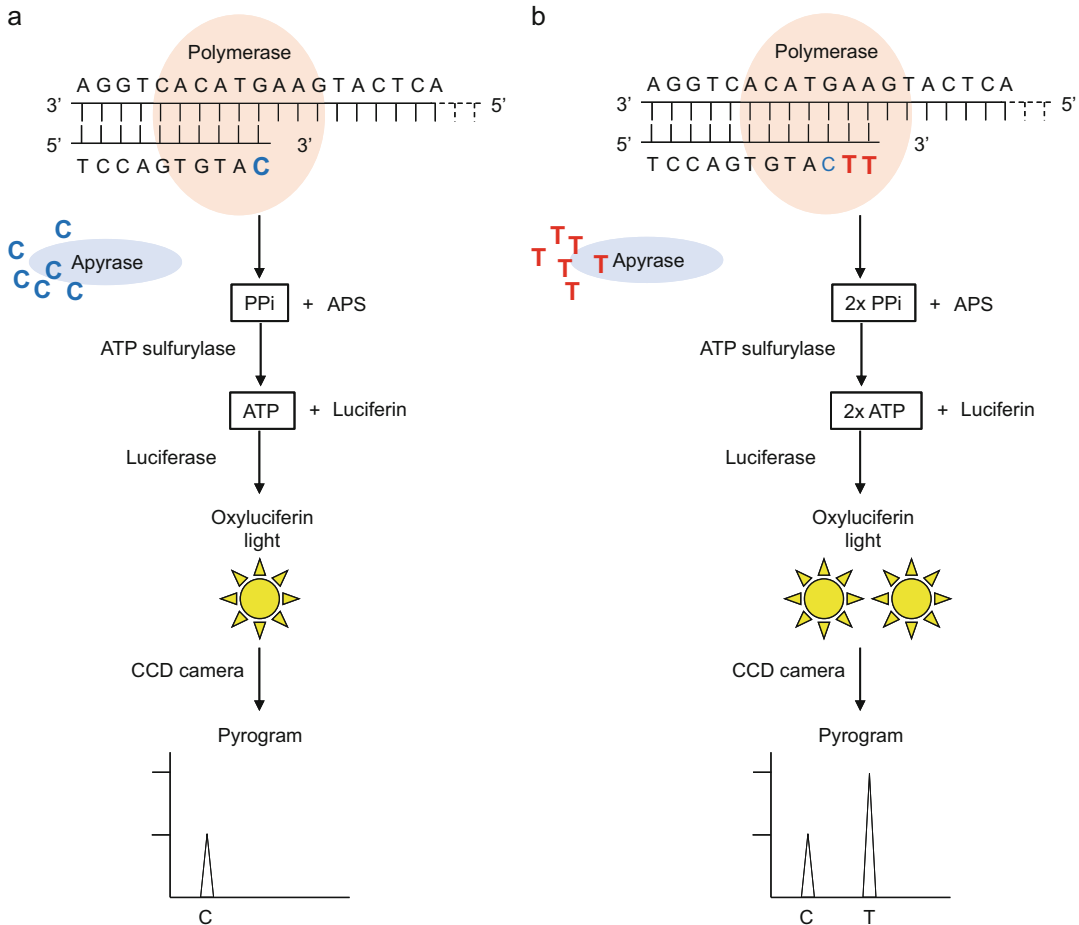
In recent years, there has been rapid progress in genome-wide DNA methylation analysis methods due to next-generation sequencing and microarray hybridization technologies such as whole genome bisulfite sequencing, reduced representation bisulfite sequencing, and DNA methylation microarrays. These methods allow us to identify DNA methylation alterations that may be present in common diseases besides cancer, including neurological diseases, metabolic disorders, and autoimmune diseases [8, 9]. DNA methylation at certain sites could serve as a biomarker for disease diagnosis, prognosis, treatment selection, and onset risk assessment [8].

Although there are various common genome-wide DNA methylation analysis methods, data normalization, bias correction, batch effect determination, and selection of appropriate statistical methods for detecting alterations in DNA methylation remain complicated. Therefore, the results of genome-wide DNA methylation analyses often require validation by other methods [10, 11]. For the methods to be suitable for examining DNA methylation biomarkers in a clinical setting, they must be simple and fast, with high reproducibility and low cost. Bisulfite pyrosequencing can fulfill all of these requirements and reduce complications.

In the sections that follow, we describe the principles of pyrosequencing and detail a bisulfite pyrosequencing protocol based on our experience and the PyroMark Q24 User Manual.

### **1.1 Principles of Pyrosequencing Technology**

Pyrosequencing is a DNA sequencing-by-synthesis method that depends on the sequential addition and incorporation of nucleotides during DNA polymerase primer extension and can analyze DNA sequences quantitatively [11–13]. As the polymerase extends a sequencing primer annealed to a single-strand DNA template, a light signal is generated through an enzymatic cascade reaction initiated by the incorporation of a nucleotide (Fig. 1a). This enzymatic cascade reaction involves four enzymes, including Klenow DNA polymerase, ATP sulfurylase, luciferase, and apyrase, as well as two substrates besides the synthetic nucleotides: adenosine 5'-phosphosulfate (APS) and luciferin. The pyrosequencing machine dispenses the complementary nucleotides iteratively to the template based on a pre-programmed sequence, facilitating



**Fig. 1** Flow chart showing the principles of pyrosequencing. **(a)** An example where cytosine (C) is incorporated during primer extension. The single-strand DNA template annealed with the pyrosequencing primer is mixed with three enzymes, ATP sulfurylase, luciferase, and apyrase, and two substrates, APS and luciferin. When C is incorporated by DNA polymerase as a complement to the guanine of the template, PPI is released. The PPI is then converted to ATP in the presence of APS by ATP sulfurylase. With ATP, the luciferin is converted to oxyluciferin and light by luciferase. The light is detected by a CCD camera, and its intensity is recorded as a peak on the pyrogram. Unincorporated C as well as a small amount of ATP is degraded by apyrase before the next nucleotide is dispensed. **(b)** Subsequently, thymine (T) is incorporated during primer extension. When the T is incorporated by DNA polymerase as a complement to the adenine of the template, PPI is released. Twice the amount of ATP is generated because of the presence of two consecutive adenine residues in the template. This produces light that is twice as intense, and in the pyrogram, the peak height of T is twice the peak height of the previously incorporated C **(a)**. Unincorporated T as well as a small amount of ATP is degraded by apyrase before the next nucleotide is dispensed. These reactions are repeated following each dispensed nucleotide. APS adenosine 5'-phosphosulfate, ATP adenosine triphosphate, CCD charge-coupled device, PPI pyrophosphate

the extension of the sequencing primer. Each nucleotide incorporation event by Klenow DNA polymerase is accompanied by the release of pyrophosphate (PPi) proportional to the amount of nucleotide incorporated (Fig. 1a, b). The PPi is then converted into ATP quantitatively by the ATP sulfurylase using APS. With the generated ATP, luciferase converts luciferin to oxyluciferin, which results in light emission proportional to the amount of ATP. The light is detected by a charge-coupled device (CCD) camera and is represented as a peak on the pyrogram. The height of each peak thus reflects the amount of nucleotide incorporated (Fig. 1a, b). For synthetic nucleotides, dATP $\alpha$ S is incorporated efficiently by DNA polymerase, and is used in pyrosequencing as an analogue of dATP, which can participate in the luciferase reaction. The unincorporated nucleotides, as well as the small amount of remaining ATP, are completely degraded by apyrase before the next nucleotide is added according to the pre-programmed sequence (Fig. 1a, b). Through this iterative process, the sequence information of a target region can be determined quantitatively in real time.

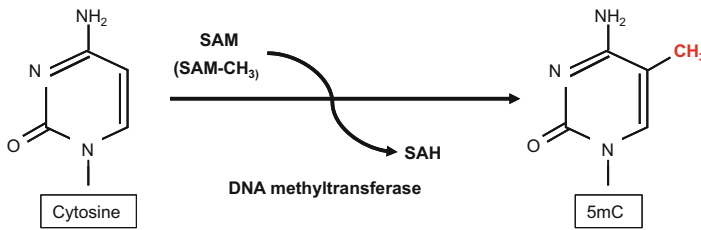
## **1.2 Principles of Bisulfite Conversion**

DNA methylation performed by DNMTs predominantly targets C at CpG sites in mammals. SAM, which is an intermediate product of methionine metabolism, is used as a methyl group donor in this process. DNMTs transfer the methyl group from SAM to the fifth carbon of C, forming 5mC (Fig. 2a) [13]. Bisulfite conversion of genomic DNA is a commonly used method for DNA methylation analysis [14, 15]. In the bisulfite conversion reaction, unmethylated C in a single strand of DNA is converted to cytosine sulfonate via sulfonation with sodium bisulfite at acidic pH and high temperatures (Fig. 2b). Next, the cytosine sulfonate undergoes hydrolytic deamination, which produces a uracil sulfonate. Finally, the uracil sulfonate is desulfonated to uracil (U) at alkaline pH (Fig. 2b). However, 5mC does not react to bisulfite treatment and remains unaltered. Thus, bisulfite conversion distinguishes between unmethylated C and 5mC with single-base resolution.

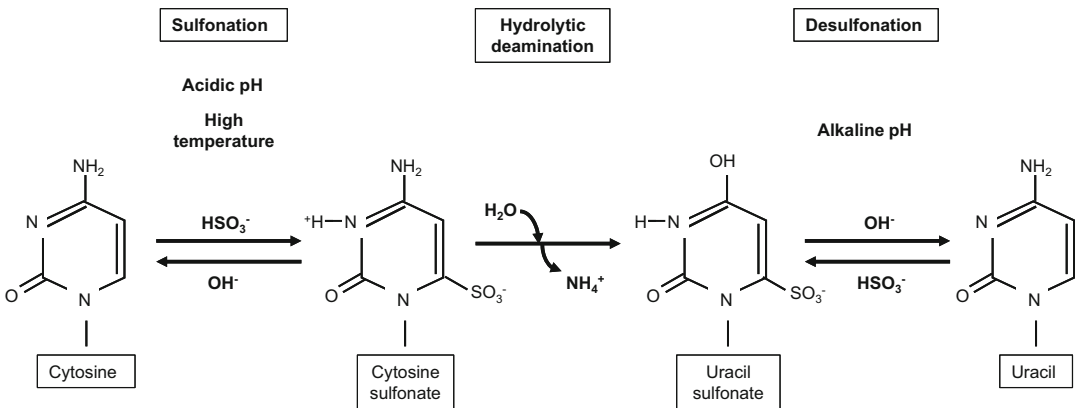
## **1.3 Overview of Bisulfite Pyrosequencing**

First, genomic DNA is treated with bisulfite, and the region containing the target CpG sites is amplified via polymerase chain reaction (PCR). One of the primer pairs used in the PCR amplification reactions is biotinylated, and the biotinylated strand of the amplicon is immobilized on streptavidin beads after denaturation with NaOH. The resulting single-strand DNA template is annealed to the primer, and the complementary strand is extended through discrete dispensation of nucleotides. As mentioned above, bisulfite treatment converts unmethylated C to U, while 5mC remains unchanged (Figs. 2b and 3). The subsequent PCR amplification converts U to thymine (T), and 5mC to C (Fig. 3). Therefore, in the pyrogram, unmethylated C and 5mC end up being shown as T and C peaks, respectively. The DNA methylation ratio (%) of each

a



b



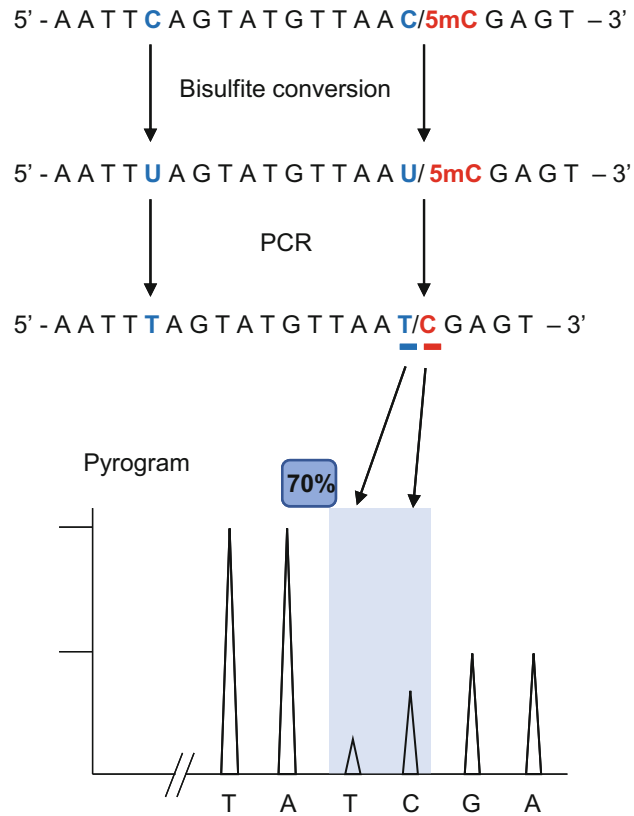
**Fig. 2** Depiction of the process of bisulfite conversion. **(a)** Methylation of cytosine (C) by DNA methyltransferase, which transfers a methyl group derived from SAM to the fifth carbon of C, forming 5-methylcytosine (5mC). **(b)** The chemistry of the bisulfite conversion reaction. First, unmethylated C in a single strand of DNA is converted into cytosine sulfonate at acidic pH and high temperatures via sulfonation with sodium bisulfite. Then, cytosine sulfonate undergoes hydrolytic deamination, producing uracil sulfonate. Finally, uracil sulfonate is desulfonated at alkaline pH, leaving uracil. However, 5mC does not react to the treatment and remains the same.  $\text{CH}_3$  methyl group, SAH S-adenosylhomocysteine, SAM S-adenosylmethionine

CpG site is then calculated based on the peak height ratio of artificial C/T SNPs (Fig. 3). The average methylation level is also calculated using the DNA methylation ratios of all measured CpG sites within the template.

## 2 Materials

### 2.1 Bisulfite Conversion

1. Genomic DNA to be analyzed.
2. Human Methylated & Non-methylated DNA Set (Zymo Research; *see Note 1*).
3. EZ DNA Methylation Kit (Zymo Research; available from several suppliers).
4. Sterilized deionized distilled water ( $\text{ddH}_2\text{O}$ ) for PCR, primer dilution, and elution of bisulfite-converted DNA from column.



**Fig. 3** Example depicting methylation ratio determination in bisulfite pyrosequencing. During bisulfite conversion, unmethylated cytosine (C) is converted to uracil (U), whereas methylated cytosine (5mC) remains unchanged. In the subsequent polymerase chain reaction, U and 5mC are amplified as thymine (T) and C, respectively. Hence in the pyrogram, unmethylated and methylated Cs are represented with T and C, respectively. The methylated ratio of each CpG site is calculated automatically by PyroMark Q24 software with the following formula: (%) methylation = C peak height / (C peak height + T peak height)  $\times$  100. In the illustrated case, the methylation ratio is 70%

5. Single 0.2-mL PCR tube.
6. 1.5-mL microcentrifuge tube.
7. 20- $\mu$ L filter pipette tip.
8. 200- $\mu$ L filter pipette tip.
9. 1000- $\mu$ L filter pipette tip.
10. Vortex mixer.
11. Microcentrifuge.
12. Centrifuge.
13. PCR machine.

## 2.2 PCR

1. Ex Taq DNA Polymerase Hot Start Version (Takara Bio; available from several suppliers).
2. PCR primers for amplification of bisulfite-converted DNA, one biotinylated and purified via high performance liquid chromatography.
3. 1.5-mL microcentrifuge tube.
4. Single 0.2-mL PCR tube.
5. Sterilized ddH<sub>2</sub>O.
6. 10- $\mu$ L filter pipette tip.
7. 20- $\mu$ L filter pipette tip.
8. 200- $\mu$ L filter pipette tip.
9. 1000- $\mu$ L filter pipette tip.
10. Agarose.
11. Electrophoresis buffer: 40-mM Tris-acetate, 1-mM EDTA.
12. 100-bp DNA ladder.
13. Ethidium bromide.
14. Vortex mixer.
15. Microcentrifuge.
16. PCR machine.
17. Agarose gel electrophoresis device.
18. UV transilluminator.
19. PyroMark Assay Design Software 2.0 (QIAGEN).

## 2.3 Pyrosequencing

1. 10- $\mu$ L filter pipette tip.
2. 20- $\mu$ L filter pipette tip.
3. 200- $\mu$ L filter pipette tip.
4. 1000- $\mu$ L filter pipette tip.
5. 1.5-mL microcentrifuge tube.
6. 0.2-mL PCR eight-strip tubes with separate strip caps.
7. Streptavidin Sepharose High Performance (GE Healthcare).
8. PyroMark Binding Buffer (QIAGEN).
9. PyroMark Annealing Buffer (QIAGEN).
10. Sequencing primer.
11. Sterilized ddH<sub>2</sub>O.
12. ddH<sub>2</sub>O.
13. 70% ethanol.
14. 0.2-N NaOH.
15. PyroMark Wash Buffer, 10 $\times$  (QIAGEN).

16. PyroMark Q24 Plate (QIAGEN).
17. PyroMark Q24 Cartridge (QIAGEN).
18. PyroMark Gold Q24 Reagents (QIAGEN).
19. Vortex mixer.
20. Microcentrifuge.
21. PyroMark Q24 System (QIAGEN), with instruments and software for pyrosequencing.
22. Vortex mixer, with head attachment for 0.2-mL PCR eight-strip tubes.
23. Hot plate for heating PyroMark Q24 Plate with samples.
24. PyroMark Q24 Software v2.0.8 (QIAGEN).

---

## 3 Methods

### **3.1 Design and Validation of Primers for Pyrosequencing**

The design of primer pairs is the most important step of bisulfite PCR because the methylated and unmethylated molecules of the bisulfite-converted DNA template must be amplified equally; in other words, there should be no PCR bias. The primers can be designed using commercial or free software. PyroMark Assay Design Software 2.0, which is attached to the PyroMark Q24 System, designs the sequencing primer along with the bisulfite PCR primer. Alternatively, pyrosequencing primers reported previously by other researchers could be used. We describe how to use design software and how to validate the primers for pyrosequencing.

1. Launch PyroMark Assay Design Software 2.0.
2. Select “Methylation Analysis (CpG)” from the “Assay Type” drop-down menu.
3. Paste the target genomic DNA sequence in the “Upper Strand (5' to 3')” input field. The lower strand sequence of the pasted sequence is automatically shown in the “Lower Strand (5' to 3')” field. Choose the strand for analysis by clicking on it to add a check mark.
4. Click on the “Converted Sequence” tab to view the resulting bisulfite-converted sequence. The converted Cs at non-CpG sites and Cs at CpG sites are represented as orange Ts and bold Cs, respectively (Fig. 4).
5. Select a 10- to 30-bp sequence as the region of interest by dragging the cursor. Next, right-click the highlighted sequence and go to “Target Region > Set Target Region” (Fig. 4).
6. Click the large blue play button at the top of the software window. This will initiate the assay design program (Fig. 4).



The screenshot displays the PyroMark Assay Design Software 2.0 interface. The main window shows the 'Original Sequence Editor' with a DNA sequence. The 'Converted Sequence' tab is active, showing the bisulfite-converted sequence where cytosines at non-CpG sites are orange (Ts) and cytosines at CpG sites are bold. A region of interest is highlighted in blue. On the right, the 'Primer Set 1' panel displays candidate primer sets ranked by score. The top primer set (score 67) is selected, with the forward primer (F1) circled and an arrow pointing to it, indicating it is the biotinylated primer. Below the primer sets, the 'Variable Positions' section shows the CpG site being analyzed.

**Fig. 4** Primer design window of PyroMark Assay Design Software 2.0 with a selected Assay Type of “Methylation Analysis (CpG).” The bisulfite-converted sequence is shown by clicking on the “Converted Sequence” tab. The converted cytosines (Cs) at non-CpG sites are represented with an orange thymines (Ts), and remaining Cs at CpG sites are indicated with a bold Cs. The sequence region of interest is highlighted in blue and is set by dragging the mouse over the converted sequence. After clicking the large blue play button and finishing the analysis, the candidate primer sets containing the forward and reverse PCR primers and sequencing primer are displayed in a panel on the right side of the window. The arrow with a circle indicates the biotinylated primer. The primer sets are ranked in descending order according to their score. In the illustrated example, the top primer score is 67

7. After finishing the analysis, the candidate primer sets containing the forward and reverse PCR primers and sequencing primer are shown in a panel on the right side (Fig. 4). The primer sets are ranked in descending order according to their score from 0 to 100. Theoretically, the higher the score, the better the primer set. The one represented by an arrow with a circle is the biotinylated primer (*see Note 2*).
8. Upon clicking the “Graphic View” tab, the exact position of each primer in the sequence is shown. You can see a “Pyrosequencing Assay Design Analysis Report,” which includes various information about the primer sequence and its features, by selecting “View Report” after right-clicking on a primer set in the list. The report also contains a “Sequence to Analyze” section, which includes analyzed sequence information and is necessary for making a new CpG assay file in PyroMark Q24 Software v2.0.8. Therefore, it is important to save this report file.
9. Before ordering the biotinylated primer, we strongly recommend verifying whether your bisulfite PCR primer pair without the biotin modification produces a strong and specific bisulfite

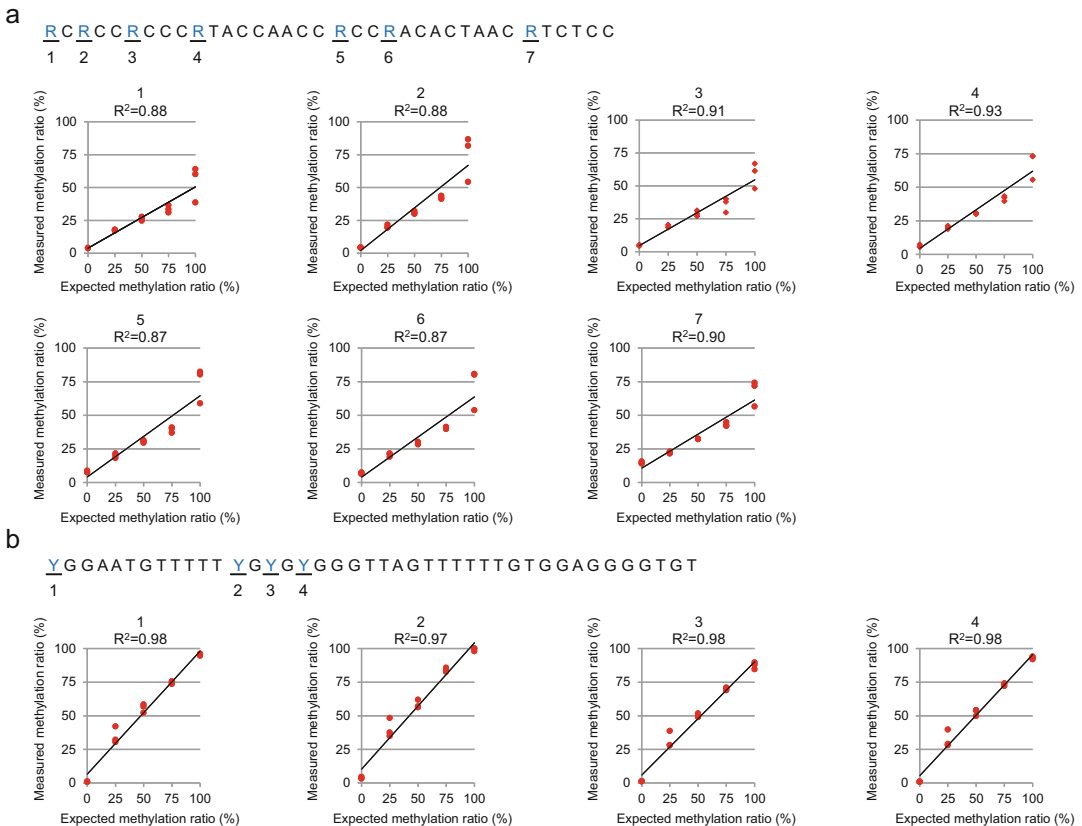
PCR product, which can help reduce costs and save time. Purchase the biotinylated primer and sequencing primer after determining the optimal PCR conditions.

10. To validate your primer set for pyrosequencing, prepare 0%, 25%, 50%, 75%, and 100% mixtures of methylated genomic DNA (500 ng) by mixing human methylated DNA (100% methylated DNA) and non-methylated DNA (0% methylated DNA; *see Note 3*). Then, perform the bisulfite conversion using this DNA.
11. Amplify bisulfite-converted DNA via PCR. After confirming the successful production of PCR products, perform the bisulfite pyrosequencing assay. The protocols for bisulfite PCR and pyrosequencing are described in the sections that follow.
12. Generate linear regression plots of each CpG site based on the pyrosequencing results and expected DNA methylation ratios. Verify that the measured methylation ratios correspond to the expected mixed methylation ratios and that the R-squared value is greater than 0.9. Use the CpG sites that meet these criteria. If most CpG sites do not fulfill the criteria, consider exploring an optimal annealing temperature that can overcome the PCR bias or redesign the primers (Fig. 5; *see Note 4*).

### **3.2 Bisulfite PCR for Pyrosequencing**

Bisulfite PCR is performed using one biotinylated primer and one non-biotinylated primer. A small amount of primer is used with a large number of cycles to deplete any free biotinylated primer that may work as an unexpected sequencing primer. A negative control is always necessary when a large number of cycles are performed.

1. Isolate genomic DNA from your samples, and perform bisulfite conversion on 500 ng of genomic DNA using the EZ DNA Methylation Kit according to the manufacturer's instructions, with two minor modifications. First, for the bisulfite reaction, use the following cycling protocol: 20 cycles of 95 °C for 30 s and 50 °C for 15 min; then leave at 4 °C. Second, elute the bisulfite-converted DNA from the column using 100 µL of water.
2. Amplify bisulfite-converted DNA via PCR. Prepare 15-µL PCR solution containing 7.5-ng bisulfite-converted DNA, 0.5 units of Ex Taq Hot Start polymerase, 1× Ex Taq buffer, 0.2-mM deoxynucleoside triphosphate, 0.2-µM biotinylated primer, and 0.2-µM non-biotinylated primer. Use the following reaction conditions: initial denaturation at 95 °C for 2 min; 45 cycles of denaturation at 95 °C for 30 s, annealing at optimized temperature for 30 s, elongation at 72 °C for 30 s, and a final extension at 72 °C for 10 min (*see Note 5*). In general, the same two PCR products are combined to make one sample because pyrosequencing requires many PCR



**Fig. 5** Graphs comparing the validation results of two pyrosequencing primer sets from a DMR in the *FAM20A* promoter. **(a)** The validation results of the first primer set with the analyzed sequence shown at the top. The primer set was designed on the reverse strand, so an R corresponds to a cytosine (C) at a CpG site and could indicate guanine (G) or adenine (A). The methylation ratios are calculated with the following formula: (%) methylation = G peak height / (G peak height + A peak height)  $\times$  100. At all CpG sites, there is a large difference between the methylation ratios measured through pyrosequencing and the expected methylation ratios based on the ratio of mixed unmethylated and methylated DNA. Furthermore, the R-squared values are below 0.9 at four of the seven examined CpG sites. Therefore, the primer should be redesigned. **(b)** Validation results of the second primer set, which was designed on the forward strand about 230 bp away from the first primer set. The analyzed sequence is shown at the top. The methylation ratios measured through pyrosequencing have a high correspondence with the expected methylation ratios based on the ratio of mixed unmethylated and methylated DNA. Moreover, the R-squared values are greater than 0.9 at all examined CpG sites. The second primer set can thus be used

products to achieve analyzable peak heights in the pyrogram. Using a low amount of PCR product results in a failed analysis with low pyrogram peaks.

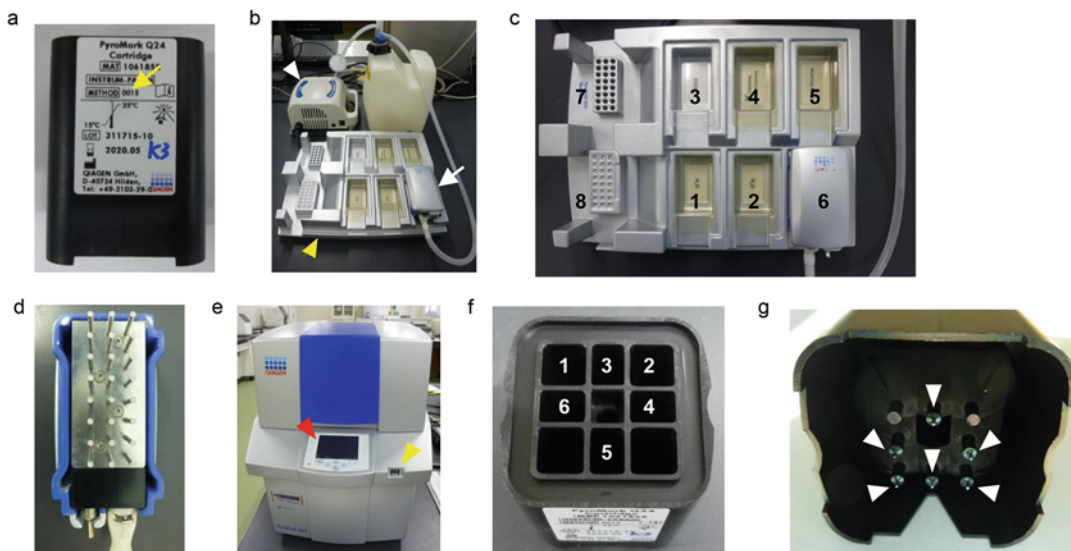
- To verify the specific amplification of the PCR products, run a 3- $\mu$ L aliquot of the PCR products with size markers on a 2% agarose gel in electrophoresis buffer. After electrophoresis, stain the gel with ethidium bromide solution and view on a UV transilluminator.

### 3.3 Bisulfite Pyrosequencing

Bisulfite pyrosequencing was performed using a PyroMark Q24 pyrosequencer according to the instructions included in the user manual.

#### 3.3.1 Pyrosequencer Preparation

1. To make a new CpG assay file, start PyroMark Q24 Software v2.0.8. Go to “File” > “New Assay” > “CpG Assay.” Paste the sequence information obtained from the primer design protocol using PyroMark Assay Design Software 2.0 into the “Sequence to Analyze” field. Click “Generate Dispensation Order.” If a bold orange T (or a bold orange A in a reverse assay) is visible in the histogram window, click it, and select either “Add Potential Bisulfite Treatment Control Before Dispensation” or “Add Potential Bisulfite Treatment Control After Dispensation.” C then may be added at the position adjacent to the bold orange T (or G in the case of a bold orange A) in the “Dispensation Order” and “Histogram” fields. The detection of an added C means bisulfite conversion failed, because the bold orange T does not represent a C at the CpG site in the genomic sequence. After adding bisulfite treatment control, select “Save as” from the “File” menu, and save the CpG assay file onto a USB device under a new file name.
2. To run the file, start PyroMark Q24 Software v2.0.8 on a computer that recognizes the USB device. Select “New Run” from the “File” menu, and next to “Instrument Method,” and choose the method number written on the PyroMark Q24 Cartridge (Fig. 6a; *see Note 6*). Add your CpG assay file to the shortcut browser by clicking “Add File Shortcut” or “Add Folder Shortcut.” Select the suitable CpG assay file in the shortcut browser, and drag the assay file to each well in the “Plate Setup” panel. Enter a sample name in the “Sample ID” column. Finally, save the run file onto the USB device under a new file name.
3. Put water, 70% ethanol, 0.2-N NaOH (denaturation buffer), and 1 × PyroMark Wash Buffer into the appropriate plastic trays on the PyroMark Q24 Vacuum Workstation (Fig. 6b, c; *see Note 7*). Turn on the vacuum pump and apply the vacuum to the Vacuum Prep Tool by opening the vacuum switch (Fig. 6b). Wash the filter probes of the Vacuum Prep Tool twice in two water trays by shaking for 10 s (Fig. 6c, d). Then, turn off the vacuum pump and place the Vacuum Prep Tool on the bench with the probes facing up. Refill the water in the water trays.
4. Turn on the PyroMark Q24 machine (Fig. 6c) and set the hot plate to 80 °C.



**Fig. 6** Photos of the PyroMark Q24 Cartridge and PyroMark Q24 System. **(a)** The front of the cartridge. The yellow arrow indicates the method number. **(b)** PyroMark Q24 Vacuum Workstation setup. The white and yellow arrowheads point to the vacuum pump and the PyroMark Q24 Vacuum Workstation, respectively. The white arrow points to the Vacuum Prep Tool. **(c)** Top view of the PyroMark Q24 Vacuum Workstation. The following solutions are added into the trays at the position with the corresponding number: positions 1 and 2, water; position 3, 70% ethanol; position 4, 0.2-N NaOH; position 5, 1× PyroMark Wash Buffer. Position 6 is “Parking” position of the Vacuum Prep Tool, and positions 7 and 8 are places to set the 0.2-mL PCR eight-strip tubes and the PyroMark Q24 Plate, respectively. **(d)** The bottom of the Vacuum Prep Tool with 24 filter probes. **(e)** PyroMark Q24 machine. The red and yellow arrowheads indicate the screen and the USB port, respectively. **(f)** The cartridge as seen from above. The following solutions are added into each compartment at the position with the corresponding number: position 1, enzyme; position 2, substrate; position 3, dATP $\alpha$ S (analogue of dATP); position 4, dCTP; position 5, dGTP; position 6, dTTP. **(g)** The bottom of the cartridge. The six white arrowheads indicate the dispensing needles. They are fragile, so touch them as little as possible

### 3.3.2 Preparation of Pyrosequencing Template

1. Add water to the bisulfite PCR product so that the total volume is 38  $\mu$ L. Transfer them into 0.2-mL PCR eight-strip tubes according to your run file.
2. Add 40- $\mu$ L PyroMark Binding Buffer and 1- $\mu$ L streptavidin beads per tube. Agitate them immediately with a vortex mixer for 10 min to allow the streptavidin beads to bind the biotinylated strand of the PCR product.
3. Add 25  $\mu$ L of 0.3- $\mu$ M pyrosequencing primer in PyroMark Annealing Buffer to each well of the PyroMark Q24 Plate according to your run file.
4. After agitating for 10 min, stop the vortex mixer and turn on the vacuum pump. Immediately transfer the 0.2-mL PCR eight-strip tubes into the worktable of the PyroMark Q24 Vacuum Workstation (Fig. 6c). Suck up the samples from the 0.2-mL PCR eight-strip tubes with the filter probes of the

Vacuum Prep Tool (Fig. 6d). If you take too long, the beads will precipitate into the bottom of the tubes, resulting in a lower PCR product recovery rate.

5. Immerse the filter probes of the Vacuum Prep Tool in a tray filled with 70% ethanol for 5 s.
6. Immerse the filter probes of the Vacuum Prep Tool in a tray filled with 0.2-N NaOH (denaturation buffer) for 5 s. The PCR product is then denatured and the biotinylated strand, which is the pyrosequencing template, remains on the beads.
7. Immerse the filter probes of the Vacuum Prep Tool in a tray filled with 1× PyroMark Wash Buffer for 10 s.
8. Place the Vacuum Prep Tool with the probes facing up and turn it and the vacuum pump off.
9. Set the PyroMark Q24 Plate in the worktable of the PyroMark Q24 Vacuum Workstation (Fig. 6c). Place the probes of the Vacuum Prep Tool on the wells of the PyroMark Q24 Plate in the correct orientation. Then, agitate the Vacuum Prep Tool to drop the beads from the probes into the wells containing the sequencing primer.
10. Transfer the Vacuum Prep Tool into the tray filled with water, and turn it and the vacuum pump on to wash the probes twice in two water trays by shaking for 10 s. Then turn off the vacuum pump and the Vacuum Prep Tool. Put the Vacuum Prep Tool in the “Parking” position of the PyroMark Q24 Vacuum Workstation (Fig. 6c).
11. To remove secondary structures in the biotinylated template strand that may impede primer annealing and extension, put the PyroMark Q24 Plate on the 80 °C hot plate for 2 min.
12. Remove the PyroMark Q24 Plate from the hot plate and cool it at room temperature for 5 min. The pyrosequencing primer will thus anneal to the biotinylated template strand.

### 3.3.3 Pyrosequencing Reaction

1. Open your run file and click “Pre Run Information” from the “Tools” menu. The required volumes of reagents, including the enzymes, substrates, and nucleotides from the PyroMark Gold Q24 Reagents box, will be shown in the window. Add each reagent to the specified compartment in the PyroMark Q24 Cartridge (Fig. 6f).
2. Open the lid of the PyroMark Q24 machine (Fig. 6e), and set the cartridge and above the prepared plate before closing the lid.
3. Insert the USB device containing your run file into the USB port of the PyroMark Q24 machine (Fig. 6e).
4. Select “Run” in the main menu onscreen and press “OK” (Fig. 6e).



5. Select your run file and press “Select.” Press “Yes” to start the run.
6. While the pyrosequencer is running, you can see the pyrogram peaks in real time on the screen.

### 3.3.4 Pyrosequencer Cleanup

1. Immediately after the run is over, clean the cartridge before the dispensing needles are clogged by salt formation (Fig. 6g). Completely fill the compartments of the cartridge with water. Sufficiently discharge the water from the dispensing needles by pressing firmly on the top of each compartment with a finger while wearing powder-free gloves. After discarding the water, dry the cartridge on lint-free paper. Store the cartridge in a plastic box with a lid to keep dust and light out (*see Note 8*).
2. Discard the used PyroMark Q24 Plate.
3. Discard the contents of the trays on the PyroMark Q24 Vacuum Workstation and clean them by rinsing with water. Dry them at room temperature.

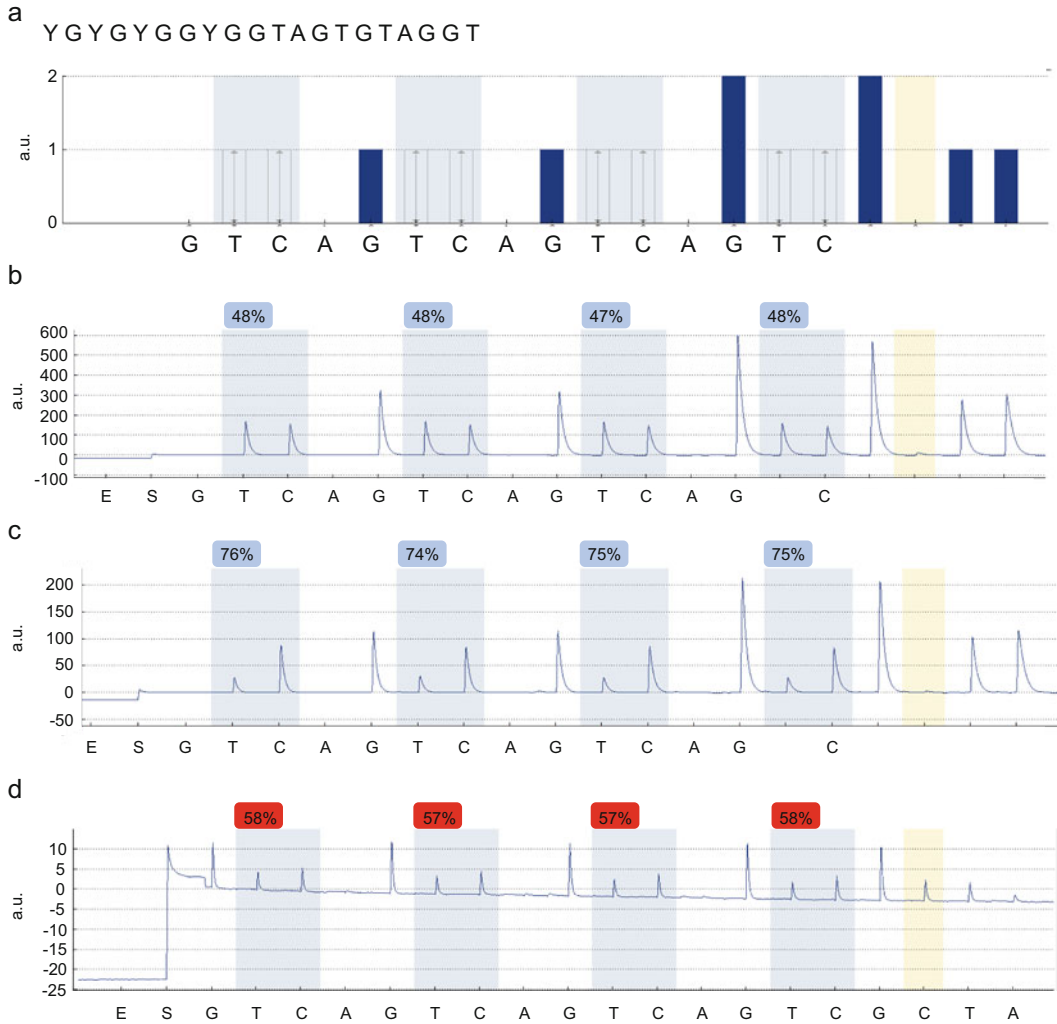
### 3.3.5 Data Analysis

1. After the run is finished, the pyrosequencing results are automatically recorded in your run file on the USB device. Open your run file on the computer with PyroMark Q24 software.
2. To analyze the results, click “Analyze All Wells” from the “Analyze” menu at the top right of the window. The analysis presents the methylation ratio (%) at each CpG site and the quality assessment based on peak height, deviations from the reference sequence pattern, and unexpected peak heights (Fig. 7a–d). The quality assessment at each CpG site is coded by color: blue indicates high sequencing quality (Fig. 7b, c); yellow indicates some unexpected deviations, resulting in intermediate sequencing quality; and red represents failure of the pyrosequencing analysis (Fig. 7d). Generally, methylation values in blue and yellow can be used in subsequent analysis.
3. For subsequent analysis, select “CpG Analysis Results” from the “Reports” menu and click “Preview.” Copy the CpG analysis results, including methylation ratios, quality assessment, and any applicable warnings at each CpG site, and paste them into a spreadsheet file.

---

## 4 Notes

1. The human non-methylated DNA was purified from the human colorectal cancer HCT116 cell line with two double knockout (DKO) DNA methyltransferase genes, *DNMT1* and *DNMT3b*. The DNA derived from HCT116 DKO cells has a low DNA methylation level, so it can be used as an



**Fig. 7** Representative pyrograms from a patient with Beckwith–Wiedemann syndrome with hypermethylation at the *H19*-DMR and a normal individual. **(a)** Predictive pyrogram generated by PyroMark Q24 software based on the analyzed *H19*-DMR sequence. The sequence indicated at the top is the analyzed sequence; Y, which corresponds to cytosine (C) at CpG sites, indicates thymine (T) or C. The purple and yellow shaded areas indicate the positions of Cs at CpG sites and the bisulfite treatment control, respectively. The *y*-axis shows the intensity of light in arbitrary units (a.u.), while the *x*-axis represents the dispensation order of nucleotides. **(b)** Pyrogram analyzed using DNA from the peripheral blood of a normal individual. In theory, the methylation ratio (%) at each CpG site within the *H19*-DMR should be 50%, because the DMR is methylated on the paternal allele and unmethylated on the maternal allele in somatic cells. Indeed, all methylation ratios are almost 50% in the normal individual. **(c)** Pyrogram analyzed using DNA from the peripheral blood of a Beckwith–Wiedemann syndrome patient with hypermethylation at *H19*-DMR. The methylation ratios range from 74% to 76%, indicating increments of methylation, or hypermethylation at *H19*-DMR. **(d)** An example of a failed pyrosequencing analysis. The main reasons for this failure were incomplete bisulfite conversion, as indicated by a peak in the yellow shaded area and low peak heights



unmethylated control. The human methylated DNA was purified after treating HCT116 DKO DNA with *M.SssI* methyltransferase, which catalyzes the methylation of all cytosines at CpG sites. This DNA can thus be used as a methylated control.

2. A score above 60 and a PCR product size of less than 250 bp are preferable. When a desired primer set cannot be obtained, check the reverse strand.
3. Methylated mouse DNA is commercially available, whereas non-methylated mouse DNA is not. If non-methylation within the target DNA region can be confirmed by bisulfite sequencing, the DNA can be used as non-methylated DNA. However, if the target region is methylated, non-methylated and methylated DNA must be made as follows: (1) amplify the whole genome from a small amount of your DNA using an appropriate kit. Since your DNA will be amplified more than 1000-fold and the amplified DNA is non-methylated, the purified whole genome amplified DNA can be used as an unmethylated control; (2) methylate half of the purified whole genome amplified DNA with *M.SssI* methyltransferase. The treated and purified whole genome amplified DNA can then be used as a methylated control. At this point, avoid using commercially available methylated DNA because there is a qualitative difference in the length of genomic DNA and whole genome amplified DNA.
4. Figure 5a, b compares two examples of validation results, one indicating the need to redesign the primers and one indicating that the bisulfite pyrosequencing works well.
5. To avoid PCR contamination, always keep the laboratory bench clean and wear fresh gloves. To prevent aerosol contamination, use filtered tips.
6. To add a new instrument method to the “Instrument Method” menu in the PyroMark Q24 software, find the instrument method file on the QIAGEN website that corresponds to the method number written on the PyroMark Q24 Cartridge. Download and save it on the computer that has PyroMark Q24 software installed. Open the software and click “Instrument Methods” from the “Tools” menu. In the “Instrument Methods” dialog box, click “Import.” Select the downloaded instrument method file and click “Open.” This sets up a new instrument method option in the PyroMark Q24 software.
7. Add each solution until slightly above the inner line of the tray.
8. Avoid touching the dispensing needles at the bottom of the cartridge as much as possible due to their fragility (Fig. 6g).

## Acknowledgments

This study was supported in part by the Japan Society for the Promotion of Science, KAKENHI grant numbers JP21K19451 (HS), JP20H03643 (HS), JP20K08183 (KH), and JP19K06451 (SH); the Japan Agency for Medical Research and Development, grant numbers JP20ek0109486 and JP20ek0109489 (HS); and the Ministry of Health, Labor and Welfare Program, grant number JP20FC1046 (HS).

## References

- Gibney ER, Nolan CM (2010) Epigenetics and gene expression. *Heredity (Edinb)* 105(1): 4–13. <https://doi.org/10.1038/hdy.2010.54>
- Hanna CW, Demond H, Kelsey G (2018) Epigenetic regulation in development: is the mouse a good model for the human? *Hum Reprod Update* 24(5):556–576. <https://doi.org/10.1093/humupd/dmy021>
- Fang H, Distèche CM, Berletch JB (2019) X inactivation and escape: epigenetic and structural features. *Front Cell Dev Biol* 7:219. <https://doi.org/10.3389/fcell.2019.00219>
- Pappalardo XG, Barra V (2021) Losing DNA methylation at repetitive elements and breaking bad. *Epigenetics Chromatin* 14(1):25. <https://doi.org/10.1186/s13072-021-00400-z>
- Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* 33(Suppl):245–254. <https://doi.org/10.1038/ng1089>
- Higashimoto K, Joh K, Soejima H (2017) Genomic imprinting syndromes and cancer. In: Kaneda A, Tsukada Y (eds) DNA and histone methylation as cancer targets. *Cancer drug discovery and development*. Humana Press, New York, pp 297–344. <https://doi.org/10.1007/978-3-319-59786-7>
- Ortiz-Barahona V, Joshi RS, Esteller M (2020) Use of DNA methylation profiling in translational oncology. *Semin Cancer Biol*. <https://doi.org/10.1016/j.semcancer.2020.12.011>
- Jin Z, Liu Y (2018) DNA methylation in human diseases. *Genes Dis* 5(1):1–8. <https://doi.org/10.1016/j.gendis.2018.01.002>
- Li Y (2021) Modern epigenetics methods in biological research. *Methods* 187:104–113. <https://doi.org/10.1016/j.ymeth.2020.06.022>
- Sun Z, Cunningham J, Slager S, Kocher JP (2015) Base resolution methylome profiling: considerations in platform selection, data pre-processing and analysis. *Epigenomics* 7(5): 813–828. <https://doi.org/10.2217/epi.15.21>
- Poulin M, Zhou JY, Yan L, Shioda T (2018) Pyrosequencing methylation analysis. *Methods Mol Biol* 1856:283–296. [https://doi.org/10.1007/978-1-4939-8751-1\\_17](https://doi.org/10.1007/978-1-4939-8751-1_17)
- Tost J, Gut IG (2007) DNA methylation analysis by pyrosequencing. *Nat Protoc* 2(9): 2265–2275. <https://doi.org/10.1038/nprot.2007.314>
- Delaney C, Garg SK, Yung R (2015) Analysis of DNA methylation by pyrosequencing. *Methods Mol Biol* 1343:249–264. [https://doi.org/10.1007/978-1-4939-2963-4\\_19](https://doi.org/10.1007/978-1-4939-2963-4_19)
- Sulewska A, Niklinska W, Kozłowski M, Minarowski L, Naumnik W, Niklinski J, Dabrowska K, Chyczewski L (2007) Detection of DNA methylation in eucaryotic cells. *Folia Histochem Cytobiol* 45(4):315–324
- Kristensen LS, Hansen LL (2009) PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. *Clin Chem* 55(8):1471–1483. <https://doi.org/10.1373/clinchem.2008.121962>



## Post-bisulfite Adaptor Tagging Based on an ssDNA Ligation Technique (tPBAT)

Fumihito Miura, Yukiko Shibata, Miki Miura, and Takashi Ito

### Abstract

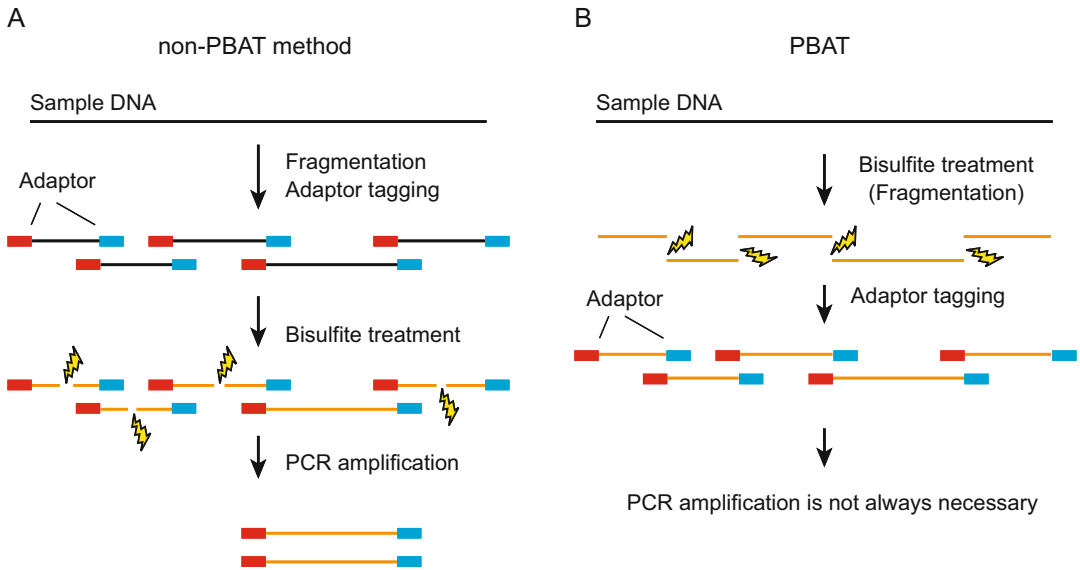
Post-bisulfite adaptor tagging (PBAT) is a concept that enables the preparation of an efficient sequencing library from bisulfite-treated DNA, and it also means the protocol implemented the concept. Although the previous PBAT or rPBAT was sensitive enough for single-cell methylome analysis, the protocol had several drawbacks owing to the repeated random priming reactions. To resolve these problems, we developed a unique single-strand DNA ligation technique, termed TACS ligation, and established a new protocol called tPBAT. With tPBAT, the data quality improved, with a longer insert and higher mapping rate than that obtained with rPBAT. In addition, paired-end sequencing and indexing were supported by the default. In this chapter, the tPBAT protocol is introduced, and a thorough description of its application to small samples is provided.

**Key words** Whole-genome bisulfite sequencing (WGBS), Methylome, Post-bisulfite adaptor tagging (PBAT), DNA methylation, Single-stranded DNA ligation, TACS ligation

---

## 1 Introduction

Currently, three major procedures are available for the measurement of 5-methylation levels of cytosines at the whole-genome scale and single-nucleotide resolution: whole-genome bisulfite sequencing (WGBS), which involves shotgun sequencing of bisulfite-treated DNA [1–6]; enzymatic methyl-seq (EM-seq), where the bisulfite reaction is replaced with mild enzymatic reactions [7, 8]; and direct detection of methylation on a single DNA molecule with a nanopore sequencer [9, 10]. When a methylome analysis is being performed with a limited amount of samples, amplification of the libraries is unavoidable. Because DNA polymerase cannot induce DNA methylation to the nascent strand, amplification causes a loss of DNA modification. Therefore, direct detection with nanopore sequencing is not suitable for samples that require amplification. Accordingly, the current options for

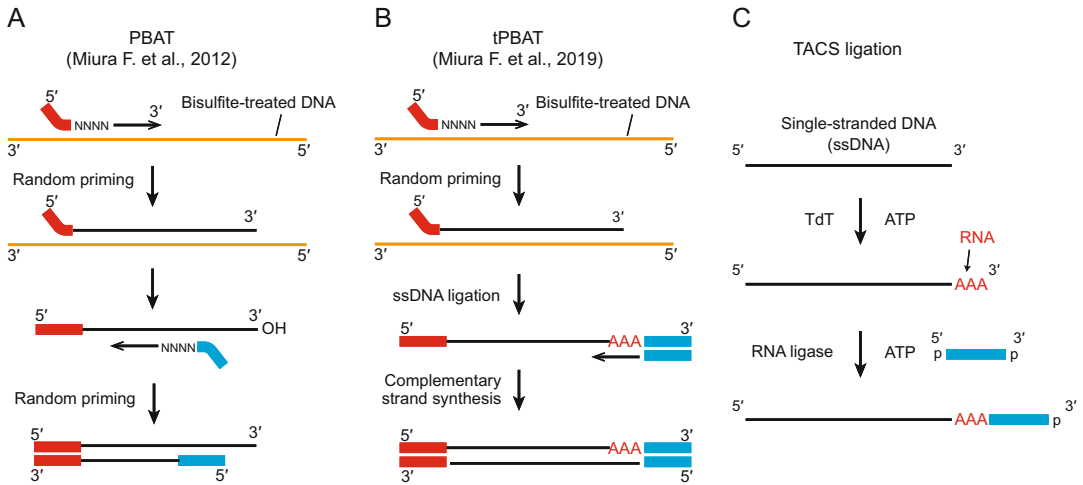


**Fig. 1** The principle of post-bisulfite adaptor tagging (PBAT). (a) Procedures before PBAT. Bisulfite treatment (BS) causes a fatal loss of library structure. (b) Adaptor tagging after BS can avoid the loss of library structure

methylome analysis of samples of limited amounts are restricted to WGBS and EM-seq. This chapter focuses on WGBS, especially for the practical application of small amounts of samples.

Owing to the destructive nature of bisulfite treatment, bisulfite-based methylome analysis would require additional considerations when preparing the sequencing library. Post-bisulfite adaptor tagging (PBAT) is a concept that enables highly efficient library preparation for shotgun bisulfite sequencing [4]. As shown in Fig. 1a, if the adaptor tagging preceded bisulfite treatment, the adaptor-tagged library molecules would be degraded during the bisulfite reaction. However, if adaptor tagging followed bisulfite treatment, such phenomena would not occur (Fig. 1b), which is the concept of PBAT. An issue must be considered when implementing PBAT. Conventional library preparation protocols adapted for double-stranded DNA (dsDNA) cannot be used for bisulfite-treated DNA (bsDNA) as it is a single-stranded DNA (ssDNA).

The initial version of the PBAT-based library preparation protocol, now called rPBAT, was implemented by employing two continuous random priming reactions [4]. As shown in Fig. 2a, the first random priming reaction was conducted using bisulfite-treated DNA as a template. The second reaction was performed using the synthesized strand as a template. Finally, a DNA molecule with two different adaptor sequences at both ends was generated. This random priming-based strategy worked well and was applied for single-cell methylome analysis by others [11]. However, random priming causes (1) shortening of the insert size (Fig. 3a);



**Fig. 2** Two PBAT protocols and TACS ligation. (a) Random priming-based PBAT (rPBAT). (b) Improved PBAT protocol (tPBAT). (c) The principle of TACS ligation

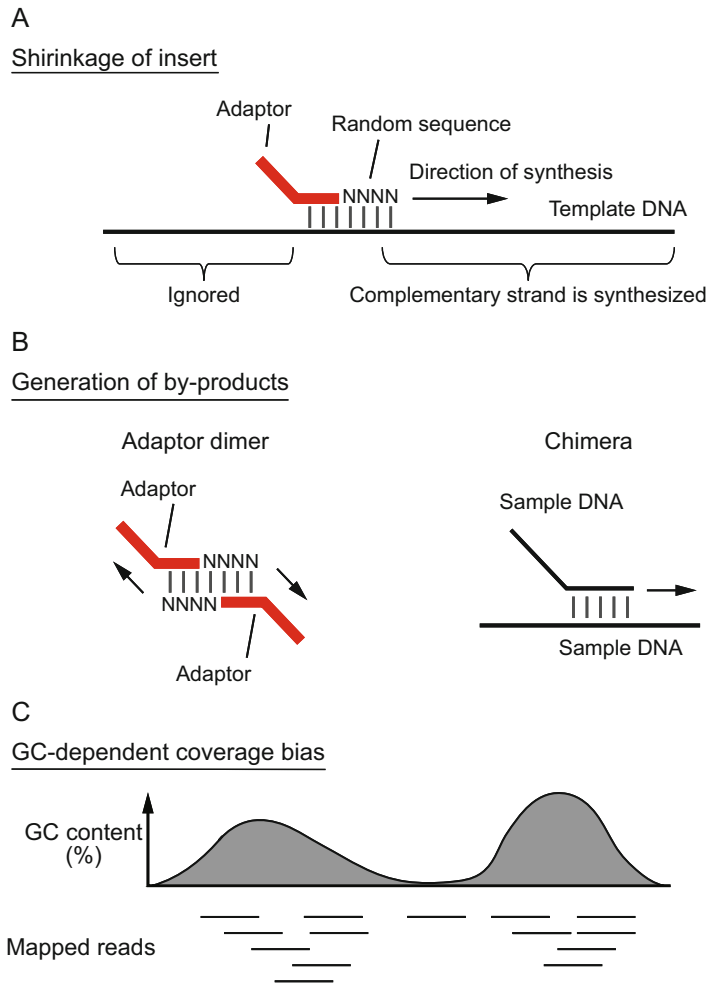
(2) production of byproducts, such as adaptor dimer and chimeric DNA (Fig. 3b); and (3) bias in mapped read coverage dependent on the GC content (Fig. 3c).

To overcome these drawbacks, we developed a new adaptor ligation technique for ssDNA, called TACS ligation (Fig. 2c), and replaced one of the two random priming reactions with TACS ligation to establish a revised PBAT protocol, termed tPBAT (Fig. 2b) [12]. This development resolved two drawbacks of rPBAT: short insert and low mapping rate (Fig. 3a, b). Accordingly, tPBAT enabled the collection of high-quality methylome data. Although the output of reads obtained from the unit amount of starting DNA did not change between rPBAT and tPBAT, tPBAT was adapted for paired-end and index sequencing by default [12].

Our original report of tPBAT provided a detailed protocol as a supplementary material. Therefore, tPBAT library preparation can be performed using this protocol. However, after developing tPBAT, we performed several methylome analyses using samples of limited amounts [13, 14]. To prepare such libraries, additional guidance had to be employed. We thought that describing these ingenuities would benefit readers that opt to conduct new methylome analysis on samples of limited amounts. Therefore, this chapter describes the tPBAT protocol and provides a thorough description of practical library preparation using samples of limited amounts.

### 1.1 Selection of the DNA Purification Method for tPBAT Library Preparation

For genomic DNA purification, commercially available kits for silica column purification can be used with the tPBAT protocol. As DNA extracted with non-silica-based procedures, such as Isogen (Nippon Gene) and Dr. GentLE (Takara Bio), occasionally led to



**Fig. 3** Three drawbacks of random priming. **(a)** Random priming causes the shrinkage of DNA information. **(b)** Random priming produces byproducts, such as adaptor dimer and chimeras. **(c)** Random priming is intrinsically biased to GC-rich sequences

issues, we do not recommend the use of these kits with the tPBAT protocol. Routinely, we use the DNeasy Blood & Tissue Kit (Qiagen), which enables reproducible results. Of note, because the inclusion of excessive amounts of RNA in the bisulfite reaction causes low library yields with the tPBAT protocol, RNA removal is strongly recommended for tissues that contain high levels of RNA, such as the liver and pancreas. The DNeasy Blood & Tissue Kit provides an easy-to-perform procedure for removing RNA, which involves a 5-min incubation at room temperature with RNase A after the protease digestion of cells.

DNA purification is not always necessary for samples with a limited number of cells. Starting from less than 10,000 cells, the cell lysate can be directly used for bisulfite treatment. For example, Shirane et al. lysed 2000–4000 cells in a 0.1% SDS solution containing proteinase K and RNase A and directly mixed the lysate with bisulfite reagent [13].

### **1.2 Sequencing Libraries of Highly Biased Nucleotide Composition**

As bisulfite treatment converts most cytosines into uracil, the WGBS read contains only few Cs. This highly biased nucleotide composition affects the base-calling of the Illumina sequencer and occasionally results in poor sequencing results [15]. Therefore, to balance the four nucleobases, Illumina recommends spiking libraries of normal nucleotide composition into the WGBS library, which should account for at least 20% of the total input. The PhiX control is widely employed for this purpose. However, the inclusion of PhiX at such high concentration inevitably results in a low sequencing cost. Two solutions have been proposed to resolve this problem.

The first solution involves spiking the high (G + C) content library. Accordingly, the (G + C) content of the PhiX control was 44%, whereas that of *K. radiotolerans* was 74%. Improved base-calling and sequencing costs were achieved by spiking a high (G + C) content library prepared from *K. radiotolerans* genomic DNA at 5% of the total reads [16]. The second solution involves mixing libraries in two different directions. Most nucleobases in bisulfite-treated DNA are A, G, and T, whereas A, C, and T are for the opposite strand. If these two strands are mixed, C can be supplemented. Although the labor for library preparation doubles with this strategy, the sequencing power can be fully exploited [12].

### **1.3 Amplification of the Sequencing Library**

The high yields of the tPBAT protocol enable amplification-free library preparation, and 50-ng or more input DNA is required for the amplification-free procedure. However, if the DNA amount is limited, whole-library amplification is required. In such cases, one should consider overamplification of the library. The library-to-read conversion rate of the current Illumina sequencing is not low. In fact, in our experiment, nearly half of the library molecules was converted to sequenced reads at maximum [12]. In addition, extensive PCR amplification results in biased coverage of the genome. Therefore, the number of PCR cycles should be carefully determined to avoid repeated sequencing of duplicate reads. In our cases, six or fewer cycles were performed, even for precious samples.

## 2 Materials

### 2.1 Commercial Kits, Enzymes, and Consumables

1. EZ DNA Methylation-Gold Kit (Zymo Research).
2. 50,000 units/mL of Klenow Fragment (3'-5' exo-) (New England Biolabs) (*see Note 1*).
3. Terminal deoxynucleotidyl transferase (TdT).
4. CircLigase II (Epicentre).
5. 20-mg/mL Proteinase K (Qiagen).
6. 10-mM rATP.
7. Unmethylated lambda DNA (Promega).
8. AMPure XP (Beckman Coulter).
9. Qubit ssDNA Assay Kit (Thermo Fisher Scientific).
10. Qubit dsDNA Assay HS Kit (Thermo Fisher Scientific).
11. Library quantitation kit (Takara Bio).
12. PrimeSTAR Max DNA polymerase (Takara Bio).

### 2.2 Solutions

1. CT Conversion Reagent: Add 900- $\mu$ L water, 300  $\mu$ L of M-Dilution Buffer, and 50- $\mu$ L M-Dissolving Buffer to a tube containing CT Conversion Reagent. Completely dissolve the CT Conversion Reagent powder via mixing at room temperature for at least 10 min. Prepare immediately before use.
2. Size cutoff solutions: 18%, 19%, and 20% PEG 400 (for 250-bp, 300-bp, and 400-bp cutoff, respectively), 10-mM Tris-HCl (pH 8.0), and 1-M NaCl (*see Note 2*).
3. Buffer B2: 3-M guanidine hydrochloride, 20% (v/v) Tween 20.
4. 2.5 $\times$  TACS buffer: 125-mM HEPES-KOH, pH 7.5, 12.5-mM MgCl<sub>2</sub>, 1.25% (v/v) Triton X-100, 50% (w/v) PEG 6000 (*see Note 3*).
5. dNTP solution (2.5 mM each of dATP, dCTP, dGTP, and dTTP).
6. Denaturing loading dye: 95% (v/v) formamide and 0.01% (w/v) bromophenol blue.

### 2.3 Oligonucleotides (OPC Grade)

1. PEA2-N4: 5'-ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT NNN N-3'.
2. PA-anti-PEA1-P: 5'-phosphate-AGA TCG GAA GAG CAC ACG TCT GAA CTC CAG TCA C-phosphate-3'.
3. Primer-3: 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACA CAC TCT TTC CCT ACA CGA CGC TCT TCC GAT CT-3'.



**Table 1**  
**Index numbers and index sequences**

Index number	Index sequence	Index number	Index sequence	Index number	Index sequence
1	CGTGAT	9	CTGATC	18	GCGGAC
2	ACATCG	10	AAGCTA	19	TTTCAC
3	GCCTAA	11	GTAGCC	20	GGCCAC
4	TGGTCA	12	TACAAG	21	CGAAAC
5	CACTGT	13	TTGACT	22	CGTACG
6	ATTGGC	14	GGAACT	23	CCACTC
7	GATCTG	15	TGACAT	25	ATCAGT
8	TCAAGT	16	GGACGG	27	AGGAAT

4. Indexing primer: 5'-CAA GCA GAA GAC GGC ATA CGA GAT [index sequence; select one of the index sequences from Table 1] GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC-3' (*see Note 4*).
5. Amp-F: 5'-AATGATACGGCGACCACCGAGATCTACAC-3'.
6. Amp-R: 5'-CAAGCAGAAGACGGCATAACGAGAT-3'.

### 3 Methods

#### 3.1 Bisulfite Treatment

This procedure is based on the manual of the EZ Methylation Gold Kit. The manual recommends up to 500 ng of DNA per reaction to obtain optimum results. The volume of the bisulfite reaction can be increased or decreased according to the purpose. For example, when we started with diluted DNA and used the same concentration of bisulfite reagent, we increased the reaction volume depending on the DNA volume for the required mass amount and carried out a thermal reaction by splitting the reaction into several tubes. Finally, the reaction mixtures in the tubes were collected in a single silica column. Using this procedure, condensation can be easily achieved. As 100-ng DNA is usually sufficient for amplification-free library preparation for more than 30× coverage of the mammalian genome, the reaction volume can be reduced to at least 1/5 of the recommended volume. We used a 1/3 volume reaction for our routine use.

1. Mix the following in a tube:

CT conversion reagent	130 $\mu$ L
DNA	100 ng ( <i>see Note 5</i> )

(continued)

Unmethylated lambda DNA	1% ( <i>see Note 6</i> )
Water	To 150 $\mu\text{L}$

- Incubate the tubes at 98 °C for 10 min and at 64 °C for 150 min.
- Add 600  $\mu\text{L}$  of M-Binding Buffer to the reaction and mix well.
- Load the entire solution onto the ZymoSpin column and spin at 10,000  $\times g$  for 1 min.
- Discard the flowthrough.
- Wash the column with 200  $\mu\text{L}$  of M-Wash.
- Add M-Desulfonation buffer (200  $\mu\text{L}$ ) to the column and incubate at room temperature for 15 min. Spin the column to remove M-Desulfonation buffer.
- Wash the column with 200  $\mu\text{L}$  of M-Wash.
- Repeat the previous wash step.
- Transfer the column to a new tube.
- Add 20–40  $\mu\text{L}$  of M-Elution buffer directly to the matrix of the column.
- Centrifuge for 30 s at full speed to elute the DNA (*see Note 7*).
- Use 1- $\mu\text{L}$  eluent to measure the amount of DNA using the Qubit ssDNA Assay Kit (*see Note 8*).

### 3.2 Random Priming

- Mix the following reagents in a PCR tube:

Bisulfite-treated DNA and Milli-Q water	40 $\mu\text{L}$ in total
10 $\times$ NEBuffer 2	5 $\mu\text{L}$
dNTP solution	4 $\mu\text{L}$
100- $\mu\text{M}$ PEA2-N4	1 $\mu\text{L}$

- Incubate at 95 °C for 3 min and 4 °C for 5 min.
- Add the following reagent to the reaction:

Klenow Fragment (3'-5' exo-)	1 $\mu\text{L}$
------------------------------	-----------------

- Incubate the mixture at 4 °C for 15 min; the temperature increases at a rate of +1 °C/min to 37 °C. After maintenance of the reaction at 37 °C for 30 min, inactivate the enzyme via heating at 70 °C for 10 min (*see Note 9*).
- Add 50- $\mu\text{L}$  AMPure XP to the reaction, hold the tube stand at room temperature for 5 min, place the tube on a magnetic stand to collect the beads, and remove the supernatant.

6. Add 200  $\mu\text{L}$  of either size cutoff solution, resuspend the beads, place the tube on a magnetic stand, and remove the supernatant (*see Note 10*).
7. Repeat the previous step once.
8. Rinse the beads with 200- $\mu\text{L}$  70% (v/v) ethanol, and elute the DNA with 12  $\mu\text{L}$  of 10-mM Tris-HCl, pH 8.5.
9. Use 1  $\mu\text{L}$  of DNA to measure the DNA amount with the Qubit dsDNA Assay HS Kit (*see Note 11*).

### 3.3 TACS Ligation

1. Mix the following in a PCR tube:

2.5 $\times$ TACS reaction buffer	10 $\mu\text{L}$
Purified DNA from the previous step	11 $\mu\text{L}$
30- $\mu\text{M}$ PA-anti-PEA1-P	1 $\mu\text{L}$
10-mM ATP	1 $\mu\text{L}$

2. Incubate at 95  $^{\circ}\text{C}$  for 5 min and 4  $^{\circ}\text{C}$  for 5 min.
3. Add the following to the tube:

TdT (15 U/ $\mu\text{L}$ )	1 $\mu\text{L}$
CircLigase II (100 U/ $\mu\text{L}$ )	1 $\mu\text{L}$

4. Incubate at 37  $^{\circ}\text{C}$  for 30 min, 65  $^{\circ}\text{C}$  for 120 min, and 95  $^{\circ}\text{C}$  for 5 min.

### 3.4 Primer Extension (1)

1. Mix the following reagents in a PCR tube:

The reaction after TACS ligation	25 $\mu\text{L}$
ddH <sub>2</sub> O	14 $\mu\text{L}$
10 $\times$ Gene Taq Universal Buffer	5 $\mu\text{L}$
2.5-mM dNTPs	4 $\mu\text{L}$
60- $\mu\text{M}$ indexing primer	1 $\mu\text{L}$
Hot Start GeneTaq (2.5 U/ $\mu\text{L}$ )	1 $\mu\text{L}$

2. Incubate at 95  $^{\circ}\text{C}$  for 3 min, 45  $^{\circ}\text{C}$  for 3 min, and 72  $^{\circ}\text{C}$  for 30 min.
3. Mix the following reagents in a PCR tube:

The reaction after the previous step	50 $\mu\text{L}$
Buffer B2	20 $\mu\text{L}$
Protease K (20 mg/mL)	5 $\mu\text{L}$

4. Incubate at 50 °C for 15 min.
5. Add 50- $\mu$ L AMPure XP and incubate for 5 min at room temperature.
6. Place the tube on a magnetic stand, wait until the beads are collected, and remove the supernatant.
7. Wash beads with 200  $\mu$ L of either size cutoff solution (*see Note 10*).
8. Repeat the previous step once.
9. Rinse beads with 200- $\mu$ L 70% (v/v) ethanol.
10. Remove the residual solution completely.
11. Add 26  $\mu$ L of 10-mM Tris-acetate to resuspend beads, place the tube on the magnetic stand to separate the beads, and transfer the supernatant to a new PCR tube.
12. Measure DNA concentration using 1  $\mu$ L of purified DNA and Qubit dsDNA HS kit (*see Note 12*).

### 3.5 Primer Extension (2)

1. Mix the following reagents in a PCR tube:

The purified DNA in the previous step	39 $\mu$ L
10 $\times$ Gene Taq Universal Buffer	5 $\mu$ L
2.5-mM dNTPs	4 $\mu$ L
60- $\mu$ M Primer-3	1 $\mu$ L
Hot Start GeneTaq	1 $\mu$ L

2. Incubate at 94 °C for 3 min, 45 °C for 5 min, and 72 °C for 30 min.
3. Add 50- $\mu$ L AMPure XP and incubate for 5 min at room temperature.
4. Place the tube on a magnetic stand and collect the beads.
5. Wash beads with 200  $\mu$ L of either size cutoff solution (*see Note 10*).
6. Repeat the previous step once.
7. Rinse beads with 200- $\mu$ L 70% (v/v) ethanol.
8. Remove the residual solution completely.
9. Add 26  $\mu$ L of 10-mM Tris-acetate to resuspend the beads, place the tube on the magnetic stand to separate the beads, and transfer the supernatant to a new PCR tube.
10. Use 1  $\mu$ L of the purified DNA to measure the concentration with Qubit dsDNA HS kit.

**3.6 Library QC**

1. Thaw the contents of the library quantitation kit at room temperature (*see* **Note 13**).
2. Calculate the number of wells required as follows (*see* **Note 14**):

$$\text{Number of wells} = (\text{Number of samples} + \text{Number of standards} + \text{Negative control} + 1) \times \text{Multiplying factor}$$

3. Prepare a master mix solution by mixing the following (*see* **Note 15**):

Reagent	Volume per well
Terra PCR Direct TB Green Premix (2×)	10 μL
5× Primer Mix	4 μL
50× ROX Reference Dye ( <i>see</i> Note X)	0.4 μL
Water	3.6 μL

4. Dispense 18 μL of the master mix into PCR tubes.
5. Dilute the libraries with 10-mM Tris-HCl (pH 0) at appropriate dilution rates (*see* **Note 16**).
6. Add 2 μL of standard, non-templated control and diluted libraries to a PCR tube containing the master mix.
7. Cap and place the PCR tubes in a real-time PCR machine.
8. Perform PCR amplification with the following program:

1 min at 95 °C	
95 °C for 10 s	35 cycles
60 °C for 15 s	
68 °C for 45 s	
Melt curve analysis	

9. Prepare electrophoresis device.
10. Mix 1 μL of PCR-amplified DNA with 10 μL of denaturing loading dye, heat denature at 70 °C for 5 min, and load 5 μL of sample onto a 6% Novex TBE-Urea Gel.
11. Run electrophoresis at 300 V until the blue dye reached two-thirds of the gel.
12. Stain the gel with SYBR Gold nucleic acid gel stain and take a photograph (*see* **Note 17**).

### 3.7 Library Amplification

1. Mix the following reagents in a PCR tube:

Reagent	Volume per well
PrimeSTAR Max Premix (2×)	25 μL
100-μM Amp-F	0.3 μL
100-μM Amp-R	0.3 μL
Library prepared Subheading 3.5	25 μL

2. Perform PCR amplification with the following program:

1 min at 95 °C	
95 °C for 10 s	Two to six cycles
55 °C for 15 s	
68 °C for 45 s	

3. Add 50-μL AMPure XP and incubate for 5 min at room temperature.
4. Place the tube on a magnetic stand and collect the beads.
5. Rinse beads with 200-μL 70% (v/v) ethanol.
6. Remove the residual solution completely.
7. Add 26 μL of 10-mM Tris-acetate to resuspend beads, place tube on the magnetic stand to separate beads, and transfer supernatant to a new PCR tube.
8. Use 1 μL of the purified DNA to measure the concentration with a Qubit dsDNA HS kit.

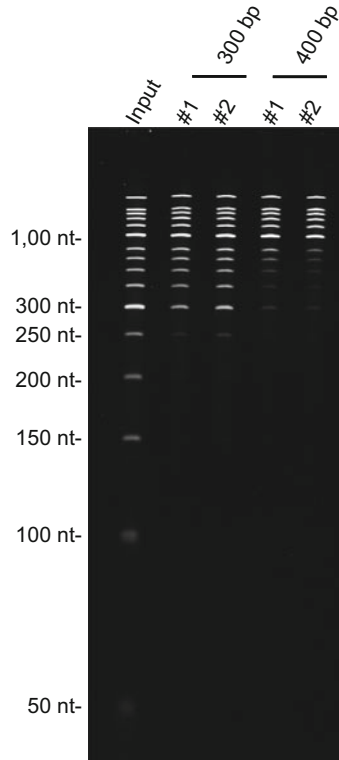
### 3.8 Sequencing

1. Adjust the concentration of the library mixture (*see Note 18*).
2. Run sequencer (*see Note 19*).

---

## 4 Notes

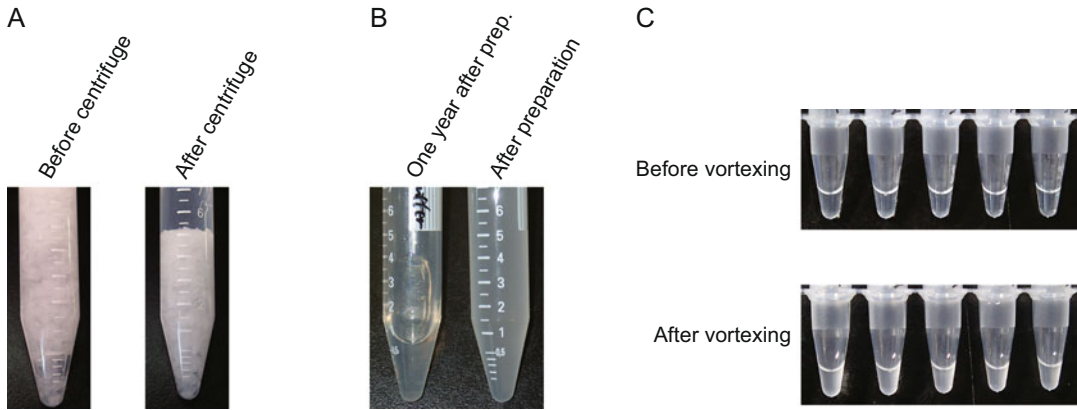
1. NEB provides a Klenow Fragment (3'-5' exo-) at two concentrations. The concentrated version (50,000 units/mL) is appropriate for this protocol.
2. The size cutoff results may vary depending on the preparation of the cutoff solutions. PEG 400 is a viscous liquid, which may cause difficulties in the accurate quantitation of liquids. Therefore, evaluating the cutoff solution at least once after each preparation is strongly recommended. An example of such evaluation is shown in Fig. 4.
3. A concentrated PEG 6000 solution is highly viscous, and dissolving a high concentration of this solution requires long incubation with heating. Overnight incubation at 50 °C is



**Fig. 4** Evaluation of the size -cutoff solution. Two different lots of size cutoff solutions of 300 and 400 bp were compared. A 50-bp DNA ladder from Takara Bio Inc. was used as the model. Slight differences in the yield of the bands were observed

sufficient to dissolve PEG 6000. As air bubbles held by the flakes of PEG 6000 sometimes cause inefficient dissolution, removing the air bubbles by brief centrifugation before heating would aid in the dissociation (Fig. 5a). After preparation, the  $2.5\times$  TACS buffer appears cloudy but becomes transparent during storage (Fig. 5b). The highly viscous  $2.5\times$  TACS buffer is difficult to dispense using the usual micropipettes. As a result, we use a positive displacement pipette, Microman E, from Gilson. The  $2.5\times$  TACS buffer is stored at room temperature. The solution appears transparent but occasionally cloudy, especially at low room temperature. The solution containing  $1\times$  TACS buffer after reaction assembly appears transparent but becomes slightly cloudy upon vortexing (Fig. 5c). In our experience, TACS ligation proceeds successfully with a cloudy reaction.

4. One of the index sequences listed in Table 1 can be inserted into the index sequence for indexing primers.

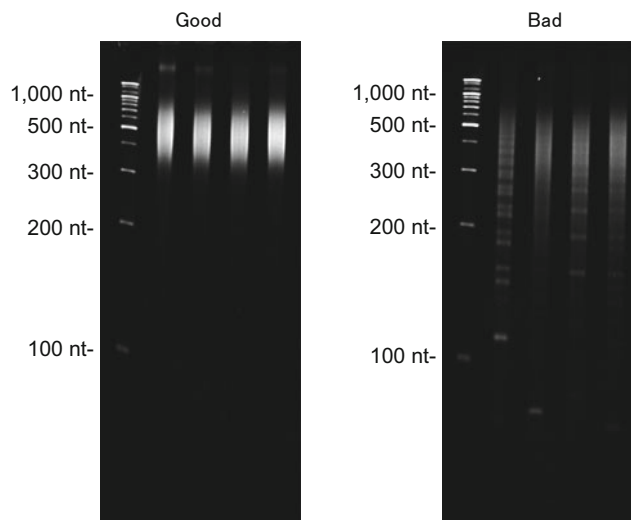


**Fig. 5** Preparation of the solution containing 50% (w/v) PEG#6000. (a) Air bubbles (left) can be removed via brief centrifugation (right). (b) Immediately after the preparation of  $2.5\times$  TACS buffer, the solution becomes cloudy (right) and then transparent during storage (left). (c) The reaction is transparent after assembly (top) but becomes cloudy after mixing (bottom)

5. As a starting material, 100 ng of DNA is recommended. Sufficient reads to cover the mammalian genome can be obtained from this DNA amount without PCR amplification. In addition, the yield of DNA can be measured at each step using Qubit-based measurements if library preparation begins with more than 100 ng of DNA.
6. In most methylome analyses, we recommend spiking  $\sim 1\%$  (w/w) unmethylated lambda DNA into the sample DNA to calculate the BS conversion rate. However, if the DNA concentration is too low for measurement, the appropriate amount of lambda DNA for spiking cannot be calculated. In that case, the inclusion of lambda DNA is not recommended.
7. Do not stop here. Proceed to the next step.
8. Approximately 70% of the input DNA was recovered from human and mouse samples. This quantitation step is essential as good library preparation cannot be expected if DNA is not detected in this step; however, if the starting DNA is less than 50 ng, the DNA concentration is beyond the detection limit. As a result, this step can be skipped.
9. This is a safe point to stop the experiment.
10. A 300-bp cutoff solution is often used. However, if the sample is precious and limited, the 250-bp cutoff solution is employed. To improve the sequencing output, we selected the 400-bp cutoff solution.
11. Typical yields of this step usually range from 40% to 80% of the input DNA. If the starting DNA is less than 50 ng, skip this step.



12. This step yields 10–30% of the input DNA. If the starting DNA is less than 50 ng, skip this step.
13. Three commercially available kits from Kapa Biosciences, Takara Bio, and Toyobo were evaluated and all were found to work well.
14. For reliable quantitation, multiplication of wells is recommended. In our experience, duplexing is sufficient for quantitation; the multiplying factor in the equation is set to 2.
15. Preparation of a master mix is strongly recommended and is effective for reproducible quantitation.
16. Usually, a thousand-fold dilution is appropriate for quantifying the tPBAT library.
17. Because the library before amplification contains many DNA byproducts, and the analysis of libraries before PCR amplification sometimes leads to the misinterpretation of size distribution, electrophoresis of PCR-amplified libraries is recommended to assess the size distribution of libraries. Denaturing gel electrophoresis is recommended because the library after the PCR plateau is difficult to analyze on a native gel. Typical electrophoresis patterns are shown in Fig. 6.



**Fig. 6** Size distribution of the amplified libraries. The amplified DNA fragments were analyzed by denaturing gel electrophoresis. Amplified DNA (1  $\mu$ L) was mixed with 10- $\mu$ L formamide and heat-denatured at 70 °C for 5 min. The sample was then loaded onto a 6% Novex TBE-Urea Gel (Thermo Fisher Scientific). After separation, the gel was stained with SYBR Gold Gel Stain (Thermo Fisher Scientific) and photographed. The successfully prepared libraries showed smear patterns on the left, whereas unsuccessful libraries showed ladder patterns

18. The concentration of the library prepared with tPBAT is sometimes low, and condensation may be required. Nonetheless, the AMPure XP-based procedure is easy to perform for this purpose. The same volume of AMPure XP is mixed with the library mixture, incubated at room temperature for 5 min, removed on a magnetic stand, and rinsed with 70% ethanol. The library mixture is eluted with an appropriate volume of 10-mM Tris-acetate. Usually, there is no need for re-quantitation using qPCR.
19. The libraries prepared with the tPBAT protocol can be successfully sequenced on Illumina sequencers, including MiSeq, NextSeq, HiSeq, and NovaSeq.

---

## Acknowledgments

This work was supported by JSPS KAKENHI through grants to F.M. (20H03243) and T.I. (17H06305) and AMED Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) through a grant to T.I. (Grant Number JP20am0101103).

## References

1. Cokus SJ, Feng S, Zhang X et al (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452(7184):215–219. <https://doi.org/10.1038/nature06745>
2. Lister R, O'Malley RC, Tonti-Filippini J et al (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133(3):523–536. <https://doi.org/10.1016/j.cell.2008.03.029>
3. Adey A, Shendure J (2012) Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. *Genome Res* 22(6):1139–1143. <https://doi.org/10.1101/gr.136242.111>
4. Miura F, Enomoto Y, Dairiki R et al (2012) Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res* 40(17):e136. <https://doi.org/10.1093/nar/gks454>
5. Raine A, Manlig E, Wahlberg P et al (2017) SPLinted Ligation Adapter Tagging (SPLAT), a novel library preparation method for whole genome bisulphite sequencing. *Nucleic Acids Res* 45(6):e36–e36. <https://doi.org/10.1093/nar/gkw1110>
6. Hui T, Cao Q, Wegrzyn-Woltosz J et al (2018) High-resolution single-cell DNA methylation measurements reveal epigenetically distinct hematopoietic stem cell subpopulations. *Stem Cell Rep* 11(2):578–592. <https://doi.org/10.1016/j.stemcr.2018.07.003>
7. Sun Z, Vaisvila R, Yan B et al (2019) Non-destructive enzymatic deamination enables single molecule long read sequencing for the determination of 5-methylcytosine and 5-hydroxymethylcytosine at single base resolution. *Genome Res*. <https://doi.org/10.1101/2019.12.20.885061>
8. Vaisvila R, Ponnaluri VKC, Sun Z et al (2021) Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Res*. <https://doi.org/10.1101/gr.266551.120>
9. Simpson JT, Workman RE, Zuzarte PC et al (2017) Detecting DNA cytosine methylation using nanopore sequencing. *Nat Methods* 14(4):407–410. <https://doi.org/10.1038/nmeth.4184>
10. Wallace EV, Stoddart D, Heron AJ et al (2010) Identification of epigenetic DNA modifications with a protein nanopore. *Chem Commun (Camb)* 46(43):8195–8197. <https://doi.org/10.1039/c0cc02864a>

11. Smallwood SA, Lee HJ, Angermueller C et al (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* 11(8):817–820. <https://doi.org/10.1038/nmeth.3035>
12. Miura F, Shibata Y, Miura M et al (2019) Highly efficient single-stranded DNA ligation technique improves low-input whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res* 47(15):e85. <https://doi.org/10.1093/nar/gkz435>
13. Shirane K, Miura F, Ito T et al (2020) NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. *Nat Genet* 52(10):1088–1098. <https://doi.org/10.1038/s41588-020-0689-z>
14. Hamazaki N, Kyogoku H, Araki H et al (2021) Reconstitution of the oocyte transcriptional network with transcription factors. *Nature* 589(7841):264–269. <https://doi.org/10.1038/s41586-020-3027-9>
15. Toh H, Shirane K, Miura F et al (2017) Software updates in the Illumina HiSeq platform affect whole-genome bisulfite sequencing. *BMC Genomics* 18(1):31. <https://doi.org/10.1186/s12864-016-3392-9>
16. Suzuki M, Liao W, Wos F et al (2018) Whole-genome bisulfite sequencing with improved accuracy and cost. *Genome Res* 28(9):1364–1371. <https://doi.org/10.1101/gr.232587.117>



## Reduced Representation Bisulfite Sequencing (RRBS)

Kazuhiko Nakabayashi, Michihiro Yamamura, Keita Haseagawa,  
and Kenichiro Hata

### Abstract

Reduced representation bisulfite sequencing (RRBS) enriches CpG-rich genomic regions using the *MspI* restriction enzyme—which cuts DNA at all CCGG sites, regardless of their DNA methylation status at the CG site—and enables the measurement of DNA methylation levels at 5% ~ 10% of all CpG sites in the mammalian genome. RRBS has been utilized in a large number of studies as a cost-effective method to investigate DNA methylation patterns, mainly at gene promoters and CpG islands. Here, we describe protocols for gel-free preparation of RRBS libraries, quality control, sequencing, and data analysis. Our protocols typically require nine cycles of polymerase chain reaction (PCR) amplification to obtain a sufficient amount of library for sequencing when 100 ng of genomic DNA is used as a starting material; moreover, it takes 3 days to complete library preparation and quality control procedures for up to eight samples.

**Key words** Reduced representation bisulfite sequencing, DNA methylation, Methylome, *MspI*, Next-generation sequencing

---

## 1 Introduction

DNA methylation is an essential epigenetic mechanism that regulates gene expression in normal development and disease. Reduced representation bisulfite sequencing (RRBS) was originally developed by Meissner et al. to examine DNA methylation patterns at GC-rich genomic regions at base-pair resolution in a quantitative manner [1, 2]. In the original RRBS protocols, the GC-rich regions were enriched by the digestion of genomic DNA using the *MspI* restriction enzyme, which targets 5'-CCGG-3', followed by end repair/A-tailing, adaptor ligation, and gel size selection. The selected regions were subjected to bisulfite conversion and amplified by PCR. Compared to whole-genome bisulfite sequencing (WGBS), which enables comprehensive DNA methylation profiling of the entire genome in principle, RRBS covers only 5% ~ 10% of the entire CpG sites in human and mouse genomes. However,

RRBS has been utilized in a large number of studies as a cost-effective method to investigate DNA methylation patterns mainly at gene promoters and CpG islands. Consequently, improved and enhanced versions of RRBS protocols have been developed by multiple groups [3, 4]. Moreover, commercial kits for RRBS library preparation are currently available from Tecan and Diagenode. Single-cell RRBS protocols have also been developed as a method devoted for methylome [5] and as part of a multi-omics technology for methylome and transcriptome [6]. In this chapter, we describe detailed protocols for the preparation and sequencing of RRBS libraries. Our library preparation protocol is a modified version of the “gel-free” RRBS protocol published by Boyle et al. in 2012 [3]. Using 100 ng of genomic DNA as a starting material, it typically requires nine PCR cycles to amplify a sufficient amount of library for subsequent next-generation sequencing. It takes 3 days to complete RRBS library preparation and their quality control procedures for up to eight samples.

---

## 2 Materials

### 2.1 Reagents

1. MspI (10 u/ $\mu$ L).
2. 10 $\times$  Buffer Tango (Thermo Scientific).
3. Klenow Fragment, exo- (5 U/ $\mu$ L).
4. dNTP mixture for RRBS containing 10-mM dATP, 1-mM dCTP, and 1-mM dGTP is prepared using a dNTP set (PCR Grade 100 mM) in the following manner: Add 20  $\mu$ L of 100-mM dATP, 2  $\mu$ L of 100-mM dCTP, and 2  $\mu$ L of 100-mM dGTP to 176  $\mu$ L of nuclease-free water, and mix well. The mixture is aliquoted and stored at  $-20^{\circ}\text{C}$ .
5. NEBNext Ultra II DNA Library Prep Kit (New England Biolabs).
6. NEBNext Multiplex Oligos for Illumina (Methylated Adaptor, Index Primers Set 1) (New England Biolabs).

The NEBNext Methylated Adaptor, whose DNA sequence is 5'-PO<sub>4</sub>- GATCGGAAGAGCACACGTCTGAACTC CAGTC U ACACTCTTCCCTACACGACGCTCTTCC GATC -s-T-3', contains methylated cytosines (instead of cytosines) and a uracil (U, underlined). “-s-” indicates a phosphorothioate bond. Prepare 150  $\mu$ L of NEBNext Methylated Adaptor (2  $\mu$ M) by adding 130  $\mu$ L of 10-mM Tris-HCl (pH 7.5) with 10-mM NaCl to 20  $\mu$ L of NEBNext Methylated Adaptor for Illumina (15  $\mu$ M), and store at  $-20^{\circ}\text{C}$ .

7. EZ Methylation Gold Kit (Zymo Research) or equivalent.
8. Agencourt AMPure XP beads (Beckman Coulter) or equivalent.

9. 80% EtOH (freshly prepared).
10. 2× KAPA HiFi Hot Start Uracil+ Ready Mix (Kapa Biosciences) or equivalent. The uracil-binding pocket of this high-fidelity DNA polymerase is inactivated so that the polymerase efficiently amplifies uracil-containing DNA.
11. High Sensitivity DNA Kit (Agilent) or equivalent.
12. MiSeq Reagent Nano Kit v2 (300-cycles) (Illumina).
13. PhiX Control v3 Library (Illumina).
14. 10-mM Tris-Cl (pH 8.5).
15. 2-N NaOH (freshly prepared from 10-N NaOH).
16. Buffer A (900-mM NaCl, 180-mM Tris-HCl, pH 7.4) [7].

## 2.2 Oligonucleotides

1. NEBNext Multiplex Oligos for Illumina (96 Unique Dual Index Primer Pairs) (New England Biolabs) (*see Note 1*).
2. PCR Primer F: 5'-AATGATACGGCGACCACCGA-3'.
3. PCR Primer R: 5'-CAAGCAGAAGACGGCATAACGA-3'.

## 2.3 Plastic Disposables

1. Microcentrifuge and PCR tubes: Low-retention tubes are recommended. The authors use MicroAmp 8-Tube Strip (0.2 mL) (N8010580, Applied Biosystems) and MicroAmp 8-Cap Strip (N8010535, Applied Biosystems).
2. Pipette tips: Low-retention tips are recommended.

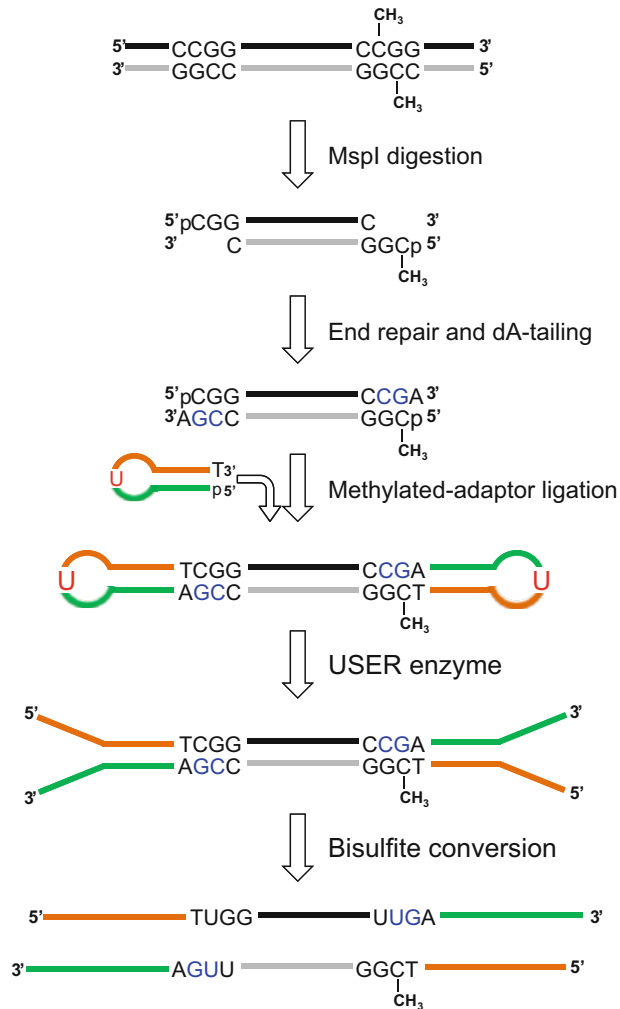
## 2.4 Equipment

1. Eight-channel micropipettes.
2. Magnet for 0.2-mL tubes (such as FastGene MagnaStand 0.2).
3. High-speed microcentrifuge for 1.5-mL tubes with a full speed over  $10,000 \times g$ .
4. Mini centrifuge to spin down 1.5-mL tubes and eight-strip tubes (0.2 mL).
5. Thermal cycler.
6. Bioanalyzer 2100 (Agilent) or equivalent.
7. The 2100 Expert software (Agilent).
8. MiSeq (Illumina).

---

## 3 Methods

Library preparation procedures are schematically presented in Fig. 1. The approximate amount of time required for each step (Fig. 2) is determined based on the assumption that up to eight samples are simultaneously subjected to RRBS library preparation. While the amounts of the reagents required for one sample are described throughout this section, master mixes including 15%



**Fig. 1** RRBS workflow. The procedures from MspI digestion of genomic DNA (Subheading 3.1) to bisulfite conversion (Subheading 3.4) are schematically depicted. Only the MspI sites at the ends of the MspI-digested fragment are depicted. The CG bases presented in blue are artificially introduced at the end repair procedure, and the C base in them must be excluded from methylation data extraction (**step 1** in Subheading 3.8). The bisulfite-converted adaptor-ligated single-strand DNA molecules depicted at the bottom are subsequently amplified using a pair of i5 and i7 PCR primers (*reagent 1* in Subheading 2.2)

additional volume are expected to be prepared when two or more samples are subjected to library preparation. Moreover, it is also recommended to simultaneously use multichannel micropipettes to handle solutions in eight-strip tubes. Thermal programs for the thermocycling block and the headed lid (*Programs A to F*) should be saved in a thermal cycler in advance.

Day	Procedure	Time (hands-on time)	Total time
Day 1	MspI digestion	150 min (20 min)	6.0 h (1 h 50 min)
	Gap filling/dA tailing	65 min (25 min)	
	AMPure beads purification	40 min (40 min)	
	Methylated-adaptor ligation	100 min (25 min)	
Day 2	AMPure beads purification	40 min (40 min)	5.5 h (2 h)
	Bisulfite conversion	240 min (60 min)	
	PCR	45 min (20 min)	
Day 3	AMPure beads purification	40 min (40 min)	1.5 h (1 h)
	Library QC (BioAnalyzer)	50 min (20 min)	

**Fig. 2** Required time for the preparation and quality check of RRBS libraries

### 3.1 MspI Digestion of Genomic DNA (Day 1, 2.5 h)

1. *Day 1.* Add 100 ng (in 25  $\mu$ L) of genomic DNA to a 0.2-mL tube, and keep the eight-tube strip on ice (*see Note 2*).
2. Prepare a master mix of 10 $\times$  Buffer Tango (3  $\mu$ L) and MspI (10 U/ $\mu$ L) (2  $\mu$ L), and keep the master mix on ice. Add 5  $\mu$ L of the master mix to the 0.2-mL tube containing genomic DNA. Mix by pipetting (*see Note 3*) and cap the eight-tube strip with an eight-cap strip.
3. Place the eight-tube strip (0.2 mL) on a thermal cycler, and run the following thermal program (*Program A*) for MspI digestion: 37  $^{\circ}$ C for 120 min; hold at 4  $^{\circ}$ C (heated lid 50  $^{\circ}$ C) (*see Note 4*).

### 3.2 End Repair and dA-Tailing (Day 1, 2 h)

1. Spin down the eight-tube strip (0.2 mL) containing MspI-digested DNA. Remove the eight-cap strip carefully without splashing the DNA solution in the tubes. Keep the eight-tube strip on ice. Add 1  $\mu$ L of dNTP mix for RRBS and 1  $\mu$ L of Klenow Fragment (exo-) to each tube and mix by pipetting (32  $\mu$ L in total). Cap the tubes with a new eight-cap strip (*see Note 5*).
2. Place the eight-tube strip (0.2 mL) on a thermal cycler, and run the following thermal program (*Program B*): 30  $^{\circ}$ C for 20 min and 37  $^{\circ}$ C for 20 min; hold at 4  $^{\circ}$ C (heated lid 50  $^{\circ}$ C).
3. Add 64  $\mu$ L (2 volumes) of AMPure XP beads to the DNA solution. Perform AMPure XP beads purification (*see Note 6*), and elute DNA with 18  $\mu$ L of 10-mM Tris-Cl (pH 8.5). Transfer 17.7  $\mu$ L of the eluted DNA solution to a new eight-tube strip (0.2 mL).

### 3.3 Adaptor Ligation (Day 1, 1.5 h; Day 2, 1 h)

1. Keep the eluted DNA in the eight-tube strip (0.2 mL) on ice. Add 2.3  $\mu$ L of NEBNext Ultra II End Prep Reaction Buffer (included in *reagent 5* in Subheading 2.1) to the eluted DNA (20.0  $\mu$ L in total) (*see Note 7*).



2. Prepare a master mix for NEBNext Ultra II Ligation Master Mix (10.0  $\mu\text{L}$ ) and NEBNext Ligation Enhancer (0.33  $\mu\text{L}$ ), both of which are included in *reagent 5* in Subheading 2.1. Mix the master mix by pipetting up and down ten times (*see Note 3*). Add 10.33  $\mu\text{L}$  of the master mix to eluted DNA and mix by pipetting up and down three times. Add 0.83  $\mu\text{L}$  of NEBNext Methylated Adaptor (2  $\mu\text{M}$ ) (*see reagent 6* in Subheading 2.1) to each tube, and mix by pipetting up and down ten times (31.16  $\mu\text{L}$  in total). Cap the eight-tube strip with a new eight-cap strip.
3. Place the eight-tube strip on a thermal cycler, and run the following thermal program (*Program C*): 20  $^{\circ}\text{C}$  for 60 min; hold at 4  $^{\circ}\text{C}$  (heated lid 50  $^{\circ}\text{C}$ ). After removing the eight-cap strip, keep the eight-tube strip on ice. Add 1  $\mu\text{L}$  of Uracil-Specific Excision Reagent (USER) Enzyme (included in *reagent 6* in Subheading 2.1 or can be purchased separately) (*see Note 8*), mix by pipetting (*see Note 3*), and cap the eight-tube strip with a new eight-cap strip. Place the tubes on a thermal cycler, and run the following thermal program (*Program D*): 37  $^{\circ}\text{C}$  for 15 min; hold at 4  $^{\circ}\text{C}$  (heated lid 50  $^{\circ}\text{C}$ ). Store the tubes containing adaptor-ligated DNA at  $-20^{\circ}\text{C}$  until the next day (*see Note 9*).
4. *Day 2*. Add 48  $\mu\text{L}$  (1.5 volumes) of AMPure XP beads to the DNA solution. Perform AMPure XP beads purification (*see Note 6*), and elute DNA with 20  $\mu\text{L}$  of 10-mM Tris-Cl (pH 8.5).

### 3.4 Bisulfite Conversion of Adaptor-Ligated DNA (Day 2, 4 h)

1. Perform bisulfite conversion of adaptor-ligated DNA using EZ Methylation Gold Kit (*see Note 10*). The thermal program (*Program E*) required for bisulfite conversion is 98  $^{\circ}\text{C}$  for 10 min and 64  $^{\circ}\text{C}$  for 150 min, and hold at 4  $^{\circ}\text{C}$  (heated lid 105  $^{\circ}\text{C}$ ) (*see Note 11*). Add 130  $\mu\text{L}$  of CT conversion reagent to the adaptor-ligated purified DNA (20  $\mu\text{L}$ ), and divide the solution into two 0.2-mL tubes (75  $\mu\text{L}$  each). Place the eight-tube strip(s) on a thermal cycler and run *Program E*. After its completion, combine the two aliquots of 75  $\mu\text{L}$  in one tube. Conduct the subsequent procedures according to the manufacturer's protocol of the EZ Methylation Gold Kit. Elute bisulfite-converted DNA with 11  $\mu\text{L}$  of M-Elution Buffer.

### 3.5 PCR Amplification (Day 2, 1 h; Day 3, 1 h)

1. Transfer 10.0  $\mu\text{L}$  of bisulfite-treated DNA to a 0.2-mL tube. Add 1.25  $\mu\text{L}$  of i7 primer (10  $\mu\text{M}$ ) and 1.25  $\mu\text{L}$  of i5 primer (10  $\mu\text{M}$ ) to the bisulfite-treated DNA, and mix briefly. Add 12.5  $\mu\text{L}$  of 2 $\times$  KAPA HiFi Hot Start Uracil+ Ready Mix and mix by pipetting (25.0  $\mu\text{L}$  in total).
2. Conduct PCR amplification using the following thermal cycling conditions (*Program F*): 98  $^{\circ}\text{C}$  for 45 s for the initial

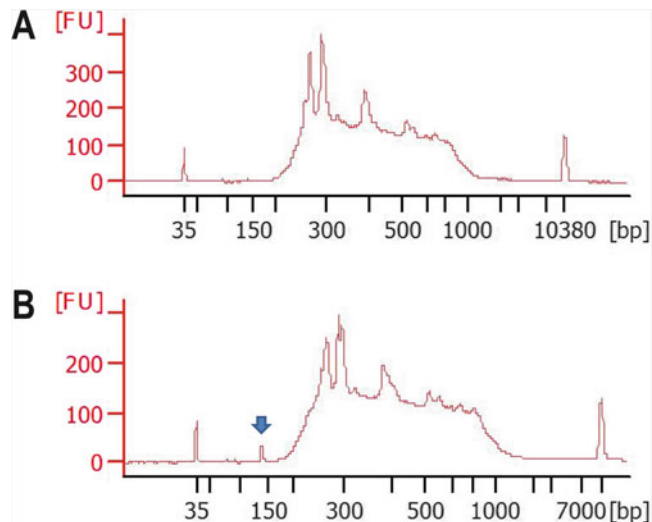
denaturation, nine cycles of 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 30 s for amplification and 72 °C for 1 min for final extension; hold at 4 °C (heated lid 105 °C). The amplified DNA samples can be kept overnight at 4 °C in a thermal cycler or kept longer at –20 °C (*see Note 9*).

- On Day 3, add 30  $\mu$ L (1.2 volumes) of AMPure XP beads to each tube (0.2 mL). Perform AMPure XP beads purification (*see Note 6*), and elute DNA with 15  $\mu$ L of 10-mM Tris–Cl (pH 8.5).

### 3.6 Library Quality Control (Day 3, 1 h, or Longer When the Second DNA Purification or an Additional PCR Amplification Is Required)

Examine the fragment size range and the molarity of each library using the Bioanalyzer 2100 instrument and the High Sensitivity DNA Kit (or an equivalent combination of a device and reagents).

- When the High Sensitivity DNA Kit is used, 1  $\mu$ L each of the purified RRBS libraries is subjected to the microcapillary-based electrophoretic analysis.
- Measure the molarity of each library (for the range 200 bp–1000 bp) using the “Region Table” function of the 2100 Expert software (*see Note 12*). Check the fragment size distribution of each library (*see Note 13*) (Fig. 3).



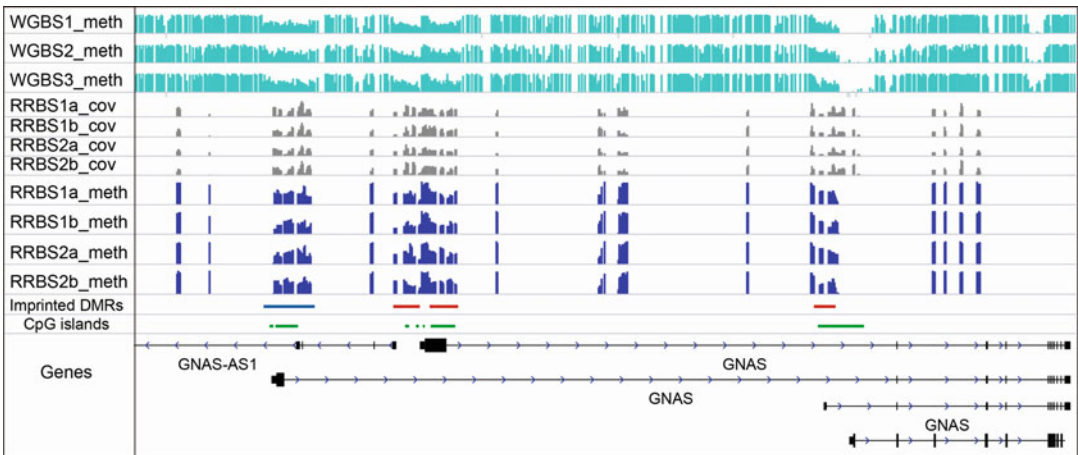
**Fig. 3** Examples of the fragment size distribution of RRBS libraries measured using the High Sensitivity DNA Kit on the Bioanalyzer 2100 equipment. Panel (a) presents an example of the libraries from which adaptor dimers were mostly removed after the first AMPure beads purification (**step 3** in Subheading 3.5). Panel (b) presents an example of the libraries that still contained a small but clearly recognizable peak of the adaptor dimers (indicated by an arrow) after the first AMPure beads purification. Such libraries must be subjected to the second round of AMPure beads purification (*see Note 13*)

### 3.7 Sequencing

1. Test sequencing on the MiSeq platform [8] and readjustment of library DNA concentration: Prepare more than 2  $\mu\text{L}$  of 2-nM solution for each library. Pool 2  $\mu\text{L}$  each of the 2-nM libraries (a total of 16  $\mu\text{L}$  in case of eight libraries). For a test sequencing using a MiSeq Reagent Nano Kit v2 (300-cycles), mix 92  $\mu\text{L}$  of 10-mM Tris-Cl (pH 8.5), 2  $\mu\text{L}$  of the 2-nM library pool, and 1  $\mu\text{L}$  of 1-nM PhiX Control v3 Library (95  $\mu\text{L}$  in total) in a 1.5-mL tube. Add 5  $\mu\text{L}$  of 2-N NaOH to the 95- $\mu\text{L}$  mixture, mix by vortexing, and leave the tube at room temperature for 5 min to denature the library DNA mixture. Add 500  $\mu\text{L}$  of ice-cold Buffer A [7], mix, and keep the solution on ice for 5 min. The final concentration of the denatured library solution (600  $\mu\text{L}$  in total), including 20% amount of spike-in PhiX Control v3 Library for color balancing of low nucleotide diversity, is 8.3 pM. Run the test sequencing on a MiSeq platform using the Generate FASTQ workflow.
2. After the completion of the MiSeq run, check the index ratios of the RRBS libraries, and recalculate the concentration of the libraries in accordance with the index ratios.
3. Conduct full-scale sequencing on the HiSeq X platform (*see Note 14*). Prepare a pool of RRBS libraries sufficient for sequencing (*see Note 15*).

### 3.8 Data Analysis (See Note 16)

1. Trim\_Galore: Trim\_Galore [10] is used for applying quality and adapter trimming to fastq files, with “--rrbs” option to also remove two additional bases (CG) that contain an un-methylated C artificially introduced in the end repair procedure during library preparation.
2. Bismark: The trimmed paired-end fastq reads are subsequently processed with Bismark [11, 12]. By using the “bismark\_genome\_preparation” command, the reference genome sequence of interest needs to be bisulfite-converted and indexed in advance. Furthermore, the *bismark* command with “--bowtie2” option is used to align the sequence reads to the reference genome, which generates a bam file and an alignment report file. The *bismark\_methylation\_extractor* and *coverage2cytosine* commands are used to generate a genome-wide cytosine report output, which contains the following seven columns: <chromosome> <position> <strand> <count methylated> <count unmethylated> <C-context> <trinucleotide context>, like “chr13 3004198 - 16 4 CG CGG.”
3. methylKit: Using a custom script, the genome-wide cytosine report output is converted to a methylation call format, which methylKit [13] accepts as an input file. The methylation call format contains the following seven columns: <chrBase> <chr> <base> <strand> <coverage> <freqC> <freqT>,



**Fig. 4** An example of RRBS data visualized using IGV. WGBS data of three samples (right, blue, methylation (meth)) and RRBS data (gray, coverage (cov); blue, methylation (meth)) of four libraries (two human cord blood samples (1 and 2) with technical replicates (a and b)) are presented for an 85-kb interval (chr20:57,405,000–57,487,000 (hg19)) in the human *GNAS* locus. The WGBS data (E050, Mobilized\_CD34\_Primary\_Cells; E112, thymus; E113, spleen) were retrieved from the NIH Roadmap Epigenomics Mapping Consortium website ([https://egg2.wustl.edu/roadmap/web\\_portal/](https://egg2.wustl.edu/roadmap/web_portal/)). Imprinted differentially methylated regions (blue, a paternally methylated region; red, maternally methylated regions) and CpG islands (green) are also depicted

like “chr13.3004198 chr13 3004198 R 20 80 20” The “meth-Read” function of methylKit is used to read information from the methylation call format for each of the multiple samples, with a minimum coverage threshold (such as 10). The “unite” function is used to retain the CpG sites with coverage from all samples and to merge reads that cover both strands of a CpG site (by specifying the “destrand” argument as true). In addition, methylKit is used to obtain descriptive statistics such as percent methylation and coverage for an individual sample, to conduct comparative analyses on a set of samples, and to search for differentially methylated bases or regions between samples or groups [14].

4. Data visualization: The united data, percent methylation, and read coverage information for individual CpG sites for a set of samples are reformatted to the igv format for data visualization using the Integrative Genomics Viewer (IGV) [15] (Fig. 4).

## 4 Notes

1. This protocol assumes full-scale sequencing of RRBS libraries by an Illumina platform that adopts patterned flow cells and, therefore, uses unique dual index primers for library amplification to minimize the effects of index hopping [17]. In addition to the NEBNext primers (*reagent 1* in Subheading 2.2), we also

use primers synthesized and purified by a reverse-phase cartridge at a custom oligo synthesis company. i5 and i7 primer sequences are 5'- AATGATACGGCGACCACCGAGATCTA CAC [XXXXXXXX] ACACTCTTTCCTACACGACGCTCT TCCGATCT -3' and 5'- CAAGCAGAAGACGGCATAACGA GAT [YYYYYYYY] GTGACTGGAGTTCAGACGTGTGCTC TTCCGATCT -3', respectively. XXXXXXXX and YYYYYYYY represent the forward sequence of Index 1 and the reverse complement sequence of Index 2, respectively.

2. When reagents and samples need to be kept cold on ice, it is recommended to place them on aluminum blocks that are placed on ice.
3. When mixing master mix or sample solutions by pipetting with a micropipette, set the volume of the micropipette to >50% of the volume of the solution to be mixed, and pipette up and down ten times. Use a multichannel micropipette when simultaneously mixing solutions in multiple tubes.
4. Stopping point. The digested genomic DNA can be stored at  $-20^{\circ}\text{C}$  for a few days.
5. To avoid contamination among samples, be careful not to splash the DNA solutions in the eight-tube strip (0.2 mL) when you remove the eight-cap strip. In addition, do not reuse the removed eight-cap strip and always use a new one when recapping the eight-tube strip. It is recommended to use a strip cap tool when capping an eight-tube strip with an eight-cap strip.
6. AMPure XP beads purification: Allow the beads to warm to room temperature before use. Add a designated volume of beads to the DNA solution, mix by pipetting, and incubate for at least 5 min. Place the eight-tube strip (0.2 mL) on a magnetic stand and leave it for 5 min (or until the solution becomes clear). Remove and discard the supernatant, add 150  $\mu\text{L}$  of 80% EtOH, and incubate at room temperature for 30 s (the first wash). Remove and discard the supernatant, add 150  $\mu\text{L}$  of 80% EtOH, and incubate at room temperature for 30 s (the second wash). Remove all visible liquid after the second wash. If the sides of the tube are wet, spin down the tube, place back the tube on the magnet, and remove the residual ethanol. Air-dry the beads for 5 min. Be careful not to overdry the beads, as this may decrease the DNA recovery. Remove the eight-tube strip from the magnetic stand. Add a designated amount of EB to the beads, mix well by pipetting, and incubate for 2 min. Place the tube on the magnet stand, leave it for 5 min (or until the solution becomes clear), and transfer the elute to a new eight-tube strip.

7. Enzymatic reactions corresponding to “End Prep” are already performed in Subheading 3.2. Only the reaction buffer is added in this step.
8. USER Enzyme is a combination of Uracil DNA glycosylase and Endonuclease VIII. By generating a single nucleotide gap at the location of the uracil residue in the NEBNext Adaptor, the enzyme converts the adaptor from its hairpin-loop form to a Y-shaped form.
9. Stopping point.
10. Use freshly prepared CT conversion reagent to consistently obtain high bisulfite conversion efficiency (>99%) and a high DNA recovery rate (>75%).
11. An optional stopping point. Although we usually conduct the subsequent procedures soon after the thermal program E is completed, the bisulfite-converted samples can be stored in the thermal cycler at 4 °C up to 20 h.
12. If the yield of the library turns out to be lower than 30 pmoles (i.e., less than 2-nM concentration with 15- $\mu$ L volume), additional PCR amplification is required to obtain a sufficient amount of library for full-scale sequencing. Transfer 12.5  $\mu$ L of the purified library to a new tube (0.2 mL), and add 1.25  $\mu$ L of Primer F (10  $\mu$ M, *reagent 2* in Subheading 2.2.), 1.25  $\mu$ L of Primer R (10  $\mu$ M, *reagent 3* in Subheading 2.2), and 15.0  $\mu$ L of 2 $\times$  KAPA HiFi Hot Start Uracil+ Ready Mix. Mix by pipetting (30.0  $\mu$ L in total). Run three cycles of PCR on the reaction mixtures using thermal Program F (in **step 2** in Subheading 3.5) with a modification of PCR cycles. Subsequently, conduct AMPure XP beads purification using 1.2 volumes (36  $\mu$ L) of beads (*see Note 6*), and elute DNA with 15  $\mu$ L of 10-mM Tris–Cl (pH 8.5). Conduct the library quality control procedure (Subheading 3.6) again.
13. If you observe a peak of adapter dimers (at approximately 140 bp in size), it must be removed through a second round of AMPure XP beads purification. Adapter dimers contain complete adapter sequences. Because of their small size, adapter dimers form clusters more efficiently than the intended library fragments at the cluster generation procedure in the Illumina sequencing. Therefore, leaving adapter dimers in your library decreases the amount of sequencing data from the intended fragments. Fill up the volume of the library DNA solution to 30  $\mu$ L by adding 10-mM Tris–Cl (pH 8.5), add 1.2 volumes of AMPure XP beads, perform AMPure XP beads purification (*see Note 6*), and elute DNA with 15  $\mu$ L of 10-mM Tris–Cl (pH 8.5). Conduct the library quality control procedure (Subheading 3.6) again.

14. It has been shown that software updates in the Illumina HiSeq platform affect whole-genome bisulfite sequencing [9]. Therefore, it is recommended to keep using the same version of the base-call software for all RRBS libraries to be analyzed together.
15. In case a sequencing service is utilized for the Illumina HiSeq X platform, the typical amount of library mix required for one lane sequencing is >100 pmoles (>20  $\mu$ L in case of 5-nM solution). Eight to ten RRBS libraries are pooled, in which 20% of PhiX Control v3 Library is added for color balancing and subjected to the paired-end sequencing (150 bp  $\times$  2) using one lane of the HiSeq X flow cell. Approximately 300 million read-pairs per lane (30–40 million read-pairs per library) are obtained.
16. Examples of raw sequence data obtained from the RRBS libraries prepared by the protocol described here are available at the BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) (accession numbers PRJNA632597 and PRJNA633097) [16].

---

## Acknowledgments

This work was supported in part by grants from the Practical Research Project for Rare/Intractable Diseases Program of the Japan Agency for Medical Research and Development (AMED), the Japan Society for the Promotion of Science KAKENHI, and Takeda Science Foundation.

## References

1. Meissner A, Gnirke A, Bell GW et al (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res* 33: 5868–5877
2. Gu H, Smith ZD, Bock C et al (2011) Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nat Protoc* 6:468–481
3. Boyle P, Clement K, Gu H et al (2012) Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biol* 13:R92
4. Garrett-Bakelman FE, Sheridan CK, Kacmarczyk TJ et al (2015) Enhanced reduced representation bisulfite sequencing for assessment of DNA methylation at base-pair resolution. *J Vis Exp* 96:e52246
5. Guo H, Zhu P, Guo F et al (2015) Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nat Protoc* 10:645–659
6. Gu H, Raman AT, Wang X et al (2021) Smart-RRBS for single-cell methylome and transcriptome analysis. *Nat Protoc* 16:4004–4030
7. Miura F, Enomoto Y, Dairiki R, Ito T (2012) Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res* 40:e136
8. Katsuoka F, Yokozawa J, Tsuda K et al (2014) An efficient quantitation method of next-generation sequencing libraries by using MiSeq sequencer. *Anal Biochem* 466:27–29
9. Toh H, Shirane K, Miura F et al (2017) Software updates in the Illumina HiSeq platform

- affect whole-genome bisulfite sequencing. *BMC Genomics* 18:31
10. [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/)
  11. Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27: 1571–1572
  12. [https://rawgit.com/FelixKrueger/Bismark/master/Docs/Bismark\\_User\\_Guide.html](https://rawgit.com/FelixKrueger/Bismark/master/Docs/Bismark_User_Guide.html)
  13. Akalin A, Kormaksson M, Li S et al (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biol* 13:R87
  14. <https://bioconductor.org/packages/release/bioc/vignettes/methylKit/inst/doc/methylKit.html>
  15. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14:178–192
  16. Nohara K, Nakabayashi K, Okamura K et al (2020) Gestational arsenic exposure induces site-specific DNA hypomethylation in active retrotransposon subfamilies in offspring sperm in mice. *Epigenetics Chromatin* 13:53
  17. MacConaill LE, Burns RT et al (2018) Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. *BMC Genomics* 19:30



# Part II

## **Analysis of the Epigenome: Histone Modification and Chromatin Structure**



## Chromatin Immunoprecipitation Sequencing (ChIP-seq) for Detecting Histone Modifications and Modifiers

Shinjiro Hino, Tetsuya Sato, and Mitsuyoshi Nakao

### Abstract

Chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) is the most widely used method for analyzing genome-wide DNA–protein interactions. Because there is considerable variation in the modes and strengths of DNA–protein interactions, chromatin immunoprecipitation (ChIP) protocols have been diversified and optimized for different needs. Here, we describe protocols for detecting histone modifications and modifiers using various crosslinking and immunoprecipitation conditions. We provide a complete ChIP-seq workflow covering sample preparation, immunoprecipitation, next-generation sequencing (NGS) library preparation, and data analyses.

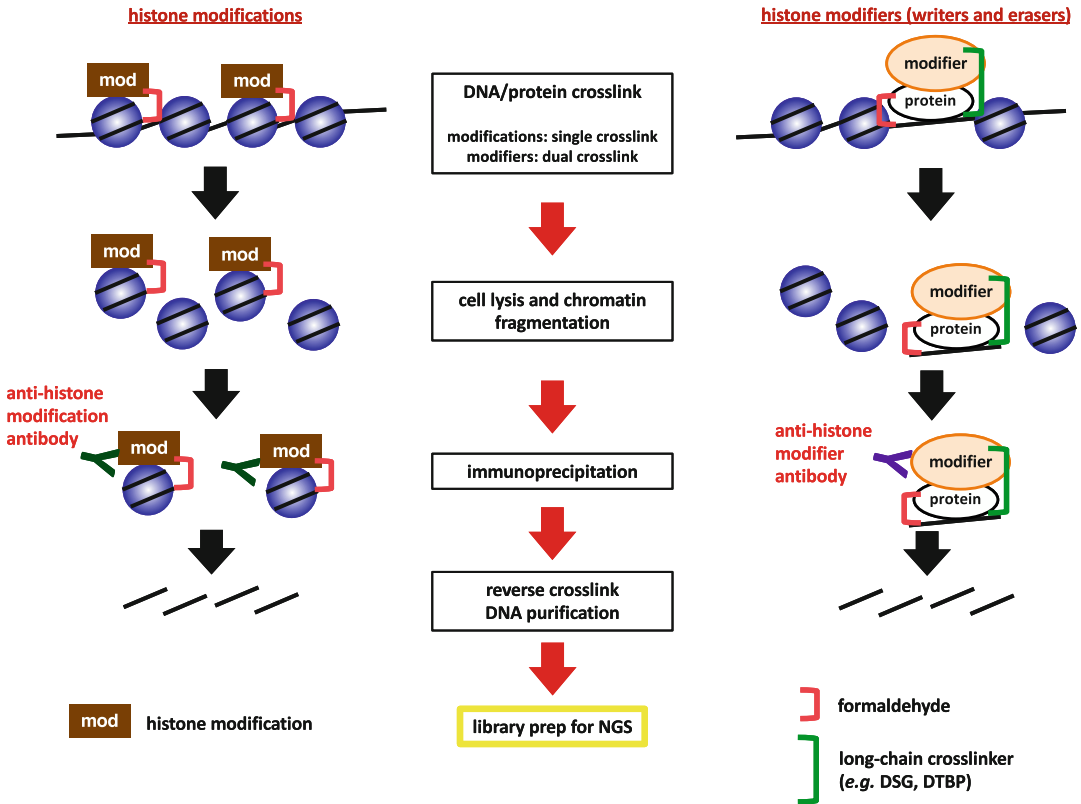
**Key words** Histone, Immunoprecipitation, Next-generation sequencing, Dual crosslink

---

### 1 Introduction

Investigating histone modifications is crucial for understanding the epigenetic regulation of the eukaryotic genome. In addition, to understand how histone modifications are dynamically regulated, studies on the genomic occupancy of histone-modifying enzymes (e.g., writers and erasers) are essential. Over the decades, chromatin immunoprecipitation (ChIP) assays have been the first-choice method for analyzing DNA–protein interactions. Advances in DNA microarray technologies, followed by next-generation sequencing (NGS) technologies, have extended the application of ChIP assays to investigate epigenomic and cis-regulatory landscapes [1–3].

In a ChIP assay, DNA–protein interactions are fixed using one or two chemical crosslinkers, followed by cell lysis and chromatin fragmentation (Fig. 1). The chromatin solution is then subjected to immunoprecipitation using an antibody specific for a histone modification or modifier. Purified DNA can be used for both PCR- and NGS-based genome-wide analyses.



**Fig. 1** Workflow of chromatin immunoprecipitation (ChIP) assay for analyzing histone modifications and modifiers

For successful ChIP assays, the use of high-quality antibodies is essential. Another key tip to obtaining reliable and reproducible results is to prepare an adequate amount of soluble chromatin fragments with an optimal DNA size. For this purpose, optimization of fixation and fragmentation conditions is necessary. To analyze histone modifications, formaldehyde fixation is routinely used, followed by DNA fragmentation using a sonicator. The fixation time and sonication strength must be optimized for each cell/tissue type. When analyzing histone modifiers, an additional crosslinking strategy is necessary, because most of them do not directly bind to DNA. Thus, long-chain crosslinkers are used along with the short-chain, formaldehyde [4]. While a buffer containing 1% SDS is typically used for assessing histone modifications, the use of a lower SDS concentration (0.1%) is recommended for studying histone modifiers since it helps maintain DNA–protein interactions.

Here, we describe ChIP protocols that allow us to assess both histone modifications and modifiers. In addition, we provide a basic workflow for obtaining ChIP-seq data and their biological interpretation.

---

## 2 Materials

The single-crosslink ChIP assay was based on the method described by Ye et al., with some modifications [5]. We combined this protocol with that of Porro and Perini (The Epigenome Network of Excellence [https://www.epigenome-noe.net/researchtools/protocol.php\\_protid=29.html](https://www.epigenome-noe.net/researchtools/protocol.php_protid=29.html)) for dual-crosslink experiments [6].

### 2.1 Chromatin Fixation and Fragmentation

1. Formaldehyde fixation solution: 30% (v/v) formaldehyde solution (formalin; *see Note 1*), 0.1-M NaCl, 1-mM EDTA, 0.5-mM EGTA, 50-mM HEPES, pH 7.5.
2. Disuccinimidyl glutarate (DSG) fixation solution: 2-mM DSG in PBS (*see Note 2*).
3. Fixation stop solution: 1.5-M glycine.
4. Proteinase inhibitor (PI) cocktail (Nacalai Tesque) (*see Note 3*).
5. Cell lysis buffer: 5-mM PIPES (pH 8.0), 85-mM KCl, and 0.5% NP-40. Add PI cocktail before use.
6. SDS lysis buffer: 50-mM Tris-HCl (pH 8.0), 10-mM EDTA, and 1% SDS. Add PI cocktail before use.
7. Picoruptor (Diagenode), a water bath sonicator (*see Note 4*).

### 2.2 Immunoprecipitation

1. Dilution buffer: 50-mM Tris-HCl, pH 8.0, 167-mM NaCl, 1.1% Triton X-100, and 0.11% sodium deoxycholate. Add PI cocktail before use.
2. Immunoprecipitation (IP) antibody: Specific antibody against the protein of interest and species-compatible control IgG from nonimmunized animals.
3. Dynabeads Protein A or G: Choose either type depending on the source species and IgG subclass of the IP antibody.
4. Magnetic rack for 1.5-mL tubes.
5. 2× RIPA buffer: 100-mM Tris-HCl (pH 8.0), 300-mM NaCl, 2-mM EDTA, 0.2% Triton X-100, 0.2% SDS, and 0.2% sodium deoxycholate. 1× RIPA buffer (low salt): Mix equal volume of 2× RIPA buffer and ultrapure water (final concentration of NaCl, 150 mM). 1× RIPA buffer (high salt): 1× RIPA buffer containing 500-mM NaCl.
6. LiCl wash buffer: 10-mM Tris-HCl (pH 8.0), 250-mM LiCl, 1-mM EDTA, 0.5% NP-40, and 0.5% sodium deoxycholate.
7. 1× TE buffer: 10-mM Tris-HCl, pH 8.0, 1-mM EDTA.
8. Direct elution buffer: 10-mM Tris-HCl (pH 8.0), 300-mM NaCl, 2-mM EDTA, and 0.5% SDS.

**2.3 DNA Purification**

1. TE/NaCl: 0.2-M NaCl in 1× TE.
2. 10-mg/mL RNase A.
3. 10-mg/mL Proteinase K.
4. Phenol/chloroform/isoamyl alcohol (25:24:1) (PCI).
5. 20-mg/mL glycogen.

**2.4 NGS Library Prep**

1. DNA quality check reagents (Agilent High Sensitivity D1000 ScreenTape).
2. Library preparation kit (NEBNext Ultra II DNA Library Prep Kit for Illumina).
3. Library barcoding kit (NEBNext Multiplex Oligos for Illumina).
4. Library quantification kit (GenNext NGS Library Quantification Kit, TOYOBO).
5. DNA low-bind 0.2- and 1.5-mL tubes.
6. Magnetic rack for 0.2-mL tubes.

---

**3 Methods**
**3.1 Preparation of Cell Suspension and Tissue Lysate**

Trypsinize the cultured adherent cells and suspend them in PBS. Alternatively, the cells can be scraped off the culture dish after fixation (Subheading 3.2.1 or 3.2.2). Typically,  $10^6$ – $10^7$  cells are used for a single IP; however, the optimal cell number varies depending on the antibody used. For preparing tissue lysates, snap freeze freshly isolated tissue sections (30–100 mg) in liquid nitrogen, and then homogenize them in PBS using a Dounce homogenizer (1–2 mL) with a loose pestle.

**3.2 Chromatin Fixation and Fragmentation**

To obtain DNA fragments with an appropriate size distribution in the soluble chromatin fraction, fixation and fragmentation conditions must be optimized for each cell/tissue type (see below for details).

**3.2.1 Histone Modifications**

1. Add 1/10 volume of fixation buffer to the cell suspension or tissue lysate and incubate at 25 °C for 5–15 min with mild mixing or rotation.
2. Add 1/10 volume of 1.5-M glycine and incubate at 25 °C for 3 min with a mild mixing or rotation.
3. Spin down and wash twice with PBS.
4. Resuspend the cell pellet in lysis buffer and incubate on ice for 15 min. Homogenize the cells ten times with a syringe equipped with a 21–23G needle. Count nuclei using a hemocytometer or by DAPI staining. Spin down the nuclei for 5 min at 4 °C.

5. Resuspend the pellet in SDS lysis buffer at a density of  $1\text{--}2 \times 10^6$  nuclei/100  $\mu\text{L}$ , and incubate on ice for 5 min.
6. Transfer the nuclei suspension to a Picoruptor tube (100–300- $\mu\text{L}$ /tube). Sonication condition: 5–30 cycles (30-sec on/30-sec off) at 4 °C.
7. Centrifuge the mixture at  $13,000 \times g$  for 5 min, at 4 °C, and collect the supernatant for the subsequent IP steps (*see Note 5*).

### 3.2.2 Histone Modifiers

Fixation and fragmentation are performed as described in Subheading 3.2.1, except for the following:

1. Dual crosslink: Prior to formaldehyde crosslink, suspend the cells/tissue lysate in 2-mM DSG, and incubate at 25 °C for 5 min.
2. Sonication: Suspend the nuclei in  $1 \times$  RIPA buffer, which contains 0.1% SDS, instead of SDS lysis buffer. The optimal sonication time may be longer (up to 50 min) than that of the single-crosslink method.

### 3.3 Immuno-precipitation

1. Pre-clear: Dilute the fragmented chromatin solution ten times with ChIP dilution buffer. Add Dynabeads Protein A or G at 20  $\mu\text{L}/\text{mL}$  and rotate at 4 °C for 1 h. Place the tube on a magnetic rack for 5 min and collect the supernatant. At this point, chromatin from  $1\text{--}4 \times 10^6$  cells will be solubilized per 1 mL.
2. Antibodies: Keep 10% of the input for subsequent DNA purification. Add antibodies specific to histone modification or modifier at 4–10  $\mu\text{g}/\text{mL}$ . A parallel sample adding an equivalent amount of control IgG should be prepared. Rotate overnight at 4 °C.
3. Beads pulldown: Add 20  $\mu\text{L}$  of Dynabeads Protein A or G and rotate at 4 °C for 2–3 h.
4. The beads should be washed five times with RIPA buffer (low salt), RIPA buffer (high salt), and LiCl buffer and twice with TE. Between each wash, mix by rotation for 3–5 min, spin down, and place on a magnetic rack.
5. Reverse crosslinking: Suspend the beads in 200  $\mu\text{L}$  of direct elution buffer. To the input samples (taken at **step 2** in Subheading 3.3), add ChIP dilution buffer to make the total volume of 192  $\mu\text{L}$ , and then add 8  $\mu\text{L}$  of 5-M NaCl. Incubate at 65 °C for at least 4 h.

### 3.4 DNA Purification

1. Add 1  $\mu\text{L}$  of RNase A, and incubate at 37 °C for 30 min.
2. Add 1  $\mu\text{L}$  of Proteinase K, and incubate at 45 °C for 1 h.

3. Add an equal volume of PCI, vortex, and centrifuge at  $13,000 \times g$  for 5 min. Collect the aqueous phase. For the back extract, add TE/NaCl to the original tube, and repeat the extraction (the aqueous phase combined will sum up to approximately 400  $\mu\text{L}$ ).
4. Add 1  $\mu\text{L}$  of glycogen and 900  $\mu\text{L}$  of 100% ethanol. Vortex and incubate at  $-20^\circ\text{C}$  for 5 min. Centrifuge at  $13,000 \times g$  for 5 min and discard the supernatant. Wash the pellet with 70–80% ethanol.
5. Suspend the pelleted DNA in 50–100  $\mu\text{L}$  of TE.

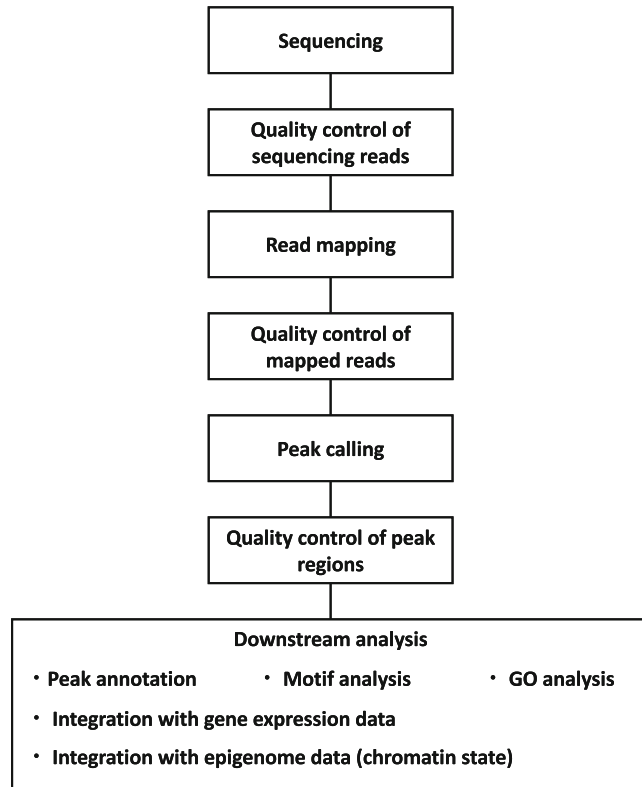
### 3.5 NGS Library Prep

1. Prior to library preparation, analyze the size distribution of DNA fragments using a TapeStation or a Bioanalyzer instrument. If neither is available, perform agarose gel electrophoresis combined with SYBR Gold staining which enhances the detection of low-concentration DNA.
2. For library preparation, use an appropriate DNA library preparation kit according to the manufacturer's instructions. We use 20–30 ng of IP DNA or 50–100 ng of input DNA for each library. For library amplification, 10–15 PCR cycles are typically used.
3. It is important to confirm that the size distributions are equivalent between the libraries generated. The amount of library DNA should be quantified using a qPCR-based method. GenNext NGS Library Quantification Kit may be used.

### 3.6 Data Analyses

The programs introduced below are compatible with Linux and Macintosh operating systems. Public cloud servers such as AWS and GCP can be used to enable the Linux-based operation of the programs. To install and execute the programs, set up an environment where conda (<https://anaconda.org/anaconda/conda>) or R (<https://www.r-project.org/>) can be used. The analytical workflow is described below (Fig. 2).

1. *Sequence read quality check (QC)*. Check the read quality of each fastq file using the FastQC program (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). MultiQC program [7] can be used to visualize multiple QC results in the same graphs or tables. Omit low-quality samples from further analyses.
2. *Read mapping*. Map the reads to a reference genome and generate a BAM file using the Bowtie [8], Bowtie2 [9], or BWA [10] programs. Bowtie2 accounts for insertions/deletions, whereas Bowtie does not. The BWA algorithm allows mismatches to a certain extent, thus making it suitable for mapping relatively long reads (approximately 150 bps). To reduce artifacts in peak detection, omit reads that align



**Fig. 2** Workflow of chromatin immunoprecipitation sequencing (ChIP-seq) data analyses

multiple times to the reference, and include only the uniquely mapped reads in the BAM file. To remove duplicate reads (produced by library PCR), execute the `rmdup` command in the SAMtools program [11].

3. *Mapped read QC.* The mapped read quality can be compared between samples based on the number of uniquely mapped reads, duplicate reads, mapping rate, and read depth. When analyzing histone modifications, generate an aggregate plot showing the accumulation of all reads in relation to the transcription start sites (TSS) and transcription end sites (TES) using the `ngs.plot` program [12]. Using the Integrated Genome Viewer (IGV) [13], the local read distribution can be visualized. The validity of the data can be confirmed based on the typical distribution patterns of individual modifications.
4. *Peak call.* Run the MACS2 program to identify enrichment peaks by scanning for regions with significant accumulation of reads over random distribution in the input sample [14]. A bed file containing peak information can be visualized on the IGV. Sharp peaks are expected for some histone modifications (e.g.,



H3K4me3 and H3K27ac) and transcription factors, whereas broad peaks are expected for other modifications (e.g., H3K27me3, and H3K36me3). The parameters of the MACS program can be adjusted based on the expected peak shape.

5. *Peak region QC.* To check the number of reads accumulated in the peaks, create aggregate plots using ngs.plot program. Evaluate the peak region quality by comparing input and ChIP samples and comparing replicate samples. Check peak similarities among samples using the Jaccard program with SAMtools.
6. *Peak annotation and biological interpretation.* Annotate the peaks by identifying neighboring genes using a web tool such as GREAT (<http://great.stanford.edu/public/html/index.php>) [15]. Gene ontology (GO) analysis of peak-associated genes enables biological interpretation of histone modification changes and their regulation by modifiers. Web tools such as the DAVID Web Serve (<https://david.ncicrf.gov/summary.jsp>) are useful for performing GO analyses (*see Note 6*).
7. *Predicting molecular mechanisms.* DNA sequence motifs that frequently appear in peak regions can be detected using HOMER [16] or MEME Suite [17] programs. Known transcription factor motifs can be identified by referring to JASPAR or similar databases [18]. Similarities with publicly deposited ChIP-seq data can be examined using the ChIP-Atlas web server (<https://chip-atlas.org/>) [19]. This analysis enables the prediction of how histone modifications are regulated and the proteins with which the modifiers cooperate.

---

## 4 Notes

1. Commercially available formaldehyde solutions typically contain 36–38% (w/v) formaldehyde. Thus, the concentration in the fixation buffer should be 10.8–11.4%, and 0.98–1.04% in the final working solution.
2. We also confirmed efficient crosslinking using 5-mM dimethyl 3,3'-dithiobispropionimidate•2HCl (DTBP) [20]. The optimal long-chain crosslinker depends on the cell type [4].
3. Protease inhibitors from other companies can also be used.
4. Other water bath sonicators (e.g., Covaris and Bioruptor) and probe-type sonicators can also be used. For some cell types, a combination of two different sonicators may be used.
5. In a pilot study, DNA fragment size should be evaluated. Add NaCl to chromatin solution at a final concentration of 200 mM, and perform reverse crosslinking at 65 °C for 2–4 h. Then, purify the DNA as described in Subheading 3.4.

Run samples on a 2% agarose gel and stain with SYBR Gold (if not available, use ethidium bromide). Typically, reproducible ChIP-seq data are obtained when the majority of fragments range from 200 to 500 base pairs. Similar size distribution should be achieved between the samples to be compared directly.

6. Long-range chromosomal interaction data, such as Hi-C data, are useful for predicting the gene regulatory potential of peak regions, especially when there are no neighboring genes that are functionally associated with the peak region.

---

## Acknowledgments

This work was supported by the following funding sources: JSPS KAKENHI Grant Numbers 20H04108 and 21 K19513 (S.H.) and 21H02686 and 20KK0185 (M.N.), Takeda Science Foundation (S.H. and M.N.), and the Japan Agency for Medical Research and Development (21gk0210029h0101 (S.H.) and JP19gm4010003 (M.N.)). We would like to thank Editage ([www.editage.com](http://www.editage.com)) for English language editing.

## References

1. Bernstein BE, Humphrey EL, Erlich RL et al (2002) Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci* 99:8695–8700. <https://doi.org/10.1073/pnas.082249499>
2. Barski A, Cuddapah S, Cui K et al (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837. <https://doi.org/10.1016/j.cell.2007.05.009>
3. Mikkelsen TS, Ku M, Jaffe DB et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448:553–560. <https://doi.org/10.1038/nature06008>
4. Zeng P-Y, Vakoc CR, Chen Z-C et al (2006) In vivo dual cross-linking for identification of indirect DNA-associated proteins by chromatin immunoprecipitation. *BioTechniques* 41:694–698. <https://doi.org/10.2144/000112297>
5. Ye SK, Agata Y, Lee HC et al (2001) The IL-7 receptor controls the accessibility of the TCRgamma locus by Stat5 and histone acetylation. *Immunity* 15:813–823. [https://doi.org/10.1016/s1074-7613\(01\)00230-8](https://doi.org/10.1016/s1074-7613(01)00230-8)
6. Anan K, Hino S, Shimizu N et al (2018) LSD1 mediates metabolic reprogramming by glucocorticoids during myogenic differentiation. *Nucleic Acids Res* 46:5441–5454. <https://doi.org/10.1093/nar/gky234>
7. Ewels P, Magnusson M, Lundin S et al (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32:3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
8. Langmead B, Trapnell C, Pop M et al (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25. <https://doi.org/10.1186/gb-2009-10-3-r25>
9. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
10. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
11. Danecsek P, Bonfield JK, Liddle J et al (2021) Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
12. Shen L, Shao N, Liu X et al (2014) ngs.plot: quick mining and visualization of next-generation sequencing data by integrating

- genomic databases. *BMC Genomics* 15:284. <https://doi.org/10.1186/1471-2164-15-284>
13. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14:178–192. <https://doi.org/10.1093/bib/bbs017>
  14. Zhang Y, Liu T, Meyer CA et al (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137. <https://doi.org/10.1186/gb-2008-9-9-r137>
  15. McLean CY, Bristor D, Hiller M et al (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28:495–501. <https://doi.org/10.1038/nbt.1630>
  16. Heinz S, Benner C, Spann N et al (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589. <https://doi.org/10.1016/j.molcel.2010.05.004>
  17. Bailey TL, Johnson J, Grant CE et al (2015) The MEME suite. *Nucleic Acids Res* 43:W39–W49. <https://doi.org/10.1093/nar/gkv416>
  18. Fornes O, Castro-Mondragon JA, Khan A et al (2020) JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 48:D87–D92. <https://doi.org/10.1093/nar/gkz1001>
  19. Oki S, Ohta T, Shioi G et al (2018) ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep* 19:e46255. <https://doi.org/10.15252/embr.201846255>
  20. Hino S, Sakamoto A, Nagaoka K et al (2012) FAD-dependent lysine-specific demethylase-1 regulates cellular energy expenditure. *Nat Commun* 3:758. <https://doi.org/10.1038/ncomms1755>



## ATAC-Seq Analysis of Accessible Chromatin: From Experimental Steps to Data Analysis

Mayu Tatara, Taiyo Ikeda, Satoshi H. Namekawa, and So Maezawa

### Abstract

Accessible chromatin often represents gene regulatory elements, including promoters and enhancers, essential for gene expression. Assay for Transposase Accessible Chromatin sequencing (ATAC-seq) is one of the most popular techniques to investigate chromatin accessibility across the genome. Here we describe, step by step, a series of optimized experimental methods and bioinformatics pipelines for ATAC-seq analysis. As an example, we present an analysis of murine spermatogenic cells: a method to isolate germ cells, a reaction step using Tn5 transposase to insert sequencing adapters into accessible DNA, a library preparation method for high-throughput sequencing, and bioinformatics analysis of sequencing data. Overall, we introduce a framework of ATAC-seq analysis that can be applied to any cell population to identify cell-type-specific gene regulatory elements and their cis-regulatory networks.

**Key words** ATAC-seq, Germ cells, Accessible chromatin, Spermatogenesis, Meiosis

---

### 1 Introduction

Eukaryotic chromatin is packaged in an array of nucleosomes; each consists of 8 histone proteins and approximately 146 base pairs of DNA [1, 2]. Gene expression requires an accessible DNA region to which transcription factors and machinery must directly bind to function. Thus, gene regulatory elements—enhancers and promoters—of active genes often locate on “open” and accessible chromatin regions. In contrast, transcriptionally repressed genes are located within “closed” chromatin regions that are refractory to the binding of transcriptional regulators [3–6]. These open or closed chromatin states are often altered to define cell-type-specific gene expression.

Genome-wide profiling of chromatin accessibility is a powerful approach to elucidate a repertoire of gene regulatory elements that underpin the gene expression program of given cells. Various assays have been developed to analyze chromatin accessibility, including

Assay for Transposase Accessible Chromatin sequencing (ATAC-seq) [7, 8], DNase I hypersensitive sites sequencing (DNase-seq) [9–11], and Formaldehyde-Assisted Isolation of Regulatory Elements sequencing (FAIRE-seq) [12]. To date, among these assays, ATAC-seq has become the most popular technique because it requires simple experimental steps, fewer starting materials, and lower sequencing costs due to a higher signal-to-background ratio. Briefly, ATAC-seq utilizes a genetically engineered hyperactive Tn5 transposase that simultaneously cuts accessible chromatin regions and tags with sequencing adaptors. This simple enzymatic reaction and the subsequent direct library construction step maximize sensitivity to detect open chromatin genome-wide, enabling analysis of rare cell populations [13–15].

Recently, an improved ATAC-seq protocol, “Omni-ATAC,” has been developed [16]. It can be applied to a broad range of cell lines, tissues, and archival frozen samples without customized optimization. The addition of several detergents such as NP-40, Tween-20, and digitonin to both cell lysis and reaction buffers improves permeabilization across a wide array of cell types and removes mitochondria DNA contamination from the transposition reaction, resulting in substantial improvement of a signal-to-background ratio and reduction of sequencing costs. Of note, Omni-ATAC is compatible with lower-input materials (approx. 500 cells), which enables analyses of unmanageable cell lines, rare primary cells, and clinically relevant frozen tissues. ATAC-seq has been further developed to single-cell open chromatin analysis that overcomes the issue of cellular heterogeneity [17, 18].

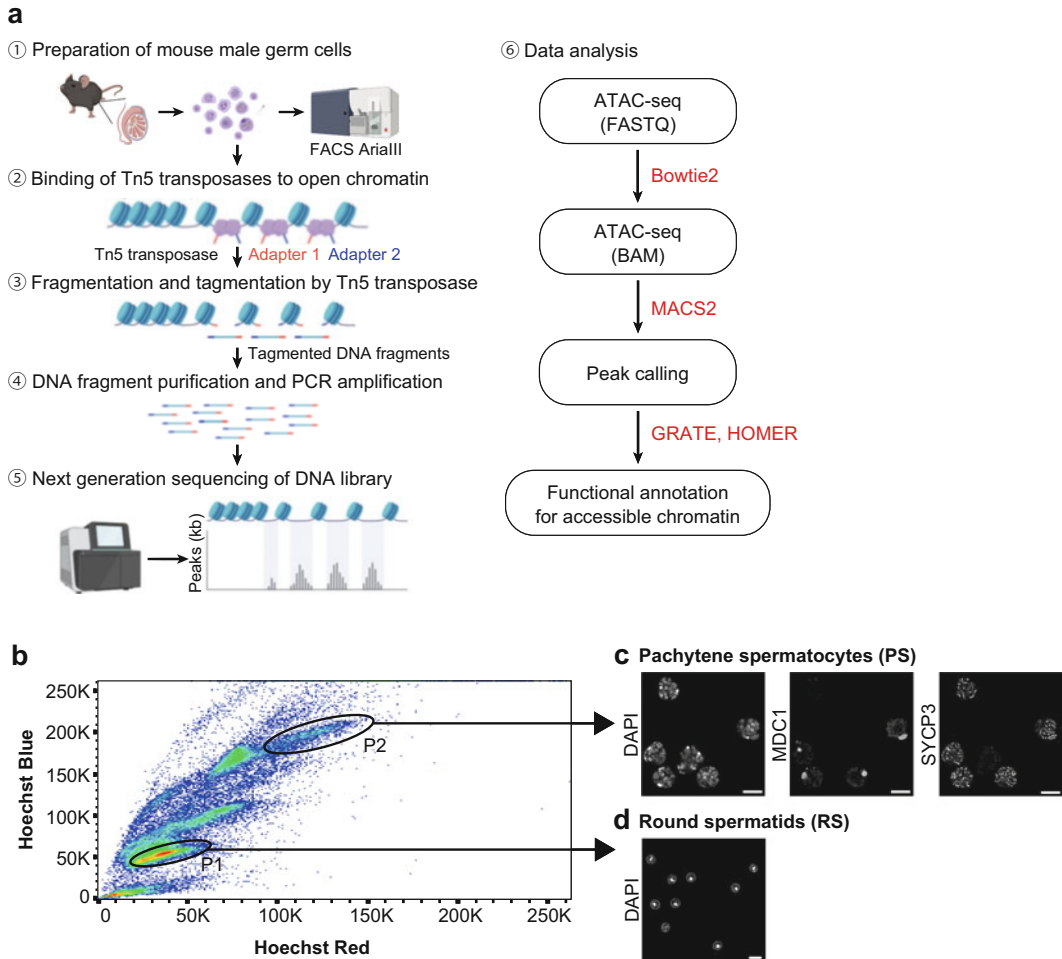
Previously, we performed ATAC-seq [16] analysis of murine spermatogenic cells and reported dynamic reorganization of open chromatin at the mitosis-to-meiosis transition [13]. Drawing on this study, we further improved the sensitivity to detect accessible chromatin. Here, we describe our optimized experimental methods and bioinformatics pipelines for Omni-ATAC analysis step-by-step (Fig. 1a). These methods can be applied to any context other than spermatogenesis and allow users to detect cell-type-specific gene regulatory elements and their cis-regulatory networks.

---

## 2 Materials

### 2.1 Reagents for Preparation of Mouse Male Germ Cells

1. DMEM (WAKO, Cat #043-30085).
2. Collagenase type IV (Worthington, Cat #CLS4).
3. Trypsin (Sigma, Cat #T4799).
4. DNase I (WAKO, Cat #047-26771).
5. Fetal bovine serum (FBS) (Thermo Fisher, Cat #10437-028).



**Fig. 1** Workflow of the Omni-ATAC method, bioinformatics pipelines, and FACS analysis of adult murine testicular cells based on Hoechst 33342. **(a)** The workflow of the method and bioinformatics pipelines for the Omni-ATAC analysis. The panel was created with BioRender (<http://biorender.com/>). **(b)** Hoechst profile of testicular cells. Gating on 1C and 4C testicular populations based on the Hoechst Blue/Hoechst Red contour plot. P1 represents the round spermatids (RS) population; P2 represents the pachytene spermatocyte (PS) population. **(c)** Immunofluorescence of sorted pachytene spermatocytes and round spermatids. Confirmation of PS by immunostaining for SYCP3 and MDC1 (upper panel). Representative DAPI image showing typical nuclear morphology of RS (lower panel). Scale bars: 10  $\mu$ m

## 2.2 *Tn5* Transposase

ATAC-seq libraries can be prepared with a commercially available kit (Nextera XT, Illumina) or in-house Tn5 as described previously [7, 19].

## 2.3 Other Reagents

1. Tris(hydroxymethyl)aminomethane (STAR, Cat #RSP-THA500G).
2. Sodium chloride.
3. Magnesium chloride.

4. Nuclease-free H<sub>2</sub>O.
5. NP-40.
6. Tween-20.
7. Digitonin (Promega, Cat #G9441).
8. 1× D-PBS.
9. Tagment DNA Enzyme 1 (TDE1) (Illumina, Cat #15027865).
10. Tagment DNA Buffer (Illumina, Cat #15027866).
11. MinElute Reaction Cleanup Kit (Qiagen, Cat #28204).
12. Primers (complete list of primers available in Sup. Table 1 of [7]).
13. NEBNext Ultra II Q5 Master Mix (NEB, Cat #M0544L).
14. SYBR Green I (Thermo Fisher, Cat #S7563).
15. Agencourt AMPure XP magnetic beads (Beckman Coulter, Cat #A63880).
16. 80% ethanol (made fresh).
17. Qubit dsDNA HS Assay Kit (Thermo Fisher, Cat #Q32851).
18. FastGene Gel/PCR Extraction Kit (Nippon Genetics, Cat #FG-91202).
19. MultiNA DNA-2500 Reagent Kit (Shimadzu, P/N 292-27912-91).

## 2.4 Buffers

1. ATAC-Resuspension Buffer (RSB).

RSB prepared here is used for both Lysis Buffer and Wash Buffer as below. RSB can be stored at room temperature long-term.

Reagent	Final concentration	Volume for 50 mL
1-M Tris-HCl, pH 7.4	10 mM	500 µL
5-M NaCl	10 mM	100 µL
1-M MgCl <sub>2</sub>	3 mM	150 µL
Nuclease-free H <sub>2</sub> O	–	49.25 mL

2. Lysis Buffer.

Reagent	Final concentration	Volume for 50 µL
RSB	–	48.5 µL
10% NP-40	0.1%	0.5 µL
10% Tween-20	0.1%	0.5 µL
1% digitonin	0.01%	0.5 µL

## 3. Wash Buffer.

Reagent	Final concentration	Volume for 1 mL
RSB	–	990 $\mu$ L
10% Tween-20	0.1%	10 $\mu$ L

4. 2 $\times$  Tagment Buffer (TD).

Reagent	Final concentration	Volume for 1 mL
1-M Tris-HCl, pH 7.4	20 mM	2 mL
1-M MgCl <sub>2</sub>	10 mM	1 mL
Dimethyl formamide	20%	20 mL
Nuclease-free H <sub>2</sub> O	–	Bring up to 100 mL

**2.5 Equipment**

1. Low retention microtube (e.g., 1.5-mL *microtubes*, “MAXY-Mum Recovery,” Axygen).
2. Centrifuge (e.g., Kubota 3740 with swing bucket rotor (*see Note 1*)).
3. Fluorescence-activated cell sorting (FACS) instrument (e.g., BD FACSAria III, BD Biosciences).
4. Qubit fluorometer (Thermo Fisher).
5. PCR Thermal Cycler (e.g., Life ECO, NIPPON Genetics).
6. qPCR system (e.g., AriaMx Real-Time PCR system, Agilent Technologies).
7. Magnetic stand (e.g., DiaMag 0.2 mL, Diagenode).
8. General electrophoresis system for agarose gel electrophoresis.
9. Gel imaging system (e.g., iBright CL1500 imaging system, Thermo Fisher Scientific).
10. MCE-202 MultiNA (Shimadzu).

**2.6 Software**

1. Bowtie2 v2.4.4 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml/>) [20].
2. Samtools v1.3.1 (<http://www.htslib.org/>) [21].
3. deepTools v3.1.3 (<https://deeptools.readthedocs.io/en/develop/>) [22].
4. IGV v2.9.2 (<https://software.broadinstitute.org/software/igv/>) [23].
5. MACS2 v2.1.4 (<https://pypi.org/project/MACS2/2.1.4/>) [24].
6. GREAT v4.0.4 (<http://great.stanford.edu/public/html/>) [25].



7. HOMER v4.10 (<http://homer.ucsd.edu/homer/>) [26].
8. ngsplot v2.61 (<https://github.com/shenlab-sinai/ngsplot>) [27].

### 2.7 Reagent Setup

1. 1-M Tris-HCl (pH 7.4): Dissolve 60.6 g of Tris in 400 mL of ddH<sub>2</sub>O and adjust pH to 7.4 with hydrochloric acid. Make up to the final volume of 500 mL with ddH<sub>2</sub>O and sterilize by autoclave. It can be stored at room temperature (RT).
2. 5-M NaCl: Dissolve 146 g of NaCl in 400 mL of H<sub>2</sub>O. Adjust the volume to 500 mL with H<sub>2</sub>O. Dispense into aliquots and sterilize by autoclaving. Store the NaCl solution at RT.
3. 1-M MgCl<sub>2</sub>: Dissolve 101.65 g of MgCl<sub>2</sub> · 6H<sub>2</sub>O in 400-mL H<sub>2</sub>O. Adjust the volume to 500 mL using H<sub>2</sub>O. Dispense into aliquots and sterilize by autoclaving. Store the MgCl<sub>2</sub> solution at RT.
4. 10% (100×) NP-40: Mix 10 mL of NP-40 with 90 mL of ddH<sub>2</sub>O by stirring. Store the 10% NP-40 solution at 4 °C.
5. 10% (100×) Tween-20: Mix 10 mL of Tween-20 with 90 mL of ddH<sub>2</sub>O by stirring. Store the 10% Tween-20 solution at 4 °C.
6. 1% (100×) digitonin: Dilute 1:1 with ddH<sub>2</sub>O to make 1% working stock, aliquot, and store at -20 °C for up to 6 months. Do not freeze/thaw more than five times.

---

## 3 Methods

### 3.1 Preparation of Mouse Male Germ Cells

Each developmental stage of germ cells can be isolated with the previously established FACS method based on nuclear staining with Hoechst 33342 [28] as described here or DyeCycle Violet (DCV) [29].

1. Collect the adult testis in a 15-mL tube in 4-mL DMEM with penicillin and streptomycin per two testes.
2. Remove the tunica albuginea membrane.
3. Add 4-mL DMEM containing 1-mg/mL collagenase type IV per two testes.
4. Incubate at 34 °C for 20 min.
5. Wash three times with 4-mL DMEM per two testes.
6. Add 4-mL DMEM containing 2.5-mg/mL trypsin, 1-mg/mL collagenase type IV, and 20-μg/mL DNase I per two testes.
7. Incubate at 34 °C for 15 min.
8. Add 10% fetal bovine serum (FBS) to stop the enzymatic reaction.

9. Filter the cell suspension with a 70- $\mu\text{m}$  nylon cell strainer and count cells (*see Note 2*).
10. Add 6  $\mu\text{g}$  of Hoechst 33342 per  $1.0 \times 10^6$  cells and incubate at 34 °C for 10 min.
11. Resuspend with DMEM containing 10% FBS and 20- $\mu\text{g}/\text{mL}$  DNase I (approx.  $5\text{--}10 \times 10^6$  cells/mL).
12. Filter the cell suspension with the 40- $\mu\text{m}$  nylon mesh (*see Note 2*), and sort with FACSAria III based on signals of Hoechst Blue, Hoechst Red, and cell size, as described in Gaysinskaya et al. [28] (Fig. 1b, c).

### 3.2 Cell Lysis and Transposition

By our optimized Omni-ATAC analysis using commercially available Tn5 [7, 16, 30], we have successfully detected accessible chromatin in spermatogenic cells with a high signal-to-noise ratio and ascertained our previous conclusion of the dynamic reorganization of accessible chromatin at the mitosis-to-meiosis transition in mouse spermatogenesis [13]. We use live, freshly isolated 50,000–100,000 cells. Pipette and centrifuge gently to avoid damaging/lysing cells before the cell lysis step (*see Note 3*).

1. Wash 50,000 cells with 50- $\mu\text{L}$  cold  $1 \times$  PBS in a 1.5-mL microtube; centrifuge at  $500 \times g$  for 5 min.
2. Aspirate all supernatant, carefully avoiding visible cell pellets.
3. Add 50- $\mu\text{L}$  cold  $2 \times$  Lysis Buffer and pipette up and down three times.
4. Incubate on ice for 3 min.
5. Wash out lysis with 1 mL of cold Wash Buffer and invert tube three times to mix (*see Note 4*).
6. Centrifuge at  $500 \times g$  for 10 min at 4 °C.
7. Aspirate all supernatant (cytoplasm), carefully avoiding visible cell pellet (nuclei), using two pipetting steps (aspirate down to 100  $\mu\text{L}$  with a p1000 pipette and remove final 100  $\mu\text{L}$  with a p200 pipette).
8. While cells are centrifuging, make the transposition reaction mix. If Tn5 from Nextera XT is used, the reaction mix is prepared as follows: Volumes per sample of 50,000–100,000 cells:

Reagent	Final concentration	Volume for 50 $\mu\text{L}$
$2 \times$ TD buffer	–	25 $\mu\text{L}$
$1 \times$ PBS	–	16.5 $\mu\text{L}$
10% Tween-20	0.1%	0.5 $\mu\text{L}$

(continued)

Reagent	Final concentration	Volume for 50 $\mu$ L
1% digitonin	0.01%	0.5 $\mu$ L
Tn5 transposase (Tagment DNA enzyme 1)	100 nM	2.5 $\mu$ L
Nuclease-free H <sub>2</sub> O	–	5 $\mu$ L

9. Add 50  $\mu$ L of the transposition reaction mix to the cell pellet and resuspend by pipetting up and down six times.
10. Incubate the reaction at 37 °C for 30 min in PCR Thermal Cycler (*see Note 5*).

### 3.3 DNA Purification

1. Clean up reaction with a FastGene Gel/PCR Extraction Kit (*see Note 6*).
2. Elute DNA in 22- $\mu$ L elution buffer (10-mM Tris buffer, pH 8) and store –20 °C until ready to amplify. This elution typically results in ~20  $\mu$ L of product.

### 3.4 PCR Amplification (Library Generation)

1. To amplify transposed DNA fragments, combine the following in a PCR tube:

Reagent	Final concentration	Volume for 50 $\mu$ L
Transposed DNA	–	19.7 $\mu$ L
25- $\mu$ M Nextera PCR Primer 1 <sup>a</sup>	1.25 $\mu$ M	2.5 $\mu$ L
25- $\mu$ M Nextera PCR Primer 2 with barcode <sup>a</sup>	1.25 $\mu$ M	2.5 $\mu$ L
100 $\times$ SYBR Green I <sup>b</sup>	–	0.5 $\mu$ L
NEBNext Ultra II Q5 Master Mix	–	25 $\mu$ L

<sup>a</sup>Complete list of primers available in Sup. Table 1 of [7]

<sup>b</sup>10,000 $\times$  SYBR Green is diluted in 10-mM Tris buffer, pH 8, to make a 100 $\times$  working solution

2. PCR cycling parameters are as follows:
  - Cycle 1: 72 °C for 5 min
  - Cycle 2: 98 °C for 30 s
  - Cycle 3: 98 °C for 10 s
  - Cycle 4: 63 °C for 30 s
  - Cycle 5: 72 °C for 30 s
  - Repeat Cycles 3–5 five times.

3. To reduce a GC and size bias in PCR, the PCR reaction is monitored using qPCR to stop amplification prior to saturation. To run a qPCR side reaction, combine the following:

Reagent	Volume for 15 $\mu$ L
5 cycles PCR-amplified DNA <sup>a</sup>	5 $\mu$ L
Nuclease-free H <sub>2</sub> O	4.4 $\mu$ L
25- $\mu$ M Nextera PCR primer 1	0.25 $\mu$ L
25- $\mu$ M Nextera PCR primer 2 with barcode <sup>a</sup>	0.25 $\mu$ L
100 $\times$ SYBR green I <sup>b</sup>	0.1 $\mu$ L
NEBNext ultra II Q5 master mix	5 $\mu$ L

<sup>a</sup>Use 5- $\mu$ L PCR-amplified DNA in **step 2** and keep the remaining 45  $\mu$ L for **step 6**

<sup>b</sup>10,000 $\times$  SYBR Green is diluted in 10-mM Tris buffer, pH 8, to make a 100 $\times$  working solution

4. qPCR cycling parameters are as follows:

Cycle 1: 98 °C for 30 s

Cycle 2: 98 °C for 10 s

Cycle 3: 63 °C for 30 s

Cycle 4: 72 °C for 30 s

Repeat Cycles 2–4 20 times.

5. Plot R vs. cycle number. Calculate the number of additional PCR cycles that correspond to 1/4 of maximum fluorescent intensity (Fig. 2a).

- If the number of cycles to be added lies in between two cycles, the number is determined by taking the smaller integer as the number of cycles to be added.
- If two samples have similar Ct values but differ in the fluorescent intensities, calculate the number of cycles using the sample with lower fluorescent intensity.

6. Continue PCR on remaining 45  $\mu$ L of each partially amplified library for the appropriate number ( $N$ ) of cycles:

Cycle 1: 98 °C for 30 s

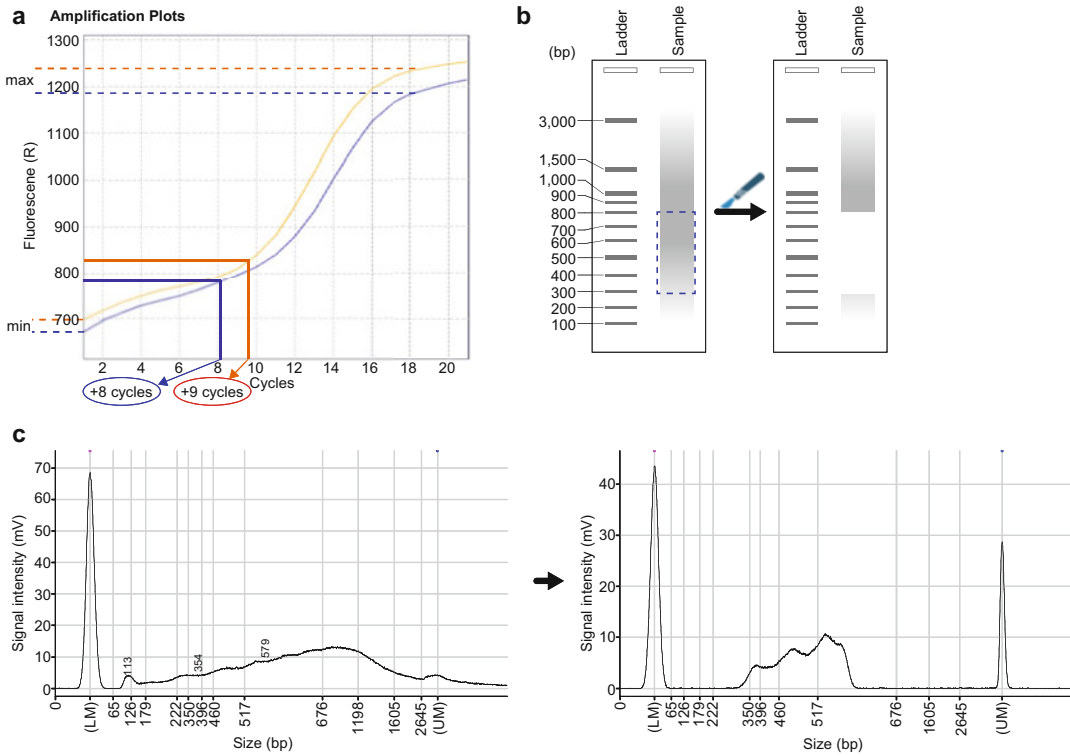
Cycle 2: 98 °C for 10 s

Cycle 3: 63 °C for 30 s

Cycle 4: 72 °C for 30 s

Cycle 5: 72 °C for 5 min

Repeat Cycles 2–4  $\times N$  times (*see Note 7*).



**Fig. 2** Determining the correct number of cycles for PCR and the size distribution of libraries. **(a)** Calculation of the additional number of PCR cycles for the final enrichment of each library. PCR amplification curves for two different libraries are shown in yellow and blue. Calculate the additional number of cycles corresponding to 1/4 of maximum fluorescent intensity. The yellow library requires eight additional amplification cycles, while the blue library requires nine cycles. **(b)** Size selection of libraries on a 1.5% agarose gel. Libraries are excised between 300 bp and 800 bp (right). The 100-bp DNA Ladder is shown on the left. **(c)** Output generated by the MultiNA Viewer software. Electropherogram of a library after purified AMPure XP magnetic beads (left panel). Electropherogram of a library after size selection on a 1.5% agarose gel. The overall size range corresponds to excised fragments shown in **(b)** (right panel)

### 3.5 Library Purification

1. Warm AMPure XP beads to RT, and vortex to resuspend.
  - For single left-sided bead purification (to remove primer dimers) and agarose gel electrophoresis followed by size selection:
    1. Transfer each PCR sample to a 200- $\mu$ L PCR tube, add 1.8 $\times$  volume AMPure beads, and pipette up and down ten times to mix thoroughly.
    2. Incubate at RT for 10 min.
    3. Place PCR tubes in a magnetic rack for 5 min.
    4. Discard supernatant.
    5. Wash beads with 200- $\mu$ L 80% EtOH (freshly made).
    6. Place PCR tubes in a magnetic rack for 1 min.

7. Discard EtOH.
  8. Leave tube on the magnetic rack with the cap open for 10 min.
  9. Ensure all EtOH is removed.
  10. Resuspend beads in 20- $\mu$ L nuclease-free H<sub>2</sub>O; pipette up and down ten times to mix thoroughly.
  11. Place PCR tubes in the magnetic rack for 1–5 min.
  12. Transfer supernatant to a new tube.
  13. Use 1  $\mu$ L of each diluted library to measure DNA concentration by Qubit.
  14. For multiplex sequencing, libraries are pooled into one tube.
  15. The pooled DNA is separated by 1.5% agarose electrophoresis. Excise the target DNA band from 300 to 800 bp with a sharp scalpel under an LED light (Fig. 2b).
  16. Purify the pooled DNA using the Gel Extraction Kit and elute in Buffer EB (or 10-mM Tris-HCl).
- (Option) For double-sided bead purification (to remove primer dimers and large >1000-bp fragments):
    1. Transfer each PCR sample to a 200- $\mu$ L PCR tube, add 0.5 $\times$  volume AMPure XP beads, and pipette up and down ten times to mix thoroughly.
    2. Incubate at room temperature for 10 min.
    3. Place PCR tubes in the magnetic rack for 5 min.
    4. Transfer supernatant to a new PCR tube.
    5. Add 1.3 $\times$  original volume AMPure XP beads; pipette up and down ten times to mix thoroughly. (This results in a final 1.8 $\times$  bead buffer: sample ratio.)
    6. Incubate at room temperature for 10 min.
    7. Place PCR tubes in the magnetic rack for 5 min.
    8. Discard supernatant.
    9. Wash beads with 200- $\mu$ L 80% EtOH (freshly made).
    10. Place PCR tubes in a magnetic rack for 1 min.
    11. Discard EtOH.
    12. Leave tube on the magnetic rack with the cap open for 10 min.
    13. Ensure all EtOH is removed.
    14. Resuspend beads in 20- $\mu$ L nuclease-free H<sub>2</sub>O; pipette EtOH over beads ten times to mix thoroughly.

15. Place PCR tubes in the magnetic rack for 1–5 min.
  16. Transfer supernatant to a new tube.
2. Store purified libraries at  $-20^{\circ}\text{C}$ .

### 3.6 Assessing Library Quality and Sequencing

1. Add 1.5  $\mu\text{L}$  of the library pool to 3  $\mu\text{L}$  of the marker of MultiNA DNA-2500 Reagent Kit (to make 1:2 dilution) or equivalent (*see* **Note 8**).
2. Run 4.5  $\mu\text{L}$  of the mixture on MultiNA to obtain electropherograms of the library.
3. Determine the size distribution and concentration of libraries (Fig. 2c).
4. Perform 100-bp single-end (SE) or 150-bp paired-end (PE) Illumina sequencing on the barcoded libraries following the manufacturer's instructions (e.g., 100-bp SE on a HiSeq 2500).

### 3.7 ATAC-seq Alignment

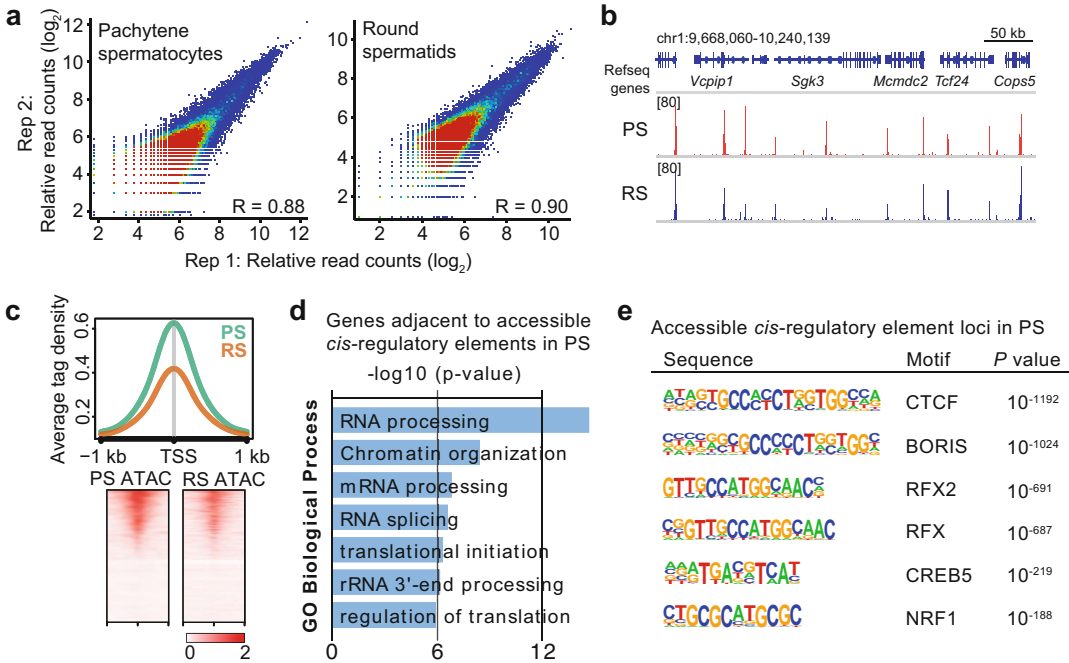
Here, we show the standard computational approaches for alignments of raw single-end reads in the FASTQ format from ATAC-seq samples to the mouse genome (GRCm38/mm10 (*see* **Note 9**)) using Bowtie2 (version 2.4.4) (*see* **Note 10**). After the alignment step, the files containing the aligned reads will be generated in the SAM format (.sam) and converted to the BAM format (.bam) using Samtools (version 1.3.1). The following commands can be executed in the command-line UNIX/Linux interface, Terminal for Mac OS X, or Cygwin for Windows.

1. A FASTA file (.fasta) containing the GRCm38/mm10 version of *Mus musculus* genome can be downloaded from the UCSC Genome Browser at <https://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/mm10.fa.gz> and decompressed.
2. Create a Bowtie2 genome index file, being saved as a "mm10\_bowtie2\_index" file.

```
$ mkdir ./genome/mm10/bowtie2-index
$ bowtie2-build -f mm10.fa mm10_bowtie2_index
```

3. Raw single-end reads of ATAC-seq data are aligned to the mouse genome (mm10) using Bowtie2 (version 2.4.4) with the option of number of threads set to 4 ( $-p 4$ ). The aligned reads to the reference genome are saved as a SAM file named "alignedFile.sam" (-S). The alignment details are stored in a file named "inputFile.log."

```
$ bowtie2 -p 4 -x ./genome/mm10/bowtie2-index/mm10_bowtie2_index -U inputATAC-seqFile.fastq.gz -S alignedFile.sam 2> inputFile.log
```



**Fig. 3** Characterization of accessible *cis*-regulatory elements using annotation tools. (a) Scatter plots show the reproducibility of ATAC-seq peaks at individual peaks between biological replicates. Each peak was identified using MACS ( $P < 1 \times 10^{-5}$ ). Enrichments of ATAC-seq reads are shown with  $\log_2$  RPKM values. The color scale indicates ATAC-seq peak density. Pearson correlation values (R) are shown. (b) Track views of enrichment of ATAC-seq read. (c) An average tag density plot and heatmaps around TSS ( $\pm 1$  kb) for the enrichment of ATAC-seq reads in PS and RS. (d) Gene ontology (GO) analysis of genes adjacent to accessible *cis*-regulatory elements in PS, using GREAT web tool (<http://great.stanford.edu/public/html/>). (e) HOMER motif analysis for putative transcription-factor-binding sites within accessible *cis*-regulatory elements in PS

- Convert aligned files from the SAM to BAM format (binary version of the SAM format) using Samtools (version 1.3.1).

```
$ samtools sort alignedFile.sam -o alignedFile.sorted.bam
```

- Remove PCR duplicates (*see Note 11*).

```
$ samtools rmdup alignedFile.sorted.bam alignedFile.sorted.rmdup.bam
```

- Index the resulting .bam files (*see Note 12*).

```
$ samtools index alignedFile.sorted.rmdup.bam
```

### 3.8 Confirmation of Reproducibility

Pearson correlations for the genome-wide enrichment of peaks corresponding to open chromatin regions among ATAC-seq library replicates were analyzed using SeqMonk (Babraham Institute) (Fig. 3a).



### 3.9 Data Visualization Through BAM Coverage Tracks

To visualize ATAC-seq signals throughout the genome, the aligned BAM file should first be converted into *the bigWig* format using the `bamCoverage` program as implemented in `deepTools` (version 3.1.3). To compare tracks between different samples, the sequencing depth in each sample is normalized with CPM (count reads per bin/million mapped reads).

1. Generate CPM-normalized coverage tracks (the *bigwig* (*bw*) format) using the `bamCoverage` program as implemented in `deepTools`.

```
$ bamCoverage --bam alignedFile.sorted.rmdup.bam -o aligned-
File.rmdup.bw -of bigwig --binSize=10 --minMappingQuality
10 --normalizeUsing CPM --numberOfProcessors=max
```

2. The converted *bigWig* files named “alignedFile.rmdup.bw” can be loaded directly into IGV to visualize the aligned reads (Fig. 3b).

### 3.10 Peak Calling with MACS2

ATAC-seq peaks in each sample are called using MACS2 v2.1.4 (Model-based Analysis of ChIP-seq) with a threshold q-value format. MACS2 reports all ATAC-seq peaks with q-value below a threshold in the BED format. The outputs of MACS2 peak calling are saved to a user-specified directory (`--outdir`) with a prefix name (`-n`, “alignedFile\_macs2”).

```
$ mkdir ./macs2
$ macs2 callpeak -t alignedFile.sorted.rmdup.bam -f BAM \
--nomodel -g mm -n alignedFile_macs2
2> macs2/alignedFile_macs2.log
```

### 3.11 Data Visualization Using `ngs.plot`

To visualize enrichment levels for ATAC-seq signals at specific genomic features or regions, those levels were calculated with `ngs.plot` (<https://github.com/shenlab-sinai/ngsplot>). `ngs.plot` is to perform a visual inspection between the enrichment of sequence alignments with respect to specific genomic features or regions. Average tag density profiles were calculated around TSSs with `ngs.plot` (Fig. 3c). A BED format file obtained in Subheading 3.10 containing three mandatory fields—chromosome, start position, and end position for the ATAC-seq peaks—was used as an input. Enrichment levels for ATAC-seq experiments were calculated for 2-kb windows promoter regions for genes ( $\pm 1$ -kb surrounding TSSs).

### 3.12 GREAT Analysis

The biological significance of genomic regions of ATAC-seq peaks can be predicted by the Genomic Region Enrichment of Annotation Tool (GREAT, version 4.0.4), which assigns biological meaning to a set of non-coding genomic regions by analyzing the annotations of the nearby genes (Fig. 3d). According to the manual

of GREAT (<http://great.stanford.edu/public/html/>), a BED format file was used as an input like Subheading 3.11.

### 3.13 De Novo Motif Analysis

Enrichment of consensus binding motifs for transcription factors on open chromatin regions can be computed by Hypergeometric Optimization of Motif EnRichment (HOMER, version 4.10). HOMER requires genomic coordinates of the BED format. The following commands and procedures are applicable for HOMER de novo motif analysis.

1. Install HOMER package with conda.

```
$ conda install -c bioconda homer
```

2. Install mouse genome and annotations for all HOMER analyses.

```
$ perl /path-to-homer-4.10-0/configureHomer.pl \
-install mm10
```

3. Run “findMotifsGenome.pl” program to find de novo motif.

```
$ findMotifsGenome.pl input.bed mm10 \
output_MOTIF -size given -mask
```

We observed the enrichment of the binding motif for the Regulatory Factor X (RFX) family that is implicated in gene regulation during spermatogenesis (Fig. 3e).

---

## 4 Notes

1. To reduce the loss of cells in any steps of centrifugation, swing-out rotors are better than fixed-angle rotors.
2. Filter cells gently because spermatocytes are fragile.
3. To compare ATAC-seq data among multiple samples, process samples simultaneously to reduce technical variation.
4. To confirm cell permeabilization, take a suitable aliquot of the cells, stain them with either DAPI or propidium iodide (PI), and observe nuclei (stained well without nuclei collapse).
5. Heat beforehand. Incubation time is critical.
6. You can use equivalent products to purify DNA.
7. The total number of PCR cycles should be under 15 when 50,000 cells are used.
8. Thaw frozen samples on ice before this step.
9. Remember the version of the reference genome. The version should match with the following analysis using the software listed above.

10. Quality control is recommended to remove low-quality reads before mapping steps.
11. PCR duplicates should be removed to avoid introducing PCR bias into data analysis.
12. To confirm that mitochondrial DNA contamination rate is under 3%, assess the number of reads mapped to the mitochondrial genome (chrM) by running “Samtools idxstats” on the sorted BAM files.

---

## Acknowledgments

We thank members of the Maezawa lab at Tokyo University of Science and Namekawa lab at UC Davis for discussions and helpful comments regarding the manuscript and Jasmine Esparza and Yasuhisa Munakata for critical reading of the manuscript. This work was supported by JSPS KAKENHI Grant Numbers JP 21H02383 and JP 21H00233, the Takeda Science Foundation, the Tokyo Biochemical Research Foundation, the Uehara Memorial Foundation to SM, and NIH grants R01 GM122776 and R35 GM141085 to SHN. We also extend acknowledgment to BioRender (<http://biorender.com/>) for the creation of images.

## References

1. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ (1997) Crystal structure of the nucleosome core particle at 2.8  resolution. *Nature* 389:251–260
2. Richmond TJ, Davey CA (2003) The structure of DNA in the nucleosome core. *Nature* 423:145–150
3. Lee CK, Shibata Y, Rao B, Strahl BD, Lieb JD (2004) Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat Genet* 36:900–905
4. Oszolak F, Song JS, Liu XS, Fisher DE (2007) High-throughput mapping of the chromatin structure of human promoters. *Nat Biotechnol* 25:244–248
5. Sheffield NC, Furey TS (2012) Identifying and characterizing regulatory sequences in the human genome with chromatin accessibility assays. *Genes* 3:651–670
6. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, Sandstrom R, Ma Z, Davis C, Pope BD et al (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature* 515:355–364
7. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10:1213–1218
8. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* 21(29):1–21.29.9
9. Song L, Crawford GE (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* 5:1–12
10. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B et al (2012) The accessible chromatin landscape of the human genome. *Nature* 489:75–82
11. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE (2008) High-resolution mapping and characterization of open chromatin across the genome. *Cell* 132:311–322
12. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD (2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements)

- isolates active regulatory elements from human chromatin. *Genome Res* 17:877–885
13. Maezawa S, Yukawa M, Alavattam KG, Barski A, Namekawa SH (2018) Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. *Nucleic Acids Res* 46:593–608
  14. Sin HS, Kartashov AV, Hasegawa K, Barski A, Namekawa SH (2015) Poised chromatin and bivalent domains facilitate the mitosis-to-meiosis transition in the male germline. *BMC Biol* 13:53
  15. Maezawa S, Sakashita A, Yukawa M, Chen X, Takahashi K, Alavattam KG, Nakata I, Weirauch MT, Barski A, Namekawa SH (2020) Super-enhancer switching drives a burst in gene expression at the mitosis-to-meiosis transition. *Nat Struct Mol Biol* 27:978–988
  16. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, Satpathy AT, Rubin AJ, Montine KS, Wu B et al (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* 14:959–962
  17. Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB, Filippova GN, Huang X, Christiansen L, DeWitt WS et al (2018) A single-cell atlas of in vivo mammalian chromatin accessibility. *Cell* 174:1–16
  18. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, Silva TC, Groeneveld C, Wong CK, Cho SW et al (2018) The chromatin accessibility landscape of primary human cancers. *Science* 362:eaav1898
  19. Picelli S, Björklund ÅK, Reinius B, Sagasser S, Winberg G, Sandberg R (2014) Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* 24:2033–2040
  20. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359
  21. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079
  22. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T (2014) DeepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42:187–191
  23. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nat Biotechnol* 29:24–26
  24. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137
  25. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28:495–501
  26. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589
  27. Shen L, Shao N, Liu X, Nestler E (2014) ngs.plot Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 15:284
  28. Gaysinskaya V, Soh IY, van der Heijden GW, Bortvin A (2014) Optimized flow cytometry isolation of murine spermatocytes. *Cytom Part A* 85:556–565
  29. Yeh YH, Hu M, Nakagawa T, Sakashita A, Yoshida S, Maezawa S, Namekawa SH (2016) Isolation of murine spermatogenic cells using a violet-excited cell-permeable DNA binding dye. *Physiol Behav* 176:139–148
  30. Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH (2016) Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol Metab* 5:233–244



## Low-Input CUT&RUN for Mouse Oocytes and Preimplantation Embryos

Ryoya Hayashi and Azusa Inoue

### Abstract

Cleavage Under Target & Release Using Nuclease (CUT&RUN) enables the detection of DNA regions that are bound by a protein of interest. This method is suitable for low-input materials because of the absence of an immunoprecipitation step. However, it sometimes fails when applying it to fragile cells, such as mouse oocytes. Here we describe our low-input CUT&RUN protocol optimized for mouse oocyte and preimplantation embryo samples in which the primary antibody and protein A-MNase binding steps are completed before the cells are bound to Concanavalin A-coated magnetic beads. This modification prevents crush of oocytes and early embryos and unwanted loss of chromatin during CUT&RUN procedures.

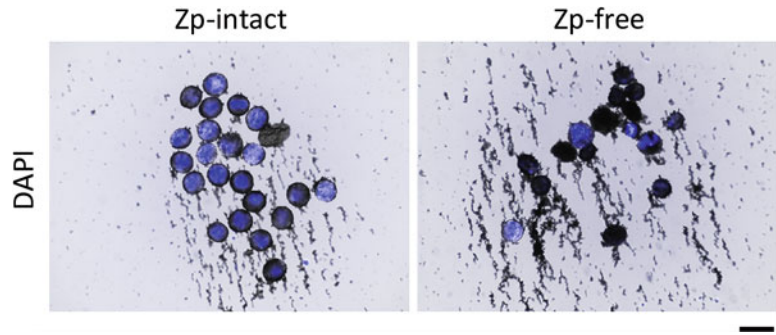
**Key words** Epigenetics, Histone modification, Oocytes, Preimplantation embryo

---

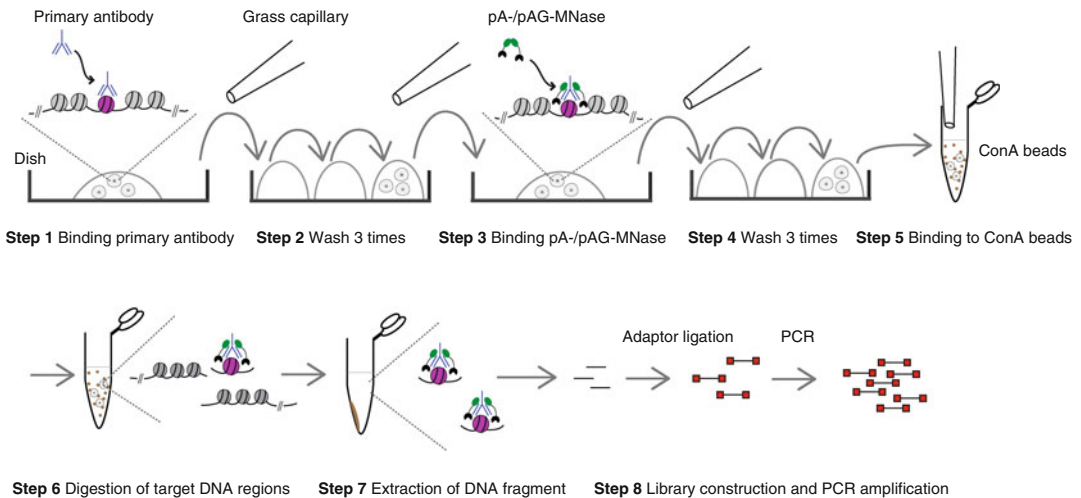
### 1 Introduction

Cleavage Under Target & Release Using Nuclease (CUT&RUN) is a chromatin profiling method in which protein A- or G-conjugated micrococcal nuclease (MNase) cleaves DNA regions that are bound by an antibody against a protein of interest [1, 2]. Purification of shorter DNA fragments followed by PCR amplification and next-generation sequencing allows genome-wide profiling of the chromatin localization of the protein of interest. This method is suitable for small numbers of *in vivo* cell materials because of the absence of an immunoprecipitation step which often causes DNA loss and high background noise.

To study epigenetic regulation in mouse oocytes and preimplantation embryos, we have used CUT&RUN for the recent few years [3–5]. However, we have realized that the amount of DNA recovered from oocytes and early preimplantation embryos varies greatly from experiment to experiment even by a protocol that stably works well for embryonic stem cells (ESCs) and late stages



**Fig. 1** Detection of DNA in zona pellucida (ZP)-intact and ZP-free oocytes after multiple washes in the standard CUT&RUN procedure. Twenty-two ZP-intact and ZP-free GV oocytes were captured by Concanavalin A-coated (ConA) beads, respectively, and washed several times with Dig-wash buffer. After ConA beads clumps were recovered by a glass capillary under a microscopy, they were stained with DAPI (blue). Note that DNA is undetectable in many of the clumps in the ZP-free oocyte sample, indicating loss of chromatin during wash steps. Scale bars, 100  $\mu$ m



**Fig. 2** Flowchart of the steps of the modified CUT&RUN method

of preimplantation embryos. By careful examination of this cause, we found that oocytes and embryos after removal of zona pellucida (ZP) (ZP-free oocytes/embryos) were fragile and frequently crushed during repeated wash steps of the CUT&RUN procedure (Fig. 1). We thought that cells with large sizes (~80  $\mu$ m in mouse oocytes) tend to be easily crushed when the cells captured by Concanavalin A-coated magnetic beads are retained in a small amount of liquid after the wash buffer is removed. Consequently, we guessed that the chromatin is leaked out and lost by wash steps. To prevent this, we modified the procedure (Fig. 2). This modified

method minimizes the variation between experiments and yields relatively constant amounts of DNA from a smaller number of oocytes and early preimplantation embryos.

---

## 2 Materials

Prepare all solutions using molecular biology grade water. To reduce the risk of DNA contamination, purchase ready-to-use reagents if commercially available. Store them in aliquots so that reagents can be discarded when contamination is found.

### 2.1 Reagents for Sampling of Oocytes and Preimplantation Embryos

1. Acidic Tyrode's solution.
2. 10% PVP (Irvine Scientific) or homemade: Dissolve PVP powder in M2 media overnight and filtrate it with 0.45- $\mu$ m pores.
3. 5-M NaCl.
4. 0.25% trypsin/0.1-mM EDTA.
5. M2 media.
6. PBS containing 0.2% BSA.
7. Cytochalasin B.
8. Colcemid.
9. 50-mm plastic dish.

### 2.2 Reagents

1. M HEPES, pH 7.5.
2. 5-M NaCl.
3. 1-M spermidine.
4. Roche complete Mini EDTA-free tablet.
5. 5% digitonin (*see Note 1*).
6. 0.5-M EDTA.
7. 0.5-M EGTA.
8. 10-mg/mL RNase A.
9. 20-mg/mL glycogen.
10. Concanavalin A-coated beads (Bangs Laboratories).
11. Primary antibody (*see Note 2*).
12. pA-/pAG-MNase (*see Note 3*).
13. 10% SDS solution.
14. Carrier RNA.
15. 20-mg/mL Proteinase K.
16. Phenol/Chloroform/Isoamyl alcohol, pH 8.0.
17. 100-mM CaCl<sub>2</sub>.

18. 3-M NaOAC.
19. 100% EtOH.
20. SPRiselect (Beckman).
21. NEBNext Ultra II DNA Library Prep Kit (NEB).
22. NEBNext Multiplex Oligos for Illumina (NEB).

### 2.3 Buffers

1. Binding buffer: 20-mM HEPES, pH 7.5, 10-mM KCl, 1-mM CaCl<sub>2</sub>, 1-mM MnCl<sub>2</sub>. Prepare 20 mL per experiment; store at 4 °C for 6 months.
2. Wash buffer: 20-mM HEPES, pH 7.5, 150-mM NaCl, 0.5-mM spermidine, one mini Roche complete EDTA-free tablet. Prepare 10 mL per experiment; store at 4 °C for 1 week.
3. Dig-wash buffer: 0.02% digitonin in wash buffer. Prepare 1.5 mL per sample; use within a day (*see Note 1*).
4. Antibody incubation buffer: 2-mM EDTA in Dig-wash buffer. Prepare 100 µL per sample; use within a day.
5. 10× STOP buffer: 1700-mM NaCl, 100-mM EDTA, 20-mM EGTA, 0.02% digitonin, 250-µg/mL RNase A, 500-µg/mL glycogen. Prepare 200 µL per ten samples; store at 4 °C for 1 week.
6. Elution buffer: 10-mM Tris-HCl, pH 8.5.

---

## 3 Methods

Keep environment clean to avoid contamination of DNA. Wipe pipettes, pens, and tube racks with DNA AWAY (Thermo Fisher). Work on a plastic wrap sheet on the bench. All Subheadings 3.1, 3.2, and 3.3 are performed on disposable plastic dish, and a glass capillary attached to a mouth pipette is used for transferring cells (*see Note 4*).

### 3.1 Removal of Zona Pellucida and Polar Bodies (See Note 5)

1. Remove zona pellucida (ZP) by brief treatment of acidic Tyrode's solution supplemented with 0.5% PVP and 50-mM NaCl (this supplementation allows gentle removal of ZP and reduces the frequency of cell lysis during the procedures), and wash three times in M2 media. Note that metaphase II (MII)-stage oocytes are melted in Ab buffer and cannot be handled by a glass capillary. Therefore, ZP should not be removed from MII oocytes.
2. When collecting one-cell stage zygotes, incubate ZP-free zygotes in M2 medium containing 5-µg/mL cytochalasin B and 0.1-µg/mL colcemid for 10–15 min at 37 °C. After incubation, remove polar bodies (PBs) by pipetting with a narrow



glass capillary. After PBs are removed, transfer the zygotes back to a culture medium without cytochalasin B and colcemid, and incubate at least 30 min so that cytoskeleton of zygotes is restored. This restoration step is critical for not being melted in Ab buffer.

3. When collecting two-cell and four-cell embryos, incubate ZP-free embryos in 0.25% trypsin/1-mM EDTA supplemented with 0.5% PVP and 50-mM NaCl for ~1 min at room temperature. Remove polar bodies by pipetting with a narrow glass capillary and wash three times in M2 media.
4. (Optional) Wash oocytes/embryos three times in 200- $\mu$ L drops of PBS containing 0.2% BSA. From this step, use new petri dish (*see Note 6*).
5. Transfer and wash oocytes/embryos two times in 200- $\mu$ L drops of Ab buffer with a thick glass capillary (a diameter of >200  $\mu$ m). Handle the cells gently from this step, because they are fragile in Ab buffer.

### **3.2 Bind Primary Antibody**

1. Transfer oocytes/embryos in a 100- $\mu$ L drop of Ab buffer containing 1  $\mu$ L of primary antibody (*see Note 7*).
2. Close the lid of the petri dish and incubate overnight at 4 °C (*see Note 8*).

### **3.3 Bind pA-/pAG-MNase**

1. Wash oocytes/embryos twice in 200- $\mu$ L drops of Dig-wash buffer. Use a newly prepared Dig-wash buffer on the day of use (*see Note 9*).
2. Close the lid and incubate for 20 min at room temperature.
3. Transfer oocytes/embryos in a 100- $\mu$ L drop of Dig-wash buffer containing 500 ng/ $\mu$ L of pA-/pAG-MNase.
4. Close the lid and incubate for 3 h at 4 °C.
5. Wash oocytes/embryos twice in 200- $\mu$ L drops of Dig-wash buffer.
6. Close the lid and incubate for 20 min at room temperature.

### **3.4 Prepare Magnetic Beads**

1. Thoroughly resuspend Bio-Mag Plus Concanavalin A-coated beads (ConA beads). Transfer 10  $\mu$ L of ConA beads per sample to 1 mL of binding buffer in a 1.5-mL tube and mix by vortexing.
2. Place the tube on a magnetic stand and wash once in 1 mL of binding buffer.
3. Place the tube on a magnetic stand and remove the supernatant. Add 50  $\mu$ L (per sample) of wash buffer and resuspend ConA beads by vortexing. Split 50  $\mu$ L into new 1.5-mL tubes ( $\times$  the number of samples).

### 3.5 Digestion

1. Transfer oocytes/embryos in the tube prepared at Subheading 3.4, step 3 and incubate for 10 min at room temperature (*see Note 10*).
2. Place the tube on a magnetic stand and remove the supernatant (*see Note 11*). Add 180  $\mu\text{L}$  of ice-cold Dig-wash buffer and mix by vortexing at weak setting for a few seconds.
3. Incubate for 2 min on ice.
4. Add 3.6  $\mu\text{L}$  of 100-mM  $\text{CaCl}_2$  and mix by vortexing at weak setting for 5–10 s. Place the tubes on ice.
5. Incubate for 20 min on ice.
6. Stop the digestion reaction by adding 20  $\mu\text{L}$  of 10 $\times$  STOP and vortexing at weak setting for 5–10 s.
7. Stand on ice for 5 min.

### 3.6 DNA Purification

1. Incubate the tubes at 37 °C for 20 min at 1000 rpm.
2. Place the tube on a magnetic stand, and transfer the supernatant to a new 1.5-mL LoBind tube.
3. Add 2  $\mu\text{L}$  of 10% SDS and 2.5  $\mu\text{L}$  of 20-mg/mL Proteinase K, and mix by tapping.
4. Incubate at 60 °C for 45 min followed by 72 °C for 20 min.
5. Add 20-ng carrier RNA.
6. Transfer the sample into a phase-lock tube, add 200  $\mu\text{L}$  of Phenol/Chloroform/Isoamyl alcohol (pH 8.0), and mix by vortexing at maximum setting for 15 s.
7. Stand the tube at room temperature for 3 min and centrifuge for 5 min at 16,000  $\times g$  at room temperature.
8. Take the supernatant to a new 1.5-mL LoBind tube, and add 20  $\mu\text{L}$  of 3-M NaOAc, 2  $\mu\text{L}$  of 20-mg/mL glycogen, and 750  $\mu\text{L}$  of 100% EtOH. Mix by vortexing at maximum setting for 15 s.
9. Incubate overnight at –20 °C (*see Note 12*).
10. Centrifuge the samples for 30 min at 16,000  $\times g$  at 4 °C and remove the supernatant. Wash the pellet once with 800  $\mu\text{L}$  of 80% EtOH (*see Notes 13 and 14*).
11. Centrifuge for 10 min at 16,000  $\times g$  at 4 °C and completely remove the supernatant (*see Note 15*). Dry in air for 5–10 min until the pellet is turned to semitransparent.
12. Dissolve the precipitates in 50- $\mu\text{L}$  water thoroughly by tapping (*see Note 16*).

### **3.7 End Repair and Adaptor Ligation (NEBNext Ultra II DNA Library Prep Kit)**

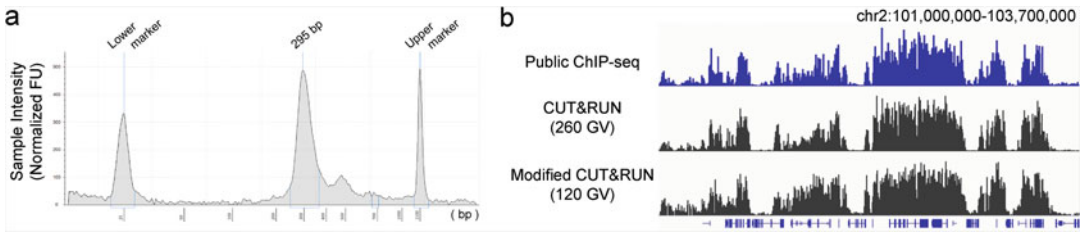
1. Transfer the sample to a PCR tube. Add 7.0  $\mu\text{L}$  of end repair buffer (*see Note 17*) and 3.0  $\mu\text{L}$  of End prep enzyme. Mix thoroughly by pipetting 15–20 times.
2. Incubate at 20 °C for 30 min followed by 65 °C for 30 min.
3. Add 2.5  $\mu\text{L}$  of NEB adaptor (1:12 diluted), 1.0  $\mu\text{L}$  of ligation enhancer, and 30  $\mu\text{L}$  of ligase master mix. Pipette the entire volume up and down 15–20 times to mix thoroughly.
4. Incubate at 20 °C for 2 h.
5. Add 3.0  $\mu\text{L}$  of USER enzyme, and mix thoroughly by pipetting 15–20 times.
6. Incubate at 37 °C for 15 min.

### **3.8 Clean Up the Adaptor-Ligated DNA**

1. Transfer the DNA sample to a new 1.5-mL LoBind tube containing  $\times 1.8$  (173.7  $\mu\text{L}$ ) of SPRI beads. Vortex thoroughly at maximum setting for 10–20 s, spin briefly, and leave it for 5 min at room temperature.
2. Place the tube on a magnetic stand for 5 min and remove the supernatant.
3. Add 180  $\mu\text{L}$  of freshly prepared 80% ethanol and stand for  $>30$  s. Then, remove the supernatant. Repeat three times.
4. Remove the supernatant completely (*see Note 18*); air-dry the beads for  $\sim 5$  min (*see Note 19*). Add 25  $\mu\text{L}$  of elution buffer (EB), and mix thoroughly by vortexing at maximum setting for 10 s.
5. Spin briefly and leave it for 5 min. Place the tube on a magnetic stand for 1 min and take the supernatant to a new PCR tube.

### **3.9 PCR Amplification and Library Purification**

1. (Optional) Determine the PCR cycle used in the next step. Dilute 0.5  $\mu\text{L}$  of the sample in 0.5-mL water ( $\times 1/1000$  dilution). Keep the rest of the sample ( $\sim 24.5$   $\mu\text{L}$ ) on ice. Amplify with Illumina universal primer and Illumina indexed primer (any index is allowed) according to a standard qPCR protocol. Use 1  $\mu\text{L}$  of the 1/1000 diluted sample as a template in a 10- $\mu\text{L}$  reaction (e.g., DW 3.4  $\mu\text{L}$ , primer mix 0.6  $\mu\text{L}$ , template 1  $\mu\text{L}$ ,  $2\times$ Mix 5  $\mu\text{L}$ ). The optimal PCR cycle is [Ct value  $-11$ ] (*see Note 20*).
2. Amplify the rest of the samples with indexed primers in a 52- $\mu\text{L}$  reaction: Primer mix 2  $\mu\text{L}$ , template 24.5  $\mu\text{L}$ , and  $2\times$  KAPA HiFi Mix 26  $\mu\text{L}$ . PCR condition; 95 °C, 2'; (98 °C, 20''; 60 °C, 30''; 72 °C, 60''); repeat for total [Ct value  $-11$ ]; 72 °C, 5'; 12 °C forever.



**Fig. 3** Evaluation of the library quality. **(a)** An example of the size distribution of DNA fragments measured by TapeStation. **(b)** Genome browser view of H3K27me3 distribution in GV oocytes in ChIP-seq [6], a standard CUT&RUN method [5], and the modified CUT&RUN method

3. After PCR amplification, add  $\times 0.9$  (46.8  $\mu\text{L}$ ) of SPRI beads. Then, follow the same procedure as **steps 1–5** in Subheading 3.8. At last, elute with 20  $\mu\text{L}$  of EB.
4. Check the library quality by Qubit and TapeStation (Fig. 3).

## 4 Notes

1. Dissolve digitonin powders in ultrapure water at 75  $^{\circ}\text{C}$  and make 20–50- $\mu\text{L}$  aliquots. On the day of use, incubate at 75  $^{\circ}\text{C}$  for 15 min to dissolve it completely.
2. It is critical to find an antibody suitable for low-input CUT&RUN. As positive controls, we recommend anti-H3K4me3 (Active Motif, #39159) and anti-H3K27me3 (Diagenode; #C15410069) antibodies, which work well in less than 200 oocytes.
3. When using mouse-raised antibodies, pG- or pAG-MNase should be used. Commercial products of pAG-MNase are available (Cell Signaling Technology; #40366S or Epicypher; #15-1016).
4. When handling hundreds of oocytes by a glass capillary, make sure to minimize the carry-over amount of solution.
5. The presence of zona pellucida did not affect the results of H3K27me3 CUT&RUN in our trial. However, to remove polar bodies, ZP must be removed. Furthermore, ZP removal allows avoiding contamination of cumulus cells and sperm.
6. Wash samples carefully under a microscopy to make sure that no cumulus cells or polar bodies are contaminated. Also, keep environment around a microscopy clean to avoid contamination of DNA.
7. We recommend 1:100 dilution of antibody as a default. A lower concentration of antibody often yields lower amount of DNA recovered.

8. We recommend overnight at 4 °C in low-input CUT&RUN.
9. Dig-wash buffer should be prepared just before use. Occasionally, digitonin may be precipitated. This precipitate may interfere with the handling of the samples by glass capillary.
10. Minimize the carryover of buffer solutions. Make sure that no cell remains in the glass capillary.
11. Before the supernatant is discarded, observe it under a microscope to make sure there is no cell dropped from the ConA beads.
12. If the centrifugation is started after 18 h of incubation at -20 °C, the pellet is occasionally bigger than usual potentially due to salt precipitation, which causes low efficiency in library construction. Therefore, we do not leave the samples at -20 °C more than 18 h but start centrifugation on the morning.
13. Samples after replacement with 80% ethanol can be stored at -20 °C up to a week.
14. Prepare 80% ethanol freshly. Store at RT.
15. Be careful that the precipitates are fragile. Remove supernatant carefully. We use a 1-mL tip to remove most of the supernatant, and then use a fin tip to completely remove it.
16. We recommend doing the following end repair and adaptor ligation steps in a 50- $\mu$ L scale according to the commercial instruction, because library construction occasionally fails if the scale is halved. This kind of failure at half usage is only seen in CUT&RUN, but not in other library constructions in our hands, for unknown reasons.
17. Thaw the buffer at 37 °C water bath particularly at the first use, and make sure that white small precipitates are completely dissolved.
18. We use a 1-mL tip to remove most of the supernatant, and then use a fin tip to completely remove it.
19. Dry until the bead clumps lose their luster and show tiny cracks.
20. The Ct value is expected to be less than 27. Over 27 often results in failure of the experiment. Such libraries will have many PCR duplicates and yield poor-quality data. In this case, increase the number of cells and/or change the antibody. Selection of good antibodies is important for low-input CUT&RUN.

---

## Acknowledgments

We thank Dr. Steven Henikoff for providing the protein A-MNase and Drs. Weikun Xia and Wei Xie for helpful discussion on CUT&RUN. This project was partly supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Leading Initiative for Excellent Young Researchers Grant to A.I., Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) (18H02359 to A.I.), Grant-in-Aid for Scientific Research on Innovative Areas (19H05754 to A.I.), Grant-in-Aid for JSPS Fellows (20J21541 to R.H.), Japan Agency for Medical Research and Development PRIME (JP18gm6110012 and 21gm6310022h0001 to A.I.), Uehara Memorial Foundation to A.I., and RIKEN Pioneering Project “Genome Building from TADs” to A.I.

## References

1. Skene PJ, Henikoff S (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* 6: e21856. <https://doi.org/10.7554/eLife.21856>
2. Meers MP, Bryson TD, Henikoff JG, Henikoff S (2019) Improved CUT&RUN chromatin profiling tools. *eLife* 8:e46314. <https://doi.org/10.7554/eLife.46314>
3. Inoue A, Jiang L, Lu F, Zhang Y (2017) Genomic imprinting of *Xist* by maternal H3K27me3. *Genes Dev* 31:1927–1932. <https://doi.org/10.1101/gad.304113.117>
4. Inoue A, Chen Z, Yin Q, Zhang Y (2018) Maternal *Ecd* knockout causes loss of H3K27me3 imprinting and random X inactivation in the extraembryonic cells. *Genes Dev* 32:1525–1536. <https://doi.org/10.1101/gad.318675.118>
5. Mei H, Kozuka C, Hayashi R, Kumon M, Koseki H, Inoue A (2021) H2AK119ub1 guides maternal inheritance and zygotic deposition of H3K27me3 in mouse embryos. *Nat Genet* 53: 539–550. <https://doi.org/10.1038/s41588-021-00820-3>
6. Zheng H, Huang B, Zhang B, Xiang Y, Du Z, Xu Q, Li Y, Wang Q, Ma J, Peng X, Xu F, Xie W (2016) Resetting epigenetic memory by reprogramming of histone modifications in mammals. *Mol Cell* 63:1066–1079. <https://doi.org/10.1016/j.molcel.2016.08.032>



## Imaging Chromatin Accessibility by Assay of Transposase-Accessible Chromatin with Visualization

Yusuke Miyanari

### Abstract

Chromatin accessibility is one of the fundamental structures regulating genome functions including transcription and DNA repair. Recent technological advantages to analyze chromatin accessibility begun to explore the dynamics of local chromatin structures. Here I describe protocols for Assay of Transposase-Accessible Chromatin with Visualization (ATAC-see), which allows us to analyze subnuclear localization of accessible chromatin and quantify accessible chromatin at single-cell level.

**Key words** Chromatin accessibility, ATAC-see, Tn5 transposase, Fluorescent microscopy

---

### 1 Introduction

While most of genomic regions are occupied by nucleosome, specific genomic regions are devoid of nucleosomal structure [1]. Since naked state of these region allows DNA-binding proteins to access their target sequences directly, all defined classes of active cis-regulatory elements including enhancers, promoters, silencers, insulators, and locus control regions display high chromatin accessibility [2]. On the other hand, heterochromatic regions containing silenced genes displayed condensed and inaccessible state. Recent advances have enabled genome-wide mapping of accessible chromatin in mammalian cells. The assay for transposase-accessible chromatin with sequencing (ATAC-seq) is a method of choice to investigate chromatin accessibility [3]. In this assay, hyperactive Tn5 transposase assembled with short adaptor DNA is loaded to permeabilized nucleus and preferentially binds to naked DNA, resulting in integration of adaptor DNA into accessible chromatin due to its transposase activity. Deep sequencing of transposed DNA enables us to identify accessible chromatin in a genome-wide man-

ner. Chromatin accessibility can be also visualized under microscopy by ATAC with visualization (ATAC-see), in which fluorescent adaptor DNA is used for in situ labeling of accessible chromatin [4]. Since chromatin accessibility is dynamically regulated during cellular events including transcription and cell cycle, imaging of chromatin structures would be a breakthrough toward uncovering the functional relevance and mechanisms regulating genome functions. I describe protocols for ATAC-see analyzed by microscopy or flow cytometer.

---

## 2 Materials

### 2.1 Preparation of Recombinant Tn5

1. T7 Express lysY/Iq competent *E. coli* (NEB, C3013).
2. 1000 mL of 2YT medium: To prepare 1 L of 2YT medium, dissolve 16 g of tryptone, 10 g of yeast extract, and 5 g of NaCl in 1 L water, and subsequently autoclave it at 120 °C for 10 min.
3. 100-mg/mL ampicillin.
4. 1-M isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG).
5. LB agar plate containing 0.1-mg/mL ampicillin.
6. HEG buffer: 20-mM HEPES-KOH at pH 7.2, 0.8-M NaCl, 1-mM EDTA, 10% glycerol.
7. 10% polyethylenimine.
8. 2 $\times$  Tn5 dialysis buffer; 100-mM HEPES (pH 7.2), 0.2-M NaCl, 0.2-mM EDTA, 2-mM DTT, 20% glycerol.
9. Lysis buffer; 1 $\times$  CellLytic B (Sigma), 1-mM PMSE, 1/10000 TurboNuclease (Accelagen) in HEG buffer.
10. Cleavage buffer: 100-mM DTT in HEG buffer.
11. Chitin beads (NEB, S6651).
12. Slide-A-Lyzer™ Dialysis Cassettes 20 K MWCO, 12 mL (Thermo Fisher, 66012).
13. 80% glycerol.
14. Amicon Ultra-15 10K (Millipore, UFC901008).
15. Empty gravity flow columns (Bio-Rad, 7321010).
16. pTXB1-Tn5 (addgene, #60240).

### 2.2 ATAC-see

1. Tn5 transposase: Tn5 transposase can be commercially purchased from Diagenode or purified from *E. coli* as described below [5].
2. Oligonucleotides for adaptor DNA: ME-fw-Cy5; [CY5] AGATGTGTATAAGAGACAG, ME-rev[Phos]; [PHO]



CTGTCTCTTATACACATCT. [CY5] and [PHO] indicate chemical modification of Cy5 and phosphorylation at 5' end, respectively (*see Note 1*).

3. 5× TAPS buffer: 50-mM N-tris(hydroxymethyl)methyl-3-aminopropanesulfonic acid (pH 8.5) and 25-mM MgCl<sub>2</sub>.
4. 10× annealing buffer: 400-mM Tris-HCl (pH 7.6), 500-mM NaCl.
5. NP buffer: 0.1% IGEPAL CA-630 (NP-40) in 1× TAPS buffer.
6. Wash buffer A: 0.1% IGEPAL CA-630 (NP-40) in PBS.
7. Cross-link buffer1: 0.5% formaldehyde in PBS.
8. Cross-link buffer2: 4% formaldehyde in PBS.
9. DMF (dimethylformamide).
10. 4',6-Diamidino-2-phenylindole (DAPI).
11. Vectashield containing DAPI (Vector) (*see Note 2*).

### 2.3 Cell Culture

1. L511 solution: 0.01-mg/mL Laminin-511 (BioLamina, LN-511) in PBS.
2. Culture medium: DMEM with GlutaMAX (Life Technologies, 10566-024), 10% fetal calf serum, and penicillin/streptomycin.
3. 0.25% trypsin solution.
4. PBS: phosphate-buffered saline. To prepare 1-L PBS, dissolve 8-g NaCl, 0.2-g KCl, 1.15-g Na<sub>2</sub>HPO<sub>4</sub> (anhydrous), and 0.2-g KH<sub>2</sub>PO<sub>4</sub> (anhydrous) into 1-L water, and subsequently autoclave PBS at 120 °C for 10 min.
5. Glass-bottom dishes.
6. Micro Culture-Insert 4 Well (ibidi, 80406).
7. NIH3T3 cells, mouse fibroblast-like cell line (*see Note 3*).

---

## 3 Methods

### 3.1 Preparation of Recombinant Tn5 Protein

Tn5 transposase can be commercially purchased from Diagenode or Illumina. Here I show a protocol to prepare recombinant Tn5 from *E. coli* as described previously with minor modifications [5]. While Triton X-100 was used for purification of Tn5 in the original paper, it is omitted in this protocol since enzymatic activity of Tn5 is lowered in detergents such as Triton X-100 and IGEPAL CA-630 (NP-40).

1. Transform pTXB1-Tn5 into T7 Express lysY/Iq competent *E. coli* as manufacturer's instruction.

2. Streak the transformed *E. coli* onto a LB plate containing 0.1-mg/mL ampicillin. Incubate the plate at 37 °C overnight.
3. Transfer a few colonies into 2-mL of 2YT medium containing 100 mg/mL of ampicillin and incubate it at 37 °C overnight.
4. Transfer 2 mL of the culture medium to 1000 mL of 2YT containing 100 mg/mL of ampicillin. Shake the flask at 37 °C until OD600 (optical density at 600 nm) reaches to 0.6.
5. Chill the flask on ice and add IPTG at final 0.25 mM and grow the cell at 30 °C for 6 h.
6. Transfer the medium to four 250-mL bottles and centrifuge them at 4200 rpm for 10 min.
7. Discard the supernatant by decantation and apply 10 mL of cold PBS to each tube, and centrifuge it again. Subsequently aspirate the PBS (*see Note 4*).
8. Suspend the pellet in 40 mL of cold lysis buffer and transfer it into a 50-mL tube.
9. Incubate the tube at 37 °C for 5 min to lyse *E. coli* and digest genomic DNA (*see Note 5*).
10. Centrifuge the tubes at 4200 rpm for 10 min at 4 °C to precipitate cell debris.
11. Add 1000 µL of 10% polyethylenimine to the lysate and mix it with gentle agitation. Color of the lysate turns to white as bacterial nucleic acids turn to insoluble upon addition of 10% polyethylenimine.
12. Centrifuge the tubes at 4200 rpm for 30 min at 4 °C.
13. Transfer the supernatant into new 50-mL tubes containing 5 mL of chitin beads pre-equilibrated with HEG buffer.
14. Incubate the tube with gentle agitation for 2 h at 4 °C.
15. Precipitate the beads by centrifugation at 1000 rpm for 5 min at 4 °C and transfer the beads to a gravity flow empty column.
16. Wash the beads with 20 mL of cold HEG buffer three times.
17. Wash the beads with 5 mL of cold cleavage buffer.
18. After closing the column bottom, apply 10 mL of cleavage buffer.
19. Close the column top tightly and incubate it at 4 °C for 48 h with gentle agitation. Tn5 is released from beads in this step by self-cleavage activity of intein.
20. Collect the eluate into a 15-mL tube.
21. Dialyze the sample using 12-mL Dialysis Cassettes against two changes of 1 L of 2× Tn5 dialysis buffer at 4 °C overnight.

22. Check the protein concentration by measuring absorbance at 280 nm. Alternatively, samples can be analyzed by SDS-PAGE using BSA protein standards.
23. Adjust the concentration of Tn5 to 0.586 mg/mL by concentrating the sample volume using Amicon Ultra 10K.
24. Add equal volume of 80% glycerol for adjusting the concentration to 0.293 mg/mL, which corresponds to 5.5  $\mu$ M. Stock Tn5 at  $-20^{\circ}\text{C}$  until use.

### **3.2 Preparation of Tn5 Complex**

Tn5 transposase specifically assembles with adaptor DNA as a dimer to form enzymatically active transposome [6]. Fluorescent Tn5 complex assembled with Tn5 transposase and fluorescent adaptor DNA is used for ATAC-see reaction.

1. Mix 10  $\mu$ L of 100- $\mu$ M ME-fw-Cy5, 10  $\mu$ L of 100- $\mu$ M ME-rev [Phos], and 2.2  $\mu$ L of 10 $\times$  annealing buffer in a 0.2-mL PCR tube.
2. Anneal the oligonucleotides by heating the mixture at 95  $^{\circ}\text{C}$  for 2 min and gradually cool to 25  $^{\circ}\text{C}$  on thermal cycler. The final concentration of the adaptor DNA is 45  $\mu$ M. The double-stranded adaptor DNA can be stored at  $-20^{\circ}\text{C}$ .
3. Assemble Tn5 complex by mixing 50  $\mu$ L of 5.5- $\mu$ M Tn5 transposase and 6.1  $\mu$ L of 45- $\mu$ M adaptor DNA in a 1.5-mL tube.
4. Incubate the tube at room temperature for 1 h. The Tn5 complex can be stored at  $-20^{\circ}\text{C}$  until use.

### **3.3 Preparation of ATAC-see Reaction Mixture**

Tn5 complex tends to aggregate in the ATAC-see reaction mixture, which results in nonspecific signals outside of nucleus. Therefore, the reaction mixture is filtrated to remove aggregated forms before use.

1. Mix 33  $\mu$ L of H<sub>2</sub>O, 10  $\mu$ L of 5 $\times$  TAPS buffer, 5  $\mu$ L of DMF, and 2  $\mu$ L of Tn5 complex in a 1.5-mL tube to make 50  $\mu$ L of reaction mixture, which is for five ATAC-see reactions.
2. Apply 50  $\mu$ L of the reaction mix to a centrifugal 0.22- $\mu$ m filter tube, and centrifuge it at 15,000 rpm for 5 min.

### **3.4 ATAC-see for Microscopic Observation**

Subnuclear localization of accessible chromatin can be addressed by ATAC-see using fluorescent microscopy. We routinely perform ATAC-see in culture insert on a glass-bottom dish, which is useful to reduce amount of reaction mixture. The insert is also helpful to perform ATAC-see in different conditions in the same dish. All the following processes are performed by applying 10  $\mu$ L of buffer to each well of the micro chamber.

1. Place a micro culture four-well insert onto a 35-mm glass-bottom dish (*see Note 6*).

2. Apply 10  $\mu\text{L}$  of L511 solution to each well and incubate it at 4 °C overnight. To avoid evaporation, place a glass coverslip onto the micro-insert.
3. Remove L511 solution completely by aspiration (*see Note 7*).
4. Apply 10  $\mu\text{L}$  of dissociated NIH3T3 cells to each well and let them stick onto glass bottom by culturing at 37 °C under 5%  $\text{CO}_2$  for several hours (*see Note 8*).
5. Aspirate culture medium and wash the cells with PBS twice.
6. Fix the cells by applying cross-link buffer1 and incubate the dish for 10 min at room temperature (*see Note 9*).
7. Wash the cells with PBS twice.
8. Permeabilize cells by applying cold NP buffer and incubate the dish for 10 min on ice.
9. After removing NP buffer completely, apply 10  $\mu\text{L}$  of ATAC reaction mixture (*see Note 10*).
10. Incubate the dish at 37 °C for 30 min. The dish can be placed on a pre-warmed heat block or floated in a water bath for the incubation.
11. Wash the wells with wash buffer A three times.
12. Apply cross-link buffer2 and incubate the dish for 10 min at room temperature.
13. After removing the cross-link buffer2, apply 10  $\mu\text{L}$  of Vecta-shield containing DAPI.
14. Fluorescent signals can be acquired with a conventional epi-fluorescent or confocal microscopy (*see Notes 11 and 12*).

### **3.5 ATAC-see Analysis for Flow Cytometry**

Total amount of accessible chromatin per nucleus can be quantified by ATAC-see using flow cytometry. Since chromatin accessibility is dynamically changed during cell cycle, amount of genomic DNA is simultaneously stained with DAPI.

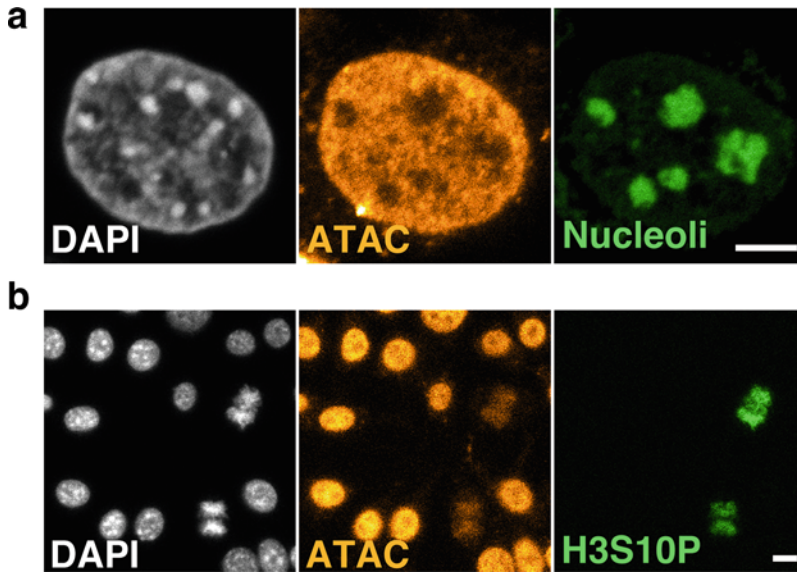
1. After PBS wash twice, treat NIH3T3 cells on a 3-cm dish with 0.5 mL of 0.5% trypsin at 37 °C for 3 min. Subsequently suspend the cells with 4 mL of culture medium and transfer it into a 15-mL tube.
2. Collect the cells by centrifugation at 1200 rpm for 2 min.
3. Suspend the cell pellet in 1 mL of PBS containing 0.5% BSA.
4. Count cell number using a hemocytometer.
5. Adjust the cell concentration to  $1 \times 10^6$  cells/mL by adding PBS containing 0.5% BSA.
6. Apply 100  $\mu\text{L}$  ( $10^5$  cells) into a PCR tube and centrifuge the tube at 1000 rpm for 2 min at 4 °C

7. Suspend the cells with 50  $\mu\text{L}$  of NP buffer containing 0.5% BSA and incubate the tube on ice for 5 min
8. Centrifuge the tube at 1000 rpm for 2 min at 4  $^{\circ}\text{C}$ .
9. Suspend the cells with 50  $\mu\text{L}$  of cold 1 $\times$  TAPS buffer containing 0.5% BSA, and subsequently centrifuge the tube at 1500 rpm for 2 min at 4  $^{\circ}\text{C}$ .
10. After removing the supernatant carefully, apply 50  $\mu\text{L}$  of ATAC-see reaction mixture (*see Note 13*).
11. Incubate the tube at 37  $^{\circ}\text{C}$  for 30 min.
12. Centrifuge the tube at 2000 rpm for 2 min at 4  $^{\circ}\text{C}$ .
13. Wash the cell pellets with 50  $\mu\text{L}$  of wash buffer A twice.
14. Centrifuge the tube at 2000 rpm for 2 min at 4  $^{\circ}\text{C}$ .
15. Suspend the cells with 50  $\mu\text{L}$  of wash buffer A containing 1- $\mu\text{g}/\text{mL}$  DAPI.
16. Incubate the tube on ice for 15 min.
17. Analyze signals for ATAC-see and DAPI using appropriate flow cytometry (*see Note 14*).

---

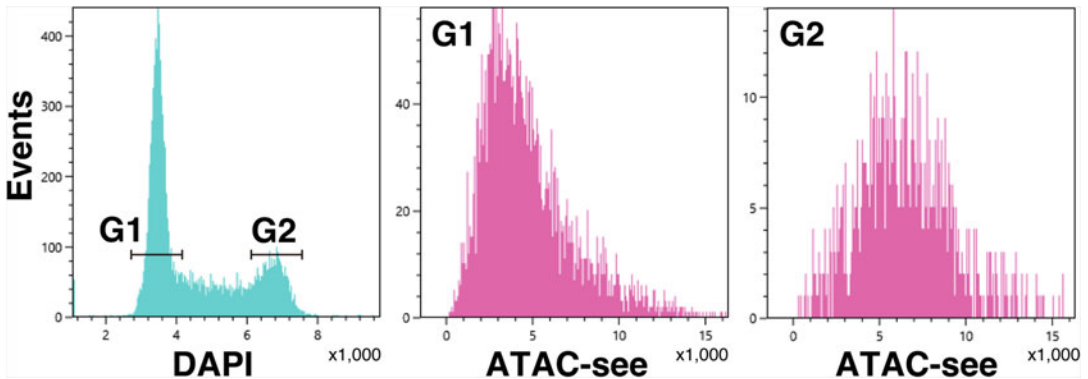
## 4 Notes

1. The other fluorophores such as Alexa488 and Cy3 are also applicable for modification on the adaptor oligonucleotide.
2. Mounting mediums from other suppliers can also be used.
3. Mouse cell line such as NIH3T3 cells is suitable for setting up the ATAC-see experiment, as their highly condensed heterochromatic chromocenters display high and low signals for DAPI and ATAC-see, respectively.
4. The *E. coli* pellet can be stored at  $-20^{\circ}\text{C}$  until use.
5. Sonication using an appropriate sonicator such as Branson Sonifier is also applicable for cell lysis.
6. While coating with Laminin-511 strongly supports cell adhesion onto a glass-bottom dish, it is not essential for ATAC-see reaction.
7. No need for washing after coating with L511 solution.
8. Cells can efficiently stick onto Laminin-511-coated glass within a few hours.
9. Since excessive cross-linking of chromatin could affect staining pattern of ATAC-see, the fixation conditions including concentration of formamide and duration of reaction should be optimized for each cell type.



**Fig. 1** Representative images of ATAC-seq. NIH3T3 cells were analyzed by ATAC-seq (orange) and stained for DAPI (gray) and Fibrillarin, a nucleolar marker (green) in (a), or histone H3 Ser10 phosphorylated (H3S10; green), a mitotic cell marker in (b). Scale bars indicate 3  $\mu\text{m}$

10. NP buffer should be carefully removed to avoid contamination of detergent, which could suppress Tn5 activity.
11. While Tn5 transposase preferentially reacts to nucleosome-free regions such as active promoters and enhancers, it also reacts to linker DNA between adjacent nucleosomes. Longer ATAC-seq reaction will result in accumulation of fluorescent signals originated from linker regions.
12. Representative images are shown in Fig. 1. Nucleoli displays weaker ATAC-seq signals compared with nucleoplasm (Fig. 1a). Heterochromatic chromocenters labeled with intense DAPI signals tend to display weak signals for ATAC-seq because of their highly condensed state. Chromatin accessibility is dynamic during the cell cycle (Fig. 1b). Nucleus at G2 phase have higher ATAC-seq signals than that at G1 phase, since the amount of genomic DNA is doubled after replication. In addition, highly condensed mitotic chromosomes display lower ATAC-seq signals compared with nucleus at interphase.
13. ATAC-seq with cross-linked cells gives cytoplasmic background signals originated from mitochondrial DNA. Therefore, cell cross-link is omitted in ATAC-seq for flow cytometry.



**Fig. 2** Representative results of ATAC-seq analyzed by flow cytometry (Sony SA3800). NIH3T3 cells were stained for ATAC-seq and DAPI. G1 and G2 phase were gated based on DAPI fluorescence (left panel). Fluorescent histograms for ATAC-seq signals at G1 and G2 phases were shown in middle and right panels, respectively

14. Representative results analyzed by flow cytometry are shown in Fig. 2. Cells at G2 phase display higher ATAC-seq signals compared with ones at G1 phase.

## References

1. Klemm SL, Shipony Z, Greenleaf WJ (2019) Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* 20(4):207–220
2. Tsompana M, Buck MJ (2014) Chromatin accessibility: a window into the genome. *Epigenetics Chromatin* 7(1):33
3. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10(12):1213–1218
4. Chen X, Shen Y, Draper W, Buenrostro JD, Litzgenburger U, Cho SW, Satpathy AT, Carter AC, Ghosh RP, East-Seletsky A, Doudna JD, Greenleaf WJ, Liphardt JT, Chang HY (2016) ATAC-seq reveals the accessible genome by transposase-mediated imaging and sequencing. *Nat Methods* 13(12):1013–1020
5. Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R (2014) Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res* 24(12):2033–2040
6. Reznikoff WS (2008) Transposon Tn5. *Annu Rev Genet* 42:269–286



## STREAMING-Tag System: Technology to Enable Visualization of Transcriptional Activity and Subnuclear Localization of Specific Endogenous Genes

Hiroaki Ohishi and Hiroshi Ochiai

### Abstract

The Spliced TetO REpeAt, MS2 repeat, and INtein sandwiched reporter Gene tag (STREAMING-tag) system enables imaging of nuclear localization as well as the transcription activity of a specific endogenous gene at sub-100-nm resolution in living cells. The use of this system combined with imaging of epigenome states enables a detailed analysis of the impact of epigenome status on transcriptional dynamics. In this chapter, we describe a method for quantifying distances between *Nanog* gene and clusters of cofactor BRD4 using the STREAMING-tag system in mouse embryonic stem cells.

**Key words** STREAMING-tag, Gene regulation, Live imaging, Transcriptional bursting, Transcriptional regulatory factor, RNAPII, Python, Trackpy, ImageJ

---

### 1 Introduction

Our bodies are composed of approximately 200 different cell types that have nearly the same genome sequence. Individual cell types exhibit distinct gene expression patterns and exhibit cell-specific features. These cell-type-specific gene expression patterns are defined by the differences in the expression levels of specific transcription and other regulatory factors. Furthermore, epigenomic states, such as DNA methylation, histone modifications, and higher-order genomic structures, can also affect the binding of transcription factors, resulting in changes in gene expression patterns [1]. In eukaryotic transcriptional processes, transcription factors bind to enhancers and recruit cofactors and RNA polymerase II (RNAPII) at promoter regions. Some transcription factors, RNAPII, and cofactors (transcriptional regulatory factors, RFs) have intrinsically disordered regions (IDRs), which are a class of polypeptides that have no static three-dimensional structure. The weak but multivalent interactions between IDRs contribute to the



formation of clusters (condensates) in the nucleus. These clusters have suggested the importance of the interaction between promoter and enhancer elements and the recruitment of RFs for efficient transcriptional initiation [2]. Once the preinitiation complex is assembled from the RF cluster and RNAPII is phosphorylated, transcription is initiated. After transcription initiation, transcription by phosphorylated RNAPII is paused 20–60 bp downstream of transcription start sites (TSS). This paused RNAPII is released by the additional phosphorylation, and processive transcription begins [3].

Transcription is a dynamic process that stochastically switches between transcriptional *ON* and *OFF* states. In the *ON* state, RNAs are continuously transcribed by RNAPII, whereas in the *OFF* state, transcription rarely occurs. This dynamic is called transcriptional bursting. This phenomenon is universally observed in various organisms and cell types [4]. The kinetics of transcriptional bursting can be described by two parameters: the burst size, which is the average number of mRNAs transcribed in a single *ON* state, and the burst frequency, which is the frequency of the *ON* state. Furthermore, the average RNA molecules can be estimated from the RNA degradation rate, in addition to these two parameters [5–7]. This indicates that gene expression levels are defined by the dynamics of transcriptional bursting. Therefore, to understand the regulatory mechanisms of gene expression, it is important to understand the regulatory mechanisms of transcriptional bursting kinetics. In mouse embryonic stem cells (mESCs), single-cell RNA sequencing and single-molecule RNA fluorescence in situ hybridization revealed that most of the expressed genes exhibited transcriptional bursting [8, 9]. Although we found that the localization of polycomb repressive complex 2 (PRC2) at the promoters and elongation factors at gene bodies are associated with regulating bursting dynamics, the contribution of these factors varies from gene to gene, suggesting that the regulatory mechanism of transcriptional bursting is essentially determined by individual genes [9]. Thus, to understand the regulatory mechanisms of transcriptional bursting of specific genes, it is necessary to analyze not only the transcriptional kinetics of individual genes but also the spatio-temporal relationships between transcriptional activity/nuclear localization at the single allele level and RF clusters that regulate the transcription of specific genes. Because the maintenance mechanism of the *OFF* state after the transition from the *ON* to the *OFF* state is largely unknown [10], we also need to understand the molecular mechanism of induction and maintenance of the *OFF* state.

The MS2 system is often used to quantify transcriptional bursting kinetics in living cells [4]. In this system, two components are used: bacteriophage-derived MS2 RNA and MS2-coat protein (MCP), which specifically binds to MS2 RNA. In this system, the

MS2 RNA-coding repeat sequence is knocked into a target gene in a cell line expressing MCP fused with a fluorescent protein (MCP-FP), enabling the fluorescent labeling of MS2 RNA derived from the gene. As the mature RNA molecules released from the transcription site rapidly diffuse into cells, a high-speed imaging system is necessary to observe them at the single-molecule level. In contrast, nascent RNAs that are continuously transcribed by RNA-Pol II and are confined to a gene locus can be observed as a spot (transcription site) with higher fluorescence intensity, compared to a single RNA spot. The transcriptional bursting kinetics of the target gene can be quantified from the intensity of the fluorescence and duration of the transcription site. It has been reported that, using the MS2 system, a higher-order genomic structure [11, 12] and RF cluster formation [13, 14] are associated with transcriptional bursting. Although these reports show an association between transcriptional activity and the distances between transcribed nascent RNAs and either enhancer or RF clusters in the ON state, a similar relationship in the OFF state remains an open question.

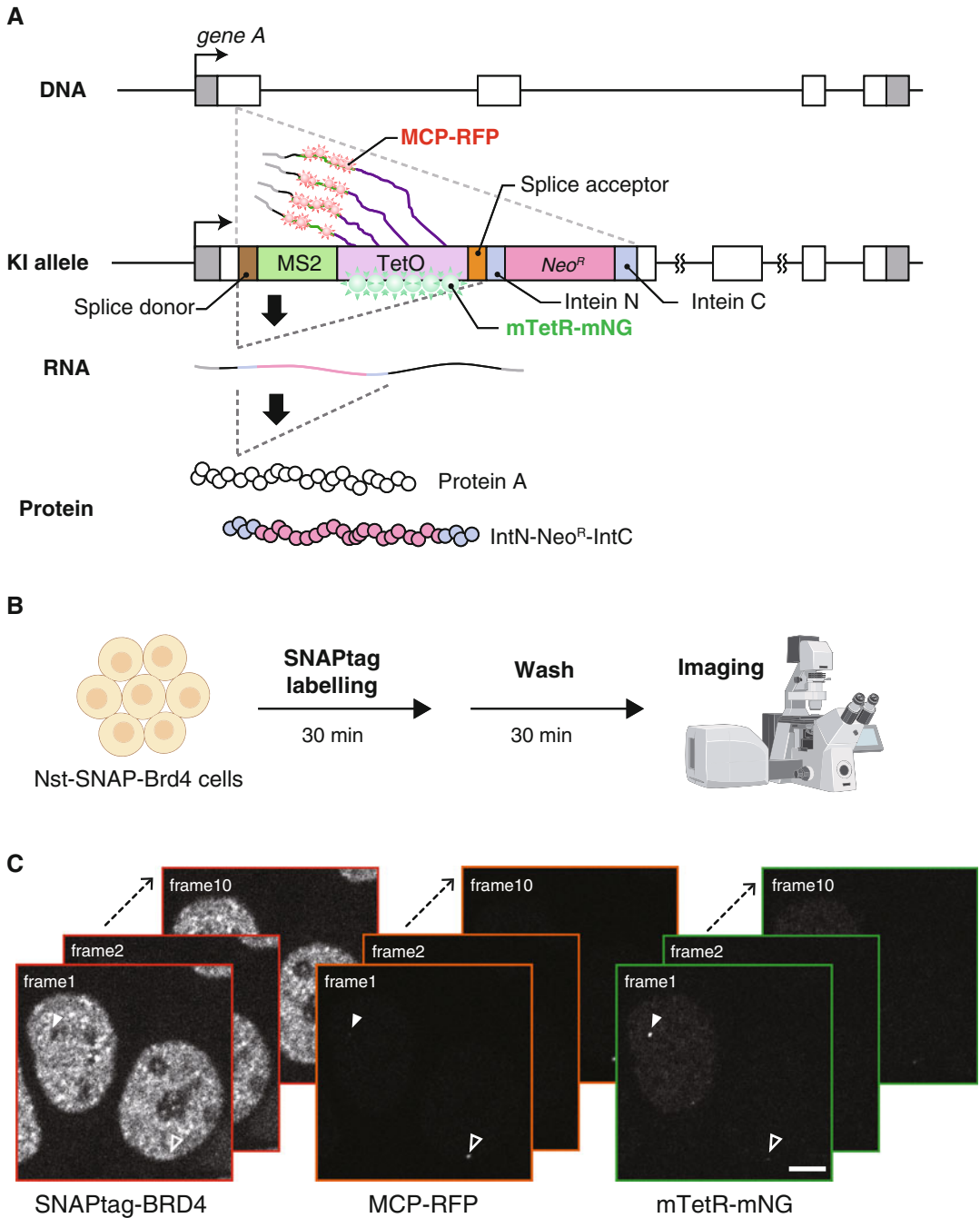
The Real-time Observation of Localization and EXpression (ROLEX) system has the potential to analyze gene loci not only in the ON state but also in the OFF state [15]. In this system, MS2 repeat sequences were knocked into a target gene. By expressing MCP fused with a red fluorescent protein (MCP-RFP), dCas9-GFP, and sgRNAs corresponding to the MS2 region in this cell line, transcriptional activity and target gene loci can be visualized. Thus, the system has the potential to investigate RF clusters that interact with genes in the OFF state. However, this system has two major problems related to image acquisition. The first problem stems from the location of the knocked-in MS2 repeat in the target gene. Because MCP bound to MS2 RNA inhibits translation [16], the MS2 repeat sequence is generally knocked into the 3' UTR. When the gene length is 6 kbp and the transcription rate is 2 kbp/min, transcriptional sites are detected by the MS2 system approximately 3 min after transcription initiation. However, it has been suggested that the transcription rate is not constant within a gene [17], and it is difficult to determine the exact time of transcription initiation or the position of the gene at the time of transcription initiation with the conventional MS2 system. Because transcription is thought to be regulated primarily at the steps of transcription initiation and pause release [18], establishing a new system for measuring transcriptional activity at the promoter-proximal site is required. Another problem is that the signal-to-noise ratio of fluorescent spots in the MS2 repeat DNA region labeled by the dCas9-GFP system was not high. This makes it difficult to determine the localization of a specific gene locus in the cell nucleus with high resolution using the ROLEX system. Therefore, to understand the spatiotemporal relationship between

RF clusters and transcriptional bursting, a novel imaging technique has been required.

To overcome these issues, we recently developed the Spliced TetO REpeAt, MS2 repeat, and INtein sandwiched reporter Gene tag (STREAMING-tag) (Fig. 1a) [19]. The STREAMING-tag includes an imaging module containing MS2 and TetO repeats flanked by splice donors (SD) and splice acceptors (SA) and a G418-selection module containing *IntN-Neo<sup>R</sup>-IntC* that encodes aminoglycoside O-phosphotransferase (APH(3')), which is a G418 resistance protein flanked by split inteins. The STREAMING-tag is knocked into a protein-coding region near the TSS; once the STREAMING-tag is transcribed, MS2 RNA can be visualized with MCP-RFP. The intra-nucleus position of the TetO repeat can be visualized by DNA binding affinity reduced mutant TetR (W43F) (mTetR) tagged with mNeonGreen (mTetR-mNG) [19]. As the imaging module is flanked by SD and SA, and APH (3') is flanked by split inteins, these modules are eliminated by splicing during RNA maturation and protein splicing, respectively. Therefore, this system is capable of selecting knock-in cells using a selection marker and imaging RNA/gene loci, with a minimal effect of the STREAMING-tag on the target gene product.

We have established an mESC line (NSt-GR) in which a STREAMING-tag was knocked into the TSS-proximal coding region of *Nanog* that shows clear transcriptional bursting [20] and is involved in ESC self-renewal, and mTetR-mNG and MCP-RFP were mildly expressed. Although the STREAMING-tag knock-in caused a slight decrease in the expression levels of *Nanog* mRNA and protein, we confirmed that transcription initiation, pause release of RNAPII, and *Nanog* gene function were not affected. To further analyze the spatial relationship between RF clusters and *Nanog* transcription, we established cell lines in which the SNAPtag was knocked into each RF coding gene (*Rpb1*, *Brd4*, *Med19*, and *Med22*) of the NSt-GR cells. Using these cell lines, we performed live imaging to quantify the transcriptional activity as well as the localization of RF clusters and *Nanog* gene locus. We found that the clusters of the RNAPII large subunit, RPB1, and the cofactor, BRD4, were significantly closer to the *Nanog* locus in the ON state than in the OFF state [19]. Single-gene imaging with STREAMING-tag system can be combined with "epigenome visualization systems" such as antibody-based mintbody [21, 22], ChromID [23], MethylRO for visualization of DNA methylation [24], and dCas9-GFP for visualization of specific regulatory elements including enhancers, which is expected to enable in-depth analysis of the impact of epigenomic modifications on transcriptional burst dynamics.

In this chapter, we introduce an image acquisition and analysis method that combines *Nanog* single-gene imaging using the



**Fig. 1** (a) Molecular structure and design of STREAMING-tag. (b) Diagram showing the workflow of SNAPtag labeling and imaging of Nst-SNAP-Brd4 cells. Nst-SNAP-Brd4 is a mouse embryonic stem cell line in which the STREAMING-tag is monoallelically knocked into the TSS-proximal region of *Nanog*, mTetR-mNG and MCP-RFP are mildly expressed, and SNAPtag is biallelically knocked into the downstream of the *Brd4* start codon. First, the SNAPtag-BRD4 protein expressed in this cell line is fluorescently labeled with the SNAPtag ligand. After washing procedures, cells are subjected to live imaging with the confocal microscope. (c) Representative acquired images of Nst-SNAP-Brd4 cells. Images are acquired in the order of the SNAPtag, RFP, and mNG channels. This step is repeated ten times with no intervals. Open and closed arrowheads indicate mTetR spots with and without MCP spots, respectively. The scale bar indicates 5  $\mu\text{m}$

STREAMING-tag system with BRD4 cluster imaging [19]. In this method, images acquired with a Nipkow-disk confocal fluorescence microscope are used; however, images acquired with a conventional point-scanning confocal microscope can be analyzed as well. From the acquired images, we determined the coordinates of target gene loci and RF clusters at sub-100-nm resolution and quantified their distances. Furthermore, it is possible to distinguish the ON and OFF states of target genes from MCP spots and to spatially investigate the cluster formation of RFs in a transcription state-dependent manner. Methods for cell line establishment are beyond the scope of this chapter; for details on such methods, please refer to reference [19].

---

## 2 Materials

### 2.1 Cell Culture of Mouse Embryonic Stem Cells

1. Nst-SNAP-Brd4 cells: A mouse embryonic stem cell line in which the STREAMING-tag is monoallelically knocked into the TSS-proximal region of *Nanog*, mTetR-mNG and MCP-RFP are mildly expressed, and SNAPtag is biallelically knocked into the downstream of the *Brd4* start codon.
2. ESM (embryonic stem cell's culture medium): Dulbecco's modified Eagle's medium supplemented with 4500-mg/mL glucose, 110-mg/L sodium pyruvate, 15% fetal bovine serum, 0.5-mM monothioglycerol solution, 1× MEM nonessential amino acids, 2-mM L-alanyl-L-glutamine solution, 1000-U/mL leukemia inhibitory factor, 20-μg/mL gentamicin, 3-μM CHIR99021, 1-μM PD0325901.
3. 60-mm cell culture dish.
4. Gelatin solution (0.1% wt/vol).
5. Trypsin-ethylenediaminetetraacetic acid (trypsin-EDTA; 0.25% wt/vol).
6. Phosphate-buffered saline (PBS).
7. PBS(+): PBS containing 100 mg/L of calcium chloride (CaCl<sub>2</sub>) and magnesium chloride (MgCl<sub>2</sub>·6H<sub>2</sub>O).
8. Laminin-511 containing solution: Laminin-511 diluted to 5 μg/mL in PBS(+).
9. Eight-well chambered cover glass (glass thickness: 0.17 mm).
10. Conical tube (15-mL volume).
11. CO<sub>2</sub> incubator (37 °C, 5% CO<sub>2</sub>).
12. Cell counter.
13. Centrifuge.
14. Water bath.

## 2.2 Multicolor Imaging of SNAPtag, mTetR, and MCP Spots

1. SNAP-Cell 647-SiR.
2. Imaging medium: FluoroBrite™ Dulbecco's modified Eagle's medium supplemented with 110-mg/L sodium pyruvate, 15% fetal bovine serum, 0.5-mM monothioglycerol solution, 1× MEM nonessential amino acids, 2-mM L-alanyl-L-glutamine solution, 1000-U/mL leukemia inhibitory factor, 20-μg/mL gentamicin, 3-μM CHIR99021, 1-μM PD0325901, VectaCell Trolox Antifade Reagent for Live Cell Imaging (1:1000).
3. Nikon Ti-2 microscope with a CSU-WI confocal unit, a 100× Nikon oil-immersion objective of 1.49 NA, and an iXon Ultra EMCCD camera, 488-, 561-, and 637-nm lasers, a stage-top microscope incubator (5% CO<sub>2</sub>; 37 °C), and an ASI MS-2000 piezo stage. In this setting, pixel size is 130 nm.
4. NIS-Elements software (Nikon, version 5.11.01).

## 2.3 Software and Modules/Packages for the Image Analysis

1. ImageJ (version: 2.3.0/1.53f, NIH, <https://imagej.nih.gov/ij/>).
2. Python 3 (version 3.9.4, <https://www.python.org>).
3. Jupyter Notebook (<https://jupyter.org>).
4. Anaconda (Python 3, <https://www.anaconda.com>).
5. Trackpy module (<http://soft-matter.github.io/trackpy/v0.5.0/>).

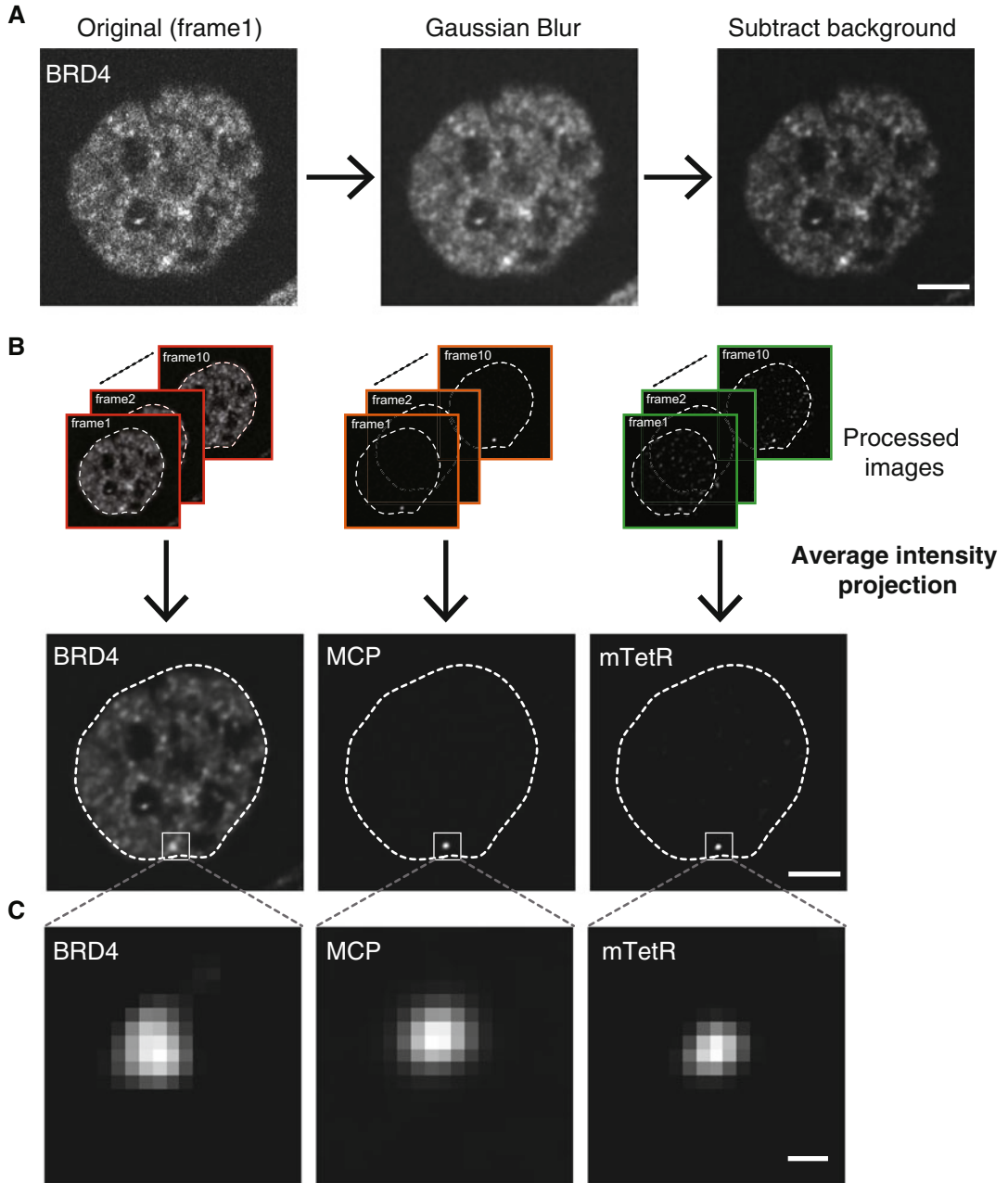
---

## 3 Methods

As described in this chapter, we used the Nst-SNAP-Brd4 cell line in which the STREAMING-tag is monoallelically knocked into the TSS-proximal region of *Nanog* and SNAPtag is biallelically knocked into the downstream of the *Brd4* start codon [19]. We describe the culture and passaging of these cells and then introduce the image acquisition method (Fig. 1), image pre-processing using ImageJ (Fig. 2), and determination of the coordinates of BRD4 cluster, MCP, and mTetR fluorescent spots and calculation of the 2D distances between each pair of spots using Python (Fig. 3). The representative image data, ImageJ macro, and Jupyter Notebook files used in this chapter can be obtained from <https://doi.org/10.17632/9py2b8m6cv.1>.

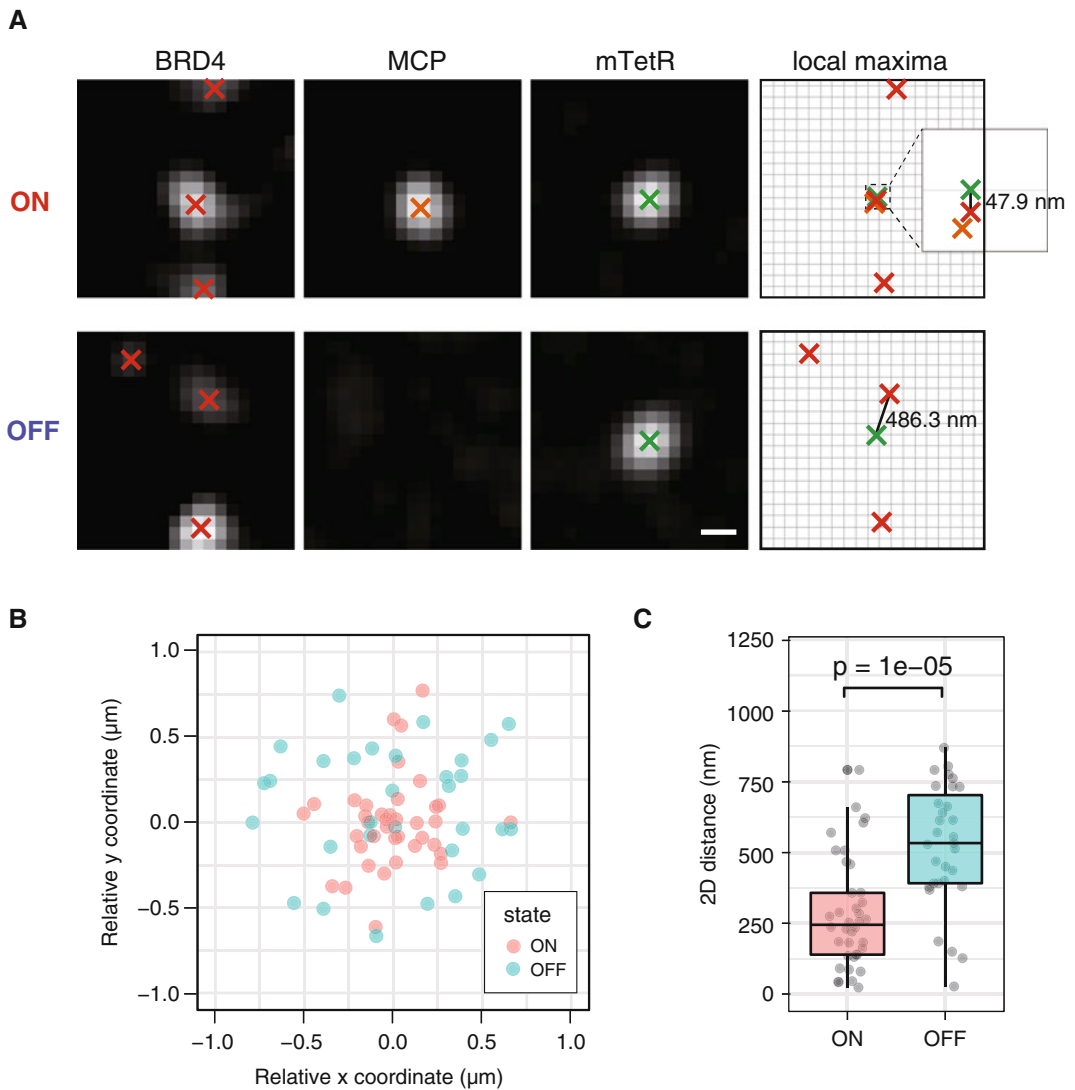
### 3.1 Cell Culture and Passage of Mouse Embryonic Stem Cells

1. Add a gelatin solution to a 60-mm culture dish. Incubate for 30 min at room temperature.
2. Thaw the cryo-preserved Nst-SNAP-Brd4 cells in a 37 °C water bath.
3. Just before the cells are completely thawed, transfer them with a pipette into a 15-mL tube containing 10 mL of pre-warmed ESM.



**Fig. 2** (a) Image processing with Gaussian blur and background subtraction. The scale bar indicates 5  $\mu\text{m}$ . (b) Average intensity projection of processed images. The scale bar indicates 5  $\mu\text{m}$ . (c) Magnified images showing colocalization of BRD4 cluster, mTetR, and MCP spots. The scale bar indicates 0.39  $\mu\text{m}$





**Fig. 3. (a)** Determination of the local maxima of the fluorescence spots by image analysis. Images show a representative gene locus of Nst-SNAP-Brd4 in the ON and OFF states. Gaussian-fitted local maxima determined by using trackpy are represented by cross marks. The scale bar indicates  $0.39 \mu\text{m}$ . **(b)** Scatter plot showing the relative xy coordinates between mTetR spots and the nearest BRD4 clusters. **(c)** Boxplot showing 2D distances between mTetR spots and the nearest BRD4 clusters.  $P$ -values correspond to the Wilcoxon rank sum test



4. Centrifuge the cells at  $190 \times g$  for 2 min at room temperature.
5. Aspirate ESM and resuspend the cells in 4-mL pre-warmed ESM.
6. Aspirate the gelatin solution from the 60-mm culture dish set up in **step 1** in Subheading 3.1, and then transfer the cells to the culture dish.
7. Incubate the cells in a CO<sub>2</sub> incubator at 37 °C under 5% CO<sub>2</sub> for 2 days.
8. Prepare a laminin-511-coated eight-well chambered cover glass by adding 130 μL of laminin-511 coating solution to each well of the cover glass. Incubate the chamber at 37 °C for 1 h.
9. Aspirate ESM from the dish and add 4 mL of pre-warmed PBS to the dish.
10. Aspirate PBS and add 500 μL of pre-warmed 0.25% trypsin-EDTA.
11. Incubate the dish in a CO<sub>2</sub> incubator for 2 min.
12. Aspirate the laminin-511 coating solution from the eight-well chambered cover glass and immediately add 200 μL of PBS(+).
13. Add 2-mL ESM to the dish with gentle pipetting to detach the cells from the dish and collect them into a 15-mL conical tube.
14. Centrifuge the cells at  $190 \times g$  for 2 min.
15. Aspirate the supernatant and resuspend the cells in 1-mL ESM. Then, count the number of cells.
16. Remove the PBS(+) from the eight-well chambered cover glass, and immediately plate  $7 \times 10^4$  cells in 200 μL of ESM (*see Note 1*).
17. Culture the cells overnight in a CO<sub>2</sub> incubator (*see next section*).

**3.2 SNAPtag  
Labeling and Image  
Acquisition by  
Confocal Microscope**

1. Aspirate ESM from the well of the eight-well chambered cover glass.
2. Add 150 μL of 300-nM SNAP-Cell 647-SiR in ESM to the well (*see Note 2*) (Fig. 1b).
3. Incubate the cells in a CO<sub>2</sub> incubator for 30 min.
4. Wash the cells with 200 μL of fresh and pre-warmed ESM three times.
5. Incubate the cells in a CO<sub>2</sub> incubator for another 30 min.
6. Aspirate ESM and add 200 μL of pre-warmed imaging medium to the well.
7. Set the eight-well chambered cover glass on the confocal microscope.

8. Open the NIS-Elements software.
9. Make two folders named “ON” and “OFF” in which to save the acquired images. One is for cells in the ON state, and the other is for cells in the OFF state. In addition, create a folder named “FOV” in these folders.
10. Focus on mTetR spots in the nucleus from confocal images of the mNG channel (*see Note 3*).
11. Check the MCP spots around the mTetR spots from the confocal image of the RFP channel; if the MCP spot is located near the mTetR spot, the gene is in the ON state; otherwise, it is in the OFF state (*see Note 4*).
12. Confirm that BRD4-SNAPtag is fluorescently labeled from confocal images of the SNAPtag channel.
13. To minimize the photobleaching effect, set the power of each laser sufficiently low such that the desired mTetR, MCP spots, and RF clusters can be recognized.
14. Acquire images in the order of the SNAPtag, RFP, and mNG channels. This step is repeated ten times with no intervals (Fig. 1c) (*see Note 5*).
15. Save images in the “ON” or “OFF” folder created in **step 9** in Subheading 3.2 according to the transcriptional state of the cells imaged.
16. Repeat **steps 10–15** in Subheading 3.2. Acquire images of more than 40 fields of view (FOVs). Ideally, at least 20 cells in each of the ON and OFF states are required for statistical analysis.

### 3.3 Pre-Processing of Acquired Images

1. Open ImageJ.
2. Make a macro file. Select “Plugin” → “New macro”.
3. Copy the following macro code and paste it in a macro window. In the [Full path] field, enter the path of the “ON” or “OFF” directory created in **step 9** in Subheading 3.2. The codes that need to be changed by the user depending on the file to be analyzed or the microscope settings are shown in red. The comment section is shown in gray.

```

1  directory_path = "[Full path]"; //Specify a full path such as
   "/Users/Username/folder1/ON".
2  File_name = "Nst-SNAP-Brd4_SNAPSHOT_ON.nd2"; // In this
   example, we have specified the file "Nst-SNAP-
   Brd4_SNAPSHOT_ON.nd2". If you want to process a different file,
   you need to change this name to the corresponding file name.
3  bg_value = 500; // This is the value for background correction. This
   value (500) needs to be changed depending on the microscope used.
   It should be about the same as the average background fluorescence
   intensity of a place where no cells and debris are contained after the
   imaging conditions are fixed.
4  pixel_x = 0.13; // Enter the length in the x-axis per pixel.
5  pixel_y = 0.13; // Enter the length in the y-axis per pixel.
6  run("Bio-Formats", "open="+directory_path+"/"+File_name+"
   color_mode=Grayscale rois_import=[ROI manager]
   view=Hyperstack stack_order=XYCZT");rename("stack"); //
   Importing Images.
7  run("Properties...", "channels=3 slices=1 frames=10
   pixel_height="+pixel_x+" → pixel_height="+pixel_y+"
   voxel_depth=1.0000000 frame=[0.00 sec]"); // Change the image
   properties. The pixel width setting and the order of the images are
   changed. The acquired image stack is a three-color time-lapse image,
   but here it is recognized as a three-color z-stack image. This enables
   the average projection of the image at a later step.
8  run("Gaussian Blur...", "sigma=1 stack"); // Apply a Gaussian blur
   filter.
9  run("Split Channels"); // Separate the image stack into channels.
10 selectWindow("C1-stack"); // Select an image for the first color
   (SNAPtag channel).
11 run("Bleach Correction", "correction=[Simple Ratio] background="+
   bg_value +"""); // Correct photobleaching.
12 run("Subtract Background...", "rolling=50 stack"); // Correct uneven
   illumination.
13 selectWindow("C1-stack"); close(); // Delete the pre-processed
   image.
14 selectWindow("C2-stack"); // Select an image for the second channel
   (RFP channel).
15 run("Bleach Correction", "correction=[Simple Ratio] background="+
   bg_value +"""); // Correct photobleaching.

```

```

16 run("Subtract Background...", "rolling=50 stack"); // Correct for
    uneven illumination.
17 run("Subtract Background...", "rolling=5 stack"); // Emphasize spots.
18 selectWindow("C2-stack"); close(); // Delete the pre-processed
    image.
19 selectWindow("C3-stack"); // Select an image for the third channel
    (mNG channel).
20 run("Bleach Correction", "correction=[Simple Ratio] background="+
    bg_value + ""); // Correct photobleaching.
21 run("Subtract Background...", "rolling=50 stack"); // Correct for
    uneven illumination.
22 run("Subtract Background...", "rolling=5 stack"); // Emphasize spots.
23 selectWindow("C3-stack"); close(); // Delete the pre-processed
    image.
24 run("Merge Channels...", "c1=DUP_C1-stack c2=DUP_C2-stack
    c3=DUP_C3-stack create ignore"); // Merge the processed three-
    color stacks.
25 run("Z Project...", "projection=[Average Intensity]"); // Average
    intensity projection
26 close("Composite"); close("Merged"); // Delete the unwanted
    images.
27 rename("Nst-SNAP-Brd4"); // Rename the image.

```

4. Select “Run” at the bottom of the macro window. In this macro, the following image processing is performed automatically: importing the images to ImageJ, changes in image properties, filtering with Gaussian blur, background subtraction, bleach correction, averaging, and so on (Fig. 2a, b). Each line in the macro executes a single process. The number of lines is shown on the left. Written after the “//” is the “comment,” which describes what processing is being carried out on each line; please refer to it.
5. After the macro finishes, an image named Nst-SNAP-Brd4 is obtained.
6. Make a new macro file. Select “Plugin” → “New macro”.
7. Copy the following macro code and paste it in the macro window:

```
makeRectangle(0, 0, 200, 200); // Change the values depending on the
position of a cell with a mTetR spot and a cell size.
setTool("rectangle");
waitForUser("Move and resize the rectangle ROI so that it contains a target
cell."); // Select a rectangular region of interest (ROI) including a cell of
interest.
Run("Duplicate...", "duplicate"); // Copy the ROI from the image.
```

8. Select “Run” at the bottom of the macro window.
9. A rectangular region of interest (ROI) appears on the image. Resize or change the location of this ROI so that it contains a single cell of interest with an mTetR spot. Then, click “OK” in the “Action required” window. The image of the ROI region is then duplicated.
10. Make a new macro file. Select “Plugin” → “New macro”.
11. Copy the following macro code and paste it in the macro window:

```
Stack.setChannel(1); // Select the SNAPtag channel.
setMinAndMax(120, 350); // Change the minimum and maximum
brightness of an image. The values may need to be changed for each
image.
makeRectangle(0, 0, 19, 19);
setTool("rectangle");
waitForUser("Move a rectangle of 19×19 pixels to the center of the
mTetR spot."); // Set a rectangular ROI so that the mTetR spot is at the
center.
run("Duplicate...", "duplicate"); // Copy the ROI from the image.
Stack.setDisplayMode("grayscale");
rename("BRD4_ON_FOV1_"); // Rename the image.
```

12. Select “Run” at the bottom of the macro window. With this macro, a new 19 × 19-pixel image named “BRD4\_ON\_FOV1\_” centered on the TetR bright spot is created (Fig. 2c).
13. Save the image file in the “FOV” folder created in **step 9** in Subheading 3.2. Select “File” → “Save As” → “Image Sequence...”. This operation saves three types of images: BRD4\_ON\_FOV1\_c001.tif, BRD4\_ON\_FOV1\_c002.tif, and BRD4\_ON\_FOV1\_c003.tif. Each corresponds to a SNAPtag, RFP, and mNG channel, respectively.
14. Repeat **steps 3–13** in Subheading 3.3 for acquired images.

### 3.4 Detection of Spots and Measurement of 2D Distances at Sub-Pixel Resolution

1. Open a Jupyter Notebook. Enter “jupyter notebook” in a terminal (macOS) or command prompt (Windows OS).
2. Create a new Python file. Select “New tab” → “Python 3”.
3. Select “File” → “Save As”. Enter the file name MiMB\_STREAMING\_BRD4.
4. Copy the following Python code, paste it in the Jupyter Notebook, and click “Run”.

```
import pims
from pims import ImageSequence
import trackpy as tp
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import math
```

5. By running this code, the tools for image analysis are imported. The following tools are used: the pims (ImageSequence) interface as to import images, the trackpy package to detect fluorescent spots, the Matplotlib library (matplotlib.pyplot interface) to plot graphs, the pandas package for data analysis and manipulation, and the numpy package/math module for mathematical functions (*see Note 6*).
6. Import images as a numpy array. Enter the following code. In the [Full path2] field, enter the path of the files saved in **step 13** in Subheading 3.3. To select all three saved files, put a \* at the end of the file name entered in Subheading 3.3, **step 12**. The [Full path2] here should thus appear as “/Users/Username/folder1/ON/FOV/BRD4\_ON\_FOV1\_\*”.

```
file_imseq = "[full path2]"
image = ImageSequence(file_imseq)
image
```

7. The following image information is displayed in the output column. In [File path information], the specified [Full path2] appears.

```
<Frames>
Source: [file path information]
Length: 3 frames
Frame Shape: (19, 19)
Pixel Datatype: uint16
```

8. Determine the local maxima of the BRD4 cluster, MCP, and mTetR spots. Copy the following Python code and paste it in the Jupyter Notebook (*see Note 7*). To analyze cells in the OFF state, delete the line starting with “df\_2”. What is written after the “#” is the “comment,” which describes what processing is being carried out on each line; please refer to it.

```
diameter = 5 # Specify the size for the target bright spots. The size should
be changed according to the pixel size of your image.
df_1 = tp.locate(image[0], diameter)
df_2 = tp.locate(image[1], diameter, topn=1) # If the target image is in
the "OFF" state, delete this line.
df_3 = tp.locate(image[2], diameter, topn=1)
```

9. By fitting a Gaussian model, the sub-pixel local maxima of the BRD4 cluster, MCP, and mTetR spots are determined. Copy the following Python code and paste it in the Jupyter Notebook. To analyze cells in the OFF state, delete the line starting with “df\_2”.

```
df_1 = tp.refine_leastsq(df_1, image[0], diameter, fit_function='gauss')
df_2 = tp.refine_leastsq(df_2, image[1], diameter, fit_function='gauss') #
If the target image is in the "OFF" state, delete this line.
df_3 = tp.refine_leastsq(df_3, image[2], diameter, fit_function='gauss')
```

10. Draw the crosses corresponding to xy coordinates on the images to confirm the centers of the BRD4 cluster, MCP, and mTetR spots. Copy the following Python code and paste it in the Jupyter Notebook (*see Note 8*) (Fig. 3a). For cells in the OFF state, delete the last six lines from “# RFP channel”.

```

# SNAPtag channel
im1 = image[0]
mean1 = im1.mean()
std1 = im1.std()
plt.imshow(im1, vmin = mean1-std1*0.7, vmax = mean1+std1*4, cmap =
'gray')
plt.scatter(df_1['x'], df_1['y'], marker="x", color = "blue", s = 250)
plt.show()

# RFP channel
im2 = image[1]
mean2 = im2.mean()
std2 = im2.std()
plt.imshow(im2, vmin = mean2-std2*0.7, vmax = mean2+std2*4, cmap =
'gray')
plt.scatter(df_2['x'], df_2['y'], marker="x", color = "blue", s = 250)
plt.show()

#mNG channel
im3 = image[2]
mean3 = im3.mean()
std3 = im3.std()
plt.imshow(im3, vmin = mean3-std3*0.7, vmax = mean3+std3*4, cmap =
'gray')
plt.scatter(df_3['x'], df_3['y'], marker="x", color = "blue", s = 250)
plt.show()

```

11. Calculate the 2D distances between the mTetR spots and the nearest BRD4 cluster. Copy the following Python code and paste it in the Jupyter Notebook. The value of 130 in the following code indicates the width and height (nm) of a pixel. Change the value depending on the image data.

```

pixel_size = 130 # Enter the length and height of the pixel.
dist_min = []
for Y, X in zip(df_1['y'], df_1['x']):
    distance = math.sqrt((Y-df_3['y'].values)**2+(X-
df_3['x'].values)**2)
    dist_min.append(distance)
min(dist_min)* pixel_size

```



12. Repeat **steps 6–11** in Subheading 3.4 for all images in the ON and OFF folders.
13. Following analysis of all images, boxplots and scatter plots are created from the obtained data (Fig. 3b, c).

---

## 4 Notes

1. The mESCs cultured on laminin-511 adhered well to the glass surface and expanded into a single layer [25]. When cells adhere well to the glass surface, the cell nuclei become flattened, increasing the possibility of observing mTetR spots in many cells on a single Z-axis confocal plane.
2. For SNAPtag labeling, the use of 30–300 nM of SNAP-Cell 647-SiR dye is sufficient [14]. SNAPtag-Janelia Fluor dyes can also be used [26].
3. When performing single-gene imaging using the STREAMING-tag system, it is recommended to first determine the location of the mTetR spot. In cells overexpressing mTetR or MCP, ectopic spots of mTetR or MCP may be observed, which may be misinterpreted as gene loci or transcriptional spots. To avoid this, first focus on cells that have only one clear mTetR spot if the STREAMING-tag is knocked in monoallelic. Then, switch to the MCP-RFP channel and check for MCP spots around the mTetR spot.
4. In this chapter, we manually determined the ON and OFF states from the MCP images. The following is a quantitative method for classifying the ON and OFF states [19]. First, quantify the average fluorescence intensity within a circle 6 pixels in diameter, centered on the local maxima of the MCP spots. When the fluorescent intensity is more than twice the average fluorescence intensity of a  $19 \times 19$ -pixel image and a distance between the MCP and mTetR spots is less than 390 nm (3 pixels), the image is classified as the “ON” state, and the other images are classified as the “OFF” state.
5. RF clusters, TetR, and MCP spots constantly diffuse in the microspace of the cell nucleus due to thermal fluctuations. However, all of them reflect the behavior of the DNA region itself (mTetR-mNG) or are bound to the DNA and RNA (SNAPtag-RFs and MCP-RFP); thus, their diffusion space is more confined than that of free molecules. In addition, there is a time lag between the capture of each fluorescence channel (e.g., even if there is no interval between channel switches, there will be a time lag equal to the exposure time associated with the capture of each channel). Therefore, spots recognized from single images taken in each channel do not necessarily

show the coordinates where the spots of each channel can exist on average. To estimate the average localized coordinates of the RF clusters, MCP, and mTetR spots, we acquired ten consecutive images for each channel. Finally, the average image can be produced to determine the average coordinates of the clusters, RF clusters, MCP, and mTetR spots.

6. If you are a beginner with the Python language and image analysis, we recommend a Jupyter Notebook. The interface is user-friendly and writing and debugging code is easy. Anaconda easily equips you to work with open-source packages and libraries written in Python on your personal computer.
7. The `trackpy.locate` function of `trackpy` can be set to the minimum fluorescence intensity, diameter size, and other parameters of the spots to be detected. Thus, these parameters can be changed according to clusters and spots.
8. If the coordinates of the cross marks do not seem to intuitively overlap with the coordinates of the spots, we recommend resetting the parameters with reference to **Note 7**.

## References

1. Andersson R, Sandelin A (2019) Determinants of enhancer and promoter activities of regulatory elements. *Nat Rev Genet* 21:71–87. <https://doi.org/10.1038/s41576-019-0173-8>
2. Bhat P, Honson D, Guttman M (2021) Nuclear compartmentalization as a mechanism of quantitative control of gene expression. *Nat Rev Mol Cell Biol* 22:653–670. <https://doi.org/10.1038/s41580-021-00387-1>
3. Cramer P (2019) Organization and regulation of gene transcription. *Nature* 573:45–54. <https://doi.org/10.1038/s41586-019-1517-4>
4. Rodriguez J, Larson DR (2020) Transcription in living cells: molecular mechanisms of bursting. *Annu Rev Biochem* 89:189–212. <https://doi.org/10.1146/annurev-biochem-011520-105250>
5. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* 4:e309. <https://doi.org/10.1371/journal.pbio.0040309>
6. Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, Simpson ML, Weinberger LS (2012) Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc Natl Acad Sci U S A* 109:17454–17459. <https://doi.org/10.1073/pnas.1213530109>
7. So L, Ghosh A, Zong C, Sepúlveda LA, Segev R, Golding I (2011) General properties of transcriptional time series in *Escherichia coli*. *Nat Genet* 43:554–560. <https://doi.org/10.1038/ng.821>
8. Shah S, Takei Y, Zhou W, Lubeck E, Yun J, Eng C-HL, Kouloua N, Cronin C, Karp C, Liaw EJ, Amin M, Cai L (2018) Dynamics and spatial genomics of the nascent transcriptome by intron seqFISH. *Cell* 174:363–376.e16. <https://doi.org/10.1016/j.cell.2018.05.035>
9. Ochiai H, Hayashi T, Umeda M, Yoshimura M, Harada A, Shimizu Y, Nakano K, Saitoh N, Liu Z, Yamamoto T, Okamura T, Ohkawa Y, Kimura H, Nikaido I (2020) Genome-wide kinetic properties of transcriptional bursting in mouse embryonic stem cells. *Sci Adv* 6:eaaz6699. <https://doi.org/10.1126/sciadv.aaz6699>
10. Nicolas D, Phillips NE, Naef F (2017) What shapes eukaryotic transcriptional bursting? *Mol Biosyst* 13:1280–1290. <https://doi.org/10.1039/c7mb00154a>
11. Fukaya T, Lim B, Levine M (2016) Enhancer control of transcriptional bursting. *Cell* 166:358–368. <https://doi.org/10.1016/j.cell.2016.05.025>
12. Chen H, Levo M, Barinov L, Fujioka M, Jaynes JB, Gregor T (2018) Dynamic interplay between enhancer-promoter topology and

- gene activity. *Nat Genet* 50:1296–1303. <https://doi.org/10.1038/s41588-018-0175-z>
13. Cho W-K, Spille J-H, Hecht M, Lee C, Li C, Grube V, Cisse II (2018) Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361:eaar4199. <https://doi.org/10.1126/science.aar4199>
  14. Li J, Dong A, Saydaminova K, Chang H, Wang G, Ochiai H, Yamamoto T, Pertsinidis A (2019) Single-molecule nanoscopy elucidates RNA polymerase II transcription at single genes in live cells. *Cell* 178:491–506.e28. <https://doi.org/10.1016/j.cell.2019.05.029>
  15. Ochiai H, Sugawara T, Yamamoto T (2015) Simultaneous live imaging of the transcription and nuclear position of specific genes. *Nucleic Acids Res* 43:e127–e127. <https://doi.org/10.1093/nar/gkv624>
  16. Stripecke R, Oliveira CC, McCarthy JE, Hentze MW (1994) Proteins binding to 5' untranslated region sites: a general mechanism for translational regulation of mRNAs in human and yeast cells. *Mol Cell Biol* 14:5898–5909. <https://doi.org/10.1128/mcb.14.9.5898-5909.1994>
  17. Liu J, Hansen D, Eck E, Kim YJ, Turner M, Alamos S, Garcia HG (2021) Real-time single-cell characterization of the eukaryotic transcription cycle reveals correlations between RNA initiation, elongation, and cleavage. *PLoS Comput Biol* 17:e1008999. <https://doi.org/10.1101/2020.08.29.273474>
  18. Bartman CR, Hamagami N, Keller CA, Giardine B, Hardison RC, Blobel GA, Raj A (2018) Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation. *Mol Cell* 73:519–532.e4. <https://doi.org/10.1016/j.molcel.2018.11.004>
  19. Ohishi H, Shimada S, Uchino S, Li J, Sato Y, Shintani M, Owada H, Ohkawa Y, Pertsinidis A, Yamamoto T, Kimura H, Ochiai H (2022) STREAMING-tag system reveals spatiotemporal relationships between transcriptional regulatory factors and transcriptional activity. *bioRxiv* 2022.01.06.472721. <https://doi.org/10.1101/2022.01.06.472721>
  20. Ochiai H, Sugawara T, Sakuma T, Yamamoto T (2014) Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. *Sci Rep* 4:7125. <https://doi.org/10.1038/srep07125>
  21. Sato Y, Mukai M, Ueda J, Muraki M, Stasevich TJ, Horikoshi N, Kujirai T, Kita H, Kimura T, Hira S, Okada Y, Hayashi-Takanaka Y, Obuse C, Kurumizaka H, Kawahara A, Yamagata K, Nozaki N, Kimura H (2013) Genetically encoded system to track histone modification in vivo. *Sci Rep* 3:2436. <https://doi.org/10.1038/srep02436>
  22. Tjalsma SJD, Hori M, Sato Y, Bousard A, Ohi A, Raposo AC, Roensch J, Saux AL, Nogami J, Maehara K, Kujirai T, Handa T, Bagés-Arnal S, Ohkawa Y, Kurumizaka H, da Rocha ST, Żylicz JJ, Kimura H, Heard E (2021) H4K20me1 and H3K27me3 are concurrently loaded onto the inactive X chromosome but dispensable for inducing gene silencing. *EMBO Rep* 22:e51989. <https://doi.org/10.15252/embr.202051989>
  23. Villaseñor R, Pfaendler R, Ambrosi C, Butz S, Giuliani S, Bryan E, Sheahan TW, Gable AL, Schmolka N, Manzo M, Wirz J, Feller C, von Mering C, Aebersold R, Voigt P, Baubec T (2020) ChromID identifies the protein interactome at chromatin marks. *Nat Biotechnol* 38:728–736. <https://doi.org/10.1038/s41587-020-0434-2>
  24. Ueda J, Maehara K, Mashiko D, Ichinose T, Yao T, Hori M, Sato Y, Kimura H, Ohkawa Y, Yamagata K (2014) Heterochromatin dynamics during the differentiation process revealed by the DNA methylation reporter mouse, MethylRO. *Stem Cell Rep* 2:910–924. <https://doi.org/10.1016/j.stemcr.2014.05.008>
  25. Domogatskaya A, Rodin S, Boutaud A, Tryggvason K (2008) Laminin-511 but not -332, -111, or -411 enables mouse embryonic stem cell self-renewal in vitro. *Stem Cells* 26:2800–2809. <https://doi.org/10.1634/stemcells.2007-0389>
  26. Grimm JB, English BP, Chen J, Slaughter JP, Zhang Z, Revyakin A, Patel R, Macklin JJ, Normanno D, Singer RH, Lionnet T, Lavis LD (2015) A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat Methods* 12:244–250. <https://doi.org/10.1038/nmeth.3256>



## Bioinformatics Pipelines for Identification of Super-Enhancers and 3D Chromatin Contacts

Akihiko Sakashita, Chikara Takeuchi, So Maezawa, and Satoshi H. Namekawa

### Abstract

Precise regulation of gene expression is integral in development. Emerging studies have highlighted that super-enhancers (SEs), which are clusters of multiple enhancers, play critical roles in regulating cell type-specific gene expression via 3D chromatin, thereby defining the cellular identities of given cells. Here we provide optimized bioinformatics pipelines to identify SEs and 3D chromatin contacts. Our pipelines encompass the processing of chromatin immunoprecipitation sequencing (ChIP-seq) data to identify SEs and the processing of genome-wide chromosome conformation capture (Hi-C) data. We can then infer long-range chromatin contacts between SEs and other genomic regions. This integrative computational approach, which can be applied to CUT&RUN and CUT&Tag, alternative technologies to ChIP-seq, allows us to identify genomic locations of SEs and their 3D genome configuration, whereby multiple SEs act in concert. We show an analysis of mouse spermatogenesis as an example of this application.

**Key words** ChIP-seq, Hi-C, Enhancer, Super-enhancer, Transcription factor-binding, Promoter–enhancer interaction, Gene regulation, Chromatin contact, Meiosis, Spermatogenesis

---

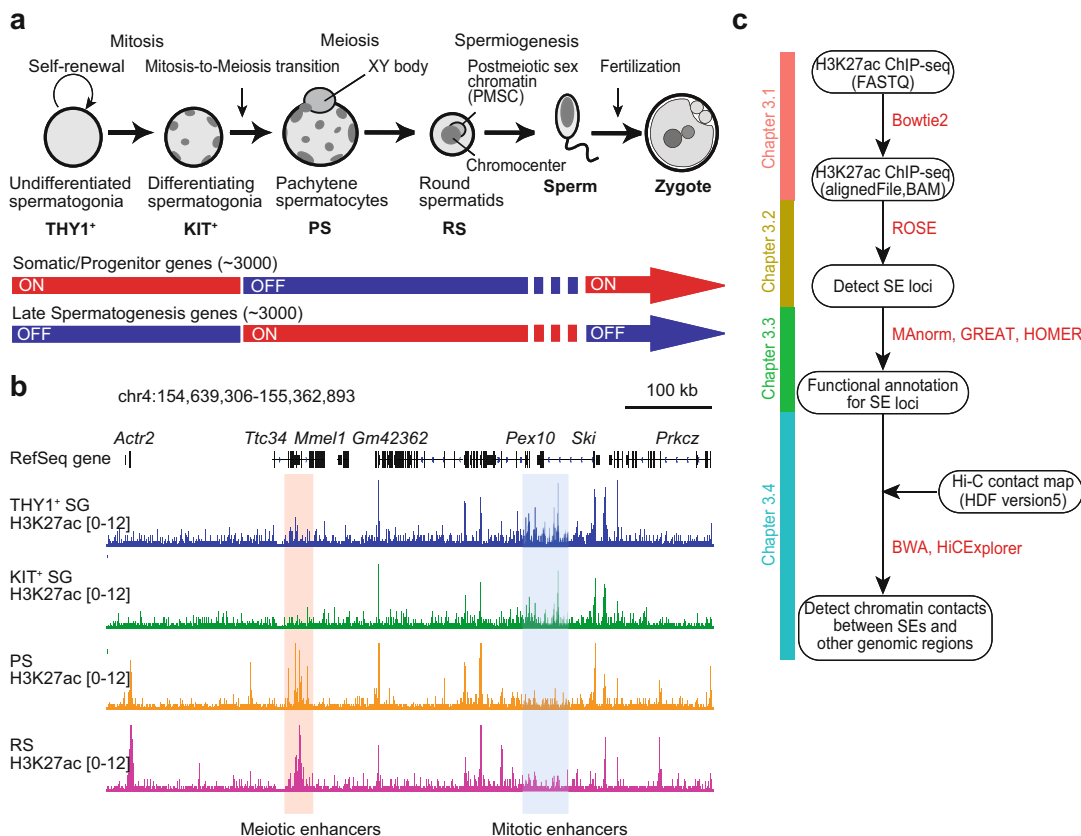
## 1 Introduction

Enhancers are *cis*-regulatory elements that orchestrate gene expression in stage- and cell type-specific manners. In general, active enhancers are defined based on epigenetic signatures, including chromatin accessibility and histone modifications (i.e., histone H3 monomethylated at lysine 4, H3K4me1, and histone H3 acetylated at lysine 27, H3K27ac). Active enhancers often provide recognition sites for transcription factors to drive expressions of their target genes over a long distance [1–3]. Clusters of multiple enhancers in proximity are defined as super-enhancers (SEs). SEs are occupied by a high density of master transcription factors, RNA polymerase II,

and mediators, as well as high H3K27ac signals, compared to typical enhancers [4]. SEs act in the context of 3D chromatin structure; the long-range interactions between SEs and promoters of target genes are mediated through phase-separating properties, which robustly drive cell type-specific transcription [5, 6]. Emerging studies have shown that SEs play critical roles in regulating cellular identities and functions, whereas SE dysregulation is associated with increased susceptibility to certain types of cancer and autoimmune diseases [7–9]. Yet the manner in which SEs are established in a stage- and cell type-specific manner remains largely unknown.

Mouse spermatogenesis is an excellent model to study cell type-specific reorganization of SEs and SE-driven gene regulation. This is because genome-wide chromatin remodeling and gene expression changes occur at a transition from mitosis to meiosis [10–13] (Fig. 1a). Spermatogenesis is a differentiation process that ensures a continuous production of spermatozoa [14, 15] and consists of three distinct phases; mitosis, meiosis, and a postmeiotic differentiation process called spermiogenesis (Fig. 1a). In the mitotic phase, THY1<sup>+</sup> undifferentiated spermatogonia (THY1<sup>+</sup> SG), including a stem cell population, undergo self-renewal or commit to differentiation. Self-renewal of spermatogonia is essential to maintain a stem cell pool throughout life in males. In contrast, in response to retinoic acid, the population of differentiating KIT<sup>+</sup> spermatogonia (KIT<sup>+</sup> SG) enters meiosis undergoing two rounds of division to generate four haploid spermatids, which continue to differentiate into spermatozoa [16, 17]. Pachytene spermatocytes (PS), in the midst of meiotic prophase I, and postmeiotic round spermatids (RS) represent two major populations in adult testes because they stay in relatively longer periods compared to other differentiating cell types in spermatogenesis. Of note, enhancers are dynamically reprogrammed during spermatogenesis; mitotic enhancers are resolved when cells enter meiosis, while meiotic enhancers specifically emerge at this transition [12] (Fig. 1b).

Here, drawing on our initial study reporting the identification of SEs in spermatogenesis [12], we introduce our optimized bioinformatics pipelines to identify SEs in spermatogenesis (Fig. 1c). To gain further insight into SE-driven gene regulation, we also performed an integrative analysis based on our Hi-C data of PS and RS [18]. We found that meiotic and postmeiotic SEs profoundly interact with other genic regions. Our analytics pipelines can be applied to any context beyond spermatogenesis, thus allowing users to identify SEs and 3D chromatin contacts. Of additional note, our analytics pipelines can be applied to CUT&RUN and CUT&Tag, which are new alternative technologies to ChIP-seq [19, 20].



**Fig. 1** Genome-wide remodeling of the active enhancers during spermatogenesis. **(a)** Schematic of mouse spermatogenesis and transcriptional changes. Spermatogenesis can be simplified into three distinct stages: mitotic (THY1<sup>+</sup>, undifferentiated spermatogonia; KIT<sup>+</sup>, differentiating spermatogonia), meiotic (PS, pachytene-stage spermatocyte), and postmeiotic (RS, round spermatid). RS further undergoes spermiogenesis to develop into mature sperm that can swim and fertilize oocytes. The transcriptome of the male germline drastically changed at the mitosis-to-meiosis transition [10, 11]. About 3000 spermatogenesis-specific genes are activated, while approximately 3000 genes expressed in both somatic lineage and spermatogonial progenitor cells are largely suppressed during late spermatogenesis, i.e., in PS and in RS. **(b)** Representative track views of H3K27ac ChIP-seq enrichments. During spermatogenesis, extensive remodeling of enhancer marks takes place at mitosis-to-meiosis transition. Mitotic and meiotic-specific enhancers are highlighted in blue and red. **(c)** The workflow of bioinformatics pipelines for identification of SEs and chromatin contacts between SEs and other genomic regions. The specific web tools and software used for key steps are highlighted in red

## 2 Materials

### 2.1 Computer/Hardware

1. A high-capacity computer with at least 32–64 Gb of RAM (depending on the genome size of interest).

### 2.2 Software

1. Bowtie2 v2.4.4 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml/>) [21].
2. Samtools v1.3.1 (<http://www.htslib.org/>) [22].

3. deepTools v3.1.3 (<https://deeptools.readthedocs.io/en/develop/>) [23].
4. IGV v2.9.2 (<https://software.broadinstitute.org/software/igv/>) [24].
5. MACS2 v2.1.4 (<https://pypi.org/project/MACS2/2.1.4/>) [25].
6. ROSE ([http://younglab.wi.mit.edu/super\\_enhancer\\_code.html](http://younglab.wi.mit.edu/super_enhancer_code.html)) [4, 26].
7. bedtools v2.30.0 (<https://bedtools.readthedocs.io/en/latest/>) [27].
8. MAnorm v1.1.4 (<https://github.com/shao-lab/MAnorm>) [28].
9. GREAT v4.0.4 (<http://great.stanford.edu/public/html/>) [29].
10. HOMER v4.10 (<http://homer.ucsd.edu/homer/>) [30].
11. Burrows–Wheeler Aligner (BWA v0.7.15, <http://bio-bwa.sourceforge.net/>) [31].
12. HiCEXplorer v3.6 (<https://hicexplorer.readthedocs.io/en/latest/>) [32].

### 2.3 NGS Data from Other Sources

Next-generation sequencing (NGS) data are available at Gene Expression Omnibus (GEO) under the following accession codes. Raw reads were downloaded from the Sequence Read Archive (SRA) and converted to the FASTQ format (.fastq).

1. GSE130652 (for H3K27ac ChIP-seq data of THY1<sup>+</sup> and KIT<sup>+</sup> SG) [12].
2. GSE107398 (for H3K27ac ChIP-seq data of PS and RS) [33].
3. GSE119805 (for Hi-C data of PS and RS) [18].

---

## 3 Methods

### 3.1 Processing H3K27ac ChIP-seq Data

A number of computational and statistical tools have been developed and proposed as open sources for processing ChIP-seq reads. In general, the first step of ChIP-seq analysis starts with the alignment of sequence reads to the reference genome. Here, we show standard computational approaches to align raw single-end reads in the FASTQ format from H3K27ac ChIP-seq samples with the mouse genome (GRCm38/mm10) using Bowtie2 (version 2.4.4) [21] with one mismatch permitted (`-N 1`) option. After the alignment step, the files containing the aligned reads will be generated in the SAM format (.sam) and converted to the BAM format (.bam) using Samtools (version 1.3.1) [22]. To visualize H3K27ac ChIP-seq signals throughout the genome, the Java-based Integrative Genome Viewer (IGV, version 2.9.2) [24] browser is used. To visualize the BAM coverage in IGV (i.e., to display the read depth for each position in the genome), the aligned BAM file should first



be converted into the bigWig format using the bamCoverage program as implemented in deepTools (version 3.1.3) [23]. The converted bigWig files are loaded directly into IGV to visualize the aligned reads (Fig. 1b). The following commands can be executed in the command-line UNIX/Linux interface, Terminal for Mac OS X, or Cygwin for Windows.

### 3.1.1 ChIP-seq Alignment

1. Download and decompress a FASTA file (.fasta) containing the GRCm38/mm10 version of *Mus musculus* genome from the UCSC Genome Browser at <https://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/mm10.fa.gz>.
2. Create a Bowtie2 genome index file, saved as a “mm10\_bowtie2\_index” file.

```
$ mkdir ./bowtie2-index
$ bowtie2-build -f mm10.fa mm10_bowtie2_index
```

3. Align raw single-end reads of H3K27ac ChIP-seq data to the mouse genome using Bowtie2, setting the option to very-sensitive, one mismatch permitted (-N 1), and the number of threads to 8 (-p 8). The aligned reads to the reference genome are saved as a SAM file named “alignedFile.sam” (-S).

```
$ bowtie2 -p 8 --very-sensitive -N 1 \
-x ./genome/mm10/bowtie2-index/mm10_bowtie2_index -U
inputChIP-seqFile.fastq -S alignedFile.sam 2> inputChIP-seq-
File.log
```

4. Convert aligned files from the SAM to BAM format (binary version of the SAM format) using Samtools.

```
$ samtools sort alignedFile.sam -o alignedFile.sorted.bam
```

5. Index the resulting .bam files.

```
$ samtools index alignedFile.sorted.bam
```

### 3.1.2 Data Visualization Through BAM Coverage Tracks

The next step in the processing is to generate bam coverage tracks for visualization of ChIP-seq data. We recommend using deepTools [23], a suite of Python tools, to convert the BAM to bigWig format (*see Note 1*).

1. Generate reads per million mapped reads (RPKM)-normalized coverage tracks (the bigWig (.bw) format) using the bamCoverage program as implemented in deepTools.

```
$ bamCoverage --bam ./alignedFile.sorted.bam \
-o ./alignedFile.bw -of bigwig --binSize=10 \
```



```
--minMappingQuality 10 --normalizeUsing RPKM \
--numberOfProcessors=max
```

2. Load the converted bigWig files named “alignedFile.bw” directly into IGV to visualize the aligned reads (Fig. 1b).

### 3.2 Identification of Cell Type-Specific Enhancers and Super-Enhancers

In contrast with typical enhancers, SEs comprise large clusters of typical enhancers (Fig. 2a). SEs can be identified based on H3K27ac ChIP-seq data analysis. This section describes a standard procedure to identify SEs using the ROSE (Rank-Ordering of Super-Enhancer) algorithm [4, 26], which detects enhancers that are ranked based on the H3K27ac ChIP-seq signal intensity. The following commands and procedures are applicable to identify SEs.

#### 3.2.1 Peak Calling with MACS2 and Identification of Active Enhancers

Prior to the identification of SEs, H3K27ac ChIP-seq peaks in each sample are called using MACS2 v2.1.4 (Model-based Analysis of ChIP-Seq) [25] with a threshold q-value =  $10^{-2}$ . MACS2 reports all H3K27ac peaks with a q-value below the threshold in the BED format (Fig. 2b) (*see Note 2*). The outputs of MACS2 peak calling are saved to a user-specified directory (--outdir) with a prefix name (-n, “alignedFile\_macs2”).

```
$ mkdir ./macs2
$ macs2 callpeak -t ./alignedFile.sorted.bam -f BAM -g mm -q 0.01 \
--outdir macs2 -n alignedFile_macs2 \
2> macs2/alignedFile_macs2.log
```

After peak calling with MACS2, in Microsoft Excel, convert the narrowPeak file into a GFF (General Feature Format) file (*see below*), because ROSE only accepts the GFF for peak files. The converted GFF file is saved as a “alignedFile\_macs2.gff” file (*see Note 3*).

Column 1: Chromosome

Column 2: Unique Peak ID

Column 3: Chromosome (same as Column 1)

Column 4: Start

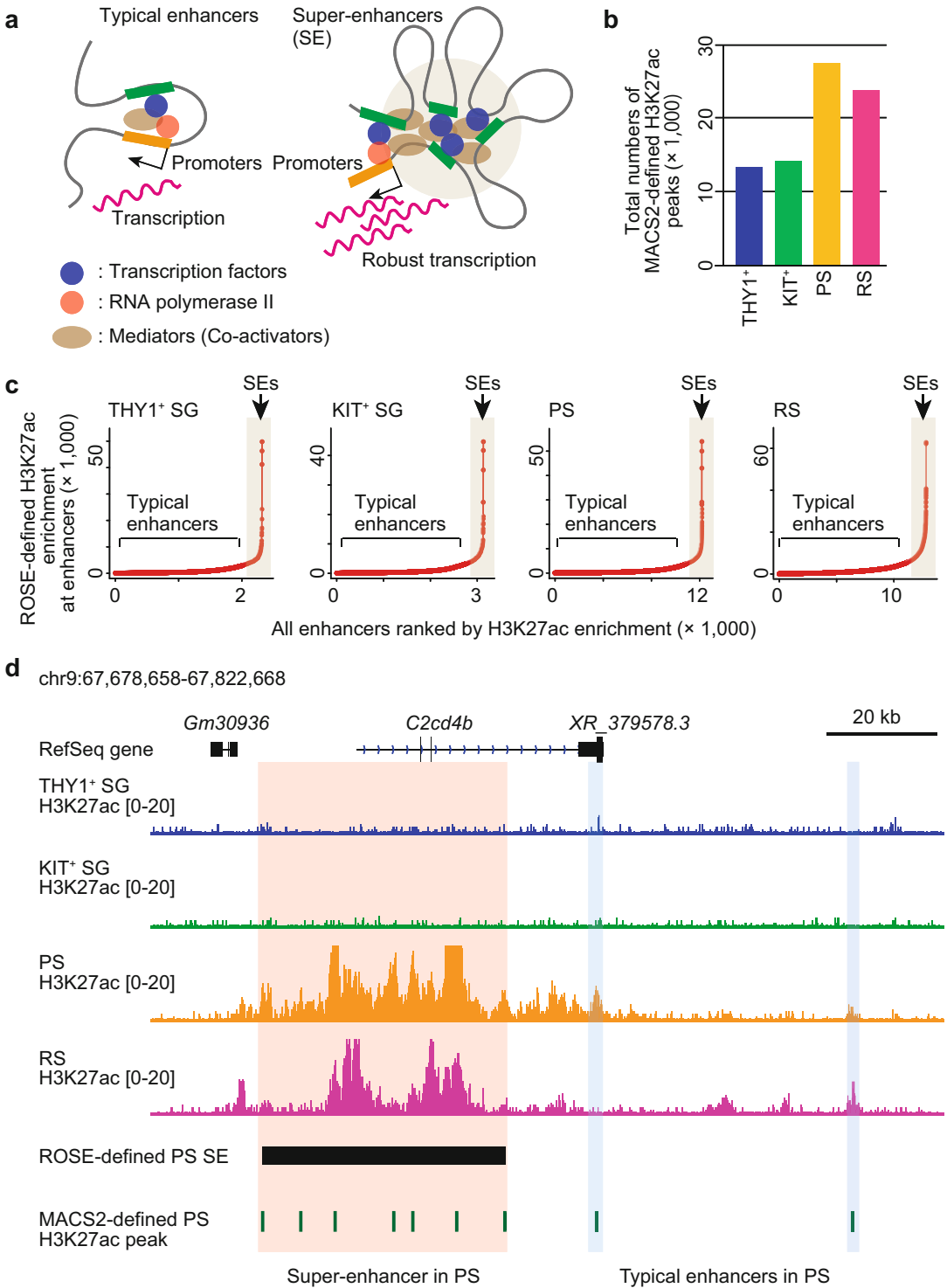
Column 5: End

Column 6: Score

Column 7: Strand (“+” or “-”)

Column 8: Fold enrichment in a given peak compared to background

Column 9: Unique Peak ID (same as Column 2)



**Fig. 2** Identification of super-enhancers during spermatogenesis. **(a)** Schematic of difference between typical enhancers (left) and super-enhancers (SE, right). SEs are large clusters of enhancers spanning multiple kilobases. Key transcription factors bind to target enhancer loci and concentrate high density of mediators/co-

### 3.2.2 Super-Enhancer Calling

1. Download all required files for the ROSE algorithm, including Python scripts and annotation files, from the ROSE code Bit-Bucket ([https://bitbucket.org/young\\_computation/rose/src/master/](https://bitbucket.org/young_computation/rose/src/master/)).
2. Organize all input files into one directory, including Python scripts, annotation directory, the GFF file named “alignedFile\_macs2.gff,” and the BAM file named “alignedFile.sorted.bam.”
3. Run the ROSE program and set the options to the exclusion of gene promoters, -1 kb to +1 kb around transcription start sites (TSSs, -t 1000), and a maximum distance between two candidate enhancers that can be stitched together, to 12.5 kb (default). The outputs of ROSE analysis are saved to a user-specified directory (-o).

```
$ mkdir ./ROSE_result_TSS-1kb
$ python ROSE_main.py -g mm10 -i alignedFile_macs2.gff \
-r alignedFile.sorted.bam -o ROSE_result_TSS-1kb -t 1000
```

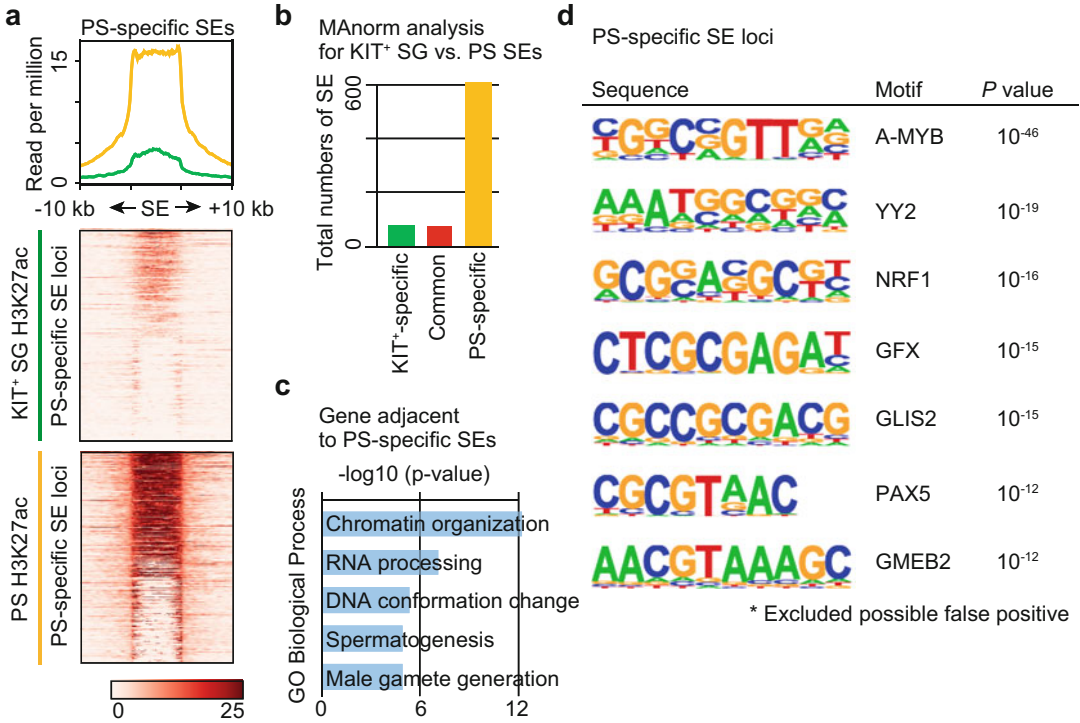
4. The ROSE program automatically calculates an inflection point and defines SEs. As shown in Fig. 2c, the visualized plots enable clear definition of inflection points to distinguish SEs from all typical enhancers in each spermatogenic cell type. Enhancer entities enriched with an H3K27ac signal above the inflection point are defined as SEs. After an SE calling with the ROSE algorithm, the defined SE regions are saved as a BED file named “\*\_Gateway\_SuperEnhancers.bed,” which can be loaded to IGV directly and visualized (Fig. 2d). In total, by using the ROSE algorithm with the above parameters, we identified 68 SEs in THY1<sup>+</sup> SG, 148 SEs in KIT<sup>+</sup> SG, 679 SEs in PS, and 786 SEs in RS (Fig. 2c). This result is largely consistent with our previous identification of SEs in spermatogenesis [12].

### 3.3 Predicted Functions and Biological Significance of Super-Enhancers

Having observed that many SEs emerged upon entry into meiosis onward (Fig. 2c, d), we sought to determine how genome-wide remodeling of SEs leads to cell type-specific binding of transcription factors and regulates the expression of adjacent genes at the mitosis-to-meiosis transition. To this end, we performed the



**Fig. 2** (continued) activators and RNA polymerase II, via long-range chromatin loops to drive robust transcriptions of target genes. **(b)** Numbers of H3K27ac ChIP-seq peaks in spermatogenesis, detected by MACS2 algorithm. **(c)** Identification of SEs based on H3K27ac ChIP-seq data using ROSE algorithm. Y-axis indicates normalized H3K27ac ChIP-seq enrichment, and x-axis indicates the rank of all enhancers based on values given on the y-axis. ROSE calculates inflection points and creates cutoffs to separate SEs (highlighted in gray) from typical enhancers. **(d)** Representative track views of ROSE-defined SE (highlighted in red) and typical enhancers (highlighted in blue) in PS



**Fig. 3** Characterization of regulatory features of super-enhancers using annotation tools. **(a)** An average tag density plot and heatmaps around PS-specific SE ( $\pm 10$  kb) for H3K27ac ChIP-seq enrichments in KIT<sup>+</sup> SG and PS. **(b)** MANorm analysis for SE at mitosis-to-meiosis transition to determine cell type-specific and common SEs between KIT<sup>+</sup> SG and PS stage. **(c)** Gene ontology (GO) analysis of genes adjacent to PS-specific SE, using GREAT web tool (<http://great.stanford.edu/public/html/>). **(d)** HOMER motif analysis for putative transcription factor-binding sites within PS-specific SE loci

following analyses: MANorm [28], GREAT [29], and motif analysis [30] to identify associated functional annotations with SEs, including binding motifs for transcription factors and gene ontology. We compared two stages: a representative mitotic stage (KIT<sup>+</sup> SG) and a representative meiotic stage (PS) (Fig. 3).

3.3.1 MANorm Analysis

To identify and extract PS-specific SE loci, we compared SE loci between KIT<sup>+</sup> SG and PS using MANorm (version 1.1.14) [28]. MANorm is a robust model for quantitative comparison of ChIP-seq peaks (or NGS reads of similar methods) and detects both unique and common features between two biological samples. For the first step of the MANorm analysis, we generated a read file in the BED format from each BAM file named “alignedFile.sorted.bam,” using the bamToBed program as implemented in bedtools (version 2.30.0) [27]. Next, we performed MANorm analysis using the generated read file in the BED format named “reads\_file.bed” and another BED file named “\*\_Gateway\_SuperEnhancers.bed”

which contains the SE loci information from the ROSE algorithm (from **step 4** in Subheading 3.2.2). The following commands and procedures are applicable for MAnorm analysis.

1. Generate “reads\_file.bed” files from “alignedFile.sorted.bam” files of KIT<sup>+</sup> SG and PS.

```
$ bedtools bamtobed -i (KIT+ SG)_alignedFile.sorted.bam >
alignedFile_read_file_1.bed
$ bedtools bamtobed -i (PS)_alignedFile.sorted.bam > aligned-
File_read_file_2.bed
```

2. Run the MAnorm program. The outputs of the MAnorm analysis are saved to a user-specified directory (-o).

```
$ conda install -c bioconda manorm
$ mkdir ./KIT-PS_SE
$ manorm --p1 (KIT+ SG)_Gateway_SuperEnhancers.bed \
--p2 (PS)_Gateway_SuperEnhancers.bed \
--r1 (KIT+ SG)_alignedFile_read_file_1.bed \
--r2 (PS)_alignedFile_read_file_2.bed \
-o ./KIT-PS_SE
```

3. After MAnorm analysis, a “KIT-PS\_SE\_all\_MAValues.xls” file will be generated in a user-specified directory. This file contains the “Peak\_Group” column, which indicates whether each SE in the row is unique or common to the KIT<sup>+</sup> SG and PS stage. In total, 607 SEs could be defined as “PS-unique/specific SEs” (Fig. 3a, b). In Microsoft Excel, convert “PS-unique/specific SEs” file into a BED format file (PS-specific\_SE\_loci.bed), to visualize the average H3K27ac signals and heatmaps from KIT<sup>+</sup> SG and PS ChIP-seq data (alignedFile.bw), using the deepTools.
4. To analyze H3K27ac ChIP-seq profiles around PS-unique/specific SE loci (SE ±10 kb, setting options to -a 10000, -b 10000) in KIT<sup>+</sup> SG and PS, we use the computeMatrix and plotHeatmap programs as implemented in deepTools [23]. To this end, we prepare a BED file with the PS-specific SE loci (from **step 3** in Subheading 3.3.1) and bigWig files with the genomic H3K27ac signals (from **step 2** in Subheading 3.1.2) because the computeMatrix program only accepts multiple score files (in the bigWig format, -S) and multiple region files (in the BED format, -R). Then, create a file containing multiple data matrices (.mat.gz), which is required for plotting average ChIP-seq signals and heatmaps with plotHeatmap program. The output is saved to the current directory with a user-specified name (-o). Notably, PS-specific SEs tend to consist of large and broad H3K27ac enrichment in a meiosis-specific

manner [12] (Fig. 3a). The following commands and procedures are applicable for the visualization of ChIP-seq signals around SE loci.

```
$ computeMatrix scale-regions -S (KIT+ SG)_alignedFile.bw \
(PS)_alignedFile.bw -R PS-specific_SE_loci.bed \
-a 10000 -b 10000 \
-o H3K27ac_enrichment_in_PS-specific_SE.mat.gz
$ plotHeatmap -m ./H3K27ac_enrichment_in_PS-specific_SE.mat.
gz \
--perGroup --colorMap Reds \
-o H3K27ac_enrichment_in_PS-specific_SE.pdf
```

### 3.3.2 GREAT Analysis

Once cell type-specific SEs are identified throughout the genome, the biological significance of those loci can be predicted by analysis using Genomic Regions Enrichment of Annotations Tool (GREAT, version 4.0.4) [29], which associates SE loci with the ontology of their putative target genes, adjacent to SE loci. To perform GREAT analysis, we pasted SE loci in the BED format to <http://great.stanford.edu/public/html/> and selected a species assembly as “Mouse: GRCm38.” In line with our expectation, “Spermatogenesis”-related terms are enriched in a gene set adjacent to the PS-specific SE loci (Fig. 3c).

### 3.3.3 De novo Motif Analysis

Enrichment of consensus-binding motifs for transcription factors on SE loci can be computed by Hypergeometric Optimization of Motif EnRichment (HOMER, version 4.10) [30], de novo motif analysis which predicts transcriptional regulators working on SE loci. To discover motifs, HOMER requires genomic coordinates of interest in the BED format (PS-specific\_SE\_loci.bed). After running a command “findMotifsGenome.pl,” HOMER would create a user-specified output directory (PS-specific\_SE\_MOTIF) with a variety of output files, including the result of de novo motif analysis, written to the HTML format. The following commands and procedures are applicable for HOMER de novo motif analysis.

1. Install HOMER package with conda.

```
$ conda install -c bioconda homer
```

2. Install mouse genome and annotations for all HOMER analyses.

```
$ perl /path-to-homer-4.10-0/configureHomer.pl \
-install mm10
```

3. Run “findMotifsGenome.pl” program to find *de novo* motif within PS-specific SE loci.

```
$ findMotifsGenome.pl PS-specific_SE_loci.bed mm10 \
PS-specific_SE_MOTIF -size given -mask
```

Among the PS-specific SE loci, we observed the most significant enrichment of binding motif for A-MYB (also known as MYBL1), a male germline-specific transcription factor that drives spermatogenesis-related gene expression from the meiotic phase onward [34, 35] (Fig. 3d).

### 3.4 Inferring Long-Range Super-Enhancer Interactions in Meiosis Based on the Hi-C Dataset

High-order chromatin regulates SE activities, shaping cell type-specific gene expression patterns and cellular functions. Recent advances in Hi-C analysis enabled us to systematically interrogate the genome-wide interactions between many *cis*-regulatory elements, such as enhancers and gene promoters, which are distal on the linear genome map but close to each other in cell nuclei. In this section, we exemplify an approach to dissect the long-range chromatin interactions of SEs with examples of PS and RS stages, using a Hi-C dataset. Hi-C data were obtained from our published work [18]. For the best alignment of Hi-C data to reference genome, we recommend using Burrows–Wheeler Aligner (BWA, version 0.7.15) [31] (*see Note 4*). Paired-end read1 (R1) and read2 (R2) FASTQ files of Hi-C data are iteratively aligned to the mouse genome (GRCm38/mm10) using the bwa mem algorithm with optimal parameters (*see below*). Since the published Hi-C libraries of PS and RS were prepared by digestion with Hind III restriction enzyme [18], we then generated a “restrictionCutFile (Hind III)” file in the BED format, which contains information of all restriction cut places with Hind III, using the hicFindRestSites program. This file is used to build Hi-C contact matrices at 10 kb and 100 kb resolutions using the hicBuildMatrix program. Including the above mentioned programs, all programs for Hi-C data analysis (such as building Hi-C contact matrix, calculations of A/B compartment, and topologically associating domains: TADs, prediction of long-range SE interactions, and visualization) are implemented in HiCExplorer (version 3.6) [32]. The following commands and procedures are applicable for HiC data analysis and inferring chromatin interactions of SEs.

#### 3.4.1 Hi-C Reads Alignment

1. Create a BWA genome index file, saved as “mm10\_bwa\_index” files.

```
$ bwa index -p mm10_bwa_index mm10.fa
```

- Individually align raw paired-end FASTQ files from Hi-C in PS and RS samples (Hi-C\_R1.fastq and Hi-C\_R2.fastq, respectively) to the mouse genome using the bwa mem algorithm, setting the optimal parameters to matching score to 1 (-A 1), mismatch penalty to 4 (-B 4), gap extension penalty to 50 (-E 50), clipping penalty to 0 (-L 0), and a number of threads to 32 (-t 32). The individually aligned reads from R1 and R2 to the reference genome are saved as "alignedFile\_R1.sam" and "alignedFile\_R2.sam" files.

[For R1 (first mates)]

```
$ bwa mem -A 1 -B 4 -E 50 -L 0 -t 32 \
./mm10_bwa_index \
Hi-C_R1.fastq > alignedFile_R1.sam \
2> alignedFile_R1.log
```

[For R2 (second mates)]

```
$ bwa mem -A 1 -B 4 -E 50 -L 0 -t 32 \
./mm10_bwa_index \
Hi-C_R2.fastq > alignedFile_R2.sam \
2> alignedFile_R2.log
```

- Convert aligned files from *.sam* to *.bam* format using Samtools.

[For R1 (first mates)]

```
$ samtools view -Shb alignedFile_R1.sam > alignedFile_R1.bam
```

[For R2 (second mates)]

```
$ samtools view -Shb alignedFile_R2.sam > alignedFile_R2.bam
```

### 3.4.2 Building Hi-C Contact Matrices and Visualization

After the alignment of the Hi-C reads, the next step is to construct Hi-C contact matrices based on valid read pairs, in which the Hind III cut site exists between R1 and R2. Uniquely aligned, valid paired reads were only kept and assigned to restriction fragments, whereas invalid reads (i.e., reads from a single side, self-circles, and dangling ends) are removed from building the Hi-C matrix. We then generated a Hi-C matrix in PS and RS samples at 10 kb and 100 kb resolutions (*see Note 5*).

- Install HiCExplorer package with conda.

```
$ conda create --name hicexplorer hicexplorer=3.6 python=3.8
-c bioconda -c conda-forge
$ conda activate hicexplorer
```



2. Create a “restrictionCutFile” file in the BED format using the `hicFindRestSite` program, to detect Hind III cut-sites throughout the genome, being saved as “HindIII\_rest\_site\_positions.bed” file (-o).

```
$ hicFindRestSite --fasta mm10.fa --searchPattern AAGCTT \
-o HindIII_rest_site_positions.bed
```

3. Build a Hi-C matrix at 10 kb resolution, using the `hicBuildMatrix` program. The outputs of the `hicBuildMatrix` program are saved to the user-specified directory with prefix names (“alignedFile\_10kb”).

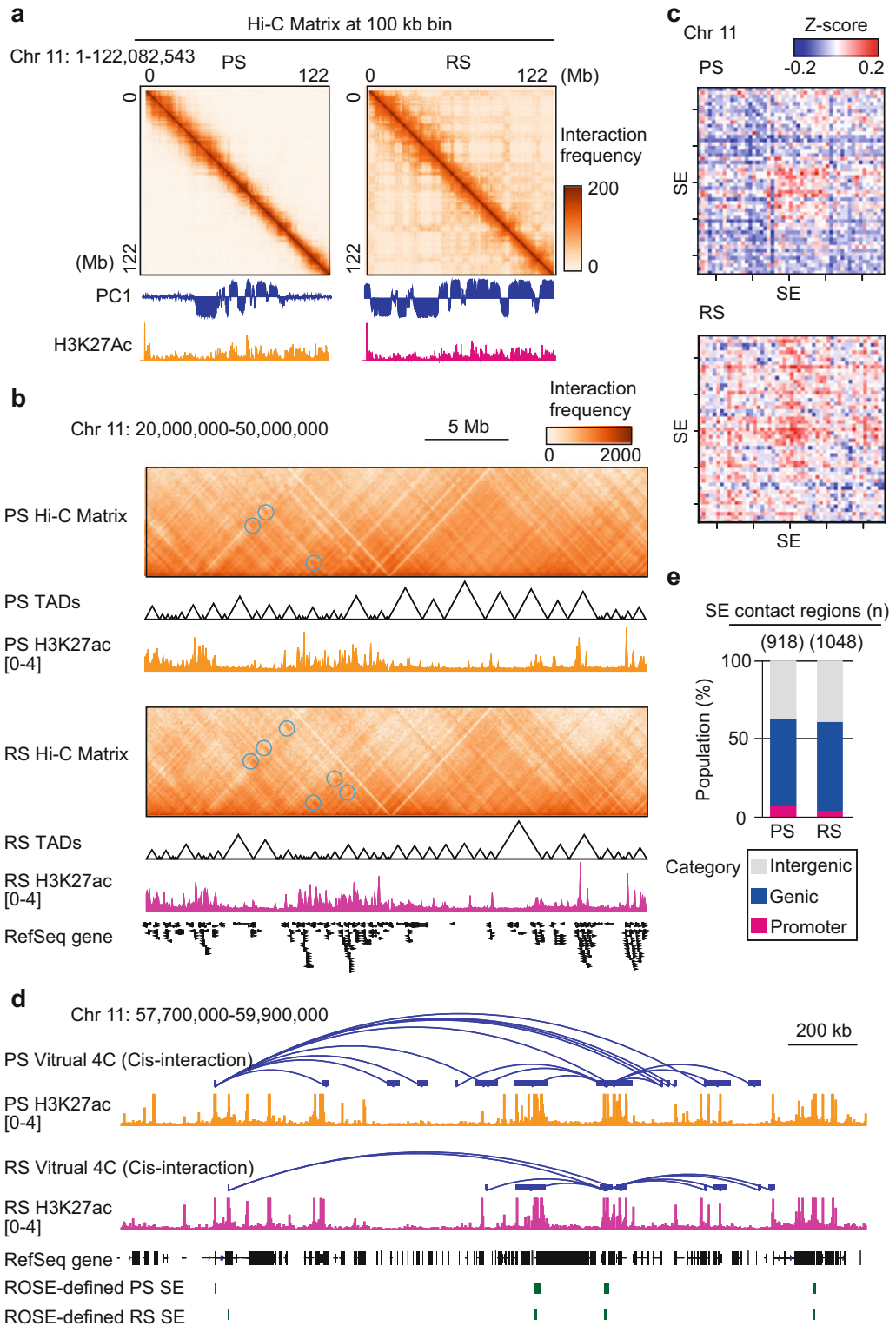
```
$ mkdir hicMatrix
$ hicBuildMatrix --samFiles alignedFile_R1.bam alignedFile_R2.bam \
--binSize 10000 \
--restrictionSequence AAGCTT \
--danglingSequence AGCT \
--restrictionCutFile HindIII_rest_site_positions.bed \
--outBam ./alignedFile_10kb_ref.bam \
--outFileName hicMatrix/alignedFile_10kb.h5 \
--QCfolder hicMatrix/alignedFile_10kb_QC \
--threads 8 \
--inputBufferSize 400000
```

4. After running the `hicBuildMatrix` program, Hi-C contact matrix is saved as “alignedFile\_10kb.h5” file in the HDF5 (Hierarchical Data Format version 5) format. Then, convert the resolution from 10 kb to 100 kb, using the `hicMergeMatrixBins` program (*see Note 6*). Output of the `hicMergeMatrixBins` program is saved as “alignedFile\_100kb.h5.”

```
$ hicMergeMatrixBins --matrix alignedFile_10kb.h5 \
--numBins 10 \
--outFileName hicMatrix/alignedFile_100kb.h5
```

5. Plot Hi-C contact matrix (100 kb resolution), using the `hicPlotMatrix` program (Fig. 4a). Output of `hicPlotMatrix` is saved to the user specified directory in the PDF format (--outFileName “alignedFile\_100kb\_matrix.pdf”).

```
$ mkdir plots
$ hicPlotMatrix --log1p --dpi 300 \
--matrix hicMatrix/alignedFile_100kb.h5 \
--chromosomeOrder chr11 \
--title " Hi-C matrix at 100 kb bin" \
--outFileName plots/alignedFile_100kb_matrix.pdf \
--colorMap Oranges
```



**Fig. 4** Identification of significant chromatin contacts and enhancer–promoter interactions from Hi-C data. (a) Heatmaps showing Hi-C contact matrices (at 100 kb bin, chromosome 11) in PS and RS. Color scale indicates

### 3.4.3 Generation of A/B Compartment Profile from Hi-C Contact Matrix

Within nuclei, genomes are spatially segregated into A and B compartments [36]. The genomic regions in the same compartments are located close to each other in nuclei, whereas A and B compartments generally do not interact. A/B compartments exhibit distinct correlations to transcriptionally active euchromatin and inactive heterochromatic regions of the genome [37]. Therefore, genomic compartmentalization into euchromatin and heterochromatin domains may contribute to the regulation of cell type-specific SE activities and gene expression patterns. In this section, we describe a standard procedure to calculate the A/B compartment profile from processed Hi-C data. A/B compartment profile is computed by the principal component analysis (PCA) of Hi-C contact matrix [37]. The principal component (PC) value can be used to determine the A/B compartment status of a given region, where positive values represent A compartment regions and negative values represent B compartment ones.

1. Prior to PCA, convert Hi-C contact matrix (alignedFile\_100kb.h5) to Pearson correlation matrix (alignedFile\_100kb\_Pearson.h5), using hicTransform program.

```
$ hicTransform --matrix hicMatrix/alignedFile_100kb.h5 \
--method pearson \
--outFileName alignedFile_100kb_Pearson.h5
```

2. Compute PCA of Pearson correlation matrix, using the hicPCA program. PC1 and PC2 values from PCA of Pearson correlation matrix would be saved as a bigWig file.

```
$ hicPCA --matrix alignedFile_100kb_Pearson.h5 \
--outFileName alignedFile_100kb_Pearson_PCA1.bw \
alignedFile_100kb_Pearson_PCA2.bw \
--format bigwig
```

---

**Fig. 4** (continued) a number of reads per bin pair. The principal component (PC) 1 values and H3K27Ac ChIP-seq data are also shown in the bottom panels. **(b)** Hi-C interaction heatmaps (at 100 kb bin, chromosome 11: 20–50 Mb) of PS and RS showing the dynamics of local interactions of genomic loci with H3K27Ac ChIP-seq data. Color scale indicates a number of reads per bin pair. The local pairwise interactions in each 100 kb bin are highlighted in the blue circle. **(c)** Aggregate Hi-C maps around pairs of SE loci on chromosome 11, in PS and RS. Interaction frequencies of each SE pair are represented in a z-score. **(d)** Representative track views of Hi-C interaction in arcs and H3K27ac ChIP-seq enrichments to show interactions between SEs and other genomic regions at 10 kb resolution. Each SE locus in PS and RS (green rectangle) was used as the viewpoint. SEs interact extensively with H3K27ac-enriched local genomic regions (blue rectangle). The significant interactions are shown as arcs ( $p < 0.05$  based on the background negative binominal distribution). **(e)** Genomic distribution of regions of physical contact with PS and RS SEs. More than half of interaction loci were annotated as gene promoters and genic regions by HOMER annotatePeaks.pl

3. After PCA, coverage track of H3K27ac ChIP-seq data (“alignedFile\_H3K27ac.bw”, from **step 2** in Subheading 3.1.2) and PC1 (or PC2 in some chromosomes) values for each 100 kb bin with positive and negative values, corresponding to A and B compartments, respectively (“alignedFile\_100kb\_Pearson\_PCA1.bw,” from **step 2** in Subheading 3.4.3), are plotted with Hi-C contact heatmap (100 kb resolution, “alignedFile\_100kb.h5,” from **step 4** in Subheading 3.4.2).

```
$ hicPlotMatrix --log1p --dpi 300 \
--matrix hicMatrix/alignedFile_100kb.h5 \
--chromosomeOrder chr11 \
--title " Hi-C matrix at 100 kb bin" \
--outFileName plots/alignedFile_100kb_matrix_w_PCA_and_H3-
K27ac_ChIP.pdf \
--colorMap Oranges \
--vMin 1 --vMax 200 \
--bigwig alignedFile_100kb_Pearson_PCA1.bw alignedFile_H3-
K27ac.bw
```

As shown in Fig. 4a, we found that genomic regions marked by the higher magnitude of H3K27ac peak are enriched in the compartment A defined by PCA, in both PS and RS stages.

#### 3.4.4 Detection of Topologically Associating Domains

The Hi-C contact matrix data (10 kb resolution) is used in this step for TAD calling. TADs are ubiquitous across the genome sequence near the diagonal in a Hi-C contact heatmap but are not observed at large genomic distances greater than a few megabases [36]. Since the genomic regions within a TAD would preferentially contact with each other, the *cis*-regulatory elements (i.e., enhancers and gene promoters) located in the same TAD could be activated simultaneously, suggesting that TAD might act as a functional unit for gene regulation. The following commands and procedures are applicable for TAD calling and visualization on local genomic loci.

1. Compute TADs based on Hi-C contact matrix (10 kb resolution, “alignedFile\_10kb.h5,” from **step 3** in Subheading 3.4.2), using the hicFindTADs program and setting the option to a method for statistical comparison and threshold to FDR,  $q < 0.01$  (--correctForMultipleTesting fdr; --thresholdComparisons 0.01), a minimum window length of 80 kb (--minDepth 80000; the value should be at least three times as large as the bin size), a maximum window length of 800 kb (--maxDepth 800000; the value should be greater than ten times the bin size), step size when moving from --minDepth to --maxDepth to 40 kb (--step 40000), and a minimum distance

between boundaries of 80 kb (--minBoundaryDistance 80000). The outputs of TAD calling are saved as a user-defined prefix name (--outPrefix “alignedFile\_10kb\_TAD”).

```
$ hicFindTADs --matrix hicMatrix/alignedFile_10kb.h5 \
--outPrefix alignedFile_10kb_TAD \
--correctForMultipleTesting fdr \
--thresholdComparisons 0.01 \
--minDepth 80000 \
--maxDepth 800000 \
--step 40000 \
--minBoundaryDistance 80000
```

In total, by using hicFindTADs program with the above parameters, we identified 3390 TADs in PS and 3415 TADs in RS (*see Note 7*).

2. After TADs calling, the hicPlotTADs program is used to visualize specific regions for the Hi-C contact heatmap in combination with pairwise point interactions within TADs and coverage tracks of H3K27ac ChIP-seq data (Fig. 4b). For visualization, we generated the following configuration file, bearing multiple tracks (*see below*), because the hicPlotTADs program only accepts .ini format.

[Example of hicPlotTADs.ini file]

```
##SE interaction
[x-axis]
fontsize=7
[hic]
file = /path-to-matrix_file/alignedFile_100kb.h5
title = alignedFile
colormap = Oranges
depth = 12000000
min_value = 1
max_value = 2000
transform = log1p
file_type = hic_matrix
[tads]
file = /path-to-TADs/alignedFile_10kb_TAD_domains.bed
file_type = domains
border_color = black
color = none
[spacer]
width = 10
[bigwig]
file = /path-to-H3K27ac.bw/alignedFile_H3K27ac.bw
title = H3K27Ac
```

```

color = red
width = 50
[gene]
file = /path-to-Refgene/mm10_genes.bed
title = Refgene
color = black
file_type = bed
width = 50

```

Then, run the following command:

```

$ hicPlotTADs --tracks hicPlotTAD_SE.ini \
--region chr11:20,000,000-50,000,000 \
-o hicPlotTAD_SE.pdf

```

#### 3.4.5 Generation of Aggregate Hi-C Map Between Pairs of Super-Enhancers

In this step, we are particularly interested in identifying SE aggregates associated with cell type-specific gene expression. To do this, we generated an aggregate Hi-C contact map centered on the SEs, in PS and RS stages, using the `hicAggregateContacts` program and the following command. Our analysis revealed that aggregate Hi-C contact maps in PS and RS clearly show the establishment of SE aggregates, in which the interactions are strengthened in RS compared to PS (Fig. 4c). We confirmed the interactions between each SE, supporting the model of SEs and SE-driven gene regulation (Fig. 2a).

```

$ hicAggregateContacts --matrix hicMatrix/alignedFile_10kb.h5 \
--BED Gateway_SuperEnhancers.bed \
--outFileName Aggregate_Contact_SE_intra-chr.pdf \
--transform z-score \
--mode intra-chr \
--chromosomes chr11 \
--colorMap bwr \
--range 200000:3000000

```

#### 3.4.6 Identification of Super-Enhancer–Promoter Interactions with Virtual 4C Analysis

Given that SEs clearly interacted with each other in both PS and RS stages (Fig. 4c), we next analyzed long-range interactions between pairs of individual SE and genic regions to investigate whether cell type-specific SEs directly regulate the expression of target genes. For this purpose, we applied virtual 4C analysis for the Hi-C contact matrix at each SE [38]. The following commands and procedures are applicable for the identification of SE-target genes in close physical proximity to the SE loci in the nuclei.

1. Compute the quality of each viewpoint (SE locus) using the `chicQualityControl` program, which considers the sparsity of the Hi-C contact frequencies. With this step, a viewpoint

would be removed if the sparsity at this viewpoint is lower than a user-defined threshold (`--sparsity`). The QC-passed SE loci are written and renamed from “new\_referencepoints.bed” to “Gateway\_SuperEnhancers\_filtered.bed,” which is used for subsequent analysis.

```
$ chicQualityControl --matrix hicMatrix/alignedFile_10kb.h5 \
-rp Gateway_SuperEnhancers.bed \
--sparsity 0.3 \
--threads 8
```

2. Compute the background model for all given viewpoints based on Hi-C contact matrix, using BED file containing filtered viewpoints and `chicViewpointBackgroundModel` program, setting the range of background model to 500 kb (`--fixateRange 500000`). Model is written and stored to user-defined file in .txt format (“alignedFile\_10kb\_model.txt”).

```
$ chicViewpointBackgroundModel \
--matrix hicMatrix/alignedFile_10kb.h5 \
--fixateRange 500000 \
--threads 8 \
-rp Gateway_SuperEnhancers_filtered.bed \
--outFileName alignedFile_10kb_model.txt
```

3. Detect all interaction points that physically contact SE loci, ranging up to 500 kb (`--range 500000 500000`), based on a background model, using the `chicViewpoint` program (*see Note 8*). After running the `chicViewpoint` program, the “InteractionFiles” folder is generated in the current directory containing all interaction points for each SE in .txt format.

```
$ chicViewpoint --matrix hicMatrix/alignedFile_10kb.h5 \
--range 500000 500000 \
-rp Gateway_SuperEnhancers_filtered.bed \
--backgroundModelFile alignedFile_10kb_model.txt
```

4. Extract only significant interaction points ( $P < 0.05$  (`-p 0.05`)) using `chicSignificantInteractions` program.

```
$ chicSignificantInteractions \
--interactionFile InteractionFiles/*.txt \
-p 0.05 \
--backgroundModelFile alignedFile_10kb_model.txt \
--range 500000 500000 \
--loosePValue 0.1 \
--threads 8
```

5. After running the `chicSignificantInteractions` program, the “targetFolder” is generated in the current directory. Each file contains significant interaction points ( $p < 0.05$ , based on the

background negative binominal distribution) for each SE in .txt format. The file name of the individual .txt file in the "targetFolder" includes genomic location (chromosome, start, end) of viewpoint (individual SE), and contains a list of significant interaction points in the BED format, consisting of three columns (chromosome, start, end). To visualize chromatin contacts between individual SE loci and other genomic regions on IGV, we generated a BEDPE file (.bedpe) containing two sets of genomic coordinates, displaying interactions between viewpoints and interaction points in arcs. Following shell script is applicable to make .bedpe file, based on files in the "targetFolder". folder.

```
$ bash ./makepe_from_targetFolder.sh > Interaction.bedpe
```

```
[makepe_from_targetFolder.sh]
```

```
#!/usr/bin/#!/usr/bin/env bash

target_files="./targetFolder/*"
prefix="alignedFile_10kb_"

for filepath in $target_files; do

regex="${prefix}(chr[0-9]*)_([0-9]*)_([0-9]*)_"
if [[ $filepath =~ $regex ]]; then
chr1=${BASH_REMATCH[1]}
start1=${BASH_REMATCH[2]}
end1=${BASH_REMATCH[3]}
fi

while IFS= read -r line; do
if [[ $line =~ ^chr[0-9]+ ]]; then
echo -e "${chr1}\t${start1}\t${end1}\t${line}\t"
fi
done < $filepath
done
```

6. Load the generated "Interaction.bedpe" files directly into IGV to visualize the interactions in arcs (Fig. 4d).
7. Assess the genomic features of the interaction point (10 kb bin) with SE using the annotatePeaks.pl program as implemented in HOMER.

```
$ annotatePeaks.pl ./Interaction_point_w_SE.bed mm10 > Interaction_point_w_SE_genomic_feats.txt
```



Overall, we found that more than half of the interaction points were annotated as gene promoters and genic regions, suggesting that SEs directly drive the robust expression of target genes (Fig. 4c).

---

## 4 Notes

1. To compare tracks between different samples, the sequencing depth in each sample should be normalized with RPKM.
2. Only the peaks not located at gene promoters ( $-1$  kb to  $+1$  kb around TSSs) can be considered as active enhancers.
3. A GFF file consists of nine columns; six of them are mandatory for the algorithm to work properly.
4. Other aligners (e.g., HiSAT2 and Bowtie2) have performed worse in our experience.
5. Owing to the read depth of our Hi-C data, higher-resolution analyses were not currently possible.
6. Since a 10-kb bin matrix is too large to plot as a heatmap, reduce the resolution from 10 kb to 100 kb.
7. The numbers of TADs are relatively high compared to our previous analysis [18]; this is likely due to different methods to align Hi-C data and create Hi-C matrices.
8. Use the same fixate range setting for the background model computation.

---

## Acknowledgments

We thank Katie Gerhardt for editing the manuscript, Yuka Kitamura for critical reading of the manuscript, all members of Siomi Lab at Keio University School of Medicine, and Namekawa Lab at UC Davis for discussions and comments on this manuscript. This work was supported by JSPS Grants-in-Aid for Early-Career Scientists and Kato Memorial Bioscience Foundation Research Grant to A.S. and NIH grants R01 GM122776 and R35 GM141085 to S. H.N.

## References

1. Visel A, Rubin EM, Pennacchio LA (2009) Genomic views of distant-acting enhancers. *Nature* 461(7261):199–205
2. Ong CT, Corces VG (2011) Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet* 12(4): 283–293
3. Jia Y, Chng WJ, Zhou J (2019) Super-enhancers: critical roles and therapeutic targets in hematologic malignancies. *J Hematol Oncol* 12(1):77
4. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH et al (2013) Master transcription factors and mediator establish

- super-enhancers at key cell identity genes. *Cell* 153(2):307–319
5. Wang X, Cairns MJ, Yan J (2019) Super-enhancers in transcriptional regulation and genome organization. *Nucleic Acids Res* 47(22):11481–11496
  6. Sabari BR, Dall’Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K et al (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361(6400):eaar3958
  7. Alam H, Tang M, Maitituoheti M, Dhar SS, Kumar M, Han CY et al (2020) KMT2D deficiency impairs super-enhancers to confer a glycolytic vulnerability in lung cancer. *Cancer Cell* 37(4):599–617 e597
  8. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA et al (2013) Super-enhancers in the control of cell identity and disease. *Cell* 155(4):934–947
  9. Peeters JG, Vervoort SJ, Tan SC, Mijnheer G, de Roock S, Vastert SJ et al (2015) Inhibition of super-enhancer activity in autoinflammatory site-derived T cells reduces disease-associated gene expression. *Cell Rep* 12(12):1986–1996
  10. Hasegawa K, Sin HS, Maezawa S, Broering TJ, Kartashov AV, Alavattam KG et al (2015) SCML2 establishes the male germline epigenome through regulation of histone H2A ubiquitination. *Dev Cell* 32(5):574–588
  11. Sin HS, Kartashov AV, Hasegawa K, Barski A, Namekawa SH (2015) Poised chromatin and bivalent domains facilitate the mitosis-to-meiosis transition in the male germline. *BMC Biol* 13:53
  12. Maezawa S, Sakashita A, Yukawa M, Chen X, Takahashi K, Alavattam KG et al (2020) Super-enhancer switching drives a burst in gene expression at the mitosis-to-meiosis transition. *Nat Struct Mol Biol* 27(10):978–988
  13. Maezawa S, Yukawa M, Alavattam KG, Barski A, Namekawa SH (2018) Dynamic reorganization of open chromatin underlies diverse transcriptomes during spermatogenesis. *Nucleic Acids Res* 46(2):593–608
  14. Fayomi AP, Orwig KE (2018) Spermatogonial stem cells and spermatogenesis in mice, monkeys and men. *Stem Cell Res* 29:207–214
  15. Matzuk MM, Lamb DJ (2008) The biology of infertility: research advances and clinical challenges. *Nat Med* 14(11):1197–1213
  16. Anderson EL, Baltus AE, Roepers-Gajadien HL, Hassold TJ, de Rooij DG, van Pelt AM et al (2008) Stra8 and its inducer, retinoic acid, regulate meiotic initiation in both spermatogenesis and oogenesis in mice. *Proc Natl Acad Sci U S A* 105(39):14976–14980
  17. O’Donnell L (2014) Mechanisms of spermiogenesis and spermiation and how they are disturbed. *Spermatogenesis* 4(2):e979623
  18. Alavattam KG, Maezawa S, Sakashita A, Khoury H, Barski A, Kaplan N et al (2019) Attenuated chromatin compartmentalization in meiosis and its maturation in sperm development. *Nat Struct Mol Biol* 26(3):175–184
  19. Skene PJ, Henikoff S (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *elife* 6:e21856
  20. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG et al (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 10(1):1930
  21. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359
  22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078–2079
  23. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T (2014) deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42(Web Server issue):W187–W191
  24. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G et al (2011) Integrative genomics viewer. *Nat Biotechnol* 29(1):24–26
  25. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE et al (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9(9):R137
  26. Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR et al (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153(2):320–334
  27. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6):841–842
  28. Shao Z, Zhang Y, Yuan GC, Orkin SH, Waxman DJ (2012) MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol* 13(3):R16
  29. McLean CY, Bristol D, Hiller M, Clarke SL, Schaar BT, Lowe CB et al (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28(5):495–501
  30. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P et al (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38(4):576–589

31. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25(14):1754–1760
32. Wolff J, Bhardwaj V, Nothjunge S, Richard G, Renschler G, Gilsbach R et al (2018) Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. *Nucleic Acids Res* 46(W1):W11–W16
33. Adams SR, Maezawa S, Alavattam KG, Abe H, Sakashita A, Shroder M et al (2018) RNF8 and SCML2 cooperate to regulate ubiquitination and H3K27 acetylation for escape gene activation on the sex chromosomes. *PLoS Genet* 14(2):e1007233
34. Bolcun-Filas E, Bannister LA, Barash A, Schimenti KJ, Hartford SA, Eppig JJ et al (2011) A-MYB (MYBL1) transcription factor is a master regulator of male meiosis. *Development* 138(15):3319–3330
35. Li XZ, Roy CK, Dong X, Bolcun-Filas E, Wang J, Han BW et al (2013) An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol Cell* 50(1):67–81
36. Dekker J, Marti-Renom MA, Mirny LA (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* 14(6):390–403
37. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950):289–293
38. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M et al (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148(3):458–472



## A Method for Detection of Somatic LINE-1 Insertions at the Single-Cell Level from Postmortem Human Brain

Miki Bundo and Kazuya Iwamoto

### Abstract

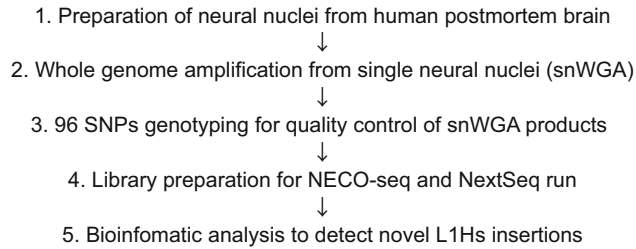
Long Interspersed Element-1 (LINE-1, L1) is a retrotransposon that has the ability to amplify its copy in the genome autonomously. L1Hs is a human-specific active subtype of L1 reported to amplify its copy in neural progenitor cells causing genomic mosaicism. This chapter describes a new method named NECO-seq (Novel Elements Concentrated-sequence) to identify the genomic locus of L1Hs insertions at the single-cell level. This protocol contains the steps of (1) preparation of neuronal cell nuclei from a postmortem human brain, (2) whole genome amplification from single neural nuclei (snWGA), (3) single nucleotide polymorphisms (SNPs) genotyping for quality control of snWGA products, (4) library preparation for next-generation sequencing to enrich the genomic locus of L1Hs insertions, and (5) bioinformatic analysis to detect novel somatic L1Hs insertions. This method can detect approximately 97% of L1Hs originally existing in reference human genome and approximately 10–20 newly inserted L1Hs copies in a neuronal cell of a postmortem human brain.

**Key words** L1, Postmortem brain, Single-cell nuclei, Whole genome amplification, Next-generation sequencing

---

## 1 Introduction

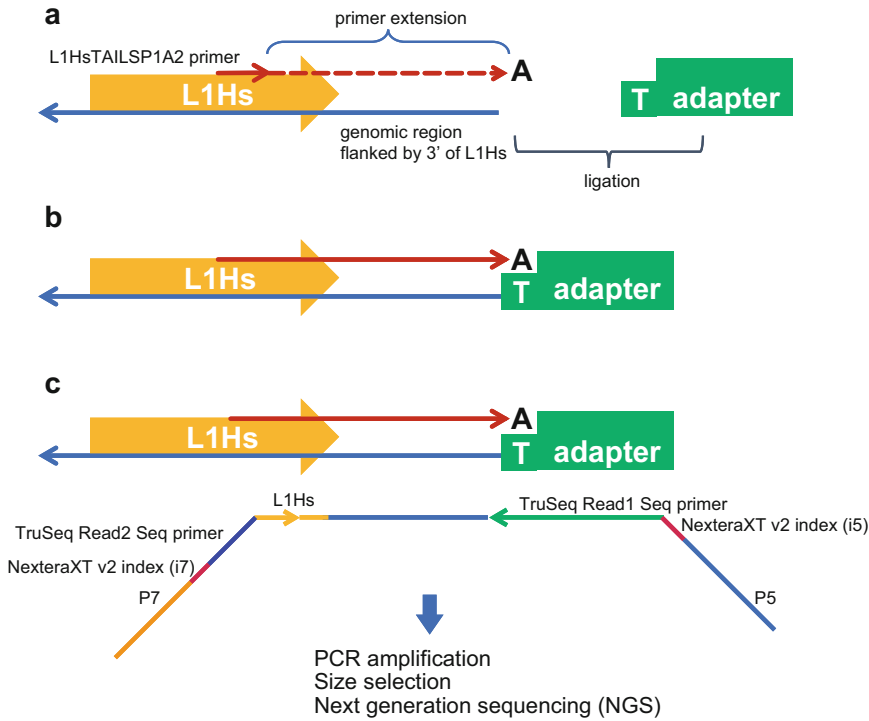
Long Interspersed Element-1 (LINE-1, L1) is one of the retrotransposons, which has approximately 600,000 copies and occupies 17% of the human genome [1]. L1Hs is the only L1 subtype that still maintains the ability of amplifying its copy in the human genome by the process called retrotransposition [2]. It has been thought that L1Hs transposable activity was strictly inhibited and insertion of L1Hs occurred rarely (one event in several dozen generations) only in germ cells or the early developmental stage. However, some reports have shown that L1Hs can amplify its copy in neural progenitor cells and the copy number of L1Hs was increased in the brain tissue compared to other tissues in the same



**Fig. 1** Scheme of NECO-seq experiment. Number of each item corresponds to the section number in the text

individuals [3]. As these L1Hs insertions in the brain should occur in the different genomic sites in each cell, the identification of novel L1Hs insertions will be required to be performed at single-cell level. This chapter describes a new method that we developed, named NECO-seq (Novel Elements Concentrated-sequence), to detect the novel L1Hs insertions from single neural cell in the human postmortem brain (Fig. 1). This method involves (1) preparation of neural nuclei from a human postmortem brain, (2) whole genome amplification from single neural nuclei (snWGA), (3) 96 SNPs genotyping for quality control of snWGA products, (4) PCR enrichment of the genomic regions containing L1Hs and its 3' side, and (5) bioinformatics to identify the novel L1Hs insertions occurring only in 1 cell.

In **step 1**, to prevent DNA loss, it is important to prepare the intact cell nuclei without giving mechanical stimulus. The cell nuclei preparation method was established in an earlier study and was appropriate for this experiment [4]. In **step 2**, snWGA is performed using a microfluidic device of C1 Single-Cell Auto Prep System (Fluidigm). This step can amplify DNA from single nuclei (approximately 6 pg) to approximately 100 ng. In this step, allele dropout (ADO) which occurs during WGA reaction can be a significant problem in single-cell analysis. ADO is a partial failure of genome amplification which causes DNA loss from one of the two alleles in any region of the genome. In **step 3**, to select the snWGA products without ADO, each snWGA product is extensively genotyped, and the heterogeneous call rates at heterogeneous SNPs of each individual are checked. In **step 4**, genomic regions containing 3' terminal of L1Hs and its 3' side are amplified by primer extension from L1Hs-specific primer followed by PCR (Fig. 2). In **step 5**, by applying a bioinformatic pipeline focusing on selecting high-confidence sequence reads and filtering the false-positive findings, novel L1Hs insertions are identified. This method can detect approximately 97% of L1Hs originally existing in a reference human genome and approximately 10–20 newly inserted L1Hs copy in a neuronal cell of a human postmortem brain.



**Fig. 2** Scheme of NECO-seq library preparation. **(a)** Primer extension from 3' terminal of L1Hs toward 3' direction is performed using fragmented snWGA products and L1HsTAILSP1A2 primer [2] by Taq polymerase. An adenine is added to the 3' end of the product. **(b)** Double-stranded adapter with a protruding thymine at 3' end is ligated to the primer extension product. **(c)** 3' of L1Hs within the genomic region flanked by 3' end of L1Hs is amplified by PCR. PCR primers which contain the sequences required for NextSeq. PCR products are size selected and run by NextSeq

## 2 Materials

### 2.1 Preparation of Neural Nuclei from Human Postmortem Brain

Refer to the method of neural cell nuclei isolation from a postmortem brain previously described [4].

### 2.2 Whole Genome Amplification from Single Neural Nuclei (snWGA)

#### 2.2.1 Equipment

1. C1 Single-Cell Auto Prep System (Fluidigm, South San Francisco, CA, USA).
2. Fluorescence microscope BZ-X710 (Keyence, Osaka, Japan) or equivalent.
3. FilterMax F5 plate reader (Molecular Devices, San Jose, CA, USA) or equivalent.

#### 2.2.2 Solutions and Reagents

1. Single-Cell PreAmp IFC, 5–10  $\mu$ m (Fluidigm, P/N 100-5757).
2. C1 Single-Cell Reagent Kit for DNA Seq\* (Fluidigm, P/N 100-7357).

\*To prevent the cell nuclei attaching to the surface of plastic vessels, some reagents in this kit are modified as follows: 30% bovine serum albumin (BSA)/PBS is added to C1 Blocking Reagent and C1 DNA Seq Cell Wash Buffer at a final concentration of 2%.

Instead of C1 Suspension Reagent, a mixture containing 12  $\mu$ L of 1% BSA/PBS and 8  $\mu$ L of C1 Suspension Reagent is used.

3. QuantiFluor dsDNA System (Promega, Madison, WI, USA, P/N E2671) or equivalent.

### **2.3 96 SNPs Genotyping for Quality Control of snWGA Products**

#### *2.3.1 Equipment*

#### *2.3.2 Solutions and Reagents*

1. BioMark HD system (Fluidigm).

1. 96.96 GT Sample/Loading & SNP Type Reagent Kit 10 IFCs (Fluidigm, P/N BMK-M10-96.96GT-SNP).
2. SNP Trace Panel (Fluidigm, P/N SKU 100-6280).

### **2.4 Library Preparation for NECO- seq and NextSeq Run**

#### *2.4.1 Equipment*

#### *2.4.2 Solutions and Reagents*

1. Covaris M220 and microTUBE-15 AFA Beads Screw-Cap (P/N 520145) (Covaris, Woburn, MA, USA).
2. Thermal cycler.
3. MultiNA (Shimadzu, Kyoto, Japan).
4. NextSeq (Illumina, San Diego, CA, USA).

1. 10 mM Tris-Cl (pH 8.5) or Buffer EB (Qiagen, Hilden, Germany, P/N 19086).
2. ExoSAP-IT Express PCR Product Cleanup Reagent (Thermo Fisher Scientific, Waltham, MA, USA, P/N 78201.1.ML).
3. L1HsTAILSP1A2 primer [2] (*see* Table 1).
4. Ex Taq (Takara, Shiga, Japan, P/N RR001A).
5. NucleoMag kit for cleanup and size selection of NGS library prep reactions (Macherey-Nagel, Düren, Germany, P/N 744970.5).
6. Tris-EDTA, pH 8.0 (TE 8.0).
7. Top strand for adapter (*see* Table 1).
8. Bottom strand for adapter (*see* Table 1).
9. T4 DNA ligase (Takara, P/N 2011A).
10. Forward primer for PCR (*see* Table 1).
11. Reverse primer for PCR (*see* Table 1).

**Table 1**  
**List of primers used in this protocol**

<b>Name</b>	<b>Sequence (5'→3')</b>
L1HsTAILSPIA2 for primer extension	GGGAGATATACCTAATGCTAGATGACAC
Top strand for adapter <sup>a</sup>	GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGA
Bottom strand for adapter <sup>a</sup>	TCTACACTCTTTCCCTACACGACGGCTCTTCCGATCT
Forward primer for PCR <sup>b</sup>	AATGATACGGCGACCCACCGAGATCTACACCTCTCTATATACACTCTTTCCCTACAC GACGCTCTTCCGATCT
Reverse primer for PCR <sup>b</sup>	CAAAGCAGAAGACGGGCATACGAGATTCGCCCTTAGTACTGGAGTTCAGACCGTGTGCTCTTCC GATCTTGACATGTACCCCTAAACTTAG

<sup>a</sup>These oligos are used for double-stranded adapters by annealing to each other

<sup>b</sup>These primers contain index sequences (with underbar) from Nextera XT Index Kit v2; forward primers and reverse primers each have i5 and i7 index (Illumina document of "Illumina Adapter Sequences" # 100000002694 v16)

Combination of i5 and i7 index can be referred to the Illumina document of "Index Adapters Pooling Guide" (# 1000000041074 v11)



12. Q5 High-Fidelity DNA Polymerase (NEB, Ipswich, MA, USA, P/N M0491).
13. MultiNA DNA-12000 kit (Shimadzu, P/N 292-36600-91).
14. NextSeq 500/550 High Output Kit v2.5, 300 cycles, or 500/550 Mid Output Kit v2.5, 300 Cycles (Illumina, P/N 20024908 or 20024905).

---

## 3 Methods

### 3.1 Preparation of Neural Nuclei from Human Postmortem Brain

Prepare neural nuclei (e.g., NeuN+ neuronal nuclei) from a human postmortem brain as previously described [4], and stop the experiment prior to the nuclei sorting step (Subheading 3.3, step 6, in the previously published protocol [4]), as nuclei would be sorted directly into integrated fluidic circuits (IFC) as described in Subheading 3.2, step 2).

### 3.2 Whole Genome Amplification from Single Neural Nuclei (snWGA)

snWGA is performed using a microfluidic device of C1 Single-Cell Auto Prep System, Single-Cell PreAmp IFC, and C1 Single-Cell Reagent Kit for DNA Seq following the manufacturer's instructions (*see Note 1*). Some reagents are slightly modified, as described in the Materials section.

1. Perform priming the Single-Cell PreAmp IFC by running "DNA seq: Prime (1791x/1792x/1793x)" script at the C1 System (*see Note 2*).
2. For neuronal nuclei loading to the cell inlet of IFC, sort 1000 of NeuN+ cell nuclei directly into the cell inlet by using the setting for 384-plate sorting following the cell sorter manufacturer's instructions (*see Note 3*).
3. Perform nuclei loading by running "DNA Seq: Cell Load (1791x/1792/1793)" script at the C1 System (*see Note 4*).
4. Confirm the IFC chambers containing only one nucleus by observing under the fluorescence microscope BZ-X710 (*see Note 5*).
5. Perform WGA reaction by running "DNA Seq: Amplify (1791x/1792/1793)" script at the C1 System.
6. Prepare positive and negative tube controls according to the manufacturer's instructions (*see Note 6*).
7. Harvest the snWGA products (approximately 3.5  $\mu$ L) to a 96-well plate with 10  $\mu$ L of C1 DNA dilution buffer in each well by following the manufacturer's instructions.
8. Quantify the snWGA products (2  $\mu$ L) using QuantiFluor dsDNA System and FilterMax F5 plate reader (*see Note 7*).

**3.3 96 SNPs  
Genotyping for Quality  
Control of snWGA  
Products**

1. To select the snWGA products that successfully amplified whole genomic regions uniformly, genotype snWGA products with 96.96 GT Sample/Loading & SNP Type Reagent Kit, 96.96 Dynamic Array IFC and SNP Trace Panel following the manufacturer's instructions (*see Note 8*). Genotype the positive and negative tube controls of the snWGA reaction from each subject simultaneously.
2. Determine the allele dropout (ADO) rate by comparing the intensities of informative SNPs with those of homozygous allele references (*see Note 9*). Avoid the use of snWGA products with severe ADO or amplification failure.

**3.4 Library  
Preparation for NECO-  
seq and NextSeq Run**

1. To prepare the double-stranded adapter, prepare a mixture containing 80  $\mu\text{L}$  of double distilled water (ddW), 10  $\mu\text{L}$  of 100  $\mu\text{M}$  of top strand, and 10  $\mu\text{L}$  of 100  $\mu\text{M}$  bottom strand (10  $\mu\text{M}$  each strand). To anneal them to each other, incubate the mixture at 65  $^{\circ}\text{C}$  for 10 min, and reduce the temperature by 0.5  $^{\circ}\text{C}$  every minute for 80 cycles using a thermal cycler to reach a final temperature of 25  $^{\circ}\text{C}$  (*see Note 10*).
2. To inactivate residual dNTPs and primers used in the former WGA reaction step, incubate 35 ng of snWGA product in 15  $\mu\text{L}$  of 10 mM Tris-Cl (pH 8.5) with 1.2  $\mu\text{L}$  of ExoSAP-IT Express PCR Product Cleanup Reagent at 37  $^{\circ}\text{C}$  for 10 min and then at 80  $^{\circ}\text{C}$  for 3 min.
3. Fragment the snWGA product with Covaris M220 and microTUBE-15 AFA Beads Screw-Cap to 700 bp with the following setting: peak power 30.0, duty factor 20.0, and cycles/burst 50, for 10 s.
4. Perform primer extension by preparing a mixture containing 7.8  $\mu\text{L}$  of ddW, 2.5  $\mu\text{L}$  of 10 $\times$  Ex Taq Buffer ( $\text{Mg}^{2+}$  plus), 2  $\mu\text{L}$  of 2.5 mM of each dNTP, 0.5  $\mu\text{L}$  of 10  $\mu\text{M}$  of L1HsTAIL-SP1A2 primer, 0.2  $\mu\text{L}$  of Ex Taq, and 12  $\mu\text{L}$  of the fragmented snWGA product. Incubate the mixture in the following order: 98  $^{\circ}\text{C}$  for 10 s, 52  $^{\circ}\text{C}$  for 30 s, and 72  $^{\circ}\text{C}$  for 1 min. Add 25  $\mu\text{L}$  of TE 8.0 to the product, and purify it with NucleoMag kit for cleanup and size selection of NGS library prep reactions without using size selection. Elute it with 10  $\mu\text{L}$  of 10 mM Tris-Cl (pH 8.5).
5. Ligate the product to the double-stranded adapter. Prepare the mixture containing 1  $\mu\text{L}$  of 10 $\times$  T4 ligase buffer, 0.5  $\mu\text{L}$  of 10  $\mu\text{M}$  double-stranded adapter, 0.5  $\mu\text{L}$  of T4 DNA ligase, and 8  $\mu\text{L}$  of purified primer extension product. Incubate the mixture at 16  $^{\circ}\text{C}$  for 1 h.
6. Perform PCR. Prepare the PCR mixture containing 17.3  $\mu\text{L}$  of ddW, 6  $\mu\text{L}$  of 5 $\times$  Q5 Reaction Buffer, 2.4  $\mu\text{L}$  of 2.5 mM each dNTPs, 1.5  $\mu\text{L}$  of 10  $\mu\text{M}$  forward primer, 1.5  $\mu\text{L}$  of 10  $\mu\text{M}$

reverse primer, 0.3  $\mu\text{L}$  of Q5 High-Fidelity DNA Polymerase, and 1  $\mu\text{L}$  of ligated DNA. The PCR conditions are as follows: 27 cycles at 98 °C for 10 s, 67 °C for 30 s, 72 °C for 30 s, and 1 cycle at 72 °C for 10 min.

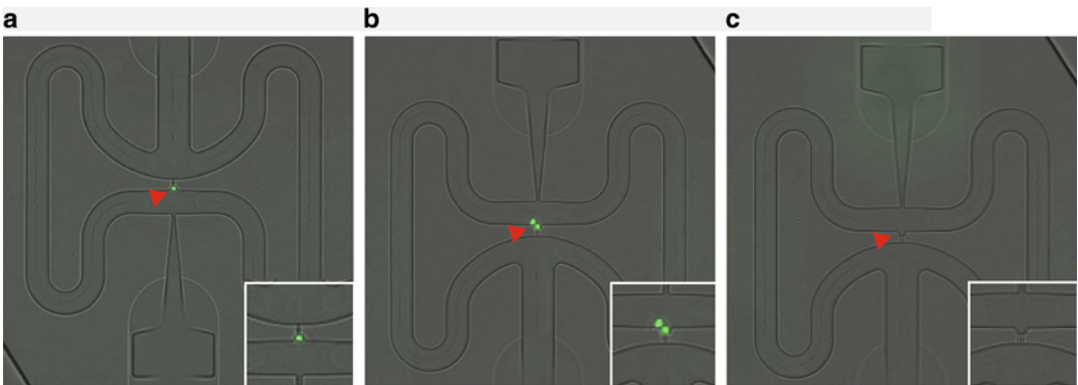
7. Add 20  $\mu\text{L}$  of TE 8.0 to 27  $\mu\text{L}$  of PCR product. Purify the PCR product with size selection from 500 to 700 bp with NucleoMag kit for cleanup and size selection of NGS library prep reactions following the manufacturer's instructions, and elute it with 20  $\mu\text{L}$  of 10 mM Tris-Cl (pH 8.5).
8. Quantify the concentration by MultiNA and MultiNA DNA-12000 kit with the smear analysis software using 2  $\mu\text{L}$  of the product (*see Note 11*).
9. Prepare the mixture of libraries with 1 nM of each library with 10 mM Tris-Cl (pH 8.5). Run the libraries on NextSeq with 500/550 High Output Kit v2.5, 300 cycles, or 500/550 Mid Output Kit v2.5, and 300 cycles with 50% PhiX Control following the manufacturer's instruction. The cluster depth of each library should be more than 1.5 million.

### **3.5 Detection of Novel L1Hs Insertion**

1. Perform QC of input FASTAQ files with FastQC v0.11.8 and TrimGalore v0.6.2 (*see Note 12*).
2. Select the reads containing 3' end of L1Hs (5'-TGCACATG TACCCTAAACTTA GAGTATAAT-3'; diagnostic G for L1Hs was underlined), by applying agrep to read 2 sequences, allowing five nucleotide mismatches (*see Note 13*).
3. Map the selected reads to the reference human genome with BWA-MEM and make a merged BAM file (*see Note 14*).
4. Calculate the cover rate of reference L1 using "Mapped as SE (read 2)" information (*see Note 15*).
5. Select the reads that contain 10 bp or more of soft-clipped sequences in either read 1 or read 2 from the merged BAM file.
6. Use the bedtools merge function with 50-bp window for regionalization of the reads.
7. Calculate the depth of each L1 insertion site by counting the number of merged reads.
8. Remove the L1 insertion sites that overlapped with reference L1 (*see Note 16*).
9. Remove the L1 insertion sites that overlapped with known non-reference L1 (*see Note 17*).
10. Identify the successful novel L1 insertion sites by introducing the following criteria: (1) In read 1, a total of 10 or more depth, and at least one read containing 10 or more T in length within the soft-clipped region, and (2) in read 2, a total of 10 or more depth, and an average conservation rate of 3' end of L1Hs sequence (5'-AGTATAATAA-3') above 80%.

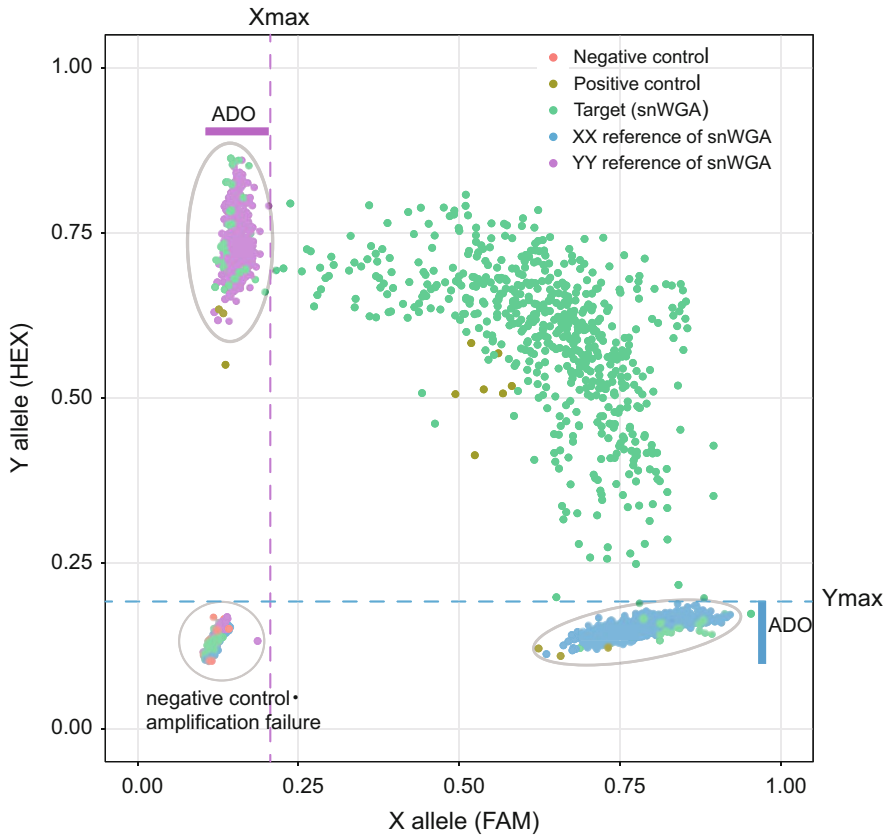
## 4 Notes

1. Refer to the document of “Using C1 to Generate Single-Cell Libraries for DNA Sequencing” (Fluidigm, P/N 100-7135 J1).
2. We recommend that the WGA reaction is performed on the same day of nuclei preparation. To avoid contamination, we recommend handling the IFC on a bench that is as clean as possible. We use KOACH T 500 (Koken) for this purpose.
3. Application note “Cell Sorting Directly to the C1 IFC” (P/N 101-5826 A2) and reagent calculation sheet “Reagent Calculations\_Sorting Into C1 IFC” (P/N 101-5840) offered from Fluidigm are also helpful.
4. Do not use “DNA Seq: Cell Load & Stain (1791x/1792/1793)” script.
5. This observation by microscopy should not take more than 1 h. Some chambers may contain two or more nuclei, while some chambers have no nuclei (Fig. 3). Use WGA products from the chambers containing one nucleus. If the nuclei loading is done well enough, then approximately 50–60% of the chambers (48–58 of 96 chambers) would contain 1 nucleus.
6. We recommend following the chapter of “(Optional) Prepare the Tube Controls” in the manufacturer’s document of “Using C1 to Generate Single-Cell Libraries for DNA Sequencing.”
7. If the snWGA is done successfully, the concentration and total amount of snWGA products should be approximately 10 ng/ $\mu$ L and approximately 120 ng, respectively.

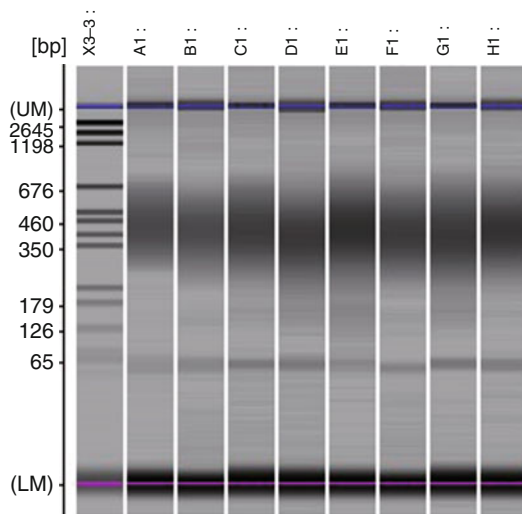


**Fig. 3** Examples of the chambers of IFC of C1. Chambers containing one (a) or two (b) cell nuclei. (c) Chamber containing no cell nuclei. snWGA products from the chamber containing only one cell nuclei (a) should be used following analysis

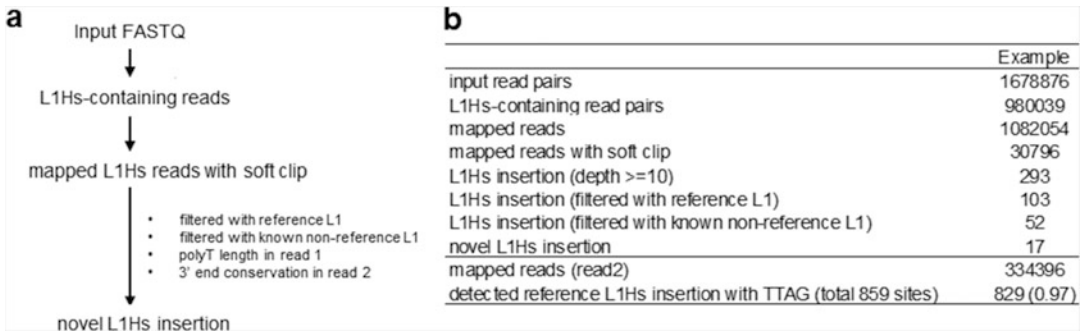
8. We recommend following the manufacturer's instruction of "SNPtrace Panel User Guide" (P/N 100-7282 B1). As the concentration of snWGA product (approximately 10 ng/ $\mu$ L) is not high enough to perform SNP typing directory, specific target amplification (STA) step is needed before genotyping. We used high-throughput SNPtrace Panel Genotyping Assays, which allows genotyping of 96 SNPs simultaneously. However, different platforms will be also available. Note that checking the quality of snWGA product is a critical step to remove the product with severe allele dropout or amplification failure [5].
9. Heterozygous SNPs are collectively regarded as informative SNPs. We recommend the use of multiple heterozygous SNPs for this purpose. In each genotyping analysis, inclusion of positive controls (non-amplified genomic DNA with XX homozygous, with YY homozygous, and that with X/Y heterozygous) and negative controls (ddW) is desirable. Signal intensities of snWGA product often deviated from those obtained in non-amplified genomic DNA. Therefore, inclusion of snWGA with XX homozygous and that with YY homozygous is highly recommended. We determined the threshold of ADO using the aggregate information of intensities of snWGAs references by calculating the intensity of maximum value of HEX of XX homozygous ( $Y_{max}$ ) and that of FAM of YY homozygous ( $X_{max}$ ) (Fig. 4). In each snWGA product, ADO rate is determined by (number of ADO SNPs)/(number of informative SNPs). We recommend the use of snWGA product with ADO <60%.
10. This preparation should be done before the NECO-seq experiment.
11. We perform this step according to the chapter of "XI. Quality control of library DNA using the MultiNA" in the document of "RamDA-seq protocol Release Version 1.1" (<https://bit.riken.jp/protocols/ramda-seq/>). Example results are shown in Fig. 5.
12. Source code of bioinformatic pipeline under the Linux environment is available upon request (Bundo et al. [6], submitted). The workflow and example result are shown in Fig. 6.
13. The quality control (QC) of the selected reads by agrep should be mapped to 3' end of LIHs using BWA-MEM. The reads mapped with quality scores of MAPQ1 or higher and those without InDel or hard clip should be selected. Reads showing completely matched sequence to 5'-TAAACTTAG-3', allowing less than 10 base soft clip, and their counterparts (read 1) should be selected.



**Fig. 4** Example of a plot of SNP genotyping (SNP17 of SNP Trace Panel, rs1876482). The green point shows the data from each snWGA product. Since the donor of the snWGA products has heterogeneous genotype in this SNP, the snWGA product with no ADO should show the high intensity of both alleles (X and Y). The snWGA products showing the intensity under Xmax or Ymax are considered to have ADO in this genomic location



**Fig. 5** Example image of NECO-seq libraries analyzed by MultiNA, a microchip electrophoresis system



**Fig. 6** (a) Workflow of bioinformatics analysis of NECO-seq data. (b) Example of the result of bioinformatic analysis. “Novel L1Hs insertion” is the number of somatic insertions of L1Hs that are not in the reference human genome (hg38) or existing L1Hs databases

14. We use the resources\_broad\_hg38\_v0\_Homo\_sapiens\_assembly38.fasta without decoy or ALT. Three categories of mapped reads will be generated: “mapped as PE,” “mapped as SE (read 1),” and “mapped as SE (read 2).”
15. Cover rate is determined by (number of detected reference L1)/(number of reference L1). In our analysis with snWGA product, the cover rate of reference L1 is typically above 0.95.
16. We use the chromosomal locations of L1s from RepeatMasker, and those of FL1-L1s, ORF2-L1s, and FLn1-L1s from L1Base2 [7].
17. We use the chromosomal locations of L1s from euL1db [8] and dbRIP [9]. In addition, from the 1000 genome project [10], chromosomal locations of L1s are provided in phase 3, and by those detected by MELT [11].

## Acknowledgments

We thank Junko Ueda and Yutaka Nakachi for their help in developing the bioinformatics tool, Emi Kiyota for her technical assistance.

## References

1. Cordaux R, Batzer MA (2009) The impact of retrotransposons on human genome evolution. *Nat Rev Genet* 10(10):691–703. <https://doi.org/10.1038/nrg2640>
2. Ewing AD, Kazazian HH Jr (2010) High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res* 20(9):1262–1270. <https://doi.org/10.1101/gr.106419.110>
3. Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, Morell M, O’Shea KS, Moran JV, Gage FH (2009) L1 retrotransposition in human neural progenitor cells. *Nature* 460(7259):1127–1131. <https://doi.org/10.1038/nature08248>
4. Bundo M, Kato T, Iwamoto K (2016) Cell type-specific DNA methylation analysis in neurons and glia. In: Karpova N (ed) *Epigenetic methods in neuroscience research*,

- neuromethods, vol 105. Humana Press, New York, NY
5. Iwamoto K, Bundo M, Ueda J, Nakano Y, Ukai W, Hashimoto E, Saito T, Kato T (2007) Detection of chromosomal structural alterations in single cells by SNP arrays: a systematic survey of amplification bias and optimized workflow. *PLoS One* 2(12):e1306. <https://doi.org/10.1371/journal.pone.0001306>
  6. Bundo M, Ueda J, Kiyota E, Nakachi Y, Kato T, Iwamoto K NECO-seq: a method for detection of somatic LINE-1 insertions at the single cell level from postmortem human brain tissue. in preparation.
  7. Penzkofer T, Jager M, Figlerowicz M, Badge R, Mundlos S, Robinson PN, Zemojtel T (2017) L1Base 2: more retrotransposition-active LINE-1s, more mammalian genomes. *Nucleic Acids Res* 45(D1):D68–D73. <https://doi.org/10.1093/nar/gkw925>
  8. Mir AA, Philippe C, Cristofari G (2015) euL1db: the European database of L1HS retrotransposon insertions in humans. *Nucleic Acids Res* 43(Database issue):D43–D47. <https://doi.org/10.1093/nar/gku1043>
  9. Wang J, Song L, Grover D, Azrak S, Batzer MA, Liang P (2006) dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum Mutat* 27(4):323–329. <https://doi.org/10.1002/humu.20307>
  10. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR (2015) A global reference for human genetic variation. *Nature* 526(7571):68–74. <https://doi.org/10.1038/nature15393>
  11. Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, Mills RE, Genomes Project C, Devine SE (2017) The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. *Genome Res* 27(11):1916–1929. <https://doi.org/10.1101/gr.218032.116>





## Solubilization of Mouse Sperm Chromatin for Sequencing Analyses Using a Chaperon Protein

Yuko Fukuda, Keishi Shintomi, Kosuke Yamaguchi, Yasuhiro Fujiwara, and Yuki Okada

### Abstract

Sperm chromatin compaction is physiologically essential for sperm to acquire the fertility. However, this unique structure composed of protamines makes us unable to solubilize the chromatin due to its resistance to sonication and enzymes usually used for chromatin fragmentation in somatic cells. Even when intense enzymatic treatment is applied, it appears to solubilize only certain portions of sperm chromatin presumably because of the heterogeneous properties. To overcome this issue, we previously developed a method to treat the sperm with recombinant nucleoplasmin, a protamine remover in fertilized embryos, followed by sonication. The nucleoplasmin treatment dramatically increased the efficiency of sperm chromatin solubilization, while a relatively large amount of recombinant nucleoplasmin was required. Here, we describe an improvement of nucleoplasmin method with a less amount of recombinant protein and a shorter reaction time.

**Key words** Mouse, Sperm, Nucleoplasmin, Nuclear decondensation, Chromatin, Histone, Solubilization

---

### 1 Introduction

Sperm chromatin is highly condensed due to the massive replacement of histones to protamines, testis-specific small basic proteins. Protamination-induced sperm chromatin condensation is essential for sperm to acquire proper motility and maintain DNA integrity. Simultaneously, the retained histones that escaped from the eviction have also been studied to identify their subnuclear and genomic localization and their biological role. Particularly in the past decade, sperm epigenome has gained more attention in the context of transgenerational epigenetic inheritance [1–4].

However, highly condensed sperm chromatin prevents its chromatin from being solubilized by biochemical methods usually applied to somatic cells. Although there is an interspecies difference

in sperm chromatin compaction, mouse sperm is the most extreme case in terms of the resistance to sonication [5] and DTT/SDS treatment [6], the latter of which was aimed to reduce disulfide bonding of inter-protamines [7]. Therefore, a relatively higher concentration of DTT (10–50 mM) and micrococcal nuclease (MNase) (2.25–7.5 units/million sperm) treatments have been applied to the genome-wide chromatin solubilization followed by next-generation sequencing analyses [8–10]. However, these treatments were demonstrated to possibly solubilize only a portion of the sperm chromatin, depending on its MNase concentration [9].

In nature, protamines are removed from paternal chromatin immediately in the ooplasm after fertilization by nucleoplasmin and its amphibian ortholog (hereafter referred to as Npm2). *Xenopus laevis* egg cell-free extract that abundantly contains Npm2 has been used extensively to study various aspects of cellular and developmental biology, including mitotic chromatid formation from *Xenopus* sperm nuclei [11, 12]. In 2015, Shintomi et al. successfully reconstituted mitotic chromatids by only six purified factors (core histones, Npm2, Nap1, FACT, topo II, and condensin I) that are originally included in the *Xenopus* egg extract [13, 14].

The protamine eviction activity of Npm2 is conserved in both *Xenopus* and mammals, and it results in efficient extraction of histones and increased sensitivity to MNase digestion [15]. Based on these observations, we utilized the Npm2-dependent protamine removal to achieve efficient solubilization of mouse sperm chromatin without MNase digestion [16]. This method enabled us to achieve genome-wide solubilization of chromatin with minimal structural barrier by nucleoprotamines. The obtained chromatin was applied to ChIP-seq analyses to identify the localization of sperm-retained histones [16]. More recently, the Npm2 method was used to assay the protamine's phosphorylation-dependent sperm chromatin decondensation [17], suggesting that more applications that accelerate sperm chromatin research potentially exist.

However, the method we published in 2018 left some scope for improvement, including a relatively high concentration of recombinant Npm2 (~500  $\mu$ M) and a long reaction time (~2 h) [16]. Here, we introduce the improved protocol with a lower concentration of recombinant Npm2 (~100  $\mu$ M) and a shorter reaction time (~15 min).

---

## 2 Materials

### 2.1 Preparation of Recombinant Nucleoplasmin

1. Expression vector for Npm2 (referred to as pET28-Npm2). A cDNA fragment encoding the C-terminally truncated version of *Xenopus laevis* Npm2 (amino acids 1-149) is cloned into the pET28a vector (Merck Millipore). The resulting vector harbors

the abovementioned cDNA fragment flanked by a hexahistidine (6xHis) tag and a thrombin cleavage site at its 5'-end [13] (*see Note 1*).

2. Competent cells of a standard  $\lambda$ DE3 lysogen *E. coli* strain for protein expression. We usually use the BL21(DE3) strain.
3. Media and antibiotics for growing the *E. coli* culture: LB liquid medium and LB agar plates supplemented with kanamycin at a working concentration of 50  $\mu$ g/mL.
4. 1.0-M IPTG (isopropyl- $\beta$ -D-thiogalactopyranoside): Dissolve powder in H<sub>2</sub>O. Make aliquots and store at  $-20^{\circ}$ C.
5. Thrombin protease.
6. Dialysis tube, molecular weight cutoff (MWCO): 12–14 KD.
7. His-A buffer: 50-mM NaH<sub>2</sub>PO<sub>4</sub>, 500-mM NaCl, pH 7.4.
8. His-B buffer: 50-mM NaH<sub>2</sub>PO<sub>4</sub>, 500-mM NaCl, 500-mM imidazole, pH 7.4.
9. His-L buffer: Mix His-A buffer and His-B buffer at a ratio of 49:1.
10. His-W buffer: Mix His-A buffer and His-B buffer at a ratio of 23:1.
11. Q-100 buffer: 20-mM Tris-HCl, 100-mM NaCl, pH 8.0.
12. Q-B buffer: 20-mM Tris-HCl, 1-M NaCl, pH 8.0.
13. KH buffer: 100-mM KCl, 20-mM HEPES-KOH, pH 7.7.
14. cOmplete Protease Inhibitor Cocktail, EDTA-free (Roche, 11873580001).
15. Ultrasonic disintegrator: Branson Sonifier 450D and a micro-tip (1/4") or equivalent.
16. Ultracentrifuge: Beckman LE-80K and a 70Ti rotor or equivalent.
17. HisTrap HP column, 5 mL (Cytiva).
18. HiTrap QXL column, 2  $\times$  1 mL (Cytiva).
19. Chromatography system: AKTA start (Cytiva) or equivalent.

## **2.2 Decondensation of Mouse Sperm Using Nucleoplasmin**

1. Medium for sperm collection and swim-out: HTF containing 3-mg/mL BSA or equivalent (*see Note 2*).
2. SLO (streptolysin O): The unit concentration is described on the product.
3. 1.0-M DTT (dithiothreitol): Dissolve the powder in acid buffer (*see Note 3*).
4. cOmplete Protease Inhibitor Cocktail, EDTA-free (Roche, 11873580001): Dissolve a tablet in PBS to make 100 $\times$  stock solution.
5. Hoechst 33342 or equivalent DNA staining dye.

6. Cell counting device (Hemocytometer, Makler™ counting chamber, etc.).
7. Ultrasonic disintegrator: Branson Sonifier 450D and a micro-tip (1/8") or equivalent.
8. Intelli-Mixer RM-2M (ELMI Tech) or equivalent placed at a 37 °C incubator.
9. Fluorescent microscope with a 40–60× objective lens to observe the DNA staining.

### **2.3 Chromatin Fragmentation of Nucleoplasmin-Treated Sperm**

1. 1% formaldehyde solution in PBS.
2. NP-40 buffer: 10-mM Tris-HCl, pH 8.0, 10-mM NaCl, 0.5% NP-40.
3. Sonication buffer: 20-mM Tris-HCl, pH 8.0, 20% glycerol, 0.5% SDS, 1× cComplete. SDS and cComplete are added immediately before use.
4. Picoruptor Sonicator (Diagenode, B01060010) and 0.65-mL microtubes for Picoruptor (Diagenode, C30010011) (*see Note 4*).

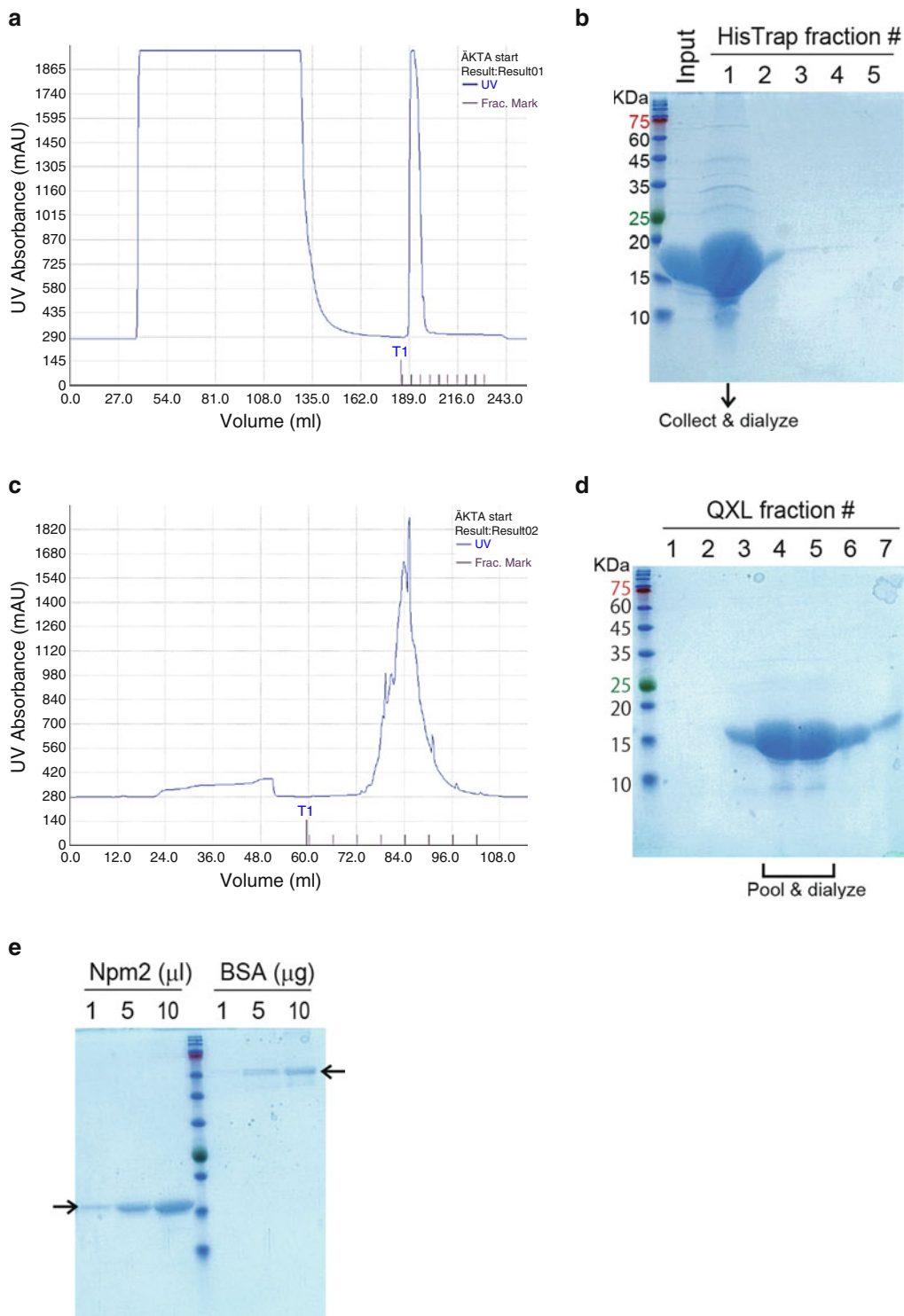
---

## **3 Methods**

### **3.1 Preparation of Recombinant Nucleoplasmin**

1. Transform the BL21 (DE3) strain of *E. coli* with the expression vector pET28-Npm2 by a standard heat-shock protocol, plate cells on an LB agar plate containing kanamycin, and allow transformants to form colonies overnight at 37 °C.
2. Inoculate a single colony of transformants into 10-m LB liquid medium containing 50-µg/mL kanamycin, and incubate the culture for 4 h at 37 °C with vigorous shaking. After incubation, scale up the culture to 1–3 L (dependently of the size of available flasks), and allow bacteria to grow under the same condition.
3. When the cell optical density (i.e., absorbance at 600 nm) reaches 0.8 (0.5–1.0 is permissive), induce expression of the recombinant a C-terminally truncated version of Npm2 (hereafter, simply referred to as Npm2) by adding IPTG at the final concentration of 0.2 mM, and continue incubation at 37 °C for another 1 h.
4. Transfer the culture to a temperature of 20 °C and continue incubation for another 12–16 h.
5. Harvest the cells by centrifugation at 4000 g for 10 min at 4 °C. After removing medium, measure the weight of cell pellet. After this step, samples and buffers are kept on ice or at 4 °C. Procedures need to be performed on ice or in a cold room unless otherwise specified. At this point, the pellets can be frozen in liquid nitrogen and kept at –80 °C until use.

6. Suspend the cell pellet in His-L buffer containing EDTA-free cOmplete protease inhibitor cocktail at a ratio of 10-mL buffer per 1-g cells; divide the suspension to 15–20 mL in each 50-mL tube pre-cooled on ice.
7. Break cells by sonication through the following program of the ultrasonic disintegrator: 15% [on 10 s, off 10 s] for 5 min (total of time is 2 min and 30 s).
8. Add lysozyme at the final concentration of 50  $\mu\text{g}/\text{mL}$  and incubate at room temperature for 5 min.
9. Repeat sonication as described in subheading 3.1, step 7; 15% [on 10 s, off 10 s] for 5 min (total of time is 2 min and 30 s).
10. Spin down lysate at 30,000 g, at 4 °C for 30 min to remove insoluble materials and collect the supernatant.
11. Load the supernatant to a HisTrap HP column that has been installed on a chromatography system and equilibrated in His-L buffer in advance that has been charged with  $\text{NiSO}_4$ .
12. After washing the column with His-W, elute the hexa-histidine-tagged Npm2 with His-B and collect eluates into tubes (Fig. 1a). The setting of a chromatography system is summarized in Table 1.
13. Analyze the eluate by SDS-PAGE (using 15% gel) followed by Coomassie Brilliant Blue (CBB) staining, and then pool the peak fractions (Fig. 1b).
14. Dialyze the pooled fraction against 1 liter of Q-100 buffer for 2 h overnight with one buffer change.
15. Measure the protein concentration by measuring  $A_{280}$  and add thrombin protease at 1/8 unit per mg protein (*see Note 5*).
16. Incubate at room temperature for 6 h to completely liberate the hexa-histidine tag. After incubation at room temperature, the sample can be kept on ice overnight until the next step.
17. Further purify the thrombin-treated protein mixture by loading it to a HiTrap QXL column. Elute the Npm2 protein by developing a linear gradient of Q-B buffer (i.e., a gradient of NaCl from 100 mM to 1000 mM) (Fig. 1c). The setting of a chromatography system is summarized in Table 1.
18. Analyze the eluate by SDS-PAGE followed by CBB staining, and then pool the peak fractions (Fig. 1d).
19. Dialyze them against 1 L of KH buffer for 2 h overnight with one buffer change.
20. Analyze the dialysate by SDS-PAGE to confirm the homogeneity of Npm2, and roughly estimate its concentration by comparing band intensities of its dilution series to those of known amounts of a standard protein (e.g., BSA) on the same gel (Fig. 1e). To know more accurate concentration, measure the absorbance of  $A_{280}$ .



**Fig. 1** Column chromatographic purification of Npm2. **(a)** Representative chromatographic pattern of 6xHis-Npm2 purification using HisTrap HP column. **(b)** CBB staining of collected fractions from a HisTrap HP column. **(c)** Representative chromatographic pattern of Npm2 purification using a HiTrap QXL column. **(d)** Image of CBB staining of collected fractions from a HiTrap QXL column. **(e)** Semi-quantification of Npm2 by SDS-PAGE followed by CBB staining

**Table 1**  
**Summary of AKTA setting (for an 800-mL bacterial culture)**

	HiTrap HP, 5 mL	HiTrap QXL column, 2 × 1 mL
Applied sample volume (mL)	100–120	30–35 (diluted with Q-100)
Equilibration (cv)	5	5
Wash (cv)	12	5
Elution mode	Isocratic	Gradient (volume: 20 cv)
Volume of each fraction (mL)	5.0	6.0

cv column volume

- Once the concentration is determined, concentrate the protein solution at 500  $\mu$ M by using Amicon Ultra centrifugal filter unit (Millipore) with the appropriate molecular weight cutoff (MWCO) (*see* **Notes 6** and **7**).

### 3.2 Isolation of Mouse Sperm

Handling of mice should be done in accordance with the regulations of the animal care and use committee of your institution.

- Collect cauda epididymis from adult (>8 weeks old) male mice.
- Make two or three small incisions in the tissue, and place them in a 1.5-mL Eppendorf tube containing 1 mL of pre-warmed appropriate medium. 1-mL medium can hold up to four pieces of epididymis.
- Incubate the tube at 37 °C for 20–30 min to allow the sperm to swim out of the tissue (*see* **Note 8**).
- Collect the medium containing sperm and wash once with 1-mL PBS.
- Count the number of sperm using cell counting device.

### 3.3 SLO Treatment of Mouse Sperm

- Spin down (6000  $g \times 5$  min) and suspend the sperm in 0.5-mL PBS at  $4\text{--}10 \times 10^7$ /mL.
- Cut the sperm tail by sonication; amplitude 10%, 10 s  $\times$  3 times. Keep the tube on ice during sonication. If necessary, check the sample under the microscope to make sure that the tails are detached.
- Spin down (6000  $g \times 5$  min) and wash the sperm with PBS once.
- Suspend the sperm in 0.5-mL PBS and add 2 units of SLO. Incubate at 37 °C for 1 h.
- Spin down (6000  $g \times 5$  min) and wash the sperm with PBS once.

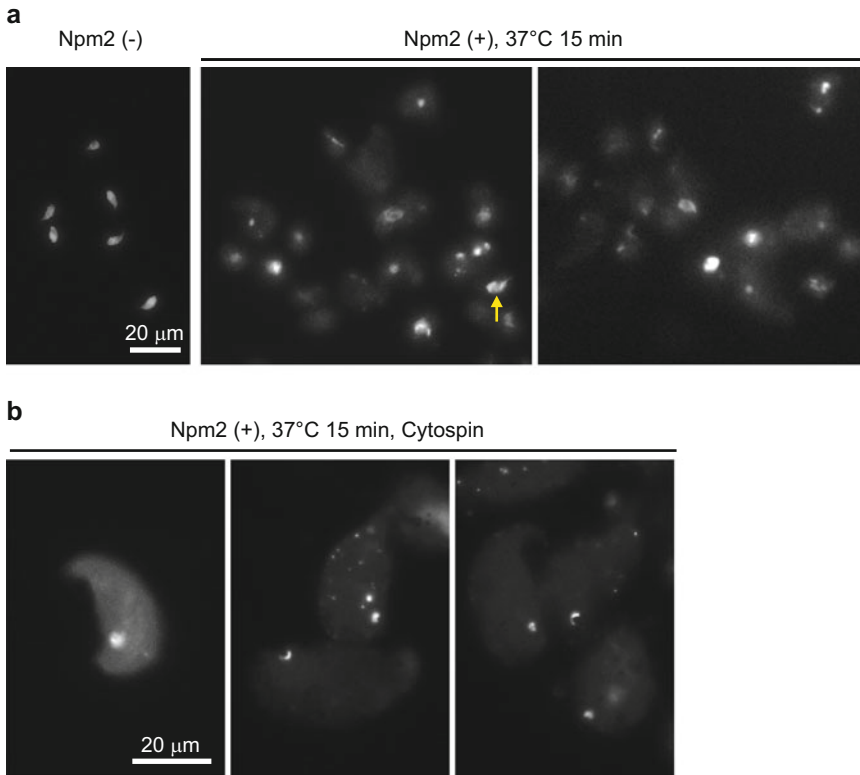
6. Suspend the pellet in 0.5-mL PBS and add 1  $\mu\text{L}$  of DTT (the final concentration is 2 mM). Incubate at 37 °C for 15 min with gentle mixing. For Intelli-Mixer RM-2M, use Mode 8F and 30 rpm.
7. Spin down (6000 g  $\times$  5 min) and wash the sperm with PBS once.
8. Divide the sperm into  $1 \times 10^7$  in a 1.5-mL Eppendorf tube and spin down.
9. Remove the supernatant. Freeze/thaw the pellet twice using liquid nitrogen and proceed to NPM treatment. If you do not proceed to the NPM treatment immediately, the pellet can be stored at  $-80$  °C. In this case, stop this step at the second freezing.

### 3.4 Nucleoplasmin Treatment of Mouse Sperm

1. Suspend the sperm pellet in 200- $\mu\text{L}$  KH buffer. Mix the suspension by pipetting until the pellet is evenly dissolved (*see Note 9*).
2. Add another 200- $\mu\text{L}$  KH buffer containing 1.25  $\mu\text{L}$  of 1-M  $\text{MgCl}_2$ , 1  $\mu\text{L}$  of 1-M DTT, 25  $\mu\text{L}$  of 3-M KCl, and 5  $\mu\text{L}$  of 100 $\times$  cOmplete Cocktail. After adding Npm2, their final concentration becomes 2.5-mM  $\text{MgCl}_2$ , 2-mM DTT, 250-mM KCl, and 1 $\times$  cOmplete, respectively. Add 100  $\mu\text{L}$  of 500- $\mu\text{M}$  Npm2 stock solution (final 100  $\mu\text{M}$ ) (*see Note 10*). The final reaction volume becomes 500  $\mu\text{L}$ .
3. Incubate the tube at 37 °C for 15 min with gentle mixing. For Intelli-Mixer RM-2M, use Mode 8F and 30 rpm.
4. Stop the reaction by placing the tube on ice soaked in water immediately.
5. Take 5  $\mu\text{L}$  of the sample and mix with 5- $\mu\text{L}$  PBS containing 10- $\mu\text{g}/\text{mL}$  Hoechst 33342.
6. After ~5-min incubation, thinly mount the sample on a glass slide, and check the morphology of the nuclei with a fluorescence microscope (Fig. 2a, b). If the nuclear swelling is insufficient, put the tube back to 37 °C, and extend the reaction in 3–5 min increments (*see Note 11*).
7. Centrifuge the tube with 9000 g  $\times$  5 min at 4 °C to remove Npm2. When the nuclear swelling is sufficient, protamines are released in supernatant, while histones are retained in pellet (= chromatin fraction) (Fig. 3a, b).

If your sample is to be subjected to chromatin immunoprecipitation sequencing (ChIP-seq), proceed the following step of chromatin fragmentation. Otherwise, process or store the samples appropriately depending on your intended experiment. If you do not fix the sample at this step, additional wash may be required to completely remove residual Npm2.

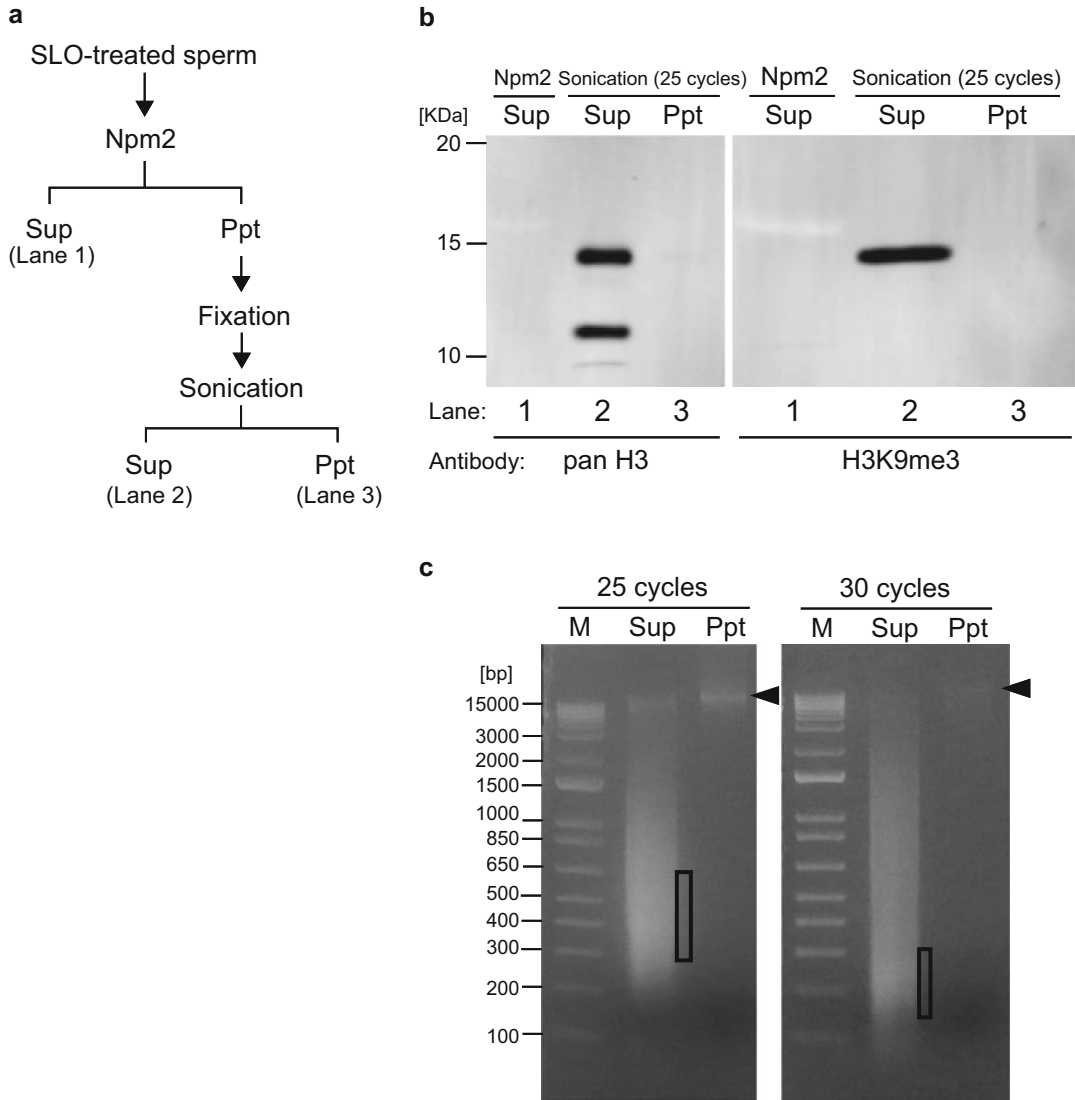




**Fig. 2** Representative nuclear morphology of Npm2-treated mouse sperm by Hoechst 33342 staining. (a) Npm2-untreated [Npm2 (–)] and Npm2-treated [Npm2 (+)] sperm thinly mounted on a glass slide. A yellow arrow indicates a sperm without swelling. (b) Sperm attached to a glass slide by cytospin after Npm2 treatment followed by 4% paraformaldehyde fixation. Note that the exposure time was adjusted for each imaging and was not standardized

### 3.5 Chromatin Fragmentation of Nucleoplasmin-Treated Sperm

1. Suspend the pellet in 1% formaldehyde (*see Note 12*). Rotate the tube for 10 min at room temperature.
2. Add 55  $\mu\text{L}$  of 2.5-M glycine in PBS to quench the reaction. Rotate the tube for 5 min at room temperature.
3. Centrifuge the sample with  $9000\text{ g} \times 5\text{ min}$  at  $4\text{ }^\circ\text{C}$  and suspend the pellet in 200  $\mu\text{L}$  of NP-40 buffer. Place the tube on ice for 1 h.
4. Centrifuge the sample with  $9000\text{ g} \times 5\text{ min}$  at  $4\text{ }^\circ\text{C}$  and suspend the pellet in 200  $\mu\text{L}$  of sonication buffer containing 0.5% SDS.
5. Divide the sample into two 0.65-mL microtubes for Picoruptor, 100  $\mu\text{L}$  each.
6. Set the samples in Picoruptor and sonicate [on 10 s, off 10 s]  $\times$   $\sim 25$  cycles (*see Note 13*).



**Fig. 3** Fragmentation and solubilization of histone-containing sperm chromatin by sonication. **(a)** Fractionation scheme of Npm2 treatment followed by sonication to release histone-containing sperm chromatin. The notations “Lane 1, 2, and 3” correspond to the Western blot shown in **(b)**. **(b)** Western blot analysis of the fractionated samples indicated in **(a)**. Anti-pan H3 (abcam, ab1791) and anti-H3K9me3 (abcam, ab8898) were used. Note that histones are efficiently released into supernatant by sonication. **(c)** DNA electrophoresis of sonicated samples using 1% TAE agarose gel. In the supernatant fractions, a major peak of 300–600 bp was observed by 25-cycle sonication and ~200 bp by 30-cycle sonication. A high-molecular-weight band (>15,000 bp, arrowheads) is constantly observed in the pellet fractions

7. Centrifuge the sample with 9000 g × 5 min at 4 °C. Histone-containing chromatin is sheared and released into supernatant (Fig. 3c). Collect supernatant for subsequent procedures such as DNA electrophoresis, Western blotting, and immunoprecipitation (*see* Note 14).

---

## 4 Notes

1. This vector is available to researchers of interest upon request to K.S. (kshintomi@riken.jp).
2. Any medium that is commonly used for *in vitro* fertilization or embryo culture can be used such as HTF, M2, TYH, etc. Because of its cost advantage, we use HTF containing 3-mg/mL BSA.
3. Activity of DTT is critical for sperm chromatin decondensation. To stabilize the activity, we usually dissolve the powder in 0.3-M Sodium acetate (pH 5.2) that is used for ethanol precipitation. Make small volume of aliquots and store at  $-20\text{ }^{\circ}\text{C}$ . Once thawed, avoid freeze/thaw. The thawed tube can be kept at  $4\text{ }^{\circ}\text{C}$  for up to 1 week. If you do not use acid buffer for solvent, freshly prepared solution is highly recommended.
4. The use of Picoruptor is strongly recommended for sonication due to its powerfulness and reproducibility. Currently, Picoruptor is no longer available and Picoruptor 2 is sold instead. If you use Picoruptor 2, you may need to adjust the output and duration.
5. For calculating protein concentration, the extinction coefficient of Npm2 (amino acids 1–149) is  $13,980\text{ M}^{-1}\text{ cm}^{-1}$ .
6. In Subheading 3.4, **step 2**, Npm2 is used at  $100\text{ }\mu\text{M}$  in a  $500\text{-}\mu\text{L}$  reaction. Since sperm pellet should be suspended in  $200\text{ }\mu\text{L}$  of buffer beforehand, the maximum volume of Npm2 that can be added is  $300\text{ }\mu\text{L}$ . Therefore, a minimum concentration of  $167\text{ }\mu\text{M}$  is required in this step.
7. Because we have confirmed that this truncated version of Npm2 forms pentamer or decamer in KH buffer by size exclusion chromatography, the MWCO-50K device can work for efficient and rapid concentration.
8. If your sperm samples possess low motility and have a problem to swim out of the tissue, consider pushing sperm out of the incised tissue with tweezers or using intratesticular sperm. In such cases, the susceptibility to Npm2 may be different, so the subsequent Npm2 treatment needs to be adjusted according to the sample properties.
9. In this step, it is important to suspend the sperm evenly. If the pellet is suspended in KH buffer containing  $\text{MgCl}_2$  or DTT from the beginning, small aggregations may occur.
10. If the concentration of your Npm2 stock solution is  $<500\text{ }\mu\text{M}$ , adjust the volume of KH buffer.
11. Heterogeneity in the size of sperm nuclei, including sperm that are not swollen at all, may be observed. In such cases, extending the incubation time will reduce the heterogeneity. On the

other hand, sperm that are already fully swollen may be disrupted by extending the incubation time, and histones may be released from chromatin.

12. To prevent excessive reaction due to residual Npm2, keep the temperature of sample low and work quickly until formaldehyde is added.
13. The number of cycles of sonication is affected not only by the degree of sperm swelling but also by the length of fixation time and temperature. Some adjustments may be needed according to your experimental design.
14. Before DNA electrophoresis, de-crosslinking to release the DNA from chromatin is required. In Fig. 3c 1% TAE agarose gel is used. For more quantitative information of the size and amount of DNA, the use of capillary electrophoresis system such as Agilent 2100 Bioanalyzer is recommended.

---

## Acknowledgments

We would like to thank Dr. Tatsuya Hirano (RIKEN) and Dr. Satoshi Kaneko (Ichikawa General Hospital, Tokyo Dental College) for sharing their knowledge. This work is supported by a grant from MEXT Grant-in-Aid (#19H05254 and #20H05939 to Y.O., #19H05755 to K.S.) and JST ERATO Chromatin Atlas (#JPMJER1901 to Y.O.).

## References

1. Bodden C, Hannan AJ, Reichelt AC (2020) Diet-induced modification of the sperm epigenome programs metabolism and behavior. *Trends Endocrinol Metab* 31:131–149
2. Le Blévec E, Muroňová J, Ray PF et al (2020) Paternal epigenetics: mammalian sperm provide much more than DNA at fertilization. *Mol Cell Endocrinol* 518:110964
3. Torres-Flores U, Hernández-Hernández A (2020) The interplay between replacement and retention of histones in the sperm genome. *Front Genet* 11:780
4. Galan C, Krykbaeva M, Rando OJ (2020) Early life lessons: the lasting effects of germline epigenetic information on organismal development. *Mol Metab* 38:100924
5. Balhorn R, Weston S, Thomas C et al (1984) DNA packaging in mouse spermatids. Synthesis of protamine variants and four transition proteins. *Exp Cell Res* 150:298–308
6. Perreault SD, Barbee RR, Elstein KH et al (1988) Interspecies differences in the stability of mammalian sperm nuclei assessed in vivo by sperm microinjection and in vitro by flow cytometry. *Biol Reprod* 39:157–167
7. Calvin HI, Bedford JM (1971) Formation of disulphide bonds in the nucleus and accessory structures of mammalian spermatozoa during maturation in the epididymis. *J Reprod Fertil Suppl* 13(Suppl 13):65–75
8. Hisano M, Erkek S, Dessus-Babus S et al (2013) Genome-wide chromatin analysis in mature mouse and human spermatozoa. *Nat Protoc* 8:2449–2470
9. Carone BR, Hung J-H, Hainer SJ et al (2014) High-resolution mapping of chromatin packaging in mouse embryonic stem cells and sperm. *Dev Cell* 30:11–22
10. Hada M, Masuda K, Yamaguchi K et al (2017) Identification of a variant-specific phosphorylation of TH2A during spermiogenesis. *Sci Rep* 7:46228
11. Lohka MJ, Masui Y (1983) Formation in vitro of sperm pronuclei and mitotic chromosomes

- induced by amphibian ooplasmic components. *Science* 220:719–721
12. Lohka MJ, Maller JL (1985) Induction of nuclear envelope breakdown, chromosome condensation, and spindle formation in cell-free extracts. *J Cell Biol* 101:518–523
  13. Shintomi K, Takahashi TS, Hirano T (2015) Reconstitution of mitotic chromatids with a minimum set of purified factors. *Nat Cell Biol* 17:1014–1023
  14. Shintomi K, Hirano T (2018) Reconstitution of mitotic chromatids in vitro. *Curr Protoc Cell Biol* 79:e48
  15. Itoh T, Ohsumi K, Katagiri C (1993) Remodeling of human sperm chromatin mediated by nucleoplasmin from amphibian eggs. *Develop Growth Differ* 35:59–66
  16. Yamaguchi K, Hada M, Fukuda Y et al (2018) Re-evaluating the localization of sperm-retained histones revealed the modification-dependent accumulation in specific genome regions. *Cell Rep* 23:3920–3932
  17. Gou L-T, Lim D-H, Ma W et al (2020) Initiation of parental genome reprogramming in fertilized oocyte by splicing kinase SRPK1-catalyzed protamine phosphorylation. *Cell* 180:1212–1227.e14

# Part III

## **Manipulation of the Epigenome: Platforms for Epigenome Editing**



## Efficient Targeted DNA Methylation with dCas9-Coupled DNMT3A-DNMT3L Methyltransferase

Pavel Bashtrykov, Nivethika Rajaram, and Albert Jeltsch

### Abstract

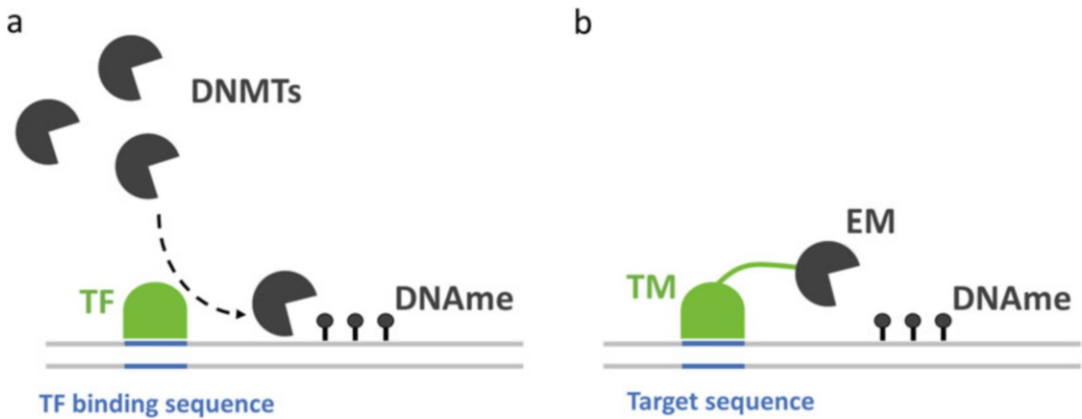
Epigenome editing is a powerful approach for the establishment of a chromatin environment with desired properties at a selected genomic locus, which is used to influence the transcription of target genes and to study properties and functions of gene regulatory elements. Targeted DNA methylation is one of the most often used types of epigenome editing, which typically aims for gene silencing by methylation of gene promoters. Here, we describe the design principles of EpiEditors for targeted DNA methylation and provide step-by-step guidelines for the realization of this approach. We focus on the dCas9 protein as the state-of-the-art DNA targeting module fused to 10×SunTag as the most frequently used system for editing enhancement. Further, we discuss different flavors of DNA methyltransferase modules used for this purpose including the most specific variants developed recently. Finally, we explain the principles of gRNA selection, outline the setup of the cell culture experiments, and briefly introduce the available options for the downstream DNA methylation data analysis.

**Key words** Epigenome editing, DNA methylation, CRISPR, dCas9, DNMT3A, SunTag

---

## 1 Introduction

The DNA encoding human genome information is stored in a cell nucleus in a form of chromatin, where the DNA is wrapped around histone proteins and organized into various defined structures by a multitude of accessory proteins and RNAs [1]. In addition to its storage function, chromatin is involved in the regulation of gene transcription, replication, and DNA repair. These fundamental cellular mechanisms are mediated by posttranslational modifications of histone proteins and DNA, which all together form the so-called epigenome of a cell [2]. One of the most well-studied epigenome signals is the methylation of DNA (DNAm) at the position 5 of the cytosine ring. The methylation reaction is catalyzed by a family of enzymes called DNA methyltransferases (DNMTs), which modify cytosines in mammals mainly in the context of CpG dinucleotides also known as CpG sites [3]. Deposition of DNAm at



**Fig. 1** Scheme of the natural recruitment of the DNMT enzyme to a target genomic locus by a TF leading to the deposition of DNA methylation (a) and the design of the corresponding EpiEditor as a single-chain protein assembled from a DNA targeting module (TM) and the editing module (EM) (b)

promoters often results in the silencing of associated genes, and the methylation of intergenic regions leads to the inactivation of repetitive elements [4]. The distribution of DNAmethylase throughout the genome is set during embryogenesis, and it is one of the key mechanisms regulating gene expression and the lineage-specific formation of cell types. The deposition of DNAmethylase is controlled by various processes, one of those is the recruitment of DNMTs by transcription factors (TFs) to selected genomic regions [5]. TFs bind to specific DNA sequence motifs present in the genome and interact with DNMTs, and the latter methylate DNA in close proximity, which results in targeted methylation (Fig. 1a).

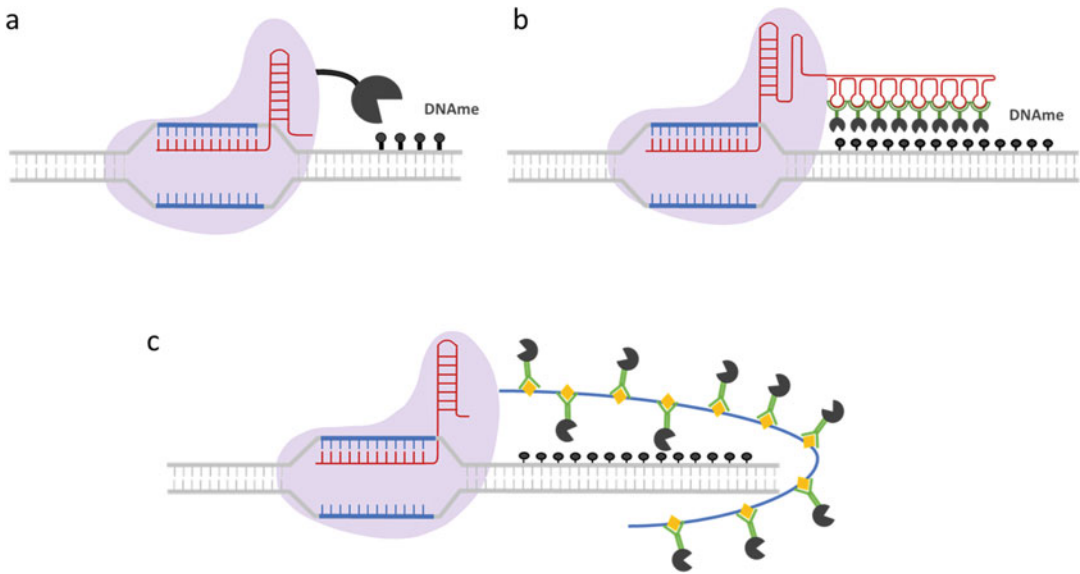
Since the mechanisms of DNA methylation are quite well understood, this knowledge inspired the development of the new field of synthetic biology termed epigenome editing, which aims to establish a desired epigenome state at a selected genomic locus [6]. This can be achieved by, for example, “writing” or “erasing” DNA methylation or in a broader sense any epigenome mark at the region of interest. Since DNA methylation is utilized in cells for long-term transcriptional repression, targeted DNA methylation can be applied for gene silencing or to study the function of genomic elements regulated by methylation. The key tools of epigenome editing are so-called EpiEditors [7], which contain two functional modules—a DNA targeting and an editing module (Fig. 1b). In its original design, the EpiEditor was a chimeric protein, which mimicked a TF complex with DNMT, where the DNA targeting module is responsible for addressing of the protein to the selected genomic region and the editing module is presented by the DNA methyltransferase enzyme [8–10]. During the time that has passed since the first applications of this method, multiple improvements have been developed, but the central design of the



bifunctional composition of EpiEditors has remained unchanged. It will be discussed in detail below.

To target an EpiEditor, its DNA-binding module should be able to recognize a specific DNA sequence present in the selected genomic region. For this purpose, TFs, zinc finger proteins (ZFPs) [11], transcription activator-like effectors (TALEs) [12], and clustered regularly interspaced short palindromic repeat (CRISPR)-associated proteins (Cas) [13] have been used. One of the key characteristics of these DNA-binding modules is specificity, and an ultimate aim is to design a DNA-binding module, which binds only the unique single target locus but nowhere else in the genome. The specificity of a DNA-binding module in part correlates with the length of the recognition sequence of the protein. TFs have usually short recognition sequences (up to ten bases), which are present in multiple copies in the large human genome; thus they cannot be used, if the targeting of a single region is required. ZFPs and TALEs are modular proteins, build by assembling of minimal units into arrays, which can lead to the generation of a relatively long recognition sequence. Genome-scale screening studies have shown that one can generate ZFPs and TALEs targeting a unique binding in the human genome [14, 15]. The most recent class of DNA-binding modules are Cas proteins, which function in a complex with RNA [13]. Interestingly, in this complex it is the RNA that reads a target DNA sequence according to the Watson–Crick rules by forming a DNA–RNA hybrid, and the Cas proteins only mediate this interaction. Natural Cas proteins are nucleases, which cleave the target DNA after binding, whereas their catalytically deactivated variants (dCas) retain only the DNA-binding property such that they can be used as targeting modules for epigenome editing [16]. The most frequently used variant is dCas9 derived from *Streptococcus pyogenes*, which recognizes a 20-nucleotide-long target DNA sequence by its bound RNA (also known as gRNA). This sequence must be followed by an NGG trinucleotide motif termed protospacer adjacent motif (PAM) sequence [17]. Thus, in total the dCas9/gRNA complex binds to 23 nucleotides at the target locus; long enough to choose a unique site in the human genome to allow single-site epigenome editing. The gRNA harbors a constant region folding into a structural scaffold, which is required for the interaction with dCas9, and a variable region, the so-called spacer, which defines the specificity to a particular genomic locus. Since the simple exchange of the spacer sequence generates a new targeting device with a different sequence specificity without tedious protein engineering, dCas proteins rapidly gained great popularity and became a default choice for designing EpiEditors.

Currently, there are three ways to connect an editing module to the dCas9/gRNA complex acting as targeting module. The first one is the generation of a single-chain protein by directly fusing



**Fig. 2** Schematic image of three types of connection of the editing module and the target module currently in use. **(a)** The direct fusion of the editing module to the dCas9 protein. **(b)** The Sirius RNA scaffold is introduced into the gRNA where it is recognized by specific proteins fused to the editing module. **(c)** dCas9 fused to the  $10\times$ SunTag provides an array of peptides which interact with a specific antibody fused to the editing module

dCas9 with the enzyme (Fig. 2a) [18]. This was historically the first implemented method. The second is the use of RNA scaffolds, such as MS2 or PP7 [19, 20]. The scaffolds are recognized by specific proteins, which can then be attached to the editing module. The scaffolds are integrated into the gRNA and may recruit up to 8 units of the binding proteins with fused enzyme, as in the case of CRISPR-Sirius (Fig. 2b) [21]. The third way is the attachment of a recruitment platform to the dCas9 protein. The two most known examples are the SunTag [22] and the concatenated coiled-coil peptide tag (CCC-tag) [23]. Both systems contain an array of peptides, each of which can bind one editing module tagged with the antibody or coiled-coil peptide, for the SunTag (Fig. 2c) and the CCC-tag, respectively. Several flavors of the SunTag have been developed, containing 5, 10, and 24 binding sites, whereas the maximum capacity of the CCC-tag is 10. These recruitment platforms increase the editing efficiency, because more than one editing enzyme is brought to the target region by one bound dCas9. Moreover, they provide more flexibility due to the long peptide tail compared to RNA scaffolds integrated into gRNA.

For DNA methylation purposes, various bacterial or mammalian DNA methyltransferases have been applied as the editing module. Human cells express three DNMTs, namely, DNMT1, DNMT3A, and DNMT3B, and the DNMT3-like protein (DNMT3L) [24]. The first one is mainly involved in the

maintenance of DNA methylation patterns, and the two other ones are the primary de novo DNA methyltransferases [25], responsible for the setting of this mark. DNMT3A possesses higher enzymatic activity than DNMT3B and therefore it is more interesting for the design of EpiEditors. DNMT3L is catalytically inactive, but it is an efficient allosteric regulator of DNMT3A and DNMT3B which stimulates their activity [26, 27]. Structural data showed that the catalytic domain of DNMT3A and the C-terminal part of DNMT3L form a linear 3L-3A-3A-3L heterotetramer, where the DNA molecule is bound by the two central DNMT3A subunits [28]. An early study showed that the catalytic domain of DNMT3A can be joined with the C-terminal part of DNMT3L by a peptide linker forming a catalytically active single-chain protein, which is the most active mammalian methyltransferase for DNA methylation editing that has been characterized by now according to our knowledge [29]. If we compare the activity of editing modules assembled from different DNA methyltransferases, one gets the following series: DNMT1 < DNMT3A < DNMT3A-DNMT3L (our unpublished data). Several studies showed that the properties of the editing module are crucial for the specificity of epigenome editing. For example, off-target DNA methylation may arise from nonspecific interaction of the DNA methyltransferase module with genomic DNA at open chromatin environments. To avoid this undesired editing, new variants with higher specificity were developed [30].

Here, we describe a step-by-step protocol for targeted DNA methylation in HEK293 cells using a dCas9-based targeting module and the optimized DNMT3A-DNMT3L fusion proteins as an editing module. The high editing efficiency is achieved by using the dCas9-10×SunTag construct in a combination with the very potent chimeric methyltransferase. The R887E mutant of DNMT3A significantly reduces potential off-target editing outside of the target region. Design of gRNAs can be conducted using one of the multiple web-based programs to assure specificity of the recognition of the target site by the dCas9/gRNA complex. Finally, one can analyze the efficiency of DNA methylation using the well-established bisulfite conversion method followed by next-generation sequencing or precipitation of methylated DNA followed by qPCR or next-generation sequencing. The first method is the “gold standard” for DNA methylation analysis. It allows for the quantitative determination of the methylation level at the CpG site resolution. This method was explained in detail previously [31]. The second group of methods is based on the ability of the methyl-CpG-binding domain (MBD) [32] or a 5-methylcytosine-specific antibody [33] to discriminate methylated DNA from unmethylated one. These assays allow to determine a relative level of DNA methylation of the target region, compared to the known highly and unmethylated loci and to the target region before

epigenome editing. In contrast to the first method, these assays are not suitable for the analysis of DNA methylation of individual CpG sites.

---

## 2 Materials

### 2.1 gRNA Cloning Using Golden Gate Assembly

1. gRNA has to be cloned into an expression vector. Multiple suitable vectors generated by the community are available at Addgene (<https://www.addgene.org/crispr/empty-grna-vectors/>). Depending on the design of the experiment, either transient gene transfer or the generation of stable cell lines is possible (*see* **Notes 1** and **2**).
2. A pair of oligonucleotides for each gRNA. Purchase the oligonucleotides from your regular supplier at a normal PCR primer quality.
3. 10× T4 DNA Ligase Reaction Buffer.
4. T4 PNK.
5. BbsI-HF<sup>®</sup>.
6. T4 DNA ligase, 2,000,000 units/mL.
7. Sterile 0.2-mL PCR Eppendorf tubes.
8. ddH<sub>2</sub>O.
9. Thermocycler.

### 2.2 Cell Culture

1. Human Embryonic Kidney cells 293 (HEK293; can be obtained from standard repositories, e.g. ATCC (<https://www.atcc.org>) or DSZM (<https://www.dsmz.de/>)).
2. Dulbecco's Modified Eagle's Medium (DMEM).
3. DMEM+: DMEM supplemented with 10% fetal bovine serum, 4-mM L-glutamine.
4. FuGENE<sup>®</sup> HD Transfection Reagent.
5. dCas9-10×SunTag vector (Addgene plasmid #174141).
6. scFv-DNMT3A(R887E)-DNMT3L (Addgene plasmid #154141).
7. Sterile 1.5-mL Eppendorf tubes.

---

## 3 Methods

### 3.1 gRNA Design

1. Select the region of interest for DNA methylation, and retrieve the genomic DNA sequence of this region, and identify potential PAM sites to define potential target sites for the EpiEditor (*see* **Note 3**).



**Fig. 3** Schematic presentation of a genomic target locus for targeted DNA methylation (green shaded area), potential gRNA targeting sequences (blue and green arrows directing from the 5' end to the PAM site), and two gRNAs with the optimal positioning (green arrows)

2. The gRNA seed sequence can be selected using one of the free web-based design tools. The most recent versions of the tools include the analysis of the gRNA sequence, which influence the binding strength at the target site and the genome-wide search for potential off-target binding sites; the latter is important for the specificity of editing. There are several sites, which offer this type of design:

<https://portals.broadinstitute.org/gppx/crispick/public>

<https://rth.dk/resources/crispr/>

<http://chopchop.cbu.uib.no/>

Analyze potential gRNAs from the list of candidates generated by the web-based tool. Choose the best candidates with the lowest off-target binding scores and an optimum binding position according to **step 1** in Subheading 3.1 (Fig. 3). Select two to four gRNAs per target gene for the first round of screening.

### 3.2 gRNA Cloning

1. The vast majority of gRNA expression vectors use Golden Gate Assembly for cloning [34], which requires oligonucleotides with overhangs compatible with the vector backbone and applies a type IIS restriction enzyme for cloning. Check thoroughly the vector documentation in order to add correct overhangs for the specified restriction enzyme. In this protocol, a home-made gRNA expression vector will be used and the cloning is assisted by the BbsI-HF enzyme.
2. Choose a target sequence from the list generated at **step 2** in Subheading 3.1; it must be 20 nucleotides long. One needs two oligonucleotides for cloning. To generate the top strand oligonucleotide, replace the first 5' nucleotide in the target sequence by G. This is necessary, because the transcription of the gRNAs is driven by the human U6 promoter that requires guanine at the first position. Add three additional nucleotides ACC to the 5' end to produce an overhang compatible with the vector backbone. The bottom strand oligonucleotide sequence



37 °C	30 min
95 °C	5 min
Ramp down to 25 °C at 1 °C/min	
8 °C	Store

4. Cloning into the backbone vector can be done using one of the ready-to-use Golden Gate Assembly Kits or by a custom-assembled reaction. Dilute the double-stranded DNA 500-fold. Assemble the reaction and incubate in the thermocycler (*see Note 4*).

	Stock concentration	Volume
H <sub>2</sub> O		Up to 20 µL
T4 DNA ligase reaction buffer	10×	2 µL
T4 DNA ligase		1 µL
BbsI-HF <sup>®</sup>		1 µL
Empty gRNA cloning vector	10 ng/µL	2 µL
Pre-annealed/diluted oligonucleotides		1.5 µL

Program for the thermocycler:

37 °C	1 min	Repeat 30 times
16 °C	1 min	
37 °C	20 min	
65 °C	20 min	

5. Add 40-µL H<sub>2</sub>O to the reaction tube and use 2 µL for transformation.
6. Pick up two to three individual clones and verify the correct insertion of the gRNA sequence by Sanger sequencing.

### 3.3 Transfection of Cells with EpiEditors

1. Cultivate HEK293 cells in DMEM+.
2. 24 h before transfection, seed 250,000 cells into six-well plates.
3. Directly before transfection replace the culture medium by 1.5-mL fresh DMEM+.
4. Conduct transfection using your reagent of choice, for example, FuGENE<sup>®</sup> HD. In this protocol, a co-transfection of three plasmids, namely, gRNA (cloned in Subheading 3.2), dCas9-

10×SunTag, and scFv-DNMT3A(R887E)-DNMT3L, will be performed (*see Note 5*). Assemble transfection mixtures using the following recipe.

DMEM	140 $\mu$ L
dCas9-10×SunTag plasmid	1000 ng
scFv-DNMT3A(R887E)-DNMT3L plasmid	500 ng
gRNA expression plasmid	80 ng
FuGENE <sup>®</sup> HD	6 $\mu$ L

Mix thoroughly by pipetting and incubate 20 min at room temperature.

Add the DNA/transfection reagent mixtures to cells and shake the plate. Keep cells in the incubator for 24 h; then replace medium with 2-mL DMEM+.

5. Harvest cells at the desired time point (*see Note 6*).
6. To enrich a population of triple-positive cells, which took up all three plasmids after transfection, FACS has to be conducted. Pellet cells by centrifugation and store at  $-80\text{ }^{\circ}\text{C}$  for further DNA methylation analysis.

---

## 4 Notes

1. In case of the transient transfection approach, we recommend to choose a vector also encoding a fluorescent protein marker, because this will allow for enrichment of a positive cell population using FACS. Alternatively, one may clone the gRNA into vectors for production of retroviruses or lentiviruses and generate stable cell lines. For viral vectors it is advisable to have an antibiotic resistance gene, which facilitates the selection of a positive population and avoids the need of the FACS device.
2. One disadvantage of the stable cell line approach is that leaky expression of the EpiEditor may confuse studies aiming to investigate the long-term stability of the edited state.
3. Consider that the dCas9/gRNA complex leaves a footprint by protecting the binding site from methylation, which is about 20–30 base pairs long and will not be methylated. The peak of methylation typically occurs 20–50 base pairs 3' of the PAM site.
4. Optionally, assemble a control reaction, which contains all reagents just leaving out the annealed oligonucleotides to check for the frequency of false-positive colonies.



5. Use only high-purity plasmid preparations for transfection (e.g., prepared using QIAGEN Plasmid Midi Kit or similar).
6. According to our experience, the DNA methylation level reaches its maximum at 4–5 days after transfection.

---

## Acknowledgments

Work in the authors' laboratory has been supported by the BW Foundation (AllEpi, ID09).

## References

1. Parmar JJ, Woringer M, Zimmer C (2019) How the genome folds: the biophysics of four-dimensional chromatin organization. *Annu Rev Biophys* 48:231–253. <https://doi.org/10.1146/annurev-biophys-052118-115638>
2. Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128:693–705. <https://doi.org/10.1016/j.cell.2007.02.005>
3. Jeltsch A (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *ChemBiochem* 3: 274–293. [https://doi.org/10.1002/1439-7633\(20020402\)3:4<274::aid-cbic274>3.0.co;2-s](https://doi.org/10.1002/1439-7633(20020402)3:4<274::aid-cbic274>3.0.co;2-s)
4. Li E, Zhang Y (2014) DNA methylation in mammals. *Cold Spring Harb Perspect Biol* 6: a019133. <https://doi.org/10.1101/cshperspect.a019133>
5. Jeltsch A, Jurkowska RZ (2014) New concepts in DNA methylation. *Trends Biochem Sci* 39: 310–318. <https://doi.org/10.1016/j.tibs.2014.05.002>
6. De Groote ML, Verschure PJ, Rots MG (2012) Epigenetic editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes. *Nucleic Acids Res* 40: 10596–10613. <https://doi.org/10.1093/nar/gks863>
7. Kungulovski G, Jeltsch A (2016) Epigenome editing: state of the art, concepts, and perspectives. *Trends Genet* 32:101–113. <https://doi.org/10.1016/j.tig.2015.12.001>
8. Carvin CD, Dhasarathy A, Friesenhahn LB, Jessen WJ, Kladde MP (2003) Targeted cytosine methylation for in vivo detection of protein-DNA interactions. *Proc Natl Acad Sci U S A* 100:7743–7748. <https://doi.org/10.1073/pnas.1332672100>
9. Li F, Papworth M, Minczuk M, Rohde C, Zhang Y, Ragozin S, Jeltsch A (2007) Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes. *Nucleic Acids Res* 35:100–112. <https://doi.org/10.1093/nar/gkl1035>
10. Xu G-L, Bestor TH (1997) Cytosine methylation targetted to pre-determined sequences. *Nat Genet* 17:376–378. <https://doi.org/10.1038/ng1297-376>
11. Wolfe SA, Nekludova L, Pabo CO (2000) DNA recognition by Cys 2 His 2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* 29: 183–212. <https://doi.org/10.1146/annurev.biophys.29.1.183>
12. Bogdanove AJ, Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. *Science* 333:1843–1846. <https://doi.org/10.1126/science.1204094>
13. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816–821. <https://doi.org/10.1126/science.1225829>
14. Kim Y, Kweon J, Kim A, Chon JK, Yoo JY, Kim HJ, Kim S, Lee C, Jeong E, Chung E, Kim D, Lee MS, Go EM, Song HJ, Kim H, Cho N, Bang D, Kim S, Kim J-S (2013) A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol* 31:251–258. <https://doi.org/10.1038/nbt.2517>
15. Liu Q, Segal DJ, Ghiara JB, Barbas CF (1997) Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci* 94:5525–5530. <https://doi.org/10.1073/pnas.94.11.5525>
16. Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene

- expression. *Cell* 152:1173–1183. <https://doi.org/10.1016/j.cell.2013.02.022>
17. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31:827–832. <https://doi.org/10.1038/nbt.2647>
  18. Stepper P, Kungulovski G, Jurkowska RZ, Chandra T, Krueger F, Reinhardt R, Reik W, Jeltsch A, Jurkowski TP (2017) Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. *Nucleic Acids Res* 45:1703–1713. <https://doi.org/10.1093/nar/gkw1112>
  19. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517:583–588. <https://doi.org/10.1038/nature14136>
  20. Shechner DM, Hacısuleyman E, Younger ST, Rinn JL (2015) Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat Methods* 12:664–670. <https://doi.org/10.1038/nmeth.3433>
  21. Ma H, Tu L-C, Naseri A, Chung Y-C, Grunwald D, Zhang S, Pederson T (2018) CRISPR-Sirius: RNA scaffolds for signal amplification in genome imaging. *Nat Methods* 15:928–931. <https://doi.org/10.1038/s41592-018-0174-0>
  22. Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159:635–646. <https://doi.org/10.1016/j.cell.2014.09.039>
  23. Lebar T, Lainšček D, Merljak E, Aupič J, Jerala R (2020) A tunable orthogonal coiled-coil interaction toolbox for engineering mammalian cells. *Nat Chem Biol* 16:513–519. <https://doi.org/10.1038/s41589-019-0443-y>
  24. Jurkowska RZ, Jurkowski TP, Jeltsch A (2011) Structure and function of mammalian DNA methyltransferases. *Chembiochem* 12:206–222. <https://doi.org/10.1002/cbic.201000195>
  25. Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99:247–257. [https://doi.org/10.1016/S0092-8674\(00\)81656-6](https://doi.org/10.1016/S0092-8674(00)81656-6)
  26. Chédin F, Lieber MR, Hsieh CL (2002) The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc Natl Acad Sci U S A* 99:16916–16921. <https://doi.org/10.1073/pnas.262443999>
  27. Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A (2005) Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *J Biol Chem* 280:13341–13348. <https://doi.org/10.1074/jbc.M413412200>
  28. Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* 449:248–251. <https://doi.org/10.1038/nature06146>
  29. Siddique AN, Nunna S, Rajavelu A, Zhang Y, Jurkowska RZ, Reinhardt R, Rots MG, Ragozin S, Jurkowski TP, Jeltsch A (2013) Targeted methylation and gene silencing of VEGF-A in human cells by using a designed Dnmt3a–Dnmt3L single-chain fusion protein with increased DNA methylation activity. *J Mol Biol* 425:479–491. <https://doi.org/10.1016/j.jmb.2012.11.038>
  30. Hofacker D, Broche J, Laistner L, Adam S, Bashtrykov P, Jeltsch A (2020) Engineering of effector domains for targeted DNA methylation with reduced off-target effects. *Int J Mol Sci* 21:502. <https://doi.org/10.3390/ijms21020502>
  31. Bashtrykov P, Jeltsch A (2018) DNA methylation analysis by bisulfite conversion coupled to double multiplexed amplicon-based next-generation sequencing (NGS). *Methods Mol Biol* 1767:367–382. [https://doi.org/10.1007/978-1-4939-7774-1\\_20](https://doi.org/10.1007/978-1-4939-7774-1_20)
  32. Jeltsch A, Broche J, Lungu C, Bashtrykov P (2020) Biotechnological applications of MBD domain proteins for DNA methylation analysis. *J Mol Biol* 432(6):18161823. <https://doi.org/10.1016/j.jmb.2019.08.020>
  33. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schübeler D (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37:853–862. <https://doi.org/10.1038/ng1598>
  34. Engler C, Kandzia R, Marillonnet S, El-Shemy HA (2008) A one pot one step precision cloning method with high throughput capability. *PLoS ONE* 3(11):e3647. <https://doi.org/10.1371/journal.pone.0003647>



## Regulation of Gene Expression Using dCas9-SunTag Platforms

Sumiyo Morita, Takuro Horii, and Izuho Hatada

### Abstract

Regulating gene expression is important for basic research and therapeutic applications. The epigenome is a record of genetic modifications such as DNA methylation and histone modifications, and epigenetic changes can play a key role in modifying gene expression. With the advent of genome editing technologies, it has become possible to manipulate the epigenome of specific genomic regions to control gene expression. In particular, CRISPR–Cas9 systems have been used widely for epigenome editing due to their high efficiency, versatility, specificity, and ease of use. Here, we describe a protocol for the upregulation of specific genes using the dCas9-SunTag system.

**Key words** dCas9, TET1, VP64, SunTag, DNA methylation, Epigenetic editing

---

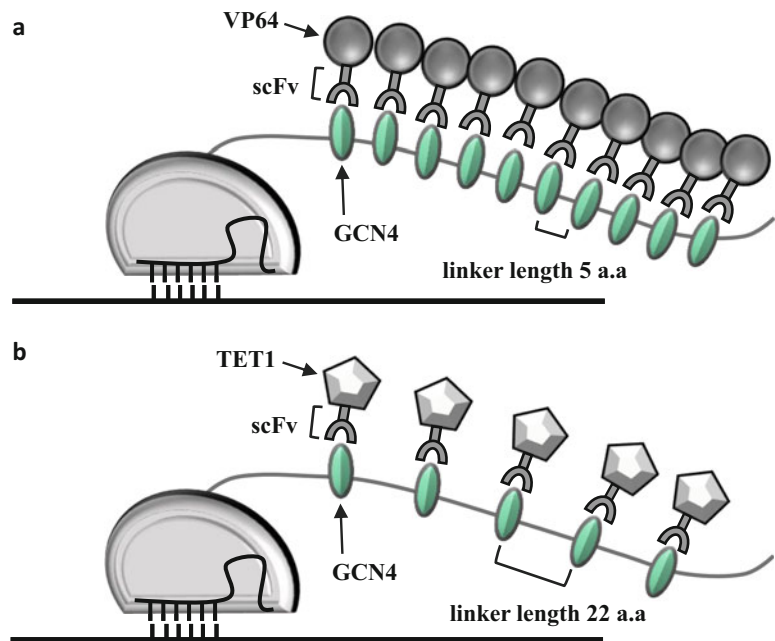
### 1 Introduction

Controlling the expression of specific genes is an extremely important tool for basic research in a wide variety of fields, including embryology, disease treatment, and drug discovery. The epigenome includes information encoded in chemical modifications of DNA and histones, such as DNA methylation and histone acetylation, and changes to the epigenome play a key role in modifying gene expression. Indeed, dysregulation of the epigenome can cause a range of diseases, such as cancer [1, 2], autoimmune disorders [3], neurological disorders [4], diabetes [5], and imprinting diseases [6]. In recent years, genome editing technologies, in particular the CRISPR–Cas9 system, have been used widely for epigenome editing due to their high efficiency, versatility, specificity, and ease of use.

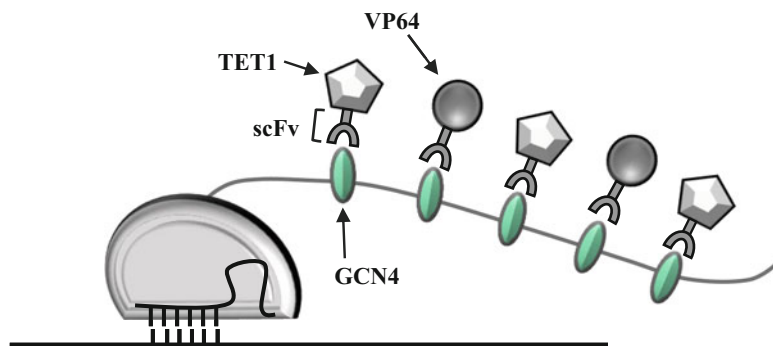
For epigenome editing, a combination of endonuclease-deficient Cas9 (dCas9) and catalytic domains of epigenetic modifiers can be used to regulate gene expression [7]. Fusions of dCas9 and a transcription activation domain (e.g., dCas9-VP64) are used

widely for the activation of endogenous genes. In this approach, expression is induced by targeting the promoter region of a gene of interest. However, because the gene activation range is low when only a single-guide RNA (sgRNA) is used, tiling of the target promoter region with several sgRNAs is often necessary to achieve robust transcriptional activation [8–10]. Consequently, it is difficult to use this format for genome-wide gain-of-function screening and gene therapy. To solve this problem, the synergistic activation mediator (SAM) system, a format that recruits multiple factors to dCas9, thereby improving the activation range, was developed recently [11, 12]. The SAM system consists of dCas9-VP64, a modified gRNA containing RNA aptamers, and a MS2 bacteriophage coat protein fused to a transcription activation domain (MS2-p65-HSF1). VP64 fused to dCas9 and MS2-p65-HSF1 recruited to dCas9 by the RNA aptamers synergistically activate gene transcription of the target gene.

In previous work, we developed a modified dCas9-SunTag format for efficient, targeted demethylation and activation of specific DNA loci using a single sgRNA [13]. SunTag is a repeated GCN4 peptide tag array that recruits multiple copies of an effector molecule fused to an anti-GCN4 peptide antibody (scFv) [14]. The original SunTag consists of ten copies of the GCN4 peptide separated by 5-amino-acid linkers (Fig. 1a). We optimized the linker length between the tags to maximize transcriptional activation by



**Fig. 1** (a) The original SunTag system, comprising ten copies of the GCN4 peptide separated by 5-amino-acid linkers. (b) The modified SunTag system, comprising 22-amino-acid linkers between the GCN4 peptides



**Fig. 2** Modified dCas9-SunTag format for simultaneous recruitment of TET1 and VP64

the effector TET1, which performs the first enzymatic steps during DNA demethylation (Fig. 1b), and found that this system was able to efficiently demethylate a target DNA sequence and increase gene expression [13].

In this section, we describe simultaneous tethering of VP64 and TET1 to a target gene using the optimized dCas9-SunTag format, resulting in synergistic activation of gene expression (Fig. 2) [15]. The procedure involves the design and construction of gRNA(s) for the target gene, transfection of cells with the gRNA(s) for transcriptional activation of the target gene(s), and quantitative RT-PCR analyses to detect expression of the target gene(s). Using the modified dCas9-SunTag format, multiple factors can be recruited to a target gene to synergistically activate transcription. The combination of TET1 and VP64 can increase the expression levels of a larger number of genes than those increased by TET1 or VP64 alone.

## 2 Materials

### 2.1 Plasmids

1. pCAG-dCas9-5×Plat2AfiD (#82560, Addgene).
2. pCAG-scFvGCN4sfGFPTET1CD (#82561, Addgene).
3. pCAG-scFvGCN4sfGFPVP64GB1 (#141417, Addgene).
4. gRNA cloning vector: an empty gRNA expression vector that can be used to create a gRNA targeting a specific sequence (#128433, Addgene).

### 2.2 Preparation of Vector

1. *BbsI* (NEB) (*see Note 1*).
2. 10-mM Tris-Cl buffer (pH 8.5).
3. LB agar with kanamycin: 1.6% agar, 1% tryptone, 0.5% yeast extract, 1% NaCl, 50- $\mu$ g/mL kanamycin (add from a 10-mg/mL stock solution after allowing the medium to cool after autoclaving).

4. LB: 1% tryptone, 0.5% yeast extract, 1% NaCl, 50-μg/mL kanamycin (add from a 10-mg/mL stock solution after allowing the medium to cool after autoclaving).
5. Competent *Escherichia coli*.
6. M13 Forward (−20): 5′-GTAAAACGACGGCCAG-3′.
7. M13 Reverse: 5′-CAGGAAACAGCTATGAC-3′.
8. DNA sequencing kit.
9. Plasmid DNA purification kit.
10. Thermal cycler.

**2.3 Transfection of Cells**

1. Lipofection reagent: Lipofectamine 2000 (Thermo Fisher).
2. Opti-MEM (Thermo Fisher).
3. Cell culture medium.

**2.4 Gene Expression Analysis**

1. RNA extraction kit.
2. RT-PCR reagents.
3. Real-time PCR machine.

**3 Methods**

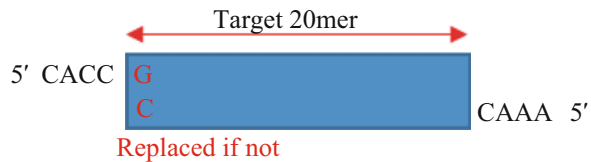
**3.1 Design and Construction of gRNA**

1. Enter the sequence encompassing the transcription start site of each target gene into CRISPRdirect (<https://crispr.dbcls.jp/>) to select a highly specific target sequence (20-mer) (*see Note 2*). If not already the case, change the 5′ nucleotide of the selected sgRNA sequence to guanine (the U6 promoter absolutely requires a guanine at the transcription start site). This target sequence is defined as **S** and the antisense complement of **S** is defined as **AS**.
2. Synthesize the following oligonucleotides:

XgRNA-S: 5′-CACCS-3′

XgRNA-AS: 5′-AAACAS-3′

Annealing of these oligonucleotides produces the following DNA:



3. To anneal the two oligonucleotides (XgRNA-S and XgRNA-AS), mix 1  $\mu\text{L}$  of  $10\times$  PCR buffer with 4.5  $\mu\text{L}$  of each oligonucleotide (100 pmol/ $\mu\text{L}$ ), and incubate at 95  $^{\circ}\text{C}$  for 5 min, 37  $^{\circ}\text{C}$  for 15 min, and then room temperature for 15 min.
4. To prepare the destination vector, incubate the gRNA cloning vector with *BbsI* at 37  $^{\circ}\text{C}$  for 3 h.
5. Purify the *BbsI*-treated vector using gel electrophoresis and a gel extraction kit. Dilute 30  $\mu\text{L}$  of 10-mM Tris-Cl buffer (pH 8.5).
6. Ligate the oligonucleotide mixture prepared in **step 3** and the *BbsI*-treated gRNA cloning vector prepared in **step 5**.
7. Transform the ligation mixture ( $\sim 1\text{--}2$   $\mu\text{L}$ ) into competent *E. coli* and spread the transformed cells onto solid medium containing kanamycin.
8. After incubation of the plate overnight at 37  $^{\circ}\text{C}$ , pick a few colonies (around four will be sufficient), and confirm the sequence by conventional DNA sequencing using the M13 Forward ( $-20$ ) and M13 Reverse primers.

TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCA  
 AGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCCTTCATATTTGCATATACGAT  
 ACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTA  
 GTACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAA  
 ATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTGAAAGTATTTTCGATTTC  
 TTGGCTTTATATATCTTGTGGAAAGGACGAAACACCgxxxxxxxxxxxxxxxxxxxxGTTTT  
 AGAGCTAGAAATAGCAAGTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGT  
 GGCACCGAGTCGGTGCTTTTTTTCTAGACCCAGCTTCTTGTACAAAGTTGGCATT

\*The target site is indicated in red.

9. Culture the colony and purify the plasmid.

### 3.2 Transfection of Cells for Transcriptional Activation of the Target Genes

This section provides an example protocol for transfection of the plasmids described in Subheading 2.1 into human A549 cells in 24-well culture dishes using Lipofectamine 2000. A vector expressing GFP alone should be used as a negative control. For drug selection, cells should also be co-transfected with a vector harboring a blasticidin S-resistance gene.

1. Mix the pCAG-dCas9-5 $\times$ Plat2AfiD, pCAG-scFvGCN4sfGFP TET1CD, and pCAG-scFvGCN4sfGFPVP64GB1 plasmids and the gRNA cloning vector at molar ratios of 1:1:1:1.5, respectively (*see* **Notes 3** and **4**).

2. Dilute 2  $\mu\text{L}$  of Lipofectamine 2000 in 50  $\mu\text{L}$  of Opti-MEM medium (1:4 ratio).
3. Dilute the plasmid DNA mixture (500 ng) in 50  $\mu\text{L}$  of Opti-MEM medium.
4. Add the diluted plasmid DNA mixture to the diluted Lipofectamine 2000 reagent, and incubate for 5 min.
5. Add the DNA/Lipofectamine 2000 mixture to 500  $\mu\text{L}$  of culture medium containing  $1 \times 10^5$  cells (*see Note 5*).
6. At 48-h post-transfection, add blasticidin S to the culture medium at a final concentration of 2  $\mu\text{g}/\text{mL}$ . After 3 days of selection, harvest the cells for quantitative RT-PCR analysis.

### 3.3 Quantitative RT-PCR Analysis

1. Isolate total RNA from the transfected cells.
2. Perform reverse transcription using the isolated total RNA as a template.
3. Measure the expression of the target gene(s) using real-time PCR. Expression levels should be normalized to the level of actin mRNA and presented as fold changes relative to GFP-transfected experimental controls (*see Note 6*).

---

## 4 Notes

1. It is possible to use the same product from another manufacturer.
2. Take care not to include the protospacer adjacent motif (PAM) sequence (NGG) in the 20-mer.
3. As an alternative to TET1CD plus VP64, TET1CD plus p65HSF1 or VP64 plus p65HSF1 can be used for transcription activation [15].
4. A combination of three or more fusion proteins (e.g., TET1CD, VP64, and p65HSF1) can also be used.
5. Detach cells using trypsin/EDTA and adjust to a density of  $1 \times 10^5$  cells per 500  $\mu\text{L}$  of culture medium.
6. To suppress gene expression, the DNA methyltransferases Dnmt3a, KRAB, and/or MeCP2 can be combined.

---

## Acknowledgments

This work was supported by the Basic Science and Platform Technology Program for Innovative Biological Medicine from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, and by the Japan Agency for Medical Research and Development (AMED).



## References

1. Timp W, Feinberg AP (2013) Cancer as a dys-regulated epigenome allowing cellular growth advantage at the expense of the host. *Nat Rev Cancer* 13:497–510
2. Chatterjee A, Rodger EJ, Eccles MR (2017) Epigenetic drivers of tumourigenesis and cancer metastasis. *Semin Cancer Biol* 51:149–159
3. Mazzone R, Zwergel C, Artico M et al (2019) The emerging role of epigenetics in human autoimmune disorders. *Clin Epigenetics* 11:34
4. Urdinguio RG, Sanchez-Mut JV, Esteller M (2009) Epigenetic mechanisms in neurological diseases: genes, syndromes, and therapies. *Lancet Neurol* 8:1056–1072
5. Rosen ED, Kaestner KH, Natarajan R et al (2018) Epigenetics and epigenomics: implications for diabetes and obesity. *Diabetes* 67: 1923–1931
6. Butler MG (2009) Genomic imprinting disorders in humans: a mini-review. *J Assist Reprod Genet* 26:477–486
7. Martella A, Fisher AI (2021) Elucidative role of CRISPR-based epigenetic modifiers and CRISPR-induced chromosome conformational changes. *CRISPR J* 4:1
8. Maeder ML, Linder SJ, Cascio VM et al (2013) CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* 10:977–979
9. Perez-Pinera P, Kocak DD, Vockley CM et al (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* 10:973–976
10. Mali P, Aach J, Stranges PB et al (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* 31: 833–838
11. Konermann S, Brigham MD, Trevino AE et al (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517:583–588
12. Breunig CT, Köferle A, Neuner AM et al (2021) CRISPR tools for physiology and cell state changes: potential of transcriptional engineering and epigenome editing. *Physiol Rev* 101:177–211
13. Morita S, Noguchi H, Horii T et al (2016) Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat Biotechnol* 34: 1060–1065
14. Tanenbaum ME, Gilbert LA, Qi LS et al (1997) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159:635–646
15. Morita S, Horii T, Kimura M et al (2020) Synergistic upregulation of target genes by TET1 and VP64 in the dCas9-SunTag platform. *Int J Mol Sci* 21:1574



## Concatenated Coiled-Coil Tag for Highly Efficient, Small Molecule-Inducible Upregulation of Endogenous Mammalian Genes

Tina Lebar and Roman Jerala

### Abstract

Regulation of epigenomic functions requires controlled site-specific alteration of epigenetic information. This can be achieved by using designed DNA-binding domains, associated with effector domains, that function as targeted transcription factors or epigenetic modifiers. These effectors have been employed to study the implications of epigenetic modifications, and sequence-specific targeting has been instrumental in understanding the effect of these modification on gene regulation. Ultimately, these tools could be used for therapeutic applications to revert the epigenetic aberrations that have been linked to various diseases. The ability to spatiotemporally control gene expression is especially important for precise regulation of the epigenomic state. In this chapter, we describe the protocol for achieving highly efficient small molecule-inducible transcriptional activation of endogenous mammalian genes, mediated by a dCas9-based system that recruits transcriptional activation domains binding to a chain of concatenated coiled-coil peptides.

**Key words** CRISPR/dCas9, Transcriptional activation, Coiled-coil peptide, Abscisic acid

---

### 1 Introduction

Transcriptional control is a prevalent mode of regulating protein expression and has received a lot of attention in terms of tool development. Over the last several decades, researchers have assembled methodologies for creating tools that enable precise, efficient control of gene expression in different organisms. Designed transcription factors and other epigenetic modifiers, that can target any selected DNA sequence, are particularly useful for control of endogenous gene expression in mammalian cells [1]. Transcriptional initiation in mammalian cells occurs via recruitment of the transcriptional machinery by clustered activation domains of the bound transcription factors. However, chromosomes can be tightly packaged, potentially limiting transcription factor access to the targeted region. Additionally, chromatin structure can differ

between cell types. The ability to robustly program transcriptional activation of selected genes within its native chromosomal context would provide a transformative tool for a plethora of applications.

By mimicking the natural cooperative recruitment process, a chimeric transcriptional activation domain, comprising multiple different activation domains (VPR), was previously shown to strongly increase the levels of transcriptional activation [2]. Other approaches employing similar principles, such as Synergistic Activation Mediator (SAM) [3] and SunTag [4], were developed by other groups of researchers. The SunTag system employs concatenated of peptide repeats, which act as a scaffold for recruitment of antibodies, genetically fused to transcriptional activation domains. The peptide epitope used in SunTag originates from the GCN4 coiled-coil peptide, which is one of the most characterized coiled-coil peptides and forms a parallel homodimer [5]. Nevertheless, hetero-specific interaction domains would facilitate more precise control of the stoichiometry. We have recently validated an orthogonal set of six de novo designed heterodimeric coiled-coil peptide pairs (NICP set) in mammalian cells [6, 7]. The heterodimeric peptides were used to construct a scaffold for recruitment of transcriptional activation domains, termed the concatenated coiled-coil tag (CCC-tag), which consistently exhibited superior transcriptional activation strength in comparison to SunTag. When used in combination with the VPR domain [2], the CCC-tag yielded one of the most potent CRISPR/dCas9-based transcriptional activator systems reported thus far, as confirmed by benchmarking to the above-described approaches [7]. Additionally, we demonstrated the extended use of the CCC-tag for amplification of light- and small molecule-inducible transcriptional regulation.

In this chapter, we describe our standard procedure for CRISPR/dCas9-based, abscisic acid-regulated transcriptional activation using the CCC-tag (Fig. 1). We outline the design and preparation of guide RNAs and describe the methods for mammalian cell line transfection and small molecule stimulation and our protocol for the analysis of endogenous gene expression levels.

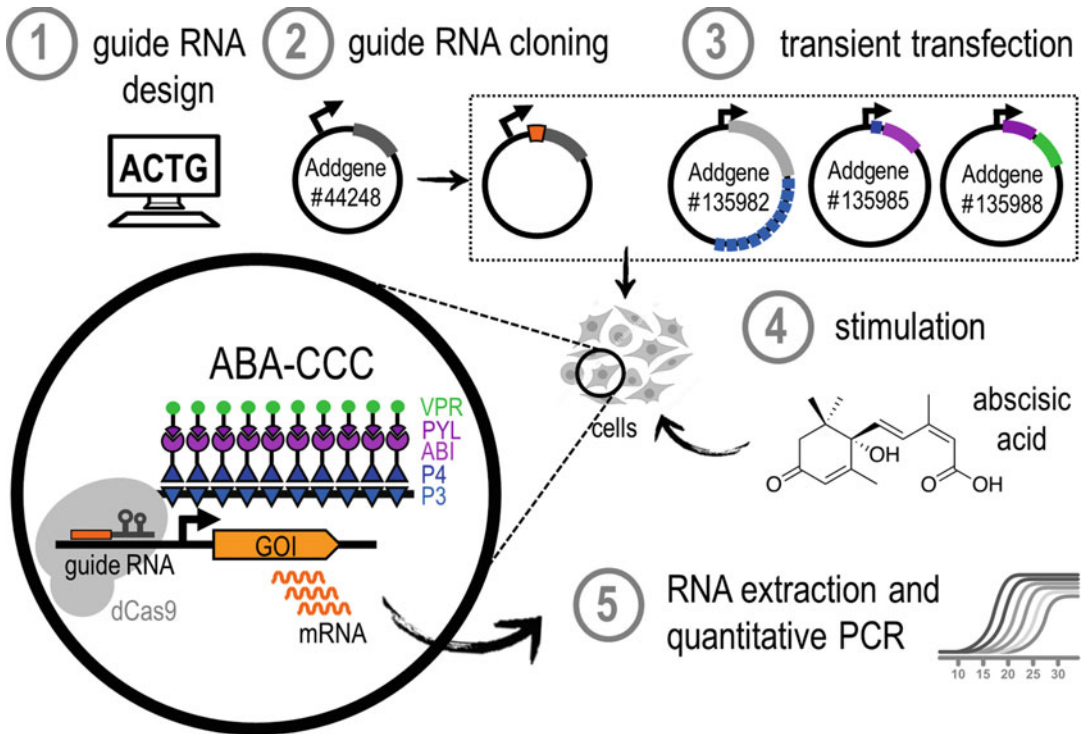
---

## 2 Materials

### 2.1 Construct Design and Preparation

#### 2.1.1 Equipment

1. PCR thermocycler.
2. Agarose electrophoresis system.
3. Transilluminator.
4. Benchtop centrifuge.
5. Nanodrop.
6. Shaker incubator for bacterial cell culture.



**Fig. 1** Step-by-step protocol for abscisic acid-inducible transcriptional activation of endogenous mammalian genes with CCC-tag. Computational guide RNA design (1) is followed by cloning of the guide RNAs into an appropriate vector (2). Guide RNA encoding constructs are transiently transfected into a selected cell line (3). Cells are stimulated by addition of abscisic acid into the media (4), followed by analysis of gene expression by quantitative PCR

### 2.1.2 Reagents

1. pgRNA-humanized lentiviral vector (Addgene plasmid no. 44248).
2. Reverse oligo for gRNA scaffold amplification (5'-GCCTAATGGATCCTAGTA-3').
3. Forward oligos for gRNA scaffold amplification (5'-GTATCCCTTGGAGAACCACCTTGTGGNNNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAGAAATAG-3').
4. BstXI and XhoI restriction enzymes (NEB).
5. NEBuffer™ r3.1.
6. High-fidelity polymerase.
7. 0.8% and 2% (w/v) agarose gels.
8. 6× loading dye (30% (v/v) glycerol, 0.25% (w/v) bromophenol blue).
9. 1-kb Plus DNA Ladder.
10. 1× concentrated TAE buffer (40-mM Tris-HCl, pH 7.6, 20-mM acetic acid, 1-mM EDTA).

11. Gel extraction kit.
12. NEBuilder<sup>®</sup> HiFi DNA Assembly Master Mix.
13. Chemically competent *E. coli* cloning strain.
14. Antibiotic-supplemented LB liquid media and LB agar plates.
15. Plasmid purification kit.
16. Sequencing oligonucleotide (5'-CAGCACAAAAGGAAACT CACC-3').

## **2.2 Mammalian Cell Culture**

### **2.2.1 Equipment**

1. Cell culture CO<sub>2</sub> incubator.
2. Laminar flow hood.
3. Benchtop centrifuge.
4. Automated cell counter.
5. Inverted fluorescent microscope.
6. Vacuum aspiration system.

### **2.2.2 Reagents**

1. HEK293T cell line (ATCC).
2. Neuro2A cell line (ATCC).
3. Dulbecco's Modified Eagle Medium (DMEM).
4. Eagle's Minimal Essential Medium (EMEM).
5. Fetal bovine serum (FBS).
6. Trypsin-EDTA solution.
7. 150-mM NaCl.
8. 7.5-mM polyethyleneimine (PEI), linear, 25,000-MW stock solution; pH 7.5.
9. Lipofectamine<sup>™</sup> LTX Reagent with PLUS<sup>™</sup> Reagent (Thermo Fisher Scientific).
10. Opti-MEM<sup>™</sup> Reduced Serum Medium (Thermo Fisher Scientific).
11. Plasmids encoding CCC-tag transcription factor constructs (Addgene plasmids no. 135982, 135983, 135985, 135988).
12. pcDNA3 plasmid vector (Thermo Fisher Scientific).
13. pgRNA-humanized lentiviral vector (Addgene plasmid no. 44248).
14. Abscisic acid.
15. Dimethyl sulfoxide (DMSO).

## **2.3 Analysis of Gene Expression**

### **2.3.1 Equipment**

1. PCR thermocycler.
2. Real-time PCR detection system.
3. Agarose electrophoresis system.
4. Transilluminator.

### 2.3.2 Reagents

1. RNA extraction kit.
2. Reverse transcription kit.
3. Real-time PCR reaction mix.
4. Oligonucleotides for amplification of genomic regions.
5. 2% (w/v) agarose gel.
6. 6× loading dye (30% (v/v) glycerol, 0.25% (w/v) bromophenol blue).
7. 100-bp DNA Ladder.
8. 1× concentrated TAE buffer (40-mM Tris-HCl, pH 7.6, 20-mM acetic acid, 1-mM EDTA).

---

## 3 Methods

### 3.1 Construct Design and Preparation

Here we describe the design and preparation of plasmids, encoding guide RNA molecules that target endogenous mammalian promoters for upregulation of gene expression using CCC-based transcription factors. We describe our routine methodology for the design of *S. pyogenes* dCas9 guide RNAs (*see Note 1*).

#### 3.1.1 Guide RNA Design

1. Core promoter regions of genes of interest (GOI) for selected organisms are found in the Eukaryotic Promoter Database (EPD). Retrieve a promoter sequence from  $-250$  to  $1$  bp relative to the transcription start site (TSS).
2. The core promoter region is searched for all potential protospacer adjacent motifs (PAM) on the template strand ( $5'$ -NGG- $3'$ ) and the coding strand ( $5'$ -CCN- $3'$ ).
3. The  $5'$  adjacent 20 nucleotides of the template strand PAM and the  $3'$  adjacent 20 nucleotides of the coding strand PAM are retrieved and checked for homopolymeric regions. All potential guide RNAs containing polymeric tracts of four or more repeats of the same nucleotide are excluded.
4. Remaining candidates are scored based on findings from Doench et al. [8]. Scores and weights are listed in Table 1. For each listed nucleotide position relative to the PAM sequence, the nucleotide score is multiplied by the weight. The guide RNA score is the sum of scores for positions  $-21$ ,  $-20$ ,  $1$ , and  $4$ .
5. Candidate guide RNAs with highest scores are used for further work (*see Note 2*).

#### 3.1.2 Plasmid Preparation

1. Design forward oligonucleotides for guide RNA amplification (*see Subheading 2.1.2*) and cloning into pgRNA-humanized lentiviral vector (Addgene plasmid no. 44248). The DNA

**Table 1**  
**Scores and weights for calculation of the guide RNA score**

	N	N	PAM			N
			N	G	G	
<i>Position</i>	-21	-20	1	2	3	4
<i>Weight</i>	0.1	0.2	0.5			0.2
<i>Nucleotide scores</i>						
<i>C</i>	0	-2	1			0
<i>G</i>	0	-2	-0.5			1
<i>A</i>	0	1	0			0.5
<i>T</i>	-1	1	-1			0.5

targeting region is represented by Ns (*see Note 3*). Different plasmid vectors can be used for guide RNA cloning; however the protocols described below should be adjusted accordingly.

2. PCR amplify individual guide RNAs using the appropriate forward oligonucleotide and reverse oligonucleotide for guide RNA amplification (*see Subheading 2.1.2*).
3. Separate the PCR product on a 2% (w/v) agarose gel in 1× TAE buffer for 30 min at 120 V. The correct bands are 154 bp in length.
4. Cut the correct bands out of the gel and use a gel extraction kit to purify the product (*see Note 4*). Measure concentration of the purified PCR product using Nanodrop.
5. Digest the pgRNA-humanized plasmid with BstXI and XhoI restriction enzymes in 1× NEBuffer™ r3.1 (*see Note 5*). 2500 ng of plasmid is used for digest in a total volume of 30 ul. The reaction is performed at 37 °C for at least 2 h. 2 µL of each enzyme is used for digestion.
6. Separate the fragments on a 0.8% (w/v) agarose gel in 1× TAE buffer for 30–45 min at 120 V. Expected lengths are 107 and 8190 bp.
7. Cut the 8190-bp band out of the gel and use a gel extraction kit to purify the product (*see Note 4*). Measure concentration of the digested vector using Nanodrop.
8. Perform the Gibson assembly reaction using the NEBuilder® HiFi DNA Assembly Master Mix (*see Note 6*). Pipette the appropriate volume of the Master Mix into a reaction tube. Add 30 ng of the digested vector and 30 ng of the PCR product (*see Note 7*). Prepare the reaction mixtures on ice.

9. Immediately after adding the PCR product, transfer the reaction mixture to 50 °C for 15 min (*see Note 8*).
10. Transform the entire reaction into a chemically competent *E. coli* cloning strain (*see Note 9*). Competent cells are thawed on ice and the appropriate volume of cells is added to the reactions. The mixture of cells and Gibson assembly reactions is incubated on ice for 10 min. Heat shock is performed at 42 °C for 30 s and mixtures are transferred on ice for additional 2 min. The entire mixture is plated on an LB plate with 100- $\mu$ g/mL carbenicillin (*see Note 10*).
11. Grow cells overnight at 37 °C (*see Note 11*). On the next day, pick colonies into liquid LB media with 100- $\mu$ g/mL carbenicillin (*see Note 10*).
12. Incubate the liquid culture overnight at 37 °C with shaking. Use a plasmid purification kit to isolate the plasmids and confirm correct assembly by Sanger sequencing.

## 3.2 Mammalian Cell Culture

Here we describe the protocols for cultivation, transfection, and stimulation of the HEK293T and Neuro2A cell lines. Any desired cell line may be used; however, cultivation conditions and construct delivery methods may vary. Other construct delivery methods can be used (e.g., plasmid electroporation, mRNA transfection, protein transfection); however, construct preparation, delivery protocols, and delivery conditions should be adjusted according to the method used.

### 3.2.1 Cell Cultivation

The HEK293T and Neuro2A cell lines are cultured at a temperature of 37 °C in an atmosphere with 5% CO<sub>2</sub>. HEK293T cells are grown in DMEM media supplemented with 10% FBS. Neuro2A cells are grown in EMEM supplemented with 10% FBS (*see Note 12*). All protocols involving cell culture should be executed in sterile conditions in a laminar flow hood.

1. Thaw and transfer cell suspension into a 15 mL conical tube. Centrifuge the cells at 300 g for 5 min and remove the supernatant by pipetting.
2. Resuspend cells in 10-mL pre-warmed media and transfer cell suspension to a T-75 culture flask (*see Note 13*). Incubate the cells at 37 °C in an atmosphere with 5% CO<sub>2</sub>. Expand cells to approximately 90% confluence (*see Note 14*). Every 2–3 days, aspirate used media and add fresh media.
3. At 90% confluence, passage the cells. Aspirate media, add 3 mL of trypsin–EDTA solution and incubate at 37 °C for 5 min. Add 6 mL of fresh media, detach cells by pipetting, and transfer cell suspension to a 15-mL conical tube.



4. Spin at 300 g for 5 min, aspirate supernatant, and resuspend cells in 10 mL of fresh media.
5. Count cells using an automated cell counter. Transfer 10  $\mu$ L of cell suspension to a microcentrifuge tube and add 10  $\mu$ L of trypan blue dye. Mix by pipetting, transfer 10  $\mu$ L to a cell counting chamber slide, and count the cells.
6. Transfer approximately  $2 \times 10^6$  cells to a new T-75 culture flask and add fresh media to a total volume of 10 mL. Repeat **steps 2–5**. Use the rest of the cells for seeding into multi-well plates or discard them.
7. Seed 1 mL of cell suspension into each well of a 12-well plate at a density of  $2 \times 10^5$  cells per well. Incubate the cells at 37 °C in an atmosphere with 5% CO<sub>2</sub> for 18–24 h.
8. Assess confluence under an inverted light microscope. At 70–90% confluence, cells are ready for transient transfection.

### 3.2.2 Transient Transfection

Transfection mixtures include combinations of plasmids, diluted in 150-mM NaCl (for HEK293T transfection) or Opti-MEM (for Neuro2A transfection). The designed guide RNA plasmids may be used for efficient constitutive activation of gene expression in combination with Addgene plasmids #135982 (dCas9:nls:10P3) and #135983 (P4:gs10:nls:VPR). For abscisic acid-inducible activation of gene expression, the designed guide RNA molecules should be used in combination with Addgene plasmids #135982 (dCas9:nls:10P3), #135985 (P4:gs10:nls:ABI), and #135988 (PYL:gs10:nls:VPR). Optimal transfection efficiency is achieved when the total DNA amount for 12-well plates is at 1400–1600 ng per well (*see Note 15*). For reliable results, transfections should be performed as at least two independent experiments and include at least two technical replicates in individual experiments. All experiments involving small molecule-inducible systems should include non-stimulated control samples (*see Note 16*).

1. Mix appropriate plasmids at equal molar ratios and dilute in appropriate solution to a total volume of 100  $\mu$ L per sample (*see Note 17*).
2. Dilute the appropriate transfection reagent to a total volume of 100  $\mu$ L per sample. For transfection of HEK293T cells, use 6  $\mu$ L of 7.5-mM PEI stock per 500 ng of DNA, diluted in 150-mM NaCl. For transfection of Neuro2A cells, use the Lipofectamine™ LTX Reagent with PLUS™ Reagent, diluted in Opti-MEM according to the manufacturer's instructions.
3. Add 100- $\mu$ L diluted transfection reagent to 100  $\mu$ L of each plasmid mixture (*see Note 18*). Incubate at room temperature for 20–30 min and gently pipette 200  $\mu$ L of the reagent–plasmid mixture to individual wells in a dropwise fashion.

4. Incubate the cells at 37 °C in an atmosphere with 5% CO<sub>2</sub>. To ensure sufficient expression of the upregulated GOI, the transfected cells should be analyzed 24 h or longer after transfection/stimulation.

### 3.2.3 Small Molecule Stimulation

1. Prepare a 5-mM stock solution of abscisic acid (ABA) in 100% DMSO.
2. Dilute ABA 1000× to a final concentration of 5 μM in an appropriate volume of media (*see Note 19*). For non-stimulated control samples, dilute the same volume of DMSO in media.
3. Remove media from wells with transiently transfected cells by aspiration or pipetting. Carefully pipette fresh media with diluted DMSO (control samples) or ABA into the wells.
4. Incubate cells at 37 °C in an atmosphere with 5% CO<sub>2</sub> for at least 24 h prior to analysis (*see Note 20*).

### 3.3 Analysis of Gene Expression

Here we describe the protocols for determination of mRNA expression levels in HEK293T and Neuro2A cells. Depending on the GOI and the intended application, Western blotting and functional assays for confirmation of protein expression and activity should be performed in addition to mRNA-level determination. Ideally, a full transcriptomic profile should be determined by NGS-based RNA sequencing to ensure that only the selected gene is upregulated.

#### 3.3.1 RNA Extraction

1. Aspirate media from wells.
2. Add 250-μL trypsin to each well and incubate at 37 °C for 5 min.
3. Add 250-μL media and detach cells from the surface by pipetting.
4. Transfer cell suspension to a 1.5-mL microcentrifuge tube and spin at 300 g for 10 min.
5. Aspirate the supernatant by pipetting and resuspend cells in 200-μL PBS.
6. Extract RNA using an RNA extraction kit following protocols provided by the manufacturer (*see Note 21*).
7. Elute RNA (*see Note 22*) and transfer the microcentrifuge tubes on ice immediately.
8. Determine the concentration and purity of RNA using Nano-drop. The ratio of absorbance at 260/280 nm should be above 2.0 for pure RNA.
9. Proceed to the reverse transcription step (*see Subheading 3.3.2*) or store the RNA at -80 °C.

### 3.3.2 Reverse Transcription

1. Perform reverse transcription using a reverse transcription kit following the protocol provided by the manufacturer (*see* **Notes 23** and **24**) in a PCR thermocycler.
2. Proceed to the quantitative PCR step (*see* Subheading **3.3.3**) or store the cDNA at  $-20^{\circ}\text{C}$ .

### 3.3.3 Quantitative PCR

For quantitative PCR we use the LightCycler<sup>®</sup> 480 System (Roche) and related products. Here we describe protocols and materials for the aforementioned system. Other instruments and reagents can be used according to the manufacturers' instructions.

1. Dilute cDNA samples  $10\times$  in ddH<sub>2</sub>O (*see* **Note 25**).
2. Plan out your reactions. For each cDNA sample, GOI and an internal control (ctrl) gene should be amplified (*see* **Note 26**). Each reaction should be performed in two to three technical replicates. For most reliable results, a negative control where ddH<sub>2</sub>O is used instead of the cDNA dilution should be included (*see* **Note 27**).
3. Each reaction contains 5  $\mu\text{L}$  of the cDNA dilution, 1  $\mu\text{L}$  of each specific oligonucleotide (*see* **Note 28**), and 7  $\mu\text{L}$  of the LightCycler<sup>®</sup> 480 SYBR Green I Master reagent (Roche).
4. To minimize pipetting errors, prepare master mixes of oligonucleotides and the LightCycler<sup>®</sup> 480 SYBR Green I Master in a microcentrifuge tube.
5. Dispense master mixes into wells of a white LightCycler<sup>®</sup> 480 Multiwell Plate (Roche). Add 5  $\mu\text{L}$  of the cDNA dilutions.
6. Run quantitative PCR using LightCycler<sup>®</sup> 480 according to the manufacturer's instructions.
7. Relative mRNA levels are determined using the formula  $2^{[(C_p(\text{GOI, mock}) - C_p(\text{ctrl, mock})) - (C_p(\text{goi, sample}) - C_p(\text{ctrl, sample}))]}$  [9].
8. Check the size and intensity of the PCR products by running them on a 2% agarose gel in  $1\times$  TAE buffer at 120 V for 30–45 min (*see* **Note 29**). A single band is expected in the cDNA samples (*see* **Note 30**).

---

## 4 Notes

1. Any other dCas9 DNA-binding domain may be used; however, guide RNA design methods may vary. Other designed DNA-binding domains, such as transcription activator-like effectors (TALEs) or zinc finger proteins (ZFPs), can be used.
2. We recommend testing multiple guide RNAs. If all tested guide RNAs exhibit low efficiency, we recommend multiplexing guide RNAs by pooling the plasmids in equal ratios.

3. For guide RNAs, targeting the template strand, the DNA targeting region is equal to the 5'-3' sequence of the coding strand. For guide RNAs, targeting the coding strand, the DNA targeting region is equal to the reverse complement of the coding strand sequence. The PAM sequence is not included in the guide RNA sequence.
4. We prefer to use the Monarch gel extraction kit (NEB); however, any gel extraction kit can be used.
5. We prefer to use NEB enzymes; however, enzymes from any manufacturer can be used. The reaction buffer should be adjusted accordingly.
6. Always perform a control reaction, where water is used instead of the guide RNA PCR product.
7. Make sure that the PCR product is added last, immediately before transferring the reaction to 50 °C.
8. Reactions can be incubated longer (up to an hour), but not shorter than 15 min.
9. We prefer to use the *E. coli* Mach1 strain (Thermo Fisher); however, other cloning strains can be used.
10. We prefer to use carbenicillin; however, ampicillin could be used for selection instead. If guide RNAs are cloned in a different plasmid vector, make sure the appropriate antibiotic is used.
11. A minimal number of colonies is expected on the control, where water was used instead of the guide RNA PCR product. If the number of colonies on the control plate is similar to the number of colonies on the reaction plates, repeat the cloning process from **step 5**.
12. Media should be stored at 4 °C. Prior to use, pre-heat the media at 37 °C for 30 min.
13. We prefer using T-75 flasks; however any format can be used.
14. To ensure that the cells don't overgrow the surface, check the flask under an inverted light microscope every 24 h.
15. The transfection can be performed at a smaller or larger format; however the transfection protocol should be adjusted and optimized.
16. If two technical replicates are used for stimulated samples, two technical replicates should be used for non-stimulated samples as well. Hence, the transfection should be performed in four technical replicates.
17. To increase the reliability and avoid pipetting of small volumes, plasmid dilutions in 150-mM NaCl or Opti-MEM can be prepared prior to mixing. We use plasmid dilutions at a concentration of 50 ng/μL.

18. Transfection mixtures can be prepared up to 1 day prior to transfection and kept at 4 °C overnight. In this case, make sure the transfection reagent is *not* added until 20–30 min before transfection.
19. The final DMSO concentration in the media shouldn't exceed 0.1%.
20. Depending on the experiment, the incubation time may be longer, but not shorter. Note that for longer experiments, the media with fresh ABA should be replaced every 24 h to avoid small molecule degradation and decrease in transcription factor activity.
21. We prefer to use the High Pure RNA Isolation kit (Roche); however, kits from other manufacturers can be used.
22. To ensure high RNA concentration, we prefer to use lower elution volumes than specified in the manufacturer's instructions (30–50 µL). If the yield is too high to accurately measure the concentration, the RNA can be further diluted in elution buffer.
23. We prefer to use the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific); however, kits from other manufacturers can be used.
24. We perform reverse transcription with random primers; however specific primers can be used.
25. Due to the different expression levels of mammalian genes, optimization of the cDNA amount in the quantitative PCR reaction may be required for amplification of specific products. In this case, prepare serial dilutions of cDNA and run a series of test reactions.
26. Internal control genes include highly expressed housekeeping genes. We prefer to use the GAPDH gene as internal control; however other genes such as SDHA, ACTB, or RRN18S can be used.
27. No band is expected in the ddH<sub>2</sub>O negative control. If bands are present in the negative control, one of the reagents might be contaminated with DNA.
28. Oligonucleotides for amplification of specific gene regions can be designed with online tools. For achieving efficient and specific amplification, testing several different primer combinations might be required. We recommend searching the literature for previously validated oligonucleotide pairs prior to designing new oligonucleotides.
29. Depending on the expression levels of the selected genes, optimization of the load volume may be required.

30. If several bands are present, the quantification of mRNA levels is not reliable, and additional optimization of the procedure might be required. Increasing the primer annealing temperature and/or addition of 2% DMSO may help.

## References

1. Brocken DJW, Tark-Dame M, Dame RT (2018) dCas9: a versatile tool for epigenome editing. *Curr Issues Mol Biol* 26:15–32
2. Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, P R Iyer E, Lin S, Kiani S, Guzman CD, Wiegand DJ, Ter-Ovanesyan D, Braff JL, Davidsohn N, Housden BE, Perrimon N, Weiss R, Aach J, Collins JJ, Church GM (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* 12:326–328
3. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F (2014) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517:583–588
4. Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159:635–646
5. Gonzalez L, Woolfson DN, Alber T (1996) Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat Struct Biol* 3:1011–1018
6. Gradišar H, Jerala R (2011) De novo design of orthogonal peptide pairs forming parallel coiled-coil heterodimers. *J Pept Sci* 17:100–106
7. Lebar T, Lainšček D, Merljak E, Aupič J, Jerala R (2020) A tunable orthogonal coiled-coil interaction toolbox for engineering mammalian cells. *Nat Chem Biol* 16:513–519
8. Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat biotechnol* 32:1262–1267
9. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  method. *Methods* 25:402–408



## Design, Construction, and Validation of Targeted Gene Activation with TREE System in Human Cells

Atsushi Kunii, Takashi Yamamoto, and Tetsushi Sakuma

### Abstract

Genome editing technologies can be diverted into artificial transcription activators. In particular, researchers have improved dCas9-based technologies by tandem-fusing or *trans*-accumulating effector domains. Previously, we developed a hierarchical effector accumulation system named “TREE,” enabling robust activation of target genes even when strongly silenced. In this chapter, we describe our protocol to design, construct, and validate the TREE-mediated target gene activation in cultured human cells.

**Key words** CRISPR–Cas9, dCas9, Transcriptional activation, Reporter assay, Real-time PCR, ddPCR

---

### 1 Introduction

Genome editing tools, such as zinc finger nuclease (ZFN), transcription activator-like effector nuclease (TALEN), and clustered regularly interspaced short palindromic repeat (CRISPR)–CRISPR-associated protein 9 (Cas9), can be diverted into platforms to recruit or accumulate effector domains of interest [1–4]. Of these, artificial transcription activators are useful for various applications, including gene function analysis, direct reprogramming of stem cells, and treating or modeling diseases [5–7].

The “first generation” of dCas9-based gene activator, namely, dCas9–VP64, reportedly often shows insufficient reactivation of strongly silenced genes [3, 8]. To solve this, researchers have developed “second-generation” systems, such as VPR, SAM, and SunTag, by fusing or accumulating multiple activation domains, although the most effective system among them depended on the tested species, cell line, and target gene [8–11]. We previously constructed a novel platform named “three-component repurposed technology for enhanced expression” (TREE), in which MS2 coat protein (MCP)-modified SunTag (22sTag) fusion





11. LB liquid medium (10-g/L tryptone, 5-g/L yeast extract, and 10-g/L NaCl in ddH<sub>2</sub>O) supplemented with 100-μg/mL ampicillin, 100-μg/mL carbenicillin, or 50-μg/mL spectinomycin (*see* **Notes 1** and **2**).
12. 37 °C microbiological incubator.
13. 37 °C shaker.
14. Plasmid miniprep kit.
15. Nucleic acid quantification system.
16. *Bp*I restriction enzyme.
17. 1× TAE electrophoresis buffer (40-mM Tris-acetate, 1-mM EDTA).
18. 1–3 w/v% agarose gel in 1× TAE.
19. Agarose gel electrophoresis system.
20. *Eco*31I or *Bsa*I-HFv2 restriction enzyme (optional; *see* **Note 1**).
21. 1 w/v% 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-Gal)/5-mM isopropyl-β-D-thiogalactopyranoside (IPTG) solution (optional; *see* **Note 1**).
22. PCR kit for colony PCR (optional; *see* **Note 1**).
23. Primers for colony PCR (optional; *see* **Note 1**).  
 Forward (CRISPR-STEP2-F): 5'-gccttttgctggccttttgctc-3'  
 Reverse (CRISPR-STEP2-R): 5'-cgggccattaccgtaagttatgt  
 aacg-3'

## **2.2 Cell Culture and Transfection**

1. Cultured human cells.
2. CO<sub>2</sub> incubator.
3. Culture medium: Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10 v/v% fetal bovine serum (FBS), 1× penicillin–streptomycin, and 1× nonessential amino acids (NEAA).
4. 100-mm tissue-coated cell culture dish.
5. Opti-MEM (Thermo Fisher Scientific) (*see* **Note 3**).
6. 0.25 w/v% trypsin–EDTA (*see* **Note 4**).
7. 96-well poly-L-lysine (PLL)-coated cell culture plate.
8. Lipofection reagent.

## **2.3 Luciferase Reporter Assay**

1. Firefly luciferase (FLuc)-cloned plasmid.
2. Human genomic DNA.
3. Gene-specific primers.
4. High-fidelity PCR kit.
5. Thermal cycler.

6. 1–3 w/v% agarose gel in 1× TAE buffer.
7. 1× TAE electrophoresis buffer.
8. Agarose gel electrophoresis system.
9. Gel-extraction kit.
10. In-Fusion HD Cloning Kit (Takara Bio).
11. XL10-Gold chemically competent bacterial cells.
12. 42 °C block incubator or water bath.
13. LB agar supplemented with 100-µg/mL ampicillin.
14. LB liquid medium supplemented with 100-µg/mL ampicillin.
15. 37 °C microbiological incubator.
16. 37 °C shaker.
17. Plasmid miniprep kit.
18. Nucleic acid quantification system.
19. *Renilla* luciferase (*RLuc*) expression vector.
20. Dual-Glo Luciferase Assay System (Promega).
21. 96-well white microplate.
22. Plate reader.

#### **2.4 cDNA Synthesis**

1. Cell lysis and RNA purification kit.
2. Nucleic acid quantification system.
3. Reverse transcription kit.

#### **2.5 Quantification of mRNA Level on Real-Time PCR System**

1. 2× SYBR<sup>®</sup> qPCR Mix.
2. 50× ROX Reference Dye.
3. Gene-specific primers.
4. PCR tube strips or PCR plate for real-time PCR.
5. Real-time PCR system.
6. Laptop with analysis software.

#### **2.6 Quantification of mRNA Level (ddPCR with Probe Method)**

1. QX200 Droplet Generator (Bio-Rad Laboratories).
2. 2× ddPCR Supermix for Probes (No dUTP) (Bio-Rad Laboratories).
3. 2× ddPCR Buffer Control Kit for Probes (Bio-Rad Laboratories).
4. 20× TaqMan Gene Expression Assay, FAM (for target gene) (Thermo Fisher Scientific).
5. 20× TaqMan Gene Expression Assay, VIC (for reference gene) (Thermo Fisher Scientific).
6. Droplet Generation Oil for Probes (Bio-Rad Laboratories).

7. DG8 Cartridge (Bio-Rad Laboratories).
8. DG8 Gasket (Bio-Rad Laboratories).
9. PCR plate.
10. Aluminum PCR sealing film.
11. PCR plate sealer.
12. Thermal cycler.
13. QX200 Droplet Generator (Bio-Rad Laboratories).
14. Droplet Reader Oil (Bio-Rad Laboratories).
15. Laptop with QuntaSoft Software.

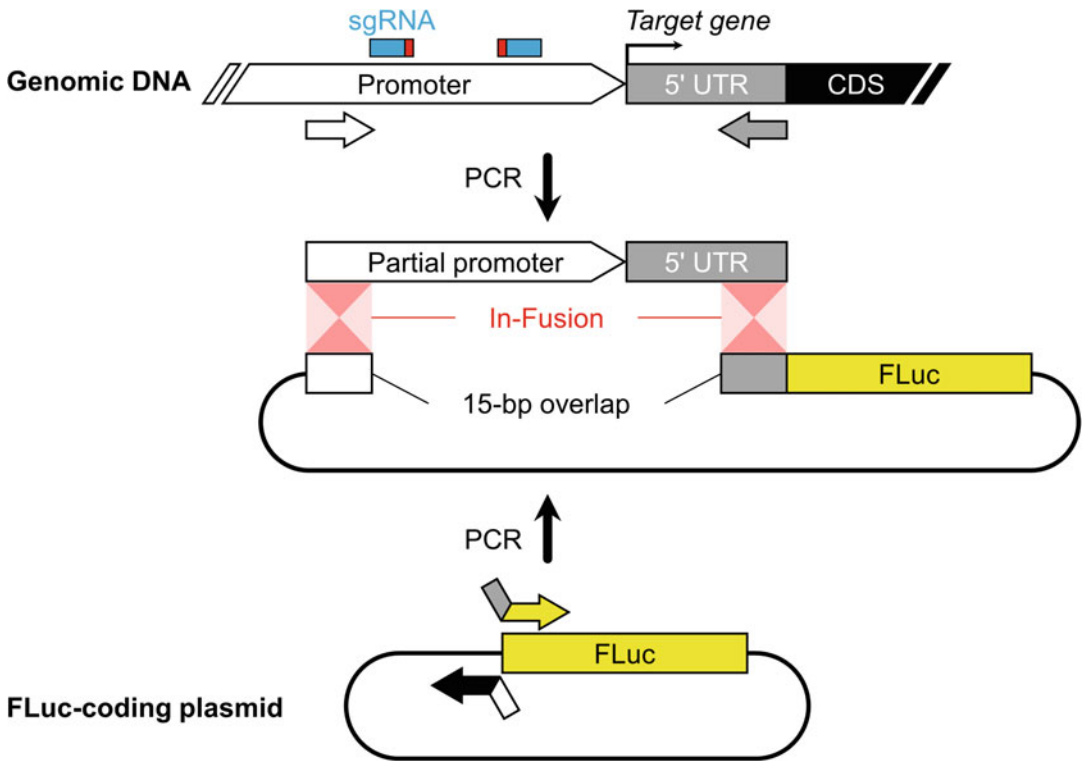
---

### 3 Methods

#### 3.1 *sgRNA Design and Vector Construction*

For transcriptional activation, sgRNAs are designed on the promoter region of their target gene. In this protocol, we describe the procedures to design *SpCas9* sgRNAs and construct an all-in-one CRISPR vector. For MCP-22sTag and scFv-effector expression vector, refer to Kunii et al. [12].

1. Obtain the sequence information of the promoter and 5'-UTR of the target gene from a website such as NCBI (<https://www.ncbi.nlm.nih.gov/>).
2. Find 5'-N<sub>21</sub>GG-3' sequences as sgRNA target sites among the promoter sequence, especially between -500 and -50 bp from the transcription start site (TSS) [15]. CRISPRdirect (<https://crispr.dbcls.jp/>) detects all considerable targets within the submitted sequence [16]. COSMID (<https://crispr.bme.gatech.edu/>) lists potential off-target sites and possibility scores of each submitted sgRNA sequence [17].
3. Modify the all-in-one CRISPR/Cas9 vector system to express dCas9-VP64 and sgRNA-MS2 [12]. Briefly, replace the sequences of normal sgRNA and Cas9 with those of sgRNA-MS2 and dCas9, respectively. Then, insert the peptide linker and VP64 sequences at the downstream of dCas9.
4. Insert sgRNA template oligo DNAs into pX330A-dCas9-VP64, in accordance with our previously reported protocol [14]. For the use of multiple sgRNAs, insert the first template into pX330A-MS2-1 × *n*-dCas9-VP64 (“*n*” represents total number of sgRNAs; 2–7) and other template(s) into pX330S-MS2-(2–7) vectors.
5. Integrate the sgRNA expression cassettes into a single pX330A-dCas9-VP64 plasmid by Golden Gate Assembly, in accordance with our previously reported protocol [14].



**Fig. 2** Schematic illustration of the construction of reporter vector by In-Fusion cloning

### 3.2 Design and Construction of Reporter Plasmid

Luciferase reporter assay, in which the activity can be analyzed by simple procedures the day after transfection, is useful for basic validation of the system and sgRNA. In this protocol, we describe the procedures to construct a reporter plasmid, in which FLuc is expressed under control of a partial promoter of the target gene, by In-Fusion cloning (Fig. 2).

1. Amplify the partial promoter region, containing all sgRNA target sites, and 5'-UTR, from genomic DNA by PCR (Fig. 2, top).
2. Amplify the FLuc-cloned plasmid by inverse PCR, to insert the promoter sequence. Add a 15-nt homologous sequence at the 5' end of each primer (Fig. 2, bottom).
3. Run the PCR products on an agarose gel and check their length by ethidium bromide staining.
4. Gel-extract DNA from the band with an appropriate size, for example, using Wizard SV Gel and PCR Cleanup System (Promega).
5. Mix the obtained DNA extract and 5× In-Fusion mix (e.g., using the following example), and incubate for 15 min at 50 °C using a thermal cyclor.

Component	Amount ( $\mu\text{L}$ )	Final concentration
5 $\times$ In-Fusion mix	0.4	1 $\times$
Insert extract	0.8	
Vector extract	0.8	
Total	2	

6. Add 2  $\mu\text{L}$  of the In-Fusion product to 20  $\mu\text{L}$  of the chemically competent bacterial cells, and incubate on ice for 10 min.
7. Incubate the transformant at 42  $^{\circ}\text{C}$  for 30 s using a block incubator or water bath.
8. Immediately cool on ice for 2 min.
9. Plate the transformant onto LB agar + ampicillin and incubate at 37  $^{\circ}\text{C}$  for 14–16 h.
10. Pick up some bacterial colonies with microtips, stick the LB + ampicillin “replica plate” with them, and put them into the colony PCR mix. For colony PCR, design a primer pair to amplify a region of several hundred base pairs across the vector and insert sides. The tables shown below present examples of colony PCR components and thermal cycling programs using SapphireAmp Fast PCR Master Mix.

Component	Amount ( $\mu\text{L}$ )	Final concentration
2 $\times$ SapphireAmp Fast PCR Master Mix	4	1 $\times$
10- $\mu\text{M}$ Primer-Forward	0.16	0.2 $\mu\text{M}$
10- $\mu\text{M}$ Primer-Reverse	0.16	0.2 $\mu\text{M}$
ddH <sub>2</sub> O	3.68	
Total	8	

Step	Temperature and duration
1	94 $^{\circ}\text{C}$ , 2 min
2	98 $^{\circ}\text{C}$ , 10 s
3	60 $^{\circ}\text{C}$ , 5 s
4	72 $^{\circ}\text{C}$ , 30 s
5	Go to step 2, 29 times
6	4 $^{\circ}\text{C}$ , $\infty$

11. Incubate the replica plate at 37 °C until the clones grow visibly.
12. Run the aliquot (~4 µL) of colony PCR products on 2–3% agarose gel and check the amplicon size by ethidium bromide staining.
13. Pick up the colonies, tested positively by colony PCR, from the replica plate. Put them into 3 mL of LB liquid medium + ampicillin and shake overnight at 37 °C.
14. Extract the plasmid from the culture using plasmid miniprep kit.
15. Confirm the sequence along the promoter, 5'-UTR, and Luciferase CDS by Sanger sequencing analysis.

### 3.3 Lipofection

In this protocol, we describe the procedures to transfect cultured human cells with plasmid vectors by a method called “reverse transfection,” using Lipofectamine LTX Reagent (Thermo Fisher Scientific) as an example. The term “reverse” describes the order of materials to be put into wells. In generalized “forward transfection,” the cells are seeded in advance, while in reverse transfection they are added after generating lipoplex. In our experience, reverse transfection with a 96-well plate showed high transfection efficiency and high cell viability, using multiple cell lines such as HEK293T, HCT116, and MCF-7 (*see Note 5*).

1. Maintain the cultured human cells in culture medium using a 100-mm tissue culture dish.
2. Prepare the appropriate concentration of cell suspension (*see Note 6*).
3. Mix 25 µL of Opti-MEM and 6 µL of plasmid vector in a 96-well culture plate (*see Notes 3 and 7*).
4. Add 25 µL of Opti-MEM and an appropriate amount of Lipofectamine LTX (*see Notes 3 and 7*).
5. Incubate at room temperature for 30 min; then add 100 µL of cell suspension.
6. Incubate the cells at 37 °C and 5% CO<sub>2</sub>.

### 3.4 Luciferase Assay

In this protocol, we describe the procedures to measure the luciferase activity of the reporter.

1. Prior to the assay, resolve Dual-Glo Luciferase Substrate in Dual-Glo Luciferase Substrate Buffer, and thaw Dual-Glo Stop & Glo Buffer.
2. Transfect  $6 \times 10^4$  cultured cells with the plasmids as follows: a total of 25–100 ng of system-expression vectors; 100 ng of the FLuc reporter vector; and 20 ng of RLuc expression vector for reference.

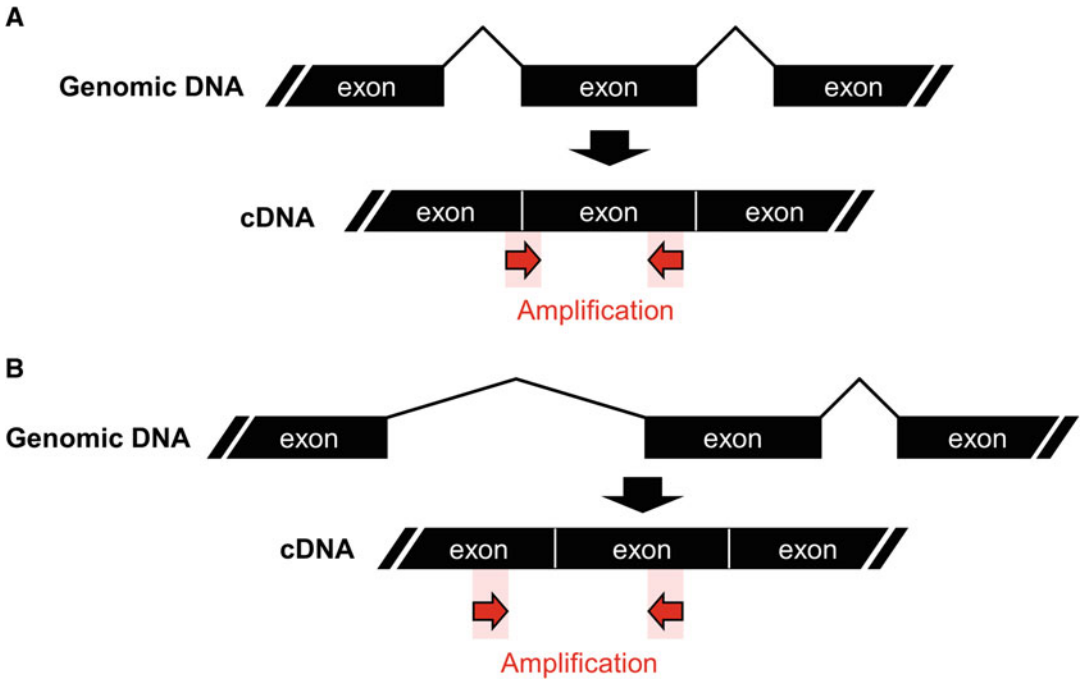
3. Incubate the cells at 37 °C and 5% CO<sub>2</sub> for 24 h.
4. Remove 75 µL of the medium.
5. Add 75 µL of Dual-Glo Luciferase Substrate and pipette to lyse the cells.
6. Incubate the lysate at room temperature for 10 min.
7. Transfer the lysate into a 96-well white microplate.
8. Measure the luminescence by FLuc reaction on a plate reader (*see Note 8*).
9. Resolve 100× Dual-Glo Stop & Glo Substrate in Dual-Glo Stop & Glo Buffer at 1× concentration.
10. Add 75 µL of 1× Dual-Glo Stop & Glo Substrate.
11. Incubate at room temperature for 10 min.
12. Measure the luminescence by the RLuc reaction on a plate reader.
13. Normalize the activity of FLuc by that of RLuc.

### 3.5 cDNA Preparation

In this protocol, we describe the procedures to synthesize cDNA for gene expression analysis at the mRNA level.

1. Transfect  $3 \times 10^4$  cultured cells with a total of 100–200 ng of system-expression vectors.
2. Incubate the cells at 37 °C and 5% CO<sub>2</sub> for 48 h.
3. Extract total RNA from the cells using a kit such as NucleoSpin RNA Plus (Takara Bio) (*see Note 9*).
4. Quantify the concentration of RNA.
5. Synthesize cDNA from the RNA. The following table shows an example of using High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) (*see Note 9*).

Component	Amount	Final concentration
Reverse Transcriptase	1 µL	
10× RT Buffer	2 µL	1×
10× Random Primers	2 µL	1×
100-mM dNTPs	0.8 µL	4 mM
Template RNA	× µL (up to 2 µg)	
Nuclease-free water	Add to 20 µL	
Total	20 µL	



**Fig. 3** Schematic illustration of the design of real-time PCR primer. (a) 3' end of the left primer anneals to the next exon. (b) The amplified region spans exons separated by a long intron

### 3.6 Real-Time PCR with SYBR Green

In this protocol, we describe the procedures to quantify the mRNA expression level of a target gene using intercalator on a real-time PCR system (*see Note 10*).

1. Design some candidate primer pairs on cDNA of the target and reference genes (*see Note 11*). To avoid detection of unremoved genomic DNA, it is desirable that one of a pair of primers should be designed across two exons (Fig. 3a), or the amplicon spans two exons separated by a long intron (Fig. 3b).
2. Select the primer pairs and optimize the annealing temperature, to avoid primer dimers and obtain high amplification efficiency (*see Note 12*). Ideally, use the same reagents and machines, if practical.
3. Prepare the templates for standard curves by two to tenfold serial dilution (*see Note 13*). If possible, establish more than five points with three replicates.
4. Prepare the real-time PCR mixture. The following table shows an example of using KOD SYBR qPCR Mix (Toyobo) (*see Note 14*).



Component	Amount	Final concentration
2× KOD SYBR qPCR mix	10 $\mu$ L	1×
50× ROX Reference Dye	0.4 $\mu$ L	1×
10- $\mu$ M primer forward	0.4 $\mu$ L	0.2 $\mu$ M
10- $\mu$ M primer reverse	0.4 $\mu$ L	0.2 $\mu$ M
cDNA	$\times$ $\mu$ L	
ddH <sub>2</sub> O	Add to 20 $\mu$ L	
Total	20 $\mu$ L	

- Run the following program on a real-time PCR system. The following program is an example of using the StepOne/StepOnePlus Real-Time PCR System (Thermo Fisher Scientific), which performs real-time PCR and melting curve analysis.

Step	Temperature and duration	Measurement
1	98 °C, 2 min	–
2	98 °C, 10 s	–
3	60 °C, 10 s	–
4	68 °C, 30 s	●
5	Go to step 2, 39 times	–
6	95 °C, 15 s	–
7	60 °C, 1 min	–
8	Slope to 95 °C (+0.3 °C/s)	●
9	95 °C, 15 s	●
10	4 °C, $\infty$	–

- Confirm that the melting curves show single peaks.
- Calculate the relative expression level in accordance with the standard curves.

### 3.7 ddPCR with TaqMan Probes

In this protocol, we describe the procedures to quantify the mRNA expression level of a target gene, using TaqMan probes on the ddPCR system (*see Note 15*). Despite the ddPCR enabling absolute quantification (i.e., analyzing the copy number of target DNA per unit volume), we recommend normalization using a reference gene.

- Prepare the ddPCR mixture as follows.

Component	Amount	Final concentration
2× ddPCR Supermix for Probes (no dUTP)	10 μL	1×
20× TaqMan Gene Expression Assay (FAM)	1 μL	1×
20× TaqMan Gene Expression Assay (VIC)	1 μL	1×
cDNA	× μL	
ddH <sub>2</sub> O	Add to 20 μL	
Total	20 μL	

- Load 20 μL of ddPCR mixtures into the sample-loading wells of a DG8 Cartridge (*see Note 16*). If the number of samples is not a multiple of 8, fill the empty wells with the following mixture instead:

Component	Amount (μL)	Final concentration
2× ddPCR Buffer Control for Probes	10	1×
ddH <sub>2</sub> O	10	
Total	20	

- Load 70 μL of Droplet Generation Oil for Probes into the oil-loading wells of a DG8 Cartridge (*see Note 17*).
- Cover the DG8 Cartridge with a DG8 Gasket.
- Set the DG8 Cartridge into the Droplet Generator.
- Confirm that the droplets have been successfully generated.
- Transfer the droplets into a 96-well PCR plate.
- Seal the PCR plate with aluminum film at 180 °C for 5 s on a PCR plate sealer.
- Perform PCR amplification accompanied by TaqMan probe digestion on C1000 Touch Thermal Cycler (Bio-Rad Laboratories), in accordance with the program shown below (*see Note 18*). The program should be run at a ramp rate of 2 °C/s.

Step	Temperature and duration
1	95 °C, 10 min
2	94 °C, 30 s

(continued)

Step	Temperature and duration
3	58 °C, 1 min
4	Go to step 2, 39 times
5	98 °C, 10 min
6	4 °C, ∞

10. Analyze the droplets on the Droplet Reader.
11. The copy numbers of the target and reference genes are automatically calculated. If necessary, align the threshold of signal intensity.

---

## 4 Notes

1. Spectinomycin, *Eco31I* or *BsaI*-HFv2, X-Gal/IPTG solution, and colony PCR reagents are required for the use of multiple sgRNAs.
2. In our experience, amplification of CRISPR vector and scFv-effector expression vector often results in a low amount of plasmid. The use of carbenicillin, instead of ampicillin, might improve the yield.
3. Serum-free medium is required for lipofection. Instead of Opti-MEM, non-supplemented DMEM can also be used.
4. TrypLE reagent (Thermo Fisher Scientific) can also be used. In our experience, TrypLE requires 1.5–2 times longer incubation than trypsin-EDTA.
5. In our experience, high reverse-transfection efficiency with high viability is not observed upon the use of 24-, 12-, and 6-well plates. To collect a larger amount of RNA, transfer the cells into a 96-well plate, and then transfer them into a larger well the next day, or pool the cells from multiple wells as a single sample.
6. This step can be performed while waiting for lipoplex formation for 30 min (**step 5** of Subheading **3.3**).
7. The amount of Lipofectamine LTX and the total mass of plasmid vectors should be optimized depending on the cell line, to achieve both high transfection efficiency and high cell viability.
8. In our experience, some reporter constructs show high basal activity, which makes it difficult to assess the activity of systems. In that case, another cell line or minimal CMV reporter should be used, as shown in Fig. 4.



**Fig. 4** Schematic illustration of the reporter vector based on minimal CMV promoter (minCMV). dCas9-based activator binds to the target region and activates the minCMV

9. Some kits can synthesize cDNA from cell lysate directly, although this results in a low concentration owing to the limit of the amount of lysate added to the reaction mixture of reverse transcription.
10. TaqMan Gene Expression Assays, which enable quantification of multiple genes in the same tube, are also useful for real-time PCR.
11. *GAPDH* is one of the most frequently used reference genes [3, 10]. As other examples, *ACTB*, *B2M*, *GUSB*, *HPRT1*, and *PPIA* have been used in studies on dCas9-based transcriptional regulation [8, 9, 18–20].
12. Prior to purchasing the candidate primer pairs, in silico prediction of primer dimer is useful. For example, Multiple Primer Analyzer (<https://www.thermofisher.com/jp/ja/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html>) predicts the base pairing of homodimer and/or heterodimer.
13. cDNA from highly expressed cells or PCR products can be used for standard curves.
14. On some real-time PCR instruments, ROX Reference Dye should be used at  $0.1 \times$  concentration.
15. Intercalator method with EvaGreen is also available. In that approach, the required ddPCR Master Mix, Buffer Control, and Droplet Generation Oil differ from those of the probe method.
16. Avoid generating bubbles. It is not necessary to load all of the 20  $\mu\text{L}$  of reaction mix because the copy number is calculated based on Poisson distribution with the ratio of positive/negative droplets.
17. This step should be performed quickly because the oil gradually volatilizes.
18. It is recommended that the temperatures of annealing and extension, which result in sufficient separation of the signal

intensities between positive and negative droplets, be optimized, although it seems that most commercially available TaqMan Gene Expression Assays can be used at 58 °C, in our experience.

---

## Acknowledgments

We thank Edanz (<https://jp.edanz.com/ac>) for editing a draft of this manuscript. This work was funded by Grant-in-Aid for JSPS Fellows (grant# 20J13292) to A.K. and Grant-in-Aid for Early-Career Scientists (grant# 19K16111) to T.S.

## References

- Kim YG, Cha J, Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* 93:1156–1160. <https://doi.org/10.1073/pnas.93.3.1156>
- Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, Meng X, Paschon DE, Leung E, Hinkley SJ, Dulay GP, Hua KL, Ankoudinova I, Cost GJ, Urnov FD, Zhang HS, Holmes MC, Zhang L, Gregory PD, Rebar EJ (2011) A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* 29:143–150. <https://doi.org/10.1038/nbt.1755>
- Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154:442–451. <https://doi.org/10.1016/j.cell.2013.06.044>
- Kunii A, Yamamoto T, Sakuma T (2020) Various strategies of effector accumulation to improve the efficiency of genome editing and derivative methodologies. *In Vitro Cell Dev Biol Anim* 56:359–366. <https://doi.org/10.1007/s11626-020-00469-y>
- Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS, Kampmann M, Weissman JS (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159:647–661. <https://doi.org/10.1016/j.cell.2014.09.029>
- Chakraborty S, Ji H, Kabadi AM, Gersbach CA, Christoforou N, Leong KW (2014) A CRISPR/Cas9-based system for reprogramming cell lineage specification. *Stem Cell Rep* 3:940–947. <https://doi.org/10.1016/j.stemcr.2014.09.013>
- Garcia-Bloj B, Moses C, Sgro A, Plani-Lam J, Arooj M, Duffy C, Thiruvengadam S, Sorolla A, Rashwan R, Mancera RL, Leisewitz A, Swift-Scanlan T, Corvalan AH, Blancafort P (2016) Waking up dormant tumor suppressor genes with zinc fingers, TALEs and the CRISPR/dCas9 system. *Oncotarget* 7:60535–60554. <https://doi.org/10.18632/oncotarget.11142>
- Chavez A, Tuttle M, Pruitt BW, Ewen-Campen B, Chari R, Ter-Ovanesyan D, Haque SJ, Cecchi RJ, Kowal EJK, Buchthal J, Housden BE, Perrimon N, Collins JJ, Church G (2016) Comparison of Cas9 activators in multiple species. *Nat Methods* 13:563–567. <https://doi.org/10.1038/nmeth.3871>
- Chavez A, Scheiman J, Vora S, Pruitt B, Tuttle M, Iyer E, Kiani S, Guzman C, Wiegand D, Ter-Ovanesyan D, Braff J, Davidsohn N, Weiss R, Aach J, Collins C, Church G (2015) Highly-efficient Cas9-mediated transcriptional programming. *Nat Methods* 12:326–328. <https://doi.org/10.1101/012880>
- Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517:583–588. <https://doi.org/10.1038/nature14136>
- Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159:635–646. <https://doi.org/10.1016/j.cell.2014.09.039>
- Kunii A, Hara Y, Takenaga M, Hattori N, Fukazawa T, Ushijima T, Yamamoto T, Sakuma T (2018) Three-component repurposed technology for enhanced expression:

- highly accumulable transcriptional activators via branched tag arrays. *CRISPR J* 1:337–347. <https://doi.org/10.1089/crispr.2018.0009>
13. Sakuma T, Nishikawa A, Kume S, Chayama K, Yamamoto T (2014) Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci Rep* 4: 5400. <https://doi.org/10.1038/srep05400>
  14. Sakuma T, Sakamoto T, Yamamoto T (2017) All-in-one CRISPR-Cas9/Foki-dCas9 vector-mediated multiplex genome engineering in cultured cells. *Methods Mol Biol* 1498:41–56. [https://doi.org/10.1007/978-1-4939-6472-7\\_4](https://doi.org/10.1007/978-1-4939-6472-7_4)
  15. Mohr SE, Hu Y, Ewen-campen B, Housden BE, Viswanatha R, Perrimon N (2016) CRISPR guide RNA design for research applications. *FEBS J* 283:3232–3238. <https://doi.org/10.1111/febs.13777>
  16. Naito Y, Hino K, Bono H, Ui-Tei K (2015) CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics* 31:1120–1123. <https://doi.org/10.1093/BIOINFORMATICS/BTU743>
  17. Cradick TJ, Qiu P, Lee CM, Fine EJ, Bao G (2014) COSMID: a web-based tool for identifying and validating CRISPR/Cas off-target sites. *Mol Ther* 3:e214. <https://doi.org/10.1038/MTNA.2014.64>
  18. Amabile A, Migliara A, Capasso P, Biffi M, Cittaro D, Naldini L, Lombardo A (2016) Inheritable silencing of endogenous genes by hit-and-run targeted epigenetic editing. *Cell* 167:219–232.e14. <https://doi.org/10.1016/J.CELL.2016.09.006>
  19. Moses C, Nugent F, Waryah CB, Garcia-Bloj B, Harvey AR, Blancafort P (2019) Activating PTEN tumor suppressor expression with the CRISPR/dCas9 system. *Mol Ther* 14:287–300. <https://doi.org/10.1016/J.OMTN.2018.12.003>
  20. Di Chen C, Zeldich E, Li Y, Yuste A, Abraham CR (2018) Activation of the anti-aging and cognition-enhancing gene *Klotho* by CRISPR-dCas9 transcriptional effector complex. *J Mol Neurosci* 64:175. <https://doi.org/10.1007/S12031-017-1011-0>

# **Part IV**

## **Manipulation of the Epigenome: Inducible System for Epigenome Editing**



## A Split CRISPR–Cpf1 Platform for Inducible Gene Activation

Takahiro Otabe, Yuta Nihongaki, and Moritoshi Sato

### Abstract

The CRISPR–Cpf1 also known as Cas12a is an RNA-guided endonuclease similar to CRISPR–Cas9. Combining the CRISPR–Cpf1 with optogenetics technology, we have engineered photoactivatable Cpf1 (paCpf1) to precisely control the genome sequence in a spatiotemporal manner. We also identified spontaneously activated split Cpf1 and thereby developed a potent dCpf1 split activator, which has the potential to activate endogenous target genes. Here we describe a method for optogenetic endogenous genome editing using paCpf1 in mammalian cells. Furthermore, we show a method for endogenous gene activation using dCpf1 split activator in mammalian cells and mice.

**Key words** Genome editing, CRISPR–Cpf1 (also known as Cas12a), crRNA, Optogenetics, T7E1, Genomic PCR, NHEJ, Gene activation, Hydrodynamic tail vein injection, In vivo gene activation

---

## 1 Introduction

Genome editing techniques based on targeted endonucleases, such as CRISPR–Cas9, are widely used in both basic sciences and biotechnological application [1, 2]. The targeted endonuclease-mediated DNA double-strand break at the target genomic region is repaired through two major DNA repair pathways: nonhomologous end joining (NHEJ) and homology-directed repair (HDR). NHEJ is an error-prone repair process, which introduces DNA insertions and deletions (indels) at the repaired target site, resulting in gene knockout by frameshift. In contrast, HDR can provide precise DNA repair. By introducing exogenous DNA donors with a point mutation or insertion sequences, genome modification with the desired mutation can be achieved by HDR-mediated genome editing.

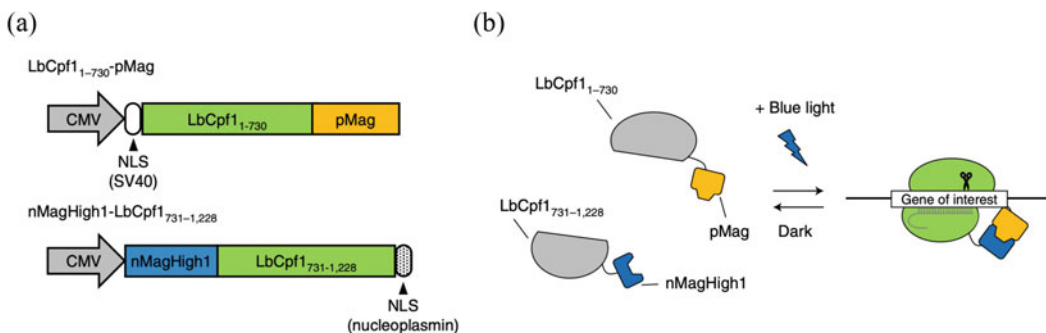
In addition to the most popular Cas9 endonuclease, CRISPR–Cpf1, a class 2 type V CRISPR-Cas RNA-guided endonuclease (known as Cas12a), also provides targeted genome editing with unique properties. First, whereas Cas9 requires both CRISPR RNA (crRNA) and trans-activating crRNA for DNA targeting, Cpf1



does single shorter crRNA. Also, Cpf1 protein can process multiple mature crRNAs from single pre-crRNA transcript, which makes it possible to simultaneously target multiple genes [3, 4]. Unlike Cas9 that introduces genome DNA cleavage next to the G-rich (NGG, N = A, G, C, T) protospacer adjacent motif (PAM) and produces blunt ends, Cpf1 cleaves genomic DNA distal to T-rich (TTTV, V = A, G, T) PAM and results in sticky ends [3].

For further expanding application of Cpf1 technology, many efforts of protein engineering have been converging on Cpf1 to generate engineered Cpf1 variants [5, 6]. We and other groups have developed inducible Cpf1 tools, which enable drug and light control of Cpf1-mediated gene manipulation with spatiotemporal resolution [7–10]. We have engineered photoactivatable Cpf1 (paCpf1), which consists of a pair of split Cpf1 fragments fused with a light-inducible dimerization system, named Magnet system [8, 11]. The split Cpf1 fragments are inactivated in the dark, and upon blue light illumination, the Magnet system is heterodimerized, subsequently resulting in active Cpf1 reconstitution to induce DNA double-strand break targeted by crRNA (Fig. 1a, b). We also have discovered spontaneously activated split Cpf1 and utilize them for developing potent dCpf1 split activator, combining them with the catalytically inactive form of Cpf1 (dCpf1) fused with transcriptional activator domain [8]. We have demonstrated that dCpf1 split activator offers potent gene activation in mammalian cells and mice (Fig. 4a).

In this chapter, we describe the protocol for optogenetic genome editing in mammalian cells using paCpf1. We also describe the protocol for gene activation in mammalian cells and living mice using dCpf1 split activator. We outline the generation of



**Fig. 1** Construction and schematic of the photoactivatable LbCpf1 (paLbCpf1). (a) Constructs for paLbCpf1. N730-fragment and C731-fragment of Cpf1 are fused with the Magnet system (pMag and nMagHigh1) to construct N730-pMag and nMagHigh1-C31, respectively. (b) paLbCpf1 is inactivated in the dark condition, and upon blue light illumination, it is activated as a result of dimerization of the paLbCpf1 fragments by the Magnet system. Active paLbCpf1 complexed with crRNA can bind to the targeted gene of interest (GOI). DNA double-strand break induced by paLbCpf1 is repaired by NHEJ or HDR in mammalian cells. (Reproduced from ref. [8] with permission from Springer Nature)

mammalian expression crRNA and crRNA array. In an optogenetic endogenous genome editing experiment, NHEJ-mediated gene editing is described with T7 endonuclease 1 (T7E1) assay for quantifying the efficiency of genome editing. We also show how to analyze endogenous gene activation in mammalian cells by quantitative rtPCR. Additionally, we introduce in vivo gene activation using dCpf1 split activator, which is detected by a bioluminescent reporter or at the endogenous gene expression level.

## 2 Materials

### 2.1 Media and Buffers

1. Lennox L Broth Base (LB) plus Amp: LB containing 100- $\mu$ g/mL ampicillin.
2. 10 $\times$  annealing buffer: 500-mM Tris-HCl (pH 7.5), 100-mM MgCl<sub>2</sub>, 10-mM dithiothreitol, 1000-mM NaCl.
3. 10 $\times$  NEB2 buffer.
4. 10 $\times$  T4PNK buffer.

### 2.2 Molecular Biology

1. *BsmB* I.
2. T7 endonuclease I.
3. DNA polymerase.
4. DNA ligase.
5. T4PNK.
6. 10-mM ATP.
7. Gel/PCR extraction kit.
8. Plasmid DNA miniprep kit.
9. Genomic DNA extraction kit.
10. RNA extraction kit.
11. Reverse transcriptase kit.
12. *E. coli*: Mach1 competent cell or other suitable competent cells for general cloning.
13. Oligonucleotides for crRNA sequence: 5'-ATAG(N)<sub>23</sub>-3' and 5'-AAAA(N)<sub>23</sub>-3'.
14. Oligonucleotides for crRNA array sequence:
  - 5'-AGATCAAGGAGCGGGAGAAAGGAACGGA-3'
  - 5'-TAGTAGAAATTCGGTTCCTTTCTCCCGCTCCTG-3',
  - 5'-ATTTCTACTAAGTG TAGATAGACAGATATTTGCA TTGAGATAAATTTCTACTAA-3',
  - 5'-ATCTACACTTAGTAGAAATTTATCTCAATGCAAA TATCTGTCTATCTACTACT-3',

5'- GTGTAGATCAGGAGGGTGACTCAGGCTAGCAA  
 ATTTCTACTAAGTGTAGAT-3',  
 5'- AAAAATCTACACTTAGTAGAAATTTGCTAGCCT  
 GAGTCACCCTCCTG-3'

15. TaqMan Gene Expression Assay IDs: *ASCL1*, Hs04187546\_g1; *IL1RN*, Hs00893626\_m1; *HBG1*, Hs00361131\_g1; *GAPDH*, Hs99999905\_m1; *Ascl1*, Mm03058063\_m1; *Gpdx*, Mm99999915\_g1.
16. 24-well clear-walled cell culture plate.
17. 96-well clear-walled cell culture plate.
18. 0.2-mL PCR tube.
19. Milli-Q water.

### 2.3 Plasmids

1. pcDNA 3.1/V5-HisA (Invitrogen).
2. CMV-N730-pMag (N-fragment of paCpfl; *see Note 1*).
3. CMV-nMagHigh1-C731 (C-fragment of paCpfl; *see Note 1*).
4. crRNA vector (*see Note 1*).
5. Pre-crRNA vector (*see Note 1*).
6. pCMV-dCpfl split activator (*see Note 1*).
7. pCMV-dCpfl-VPR (*see Note 1*).
8. GAL4-UAS luciferase reporter (*see Note 1*).

### 2.4 Cell and Transfection Reagent

1. HEK293T.
2. Lipofectamine 3000.

### 2.5 In Vivo Experiment

1. Female BALB/c mice aged 6 weeks.
2. Mouse cylinder fixator.
3. Hydrodynamic delivery solution.
4. 2-mL plastic tube.
5. Needle (27G, S· B, 3/4 inch).
6. 2.5-mL syringe.
7. Warming water.
8. 100-mM D<sub>2</sub>-luciferin: solution in Hank's balanced salt solution (HBSS).
9. Isoflurane.
10. RNAlater solution.
11. Precellys Lysing Kit CK28.

### 2.6 Instruments

1. Thermal cycler.
2. 470-nm ± 20-nm LED light source (CCS Inc.).
3. Agilent 4200 TapeStation (Agilent).

4. Step One Plus system (Thermo Fisher Scientific).
5. Lumazon bioluminescence imager (Nippon Roper).
6. Evolve 512 EMCCD camera (Photometrics).
7. Precellys Evolution homogenizer with Cryolys Evolution cooling system (Bertin Instruments).

---

### 3 Methods

#### 3.1 Generation of CRISPR RNA (crRNA) Expression Vector (Fig. 2)

1. Synthesize a pair of oligonucleotides for the crRNA sequence (*see Note 2*). The DNA oligonucleotides are resuspended at a concentration of 100  $\mu\text{M}$  in Milli-Q water.
2. Mix 1  $\mu\text{L}$  of the top DNA strand and 1  $\mu\text{L}$  of the bottom DNA strand together at 1:1 molar ratio in a 0.2-mL PCR tube along with 1  $\mu\text{L}$  of 10 $\times$  annealing buffer and 7  $\mu\text{L}$  of Milli-Q water to make the final concentration of each oligo to equal 10  $\mu\text{M}$ .
3. Place the PCR tube in a thermal cycler and start the annealing program: 95  $^{\circ}\text{C}$ , 10 min; 90–15  $^{\circ}\text{C}$ , –1  $^{\circ}\text{C}/1$  min.
4. After running the annealing program, store the sample at 4  $^{\circ}\text{C}$ .
5. Digest 1  $\mu\text{g}$  of crRNA plasmid vector with *BsmB* I enzyme at 55  $^{\circ}\text{C}$  for 60 min.
6. Purify the digested crRNA plasmid vector using Gel/PCR extraction kit.
7. DNA ligations are performed by incubating annealed oligos with an appropriately linearized crRNA plasmid vector using a DNA ligation kit.
8. Transform the ligation product into chemically competent *E. coli* cells, and plate onto standard LB plates containing 100- $\mu\text{g}/\text{mL}$  ampicillin.
9. Pick two to three colonies for isolating plasmids by plasmid DNA miniprep kit. Select plasmids with the correct target sequence by Sanger sequencing. Store at –20  $^{\circ}\text{C}$ .

#### 3.2 Generation of crRNA Array Expression Vector

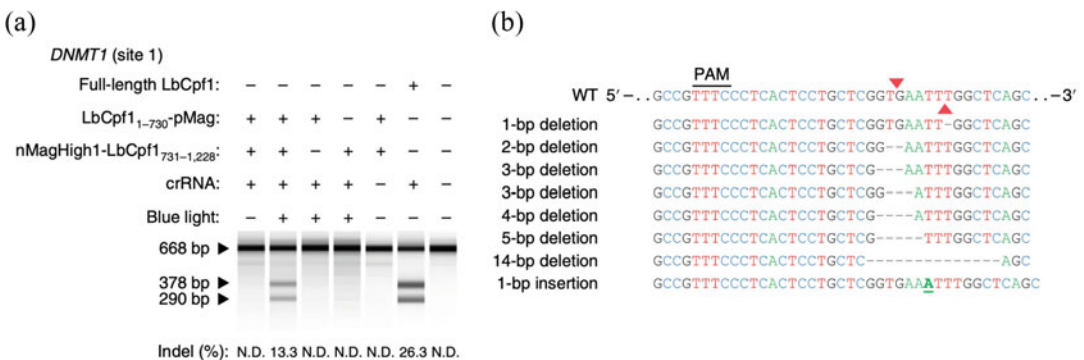
1. For crRNA array construction, synthesize oligonucleotides for one-directional annealing through their sticky ends.
2. Mix oligonucleotides in a 0.2-mL PCR tube along with 2  $\mu\text{L}$  of 10-mM ATP, 2  $\mu\text{L}$  of 10 $\times$  T4PNK buffer, 5 units of T4PNK, and 10  $\mu\text{L}$  of Milli-Q water to make the final concentration of each oligonucleotide to equal 10  $\mu\text{M}$ .
3. Place the PCR tube in a thermal cycler and start the phosphorylated and annealing program: 30 min at 37  $^{\circ}\text{C}$ ; 10 min at 95  $^{\circ}\text{C}$ ; –2.5  $^{\circ}\text{C}$  for 1 min from 90 to 15  $^{\circ}\text{C}$ .
4. Ligation, transformation, and isolation of pre-crRNA plasmid DNA are described above (*see steps 7–9* in Subheading 3.1).



- After 24 h post-transfection, incubate the cells at 37 °C in 5% CO<sub>2</sub> under successive blue light illumination or the dark for an additional 24 h. Blue light illumination is performed using a 470-nm ± 20-nm LED light source. Blue light intensity is 1.0 W/m<sup>2</sup>.
- After incubation for 24 h, isolate genomic DNA using a genomic DNA extraction kit according to the manufacturer's instructions. Isolated genomic DNA can be stored at -20 °C.

### 3.4 T7EI Assay for Quantifying Indel Mutation of Endogenous Genes

- Amplify the genomic region targeted by paCpf1 using DNA polymerase.
- Purify the genomic PCR amplicons using Gel/PCR DNA Extraction Kit following the manufacturer's protocol.
- Combine the PCR amplicons with 1 µL of 10× NEB2 buffer for restriction enzyme and Milli-Q water to make a final volume of 20 µL, and denature and reanneal the PCR products to form heteroduplex DNA (95 °C, 10 min; 90-15 °C, -1 °C/1 min).
- Mix the heteroduplexes samples with 5 units of T7 endonuclease I digestion; incubate at 37 °C for 30 min.
- For the analysis of the DNA fragments, perform agarose gel electrophoresis with Agilent 4200 TapeStation.
- The indel mutation frequency can be quantified from relative band intensities. Use the following formula to calculate the percentage of indel mutation by paCpf1:  $100 \times (1 - (1 - (b + c)/(a + b + c))^{1/2})$ , where *a* is the intensity of the uncleaved PCR product and *b* and *c* are the intensities of each (Fig. 3a).



**Fig. 3** Optogenetic genome editing of a mammalian endogenous gene by paLbCpf1. **(a)** HEK 293T cells were transfected with paLbCpf1 and crRNAs separately targeting *DNMT1* locus. This result shows that paLbCpf1 can induce indel mutations by NHEJ in a blue light-dependent manner. **(b)** Representative sequences of light-induced indel mutation in *DNMT1* locus by LbCpf1. Indel mutation frequencies were determined by the T7EI assay. Arrowheads indicate the genomic PCR product and cleaved product by T7EI. N.D. not detected, WT wild type. (Reproduced from ref. [8] with permission from Springer Nature)

7. Additionally, to identify indel mutation frequencies by paCpf1, genomic PCR samples prepared as described above (*see steps 1–2* in Subheading 3.4) were subcloned into the pcDNA3.1/V5-His A. The plasmids were isolated by plasmid DNA mini-prep kit, and then isolated plasmids were sequenced by Sanger sequencing (Fig. 3b).

### 3.5 Endogenous Gene Activation Experiments

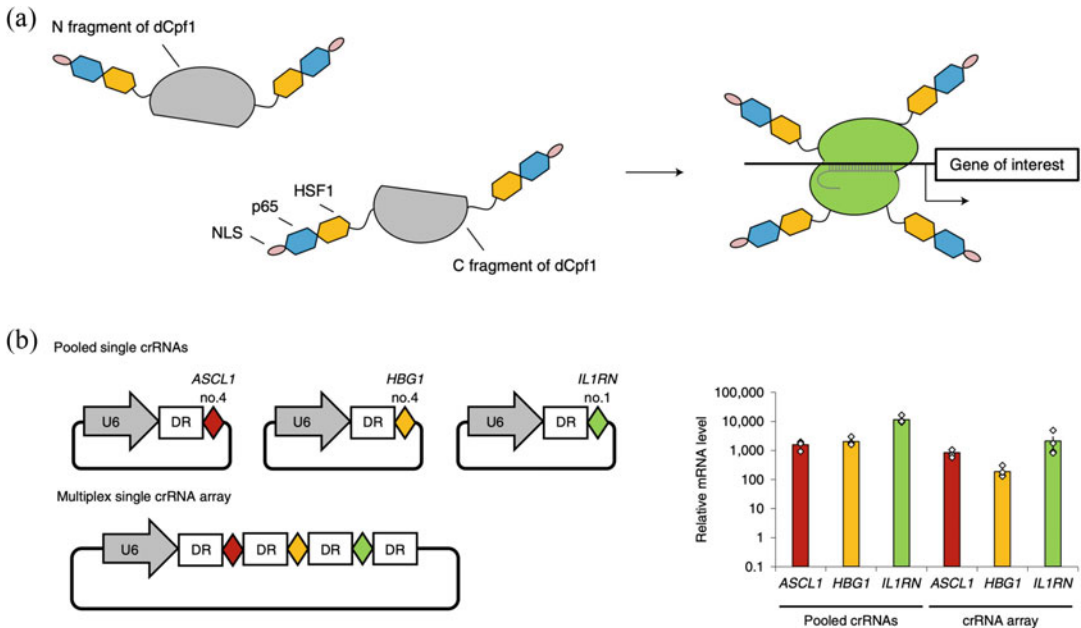
1. To assay the endogenous gene activation, seed HEK293T cells at  $2.0 \times 10^4$  cells/well in a 96-well plate, and culture the cells for 24 h at 37 °C in 5% CO<sub>2</sub>.
2. Transfect the cells with the plasmids encoding dCpf1 split activator and endogenous gene targeting crRNAs or crRNA array at a 1:1 ratio using Lipofectamine 3000 according to the manufacturer's protocols. As a negative control, transfect the empty vector. The total amount of DNA is 0.1 µg/well.
3. After 48 h post-transfection, quantitative real-time PCR (rtPCR) analysis is performed as described below.

### 3.6 Quantitative rtPCR Analysis

1. After 48 h post-transfection, extract the total RNA using an RNA extraction kit following the manufacturer's protocols.
2. Prepare the cDNA using a reverse transcriptase kit following the manufacturer's protocols.
3. For the analysis of the endogenous gene expression level, perform quantitative rtPCR using a prepared cDNA and the Step One Plus system and TaqMan Gene Expression Assay. To quantify the endogenous gene expression level of the target gene, use *GAPDH* gene as a control for endogenous gene expression (Fig. 4b).
4. To calculate relative mRNA level, use a negative control transfected with empty vector by the  $\Delta\Delta C_t$  method. TaqMan Gene Expression Assay IDs are as follows: *ASCL1*, Hs04187546\_g1; *ILIRN*, Hs00893626\_m1; *HBG1*, Hs00361131\_g1; and *GAPDH*, Hs99999905\_m1.

### 3.7 In Vivo Luciferase Reporter Activation Experiment

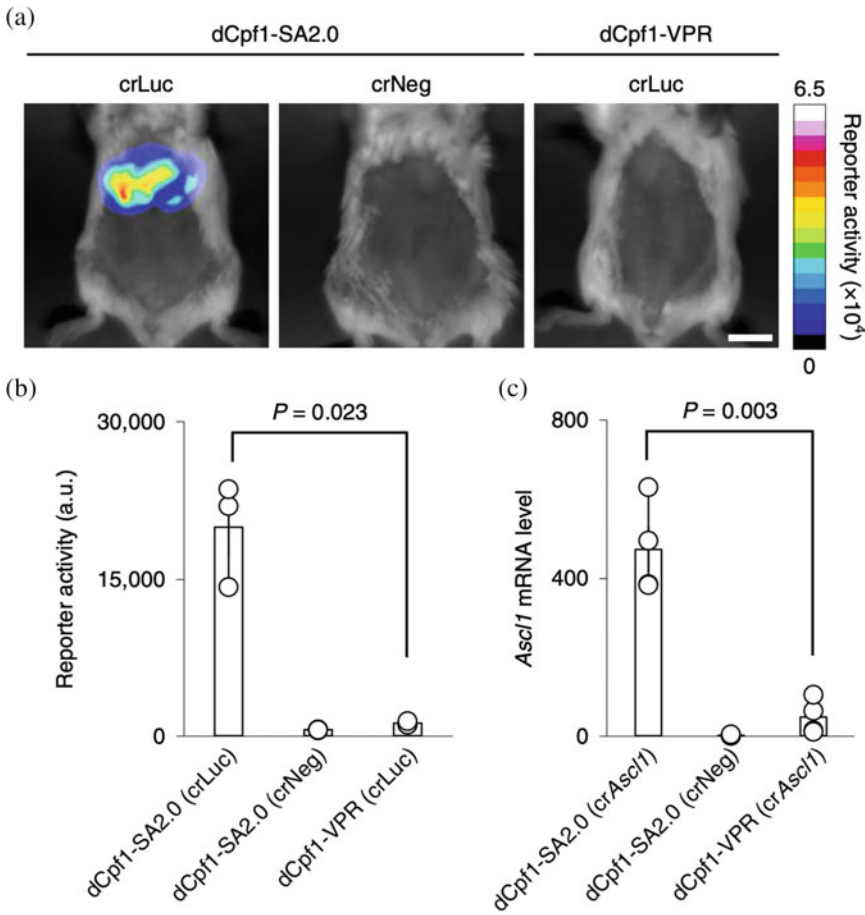
1. In an in vivo luciferase reporter experiment, use female BALB/c mice aged 6 weeks.
2. Before the hydrodynamic tail vein (HTV) injection, amplify and purify plasmid DNAs. Warm up the plasmid DNA and the delivery solution to room temperature.
3. Prepare the injection solution mix, a plasmid encoding the dCpf1 activators, the GAL4-UAS luciferase reporter, and crRNA targeting the reporter or negative crRNA targeting unrelated sequence at a 1:1:1 ratio using HTV delivery solution into the 2-mL plastic tube. The volume of the injected solution mix was 0.1 mL per weight (g), and the total amount of plasmid DNA was 75 µg per mouse.



**Fig. 4** dCpf1 split activator offers potent endogenous gene activation. **(a)** Schematic of dCpf1 split activator. Each dCpf1 fragment is fused with the transcriptional activator domain (p65-HSF1) and nuclear localization sequence (NLS). The spontaneously activated dCpf1 can have multiple activator domains at four distinct locations for reinforced target gene activation. **(b)** dCpf1 split activator enables efficient activation of endogenous *ASCL1*, *HBG1*, and *IL1RN* genes in HEK293T cells. dCpf1 split activator can simultaneously activate multiple genes using pooled single crRNAs and a single crRNA array. DR, direct repeat. This data shows that the relative messenger RNA level to the negative control is transfected with an empty vector. (Reproduced from ref. [8] with permission from Springer Nature)

4. Connect the needle (27G, S· B, 3/4 inch) to the 2.5-mL syringe and load the prepared injection solution. Require careful observation for bubbles in the syringe. Remove the bubbles by flipping the syringe and tapping the side or moving the plunger (*see Note 3*).
5. Transfer the mouse into the mouse cylinder fixator (*see Note 4*).
6. Extend the tail vein by warming the mouse using a heat source, such as warming water.
7. Inject all substances of the syringe within 4–10 s at a constant rate (*see Note 5*).
8. Stop bleeding by applying pressure to the wound.
9. To discriminate each mouse, use the ear punching method, and then the mouse will return to the cage.
10. After 20 h postinjection, the abdominal surface fur of the mouse is removed using a hair clipper and depilatory cream.





**Fig. 5** In vivo gene activation using dCpf1 split activator. **(a)** Comparison between dCpf1 split activator and dCpf1-VPR in luciferase reporter gene activation in living mice. Plasmid expressing the indicated dCpf1 activators, the GAL4-UAS luciferase, and crRNA for targeting reporter (or negative crRNA) were delivered to the liver by hydrodynamic tail vein injection. Scale bar, 1 cm. **(b)** Quantification of luciferase activities in mice liver shown in **(a)**. **(c)** Quantification of the in vivo *Ascl1* gene mRNA expression level detected by quantitative rtPCR. This data shows that the relative mRNA level to the negative control is non-transfected mice. (Reproduced from ref. [8] with permission from Springer Nature)

11. For bioluminescence detection using Lumazon bioluminescence imager and the Evolve 512 EMCCD camera, the mice are injected intraperitoneally with 200  $\mu$ L of Hank’s balanced salt solution supplemented with 100-mM <sup>3</sup>H-luciferin and then are anesthetized with isoflurane (Fig. 5a, b).

**3.8 In Vivo Endogenous Gene Activation Experiment**

1. Prepare the mouse and plasmid DNA as described above (see steps 1–2 in Subheading 3.6).
2. Prepare the injection solution mix, a plasmid encoding the dCpf1 activators, and crRNA targeting mouse *Ascl1* or negative crRNA targeting unrelated sequence at a 1:1 ratio using

HTV delivery solution into the 2-mL plastic tube. The volume of the injected solution mix was 0.1 mL per weight (g), and the total amount of plasmid DNA was 100  $\mu$ g per mouse.

3. Connect the needle (27G, S· B, 3/4 inch) to the 2.5-mL syringe and load the prepared injection solution. Require careful observation for bubbles in the syringe. Remove the bubbles by flipping the syringe and tapping the side or moving the plunger (*see Note 3*).
4. Transfer the mouse into the cylinder fixator (*see Note 4*).
5. Extend the tail vein by warming the mouse using a heat source, such as warming water.
6. Inject all substances of the syringe within 4–10 s at a constant rate (*see Note 5*).
7. Stop bleeding by applying pressure to the wound.
8. To discriminate each mouse, use the ear punching method, and then the mouse will return to the cage.
9. After 24 h postinjection, to sacrifice the mouse using cervical fracture.
10. Dissect the liver tissue and place it in the 24-well plate containing RNAlater solution to prevent RNA degradation.
11. Extract the total RNA from liver tissue using a Precellys Evolution tissue homogenizer with Cryolys Evolution cooling system and Precellys Lysing Kit CK28. After extraction of total RNA, purify the total RNA using an RNA purification kit.
12. For quantitative rtPCR, prepare the cDNA using a reverse transcriptase kit following the manufacturer's protocols.
13. For the analysis of the *in vivo* endogenous gene expression level, perform quantitative rtPCR using a prepared cDNA and the Step One Plus system and TaqMan Gene Expression Assay. To quantify the endogenous gene expression level of mice *Ascl1*, use mice *Gapdh* gene as a control for endogenous gene expression (Fig. 5c).
14. To calculate relative mRNA level, use a negative control transfected with empty vector by the  $\Delta\Delta C_t$  method. TaqMan Gene Expression Assay IDs are as follows: *Ascl1*, Mm03058063\_m1, and *Gpdh*, Mm99999915\_g1.

---

## 4 Notes

1. Contact Moritoshi Sato (moritoshisato@g.ecc.u-tokyo.ac.jp) to request plasmids expressing paCpf1, dCpf1 split activator, crRNAs, and GAL4-UAS luciferase reporter.
2. Useful websites for crRNA design are described in ref. [12].

3. To ensure successful HTV injection, carefully remove bubbles.
4. If it is difficult to administer HTV injection to the moving mouse in the cylinder fixator, before the HTV injection, the mouse is placed under anesthesia. Then, retransfer the anesthetized mouse into the cylinder fixator.
5. If the needle is inserted into the tail vein accurately, the blood from the tail vein flows back into the syringe. If that's confirmed, the solution can be injected into the tail vein without resistance.

---

## Acknowledgments

We would like to thank CREST grants (JPMJCR1653) from Japan Science and Technology Agency. This work was supported by a project grant from Kanagawa Institute of Industrial Science and Technology (KISTEC) to M.S.

## References

1. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8:2281–2308
2. Pickar-Oliver A, Gersbach CA (2019) The next generation of CRISPR-Cas technologies and applications. *Nat Rev Mol Cell Biol* 20:490–507
3. Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, Van Der Oost J, Regev A, Koonin EV, Zhang F (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163:759–771
4. Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J, Degennaro EM, Winblad N, Choudhury SR, Abudayyeh OO, Gootenberg JS, Wu WY, Scott DA, Severinov K, Van Der Oost J, Zhang F (2017) Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol* 35:31–34
5. Safari F, Zare K, Negahdaripour M, Barekati-Mowahed M, Ghasemi Y (2019) CRISPR Cpf1 proteins: structure, function and implications for genome editing. *Cell Biosci* 9:1–21
6. Paul B, Montoya G (2020) CRISPR-Cas12a: functional overview and applications. *Biom J* 43:8–17
7. Tak YE, Kleinstiver BP, Nuñez JK, Hsu JY, Horng JE, Gong J, Weissman JS, Joung JK (2017) Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. *Nat Methods* 14:1163–1166
8. Nihongaki Y, Otabe T, Ueda Y, Sato M (2019) A split CRISPR-Cpf1 platform for inducible genome editing and gene activation. *Nat Chem Biol* 15:882–888
9. Kempton HR, Goudy LE, Love KS, Qi LS (2020) Multiple input sensing and signal integration using a split Cas12a system. *Mol Cell* 78:184–191.e3
10. Wang X, Dong K, Kong D, Zhou Y, Yin J, Cai F, Wang M, Ye H (2021) A far-red light – inducible CRISPR-Cas12a platform for remote-controlled genome editing and gene activation. *Sci Adv* 2358:1–12
11. Kawano F, Suzuki H, Furuya A, Sato M (2015) Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nat Commun* 6:1–8
12. Hanna RE, Doench JG (2020) Design and analysis of CRISPR-Cas experiments. *Nat Biotechnol* 38(7):813–823

# **Part V**

## **Manipulation of the Epigenome: Epigenetically Modified Animals**



## Targeted DNA Methylation in Mouse Early Embryos

Taiga Yamazaki, Yu Hatano, Noritada Kobayashi, and Kazuo Yamagata

### Abstract

In germ cell lines, including early preimplantation embryos, centromeres and pericentromeres are known to show a marked hypomethylation pattern compared to somatic cells. Elucidation of the biological function of this region-specific DNA hypomethylation state, region-specific epigenomic manipulation is essential as an analytical method. We have applied genome editing to show that region-specific DNA methylation can be effectively introduced by a fusion protein, TALE, which recognizes pericentromeres, and SssI, a bacterial CpG methyltransferase. This makes it possible to increase the DNA methylation state of the pericentromeres, which is normally about 20%, to about 60–75%, enabling comparative analysis of the developmental processes of normal embryos with hypomethylated pericentromeres and embryos that have been epigenetically edited to be hypermethylated. In this chapter, we describe a method for introducing DNA methylation into pericentromeres of early mouse embryos by expressing TALE-SssI fusion protein and a method for detecting DNA methylation.

**Key words** Epigenome editing, DNA methylation, TALE, CpG methyltransferase, Pericentromere, Preimplantation embryos, Live-cell imaging

---

### 1 Introduction

DNA methylation is widely known as one of the epigenomic modifications involved in transcriptional regulation and maintenance of cellular differentiation and homeostasis by regulating the transcription of various genes in vivo [1, 2]. Genome-wide methylome analysis using next-generation sequencing has revealed that DNA methylation has a characteristic profile in different cell types [3]. In normal somatic cells, DNA in intergenic regions, retrotransposons, and repetitive sequences such as centromeres and pericentromeres is generally hypermethylated, whereas in cancer cells, these repetitive sequences are hypomethylated [4, 5]. To investigate the methylation status of repetitive DNA sequences in mice, we performed Southern blot analysis using DNA methylation-sensitive restriction enzymes and bisulfite sequencing in a variety of cell types and major tissues [6]. We found that in the majority of tissues, DNA is highly

methylated, whereas DNA methylation at pericentromeres and centromeres is reduced in germ line cells such as primordial germ cells, gametes, early embryos, and germ line stem cells. In addition, we compared the DNA methylation status of germ cells and somatic cells using the mCherry-MBD-NLS fluorescent probe, which can visualize DNA methylation in live cells, and found that the DNA methylation status of centromeres and pericentromeres in germ cells was very low compared to that in somatic cells at the microscopic level [7]. However, it is difficult to analyze the role of DNA hypomethylation at centromeres and pericentromeres in germ cells by simple knockout or knockdown of DNA methylation writers and erasers, and regional epigenomic manipulation is required.

In recent years, various platforms for epigenome editing of DNA methylation have been developed. Zinc finger, TALE, and dCas9 are applied as DNA-binding modules, and Dnmt3, a mammalian de novo methyltransferase, SssI CpG methyltransferase from bacteria, and TET1 protein for demethylation are used as effectors for DNA methylation modification [8, 9]. Most of the studies of epigenome editing using these combinations were performed using cultured cells, and there are limited reports of epigenome editing using mouse preimplantation embryos [10–12]. We have developed a method to introduce pericentromere-specific DNA methylation using the DNA-binding modules TALE and dCas9 and the bacterial CpG methyltransferase SssI and investigated epigenome editing of pericentromeres in early embryos [10]. The pericentromere is a genomic region consisting of more than 200,000 copies of approximately 230 nucleotide repeats adjacent to the centromere [13]. Since the pericentromere is a large genomic region that constitutes approximately 3% of the mouse genome, the introduced DNA methylation can be easily detected in live cells by mCherry-MBD-NLS, a DNA methylation detection probe. In addition to assessing the global DNA methylation status in the nucleus by fluorescence imaging, we analyzed DNA methylation at the base level using bisulfite sequencing. In this chapter, we outline the introduction of DNA methylation in early mouse embryos using TALE-SssI and the evaluation of introduced DNA methylation using fluorescence imaging and bisulfite sequencing.

---

## 2 Materials

### 2.1 Plasmid, Restriction Enzymes, and Cells

1. pcDNA3.1 TALMaj-SssI (WT) poly (A) (*see Note 1*). This plasmid encodes the pericentromere (major satellite) targeting TALE (TALMaj) obtained from pTALYM3B15 [14] and fused with bacterial CpG methyltransferase SssI [10]. pTALYM3B15 was a gift from Maria-Elena Torres-Padilla and available at Addgene.

2. pcDNA3.1 TALMaj-SssI (T313D) poly (A). This plasmid encodes enzymatically inactive TALMaj-SssI.
3. pcDNA3.1 mCherry-MBD-NLS poly (A). This plasmid encodes mCherry-MBD-NLS which binds 5-methyl cytosine (5mC) in living cells [7].
4. pcDNA3.1 histone H2B-EGFP poly (A). This plasmid encodes histone H2B fused with EGFP, which can visualize chromatin in living cells [15].
5. Restriction enzymes (*ApaI*, *XhoI*) (*see Note 2*).
6. Phenol–chloroform.
7. Ethanol.
8. Milli-Q water.
9. *E. coli* HST08 Premium Competent Cells (Takara Bio) (*see Note 3*).

## **2.2 In Vitro RNA Transcription**

1. RiboMAX™ Large Scale RNA Production System T7 (Promega).
2. Ribo m<sup>7</sup>G Cap Analog (Promega).
3. MicroSpin™ S-200 HR Columns (Cytiva).
4. Phenol–chloroform.
5. Ethanol.
6. DNase- and RNase-free water.

## **2.3 Microinjection**

1. TYH medium [16], KSOM containing amino acids (KSOMaa) medium [17], and mineral oil.
2. B6D2F1 male and female mice.
3. Hyaluronidase.
4. Micromanipulator.

## **2.4 Live-Cell Imaging and Bisulfite Sequencing**

1. Fluorescence microscopy for detecting mCherry-MBD-NLS (DNA methylation) and histone H2B-EGFP (chromatin).
2. KSOMaa containing 0.00025% polyvinyl alcohol (average molecular weight 30,000–70,000) instead of BSA and 100- $\mu$ M ethylenediaminetetraacetic acid.
3. IX-71 inverted microscope (Olympus).
4. CSU-W1 Nipkow disk confocal microscope (Yokogawa Electric).
5. iXon3-DU897E EM-CCD camera (Andor Technology).
6. Mac5000 z-motor (Ludl Electronic Products).
7. Auto x-y stage (Sigma Koki).
8. Micro-manager microscopy software (<https://micro-manager.org>).

9. TB200 Optical Power Meter (Yokogawa Electric).
10. ImageJ/Fiji software (<https://imagej.net/software/fiji/>).
11. EZ DNA Methylation-Direct Kit (Zymo Research).
12. TaKaRa EpiTaq™ HS (for bisulfite-treated DNA) (Takara Bio).
13. Primers for bisulfite PCR.
  - Major satellite F 5'- GGAATATGGTAAGAAAATT  
GAAAATTATGG -3'
  - Major satellite R 5'- CCATATTCCAAATCCTTCAATA  
TACATTTC -3'
  - Minor satellite F 5'- TAGAATATATTAGATGAGTGAGT  
TATATTG -3'
  - Minor satellite R 5'- ATTATAACTCATTAATATACAC  
TATTCTAC -3'
  - Line1 F 5'- TAGGAAATTAGTTTGAATAGGTGA  
GAGGT -3'
  - Line1 R 5'- TCAAACACTATATTACTTTAACAATTCC  
CA -3'
  - IAP F 5'- TTGATAGTTGTGTTTTAAGTGGTAAA  
TAAA -3'
  - IAP R 5'- AAAACACCACAAACCAAATCTTCTAC -3'
14. Polyacrylamide gel (5%).
15. pGEM-T Easy Vector System (Promega).
16. Competent cells (*see Note 3*).
17. LB/ampicillin plate.
18. 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal).
19. Plasmid miniprep kit.
20. Primer for sequencing pGEM-T Easy Vector.
  - T7 primer 5'-TAATACGACTCACTATAGGG-3'
21. DNA sequencer for Sanger sequencing.

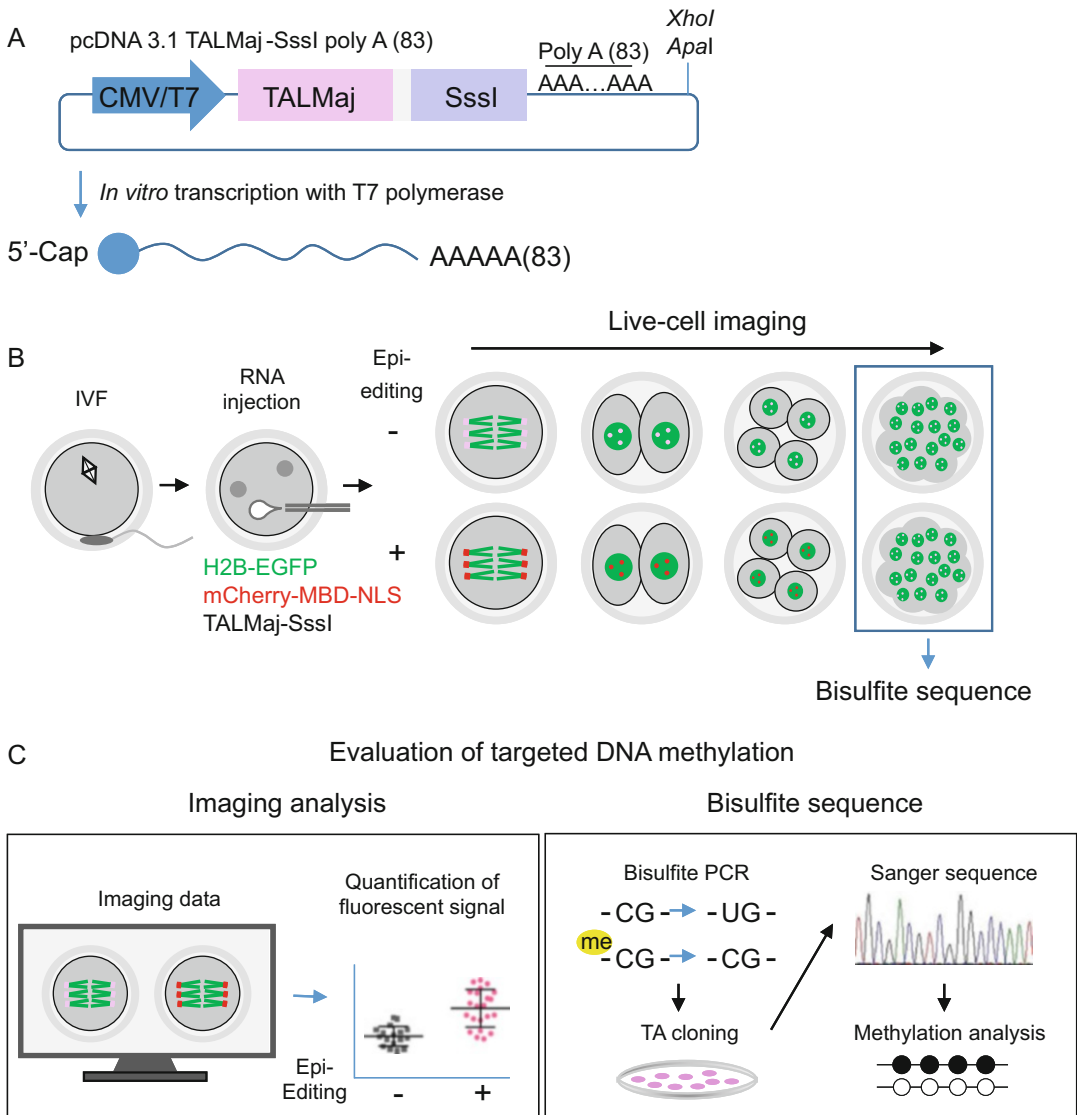
---

## 3 Methods

### 3.1 Preparation of *In Vitro* Transcribed RNA

1. Linearize plasmid DNA using restriction enzymes. Digest single restriction sites such as *Xho*I, *Xba*I, *Apa*I, and *Age*I, which are located just after 83 poly (A) sequences of plasmids (Fig. 1a). The combination of restriction enzymes and plasmids is as follows:
  - pcDNA3.1 TALMaj-SssI (WT) poly (A): *Apa*I
  - pcDNA3.1 TALMaj-SssI (T313D) poly (A): *Apa*I
  - pcDNA3.1 mCherry-MBD-NLS poly (A): *Xho*I
  - pcDNA3.1 histone H2B-EGFP poly (A): *Xho*I





**Fig. 1** Scheme of experiment for inducing targeted DNA methylation in pericentromeres with mouse preimplantation embryo. **(a)** Preparation of *in vitro* transcribed RNA encoding TALMaj-Sssl, a pericentromere targeting CpG methyltransferase. **(b)** Overview of producing epigenome editing embryos. Fertilized oocytes were injected *in vitro* transcribed RNAs of chromatin, a DNA methylation detection probe, and epigenome editing enzyme. **(c)** Analysis of targeted DNA methylation in preimplantation embryos. Imaging analysis and bisulfite sequencing were applied to quantify targeted DNA methylation in embryos

2. Purify linearized plasmids by phenol–chloroform treatment and ethanol precipitation using RNase-free reagents. Dissolve purified linear plasmids in fresh Milli-Q water.
3. Prepare RNA transcription mixture with RiboMAX™ Large Scale RNA Production System T7 and linearized plasmid template (Fig. 1a). Add Ribo m<sup>7</sup>G Cap Analog into an RNA

production mixture containing rATP, rUTP, rCTP, and rGTP (*see Note 4* for details). Use 1.5–2  $\mu\text{g}$  of plasmid template for 20- $\mu\text{L}$  volume reaction of RNA synthesis.

4. Treat the transcribed RNA with RQ-1 RNase-free DNase included in the RNA production kit.
5. Purify RNA by phenol–chloroform treatment and ethanol precipitation. Dissolve RNA with DNase- and RNase-free water.
6. Remove free rNTPs with MicroSpin™ S-200 HR Columns. First, equilibrate the column with DNase- and RNase-free water. Next, add transcribed RNA solution to the column. Then elute transcribed RNA multiple times using the above water.
7. Measure concentration of the transcribed RNA in each fraction. Mix each fraction and adjust RNA concentration to approximately 100 ng/ $\mu\text{L}$ . Confirm by electrophoresis that the appropriate size of RNA has been transcribed.

### **3.2 Preparation of Fertilized Embryos and RNA Injection**

1. Transfer cauda epididymal sperm into TYH medium and capacitate sperm for 90 min. Inseminate superovulated oocytes with capacitated sperm in TYH medium for 90 min.
2. Treat fertilized embryos with hyaluronidase to disperse cumulus cells surrounding the embryos. Select fertilized embryos that show fertilization cone and second polar body.
3. Mix RNAs encoding TALMaj-SssI, mCherry-MBD-NLS, and histone H2B-EGFP at final concentrations of 10 ng/ $\mu\text{L}$ , 1 ng/ $\mu\text{L}$ , and 5 ng/ $\mu\text{L}$ , respectively.
4. Inject RNAs into fertilized embryos using a micromanipulator (Fig. 1b). As a control, use TALMaj-SssI T313D mutant, which has no DNA methylation activity instead of TALMaj-SssI WT. Culture injected embryos in KSOMaa medium before live-cell imaging.

### **3.3 Detection of DNA Methylation and Chromatin by Live-Cell Imaging**

Global DNA methylation and chromatin can be monitored by detecting mCherry-MBD-NLS and histone H2B-EGFP in the nucleus. Fluorescent signals of mCherry-MBD-NLS and histone H2B-EGFP are detectable approximately 3–4 h after RNA injection.

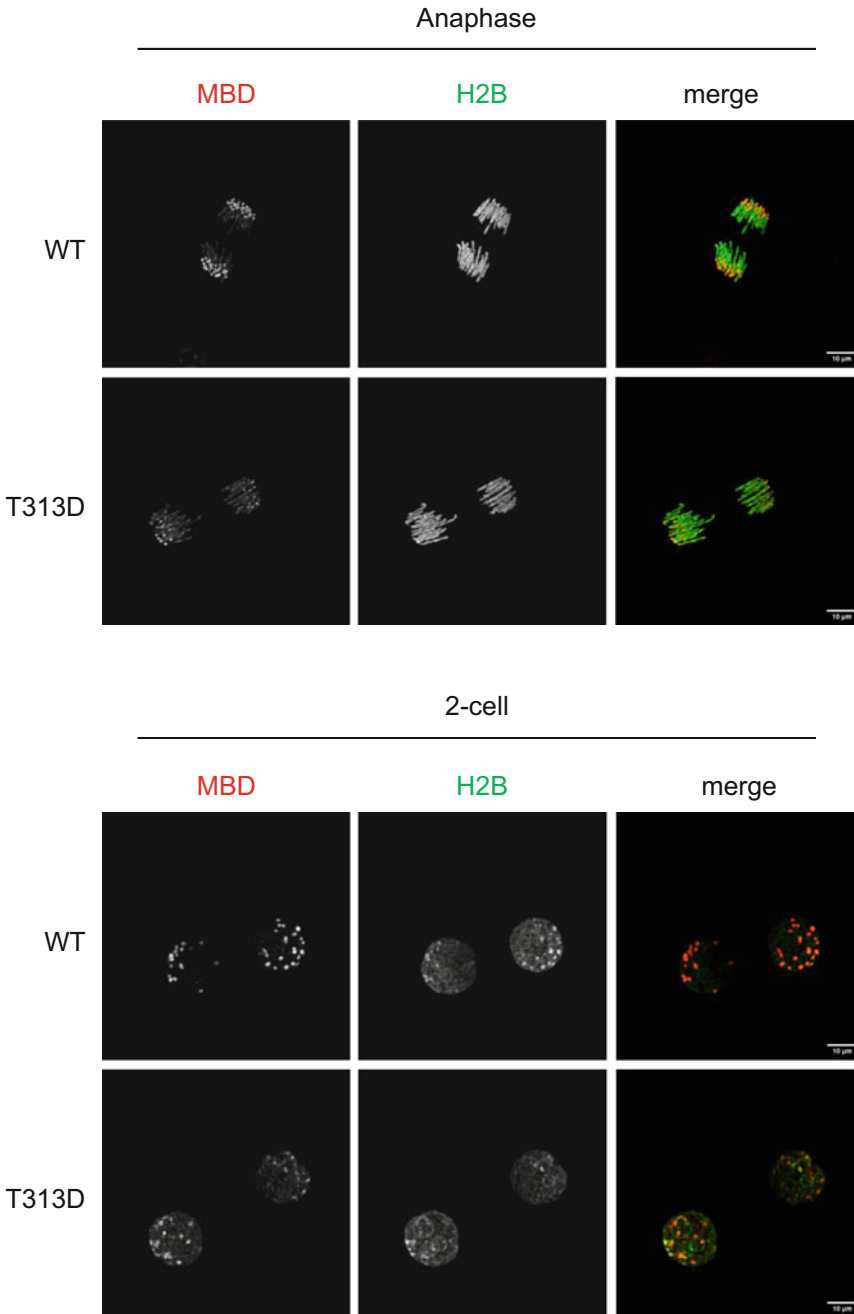
1. Use an inverted IX-71 microscope with an attached Nipkow disk confocal microscope (CSU-W1), EM-CCD camera (iXon3-DU897E), and z-motor (Mac5000). Imaging device that contained an attached auto x-y stage (XY30100T) enables simultaneous monitoring for multiple embryos. Place a set of imaging systems in a dark room maintained at 30 °C to stabilize the temperature of the whole device.

2. Transfer fertilized embryos to 5- $\mu$ L droplets of KSOMaa containing 0.00025% polyvinyl alcohol instead of BSA and 100- $\mu$ M ethylenediaminetetraacetic acid on a glass-bottomed dish. Place glass-bottomed dish in an incubation chamber set at 37 °C on the microscope stage. Introduce gas mixture of 5% O<sub>2</sub>, 6% CO<sub>2</sub>, and 89% N<sub>2</sub> into the chamber (138 mL/min).
3. Measure the power of the laser emitted from the tip of the objective using an optical power meter (TB200), and make sure the laser output is at 0.1 mW.
4. Control imaging devices by micro-manager microscopy software, and acquire 46–51 slice images in the z-axis with 1–2- $\mu$ m width for each wavelength at 488-nm laser (histone H2B-EGFP) and 561-nm laser (mCherry-MBD-NLS). Set time interval for image acquisition to 5–15 min. Acquire imaging data by time-lapse observation for approximately 1–4 days depending on the purpose (*see Note 5*).
5. Confirm the results, as follows: At the time of the first mitosis, approximately 18–22 h after fertilization, you can detect fluorescent signal of mCherry-MBD-NLS at the distal part of segregating chromosomes, which represent pericentromeres. Upregulated fluorescent signal of mCherry-MBD-NLS in the pericentromere is clearly detectable in embryos expressing enzymatically active TALMaj-SssI (WT) compared with embryos expressing enzymatically inactive TALMaj-SssI (T313D) or normal mock embryos. Also, you can detect upregulated DNA methylation in the chromocenter-like large dot pattern in the nucleus at the two-cell stage in embryo expressing TALMaj-SssI (WT) (Figs. 1c and 2).

### **3.4 Quantification of Fluorescent Signals of DNA Methylation and Chromatin**

We use ImageJ/Fiji software for measurement of fluorescence intensity.

1. Compress the Z information of the 3D image with maximum intensity projection for each fluorescence channel.
2. Determine nuclear region by binarizing the H2B-EGFP signal, and measure fluorescent intensity of mCherry-MBD-NLS inside the region.
3. Quantify nuclear heterochromatin state as a heterochromatin index because changes in DNA methylation status are likely to appear as changes in the localization pattern of MBD in nuclear heterochromatin. The heterochromatin index is defined as the coefficient of variation of signal patterns and is calculated using the following formula: heterochromatin index = SD of signal intensity / average signal intensity  $\times$  100 [7].



**Fig. 2** Snapshot of embryos' targeted DNA methylation in pericentromere. The upper panels indicate the first mitosis of embryos 20–24 h after fertilization. Fluorescent signals of mCherry-MBD-NLS (MBD) and histone H2B-EGFP (H2B) were presented. Pericentromeres can be observed in the distal part of segregating chromosomes with mCherry-MBD-NLS, and these signals were upregulated in TALMaj-Sssl (WT) expressing embryos compared with control (T313D) embryos. The lower panels indicate two-cell embryos of pericentromere-targeted DNA methylation-induced embryos

### 3.5 Evaluation of Induced DNA Methylation by Bisulfite Sequencing

Evaluation of target DNA methylation and off-target DNA methylation at the sequence level by bisulfite sequencing is possible concurrently with monitoring upregulated pericentromeric DNA methylation by live-cell imaging (Fig. 1c). Treat live-cell imaged embryos with bisulfite reaction using the EZ DNA Methylation-Direct Kit.

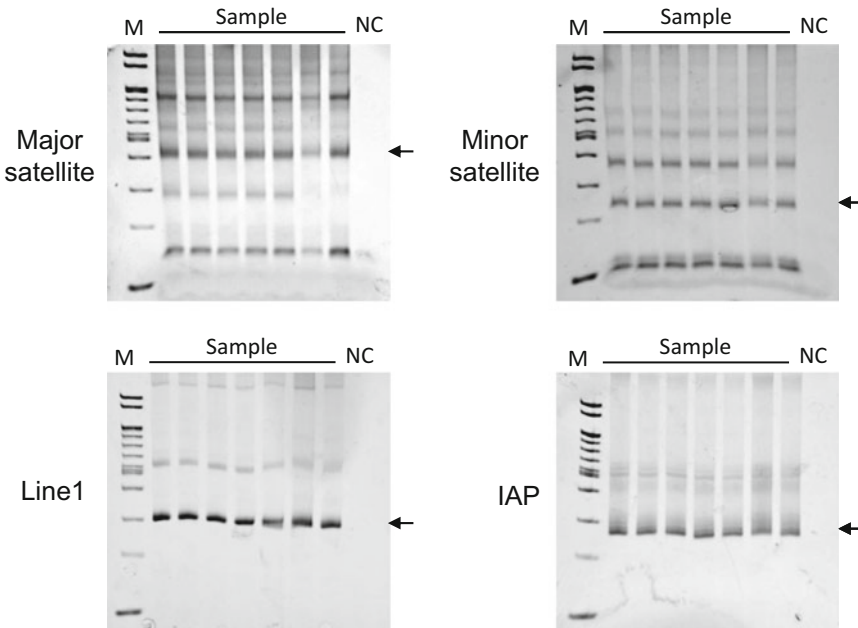
1. Pool embryos into sample digestion buffer containing proteinase K attached with the kit in a PCR tube. Place PCR tube in a thermal cycler and then perform the following program: 50 °C for 20 min. We usually use approximately 12–50 embryos in one reaction tube for bisulfite treatment.
2. After the proteinase K treatment, add the bisulfite conversion buffer to the tube.
3. Place the tube in the thermal cycler and perform the following program: 98 °C for 8 min and 64 °C for 3.5 h.
4. Purify and elute bisulfite-treated genome in 20 µL of elution buffer provided in the kit.
5. Amplify the target and off-target regions by PCR with bisulfite-treated genome.

Two microliters of bisulfite-treated genome are used as the template for each PCR. The composition of the PCR is as follows:

Components	Volume
Bisulfite-treated DNA	2 µL
10× EpiTaq PCR Buffer	2.5 µL
25-mM MgCl <sub>2</sub>	2.5 µL
dNTP mixture	3 µL
10-µM primer mixture	1 µL
TaKaRa EpiTaq HS	0.125 µL
Water	13.875 µL
Total volume	25 µL

The PCR program for amplifying the major satellite and minor satellite is as follows: 40 cycles of 95 °C for 30 s, 60 °C for 60 s, and 68 °C for 20 s. For amplifying Line1 and IAP retrotransposons, the PCR conditions are as follows: 35 cycles of 95 °C for 60 s, 60 °C for 60 s, and 72 °C for 60 s.

6. Electrophorese amplified DNA on a 5% acrylamide gel (*see* Fig. 3 and Note 6).



**Fig. 3** Gel images of bisulfite PCR. Each PCR product was electrophoresed on a 5% acrylamide gel. We used water as a negative control (NC) for the bisulfite treatment. Arrows indicate target amplicons. Note that multiple bands appear in major satellite and minor satellite PCR, because these sequences consist of tandem repeats. The marker (M) is a 100-bp DNA ladder. The target amplicons were subcloned into plasmids and sequenced

7. Extract and purify PCR fragments from acrylamide gel (Fig. 3).
8. Ligate DNA fragments directly into a T-vector using a TA-cloning kit and transform the product with competent cells (*see Note 3*).
9. Plate transformant onto LB/ampicillin plate containing 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside (X-Gal).
10. Pick white colonies and isolate plasmids using a miniprep kit (*see Note 7*).
11. Sequence plasmids using T7 primer.
12. Align sequences and analyze for DNA methylation (*see Note 8*).

---

## 4 Notes

1. We use the pcDNA poly (A) plasmid for in vitro transcription. This plasmid contains 83 poly (A) and the 3'-UTR sequences of the TRF2 gene directly under the multi-cloning site to stabilize RNA in the cytoplasm and enhance translation [18]. We provide these plasmids on request.

2. The RiboMAX™ Large Scale RNA Production System T7 does not recommend using restriction enzymes such as *ApaI*, which generate 3' overhangs. The 3' overhang can be converted to a blunt end by treating the DNA with polymerase I, a large (Klenow) fragment. However, we do not usually convert 3' overhangs into blunt ends.
3. The *E. coli* strain HST08 Premium has a  $\Delta$  *mcrA* genotype that lacks a restriction system for cytosine methylation of 5'-CpG-3' dinucleotide. This enables stable cloning efficiency of the SssI coding expression vector. When cloning pGEM vectors, the commonly used *E. coli* strain DH5 $\alpha$  is fine.
4. The volumes of 100-mM rATP, 100-mM rUTP, 100-mM rCTP, 100-mM rGTP, and Ribo m<sup>7</sup>G Cap Analog in 20- $\mu$ L reaction are 1.5  $\mu$ L, 1.5  $\mu$ L, 1.5  $\mu$ L, 0.12  $\mu$ L, and 1.5  $\mu$ L, respectively.
5. By adjusting the laser power, the tabletop incubator, and the exact expression level of the introduced fluorescent probe, we are able to construct a noninvasive observation system, and live-cell imaging up to the blastocyst can be performed without any developmental retardation or arrest.
6. Acrylamide gels rather than agarose gels are recommended for electrophoresing major satellites and minor satellite PCR products. These PCR produce PCR products of various sizes, resulting in a laddering pattern of electrophoresis. Acrylamide gels have better separation ability than agarose gels, which makes it easier to separate PCR products and pick them.
7. We usually sequence about 20 clones of plasmids with each bisulfite PCR product.
8. We use "QUMA" (<http://quma.cdb.riken.jp/top/index.html>), a web-based software, for alignment, visualization, and analysis for bisulfite sequence data sets [19]. The GenBank accession numbers of the reference sequence of the genome are as follows: major satellite (EF028077.1), minor satellite (Z22170.1), Line1 (D84391.1), and IAP (M17551.1).

---

## Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP19K05956, JP22K05562 (TY) and JP18H02357, JP18H05528 (KY). This work was also supported by a Grant-in-Aid for Research from the Kitasato University Medical Center and Kawano Masanori Memorial Public Interest Incorporated Foundation for Promotion of Pediatrics (TY).

## References

1. Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. *Nat Rev Genet* 14(3):204–220. <https://doi.org/10.1038/nrg3354>
2. Bergman Y, Cedar H (2013) DNA methylation dynamics in health and disease. *Nat Struct Mol Biol* 20(3):274–281. <https://doi.org/10.1038/nsmb.2518>
3. Lister R, Pelizzola M, Downen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271):315–322. <https://doi.org/10.1038/nature08514>
4. Wilson AS, Power BE, Molloy PL (2007) DNA hypomethylation and human diseases. *Biochim Biophys Acta* 1775(1):138–162. <https://doi.org/10.1016/j.bbcan.2006.08.007>
5. Ehrlich M (2009) DNA hypomethylation in cancer cells. *Epigenomics* 1(2):239–259. <https://doi.org/10.2217/epi.09.33>
6. Yamagata K, Yamazaki T, Miki H, Ogonuki N, Inoue K, Ogura A, Baba T (2007) Centromeric DNA hypomethylation as an epigenetic signature discriminates between germ and somatic cell lineages. *Dev Biol* 312(1):419–426. <https://doi.org/10.1016/j.ydbio.2007.09.041>
7. Ueda J, Maehara K, Mashiko D, Ichinose T, Yao T, Hori M, Sato Y, Kimura H, Ohkawa Y, Yamagata K (2014) Heterochromatin dynamics during the differentiation process revealed by the DNA methylation reporter mouse, MethylRO. *Stem Cell Reports* 2(6):910–924. <https://doi.org/10.1016/j.stemcr.2014.05.008>
8. Yamazaki T, Hatano Y, Taniguchi R, Kobayashi N, Yamagata K (2020) Editing DNA methylation in mammalian embryos. *Int J Mol Sci* 21(2):doi:10.3390/ijms21020637
9. Nakamura M, Gao Y, Dominguez AA, Qi LS (2021) CRISPR technologies for precise epigenome editing. *Nat Cell Biol* 23(1):11–22. <https://doi.org/10.1038/s41556-020-00620-7>
10. Yamazaki T, Hatano Y, Handa T, Kato S, Hoida K, Yamamura R, Fukuyama T, Uematsu T, Kobayashi N, Kimura H, Yamagata K (2017) Targeted DNA methylation in pericentromeres with genome editing-based artificial DNA methyltransferase. *PLoS One* 12(5):e0177764. <https://doi.org/10.1371/journal.pone.0177764>
11. Wei Y, Lang J, Zhang Q, Yang CR, Zhao ZA, Zhang Y, Du Y, Sun Y (2019) DNA methylation analysis and editing in single mammalian oocytes. *Proc Natl Acad Sci U S A* 116(20):9883–9892. <https://doi.org/10.1073/pnas.1817703116>
12. Horii T, Morita S, Hino S, Kimura M, Hino Y, Kogo H, Nakao M, Hatada I (2020) Successful generation of epigenetic disease model mice by targeted demethylation of the epigenome. *Genome Biol* 21(1):77. <https://doi.org/10.1186/s13059-020-01991-8>
13. Martens JH, O’Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, Jenuwein T (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J* 24(4):800–812. <https://doi.org/10.1038/sj.emboj.7600545>
14. Miyanari Y, Ziegler-Birling C, Torres-Padilla ME (2013) Live visualization of chromatin dynamics with fluorescent TALEs. *Nat Struct Mol Biol* 20(11):1321–1324. <https://doi.org/10.1038/nsmb.2680>
15. Kanda T, Sullivan KF, Wahl GM (1998) Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. *Curr Biol* 8(7):377–385. [https://doi.org/10.1016/s0960-9822\(98\)70156-3](https://doi.org/10.1016/s0960-9822(98)70156-3)
16. Toyoda Y, Yokoyama M, Hosi T (1971) Studies on the fertilization of mouse eggs in vitro I. in vitro fertilization of eggs by fresh epididymal sperm. *Jpn J Anim Reprod* 16(4):147–151. <https://doi.org/10.1262/jrd1955.16.147>
17. Biggers JD (2002) Thoughts on embryo culture conditions. *Reprod Biomed Online* 4:30–38. [https://doi.org/10.1016/s1472-6483\(12\)60009-1](https://doi.org/10.1016/s1472-6483(12)60009-1)
18. Yamagata K, Yamazaki T, Yamashita M, Hara Y, Ogonuki N, Ogura A (2005) Noninvasive visualization of molecular events in the mammalian zygote. *Genesis* 43(2):71–79. <https://doi.org/10.1002/gene.20158>
19. Kumaki Y, Oda M, Okano M (2008) QUMA: quantification tool for methylation analysis. *Nucleic Acids Res* 36(Web Server issue):W170–W175. <https://doi.org/10.1093/nar/gkn294>





## Generation of Epigenetic Disease Model Mice by Targeted Demethylation of the Epigenome

Takuro Horii, Sumiyo Morita, and Izuho Hatada

### Abstract

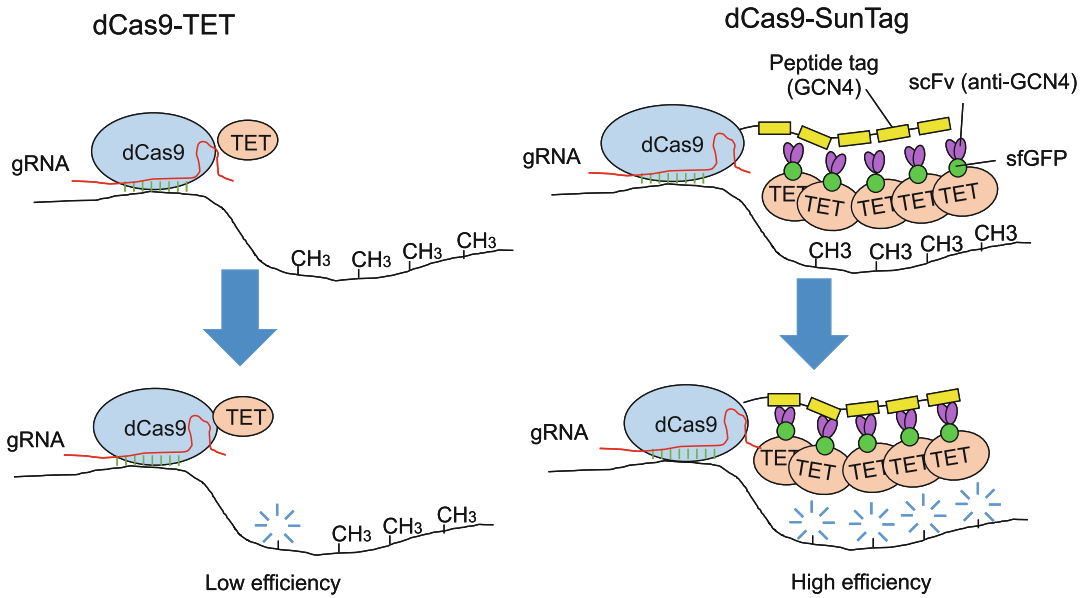
Epigenetic regulatory mechanisms play an important role in gene silencing and genome stability; therefore, epigenetic mutations cause a variety of diseases. Analysis of the epigenome by next-generation sequencers has revealed many epigenetic mutations in various diseases such as cancer, obesity, diabetes, autism, allergies, immune diseases, and imprinting diseases. Unfortunately, it has been difficult to identify the causative epigenetic mutations because there has been no method to generate animals with target-specific epigenetic mutations. However, it has become possible to generate such animals due to the recent development of epigenome editing technology. Here, we introduce the generation of epigenome-edited mice by target-specific DNA demethylation.

**Key words** Epigenome editing, Demethylation, dCas9, SunTag, TET

---

### 1 Introduction

Epigenome editing is a technology that manipulates epigenetic modifications such as DNA methylation and histone modifications in specific regions of the genome. The epigenome editing system consists of an effector module, which is an epigenetic modification enzyme, and a specific DNA sequence-binding module that can bind to specific DNA sequences. In this chapter, we introduce a system using dCas9, a nuclease-deactivated variant of Cas9, which is used for site-specific targeting in the CRISPR/Cas9 system [1–3]. For example, the hydroxymethyltransferase ten-eleven translocation (TET) 1 fused to dCas9 induces targeted demethylation of DNA around the region where the guide RNA (gRNA) binds (Fig. 1a). To improve the editing efficiency of this system, we have developed a target-specific DNA demethylation system using dCas9 and SunTag [4, 5] (Fig. 1b). SunTag is a protein scaffold containing peptide epitopes that can recruit effector domains via specific single-chain variable fragment (scFv) antibodies



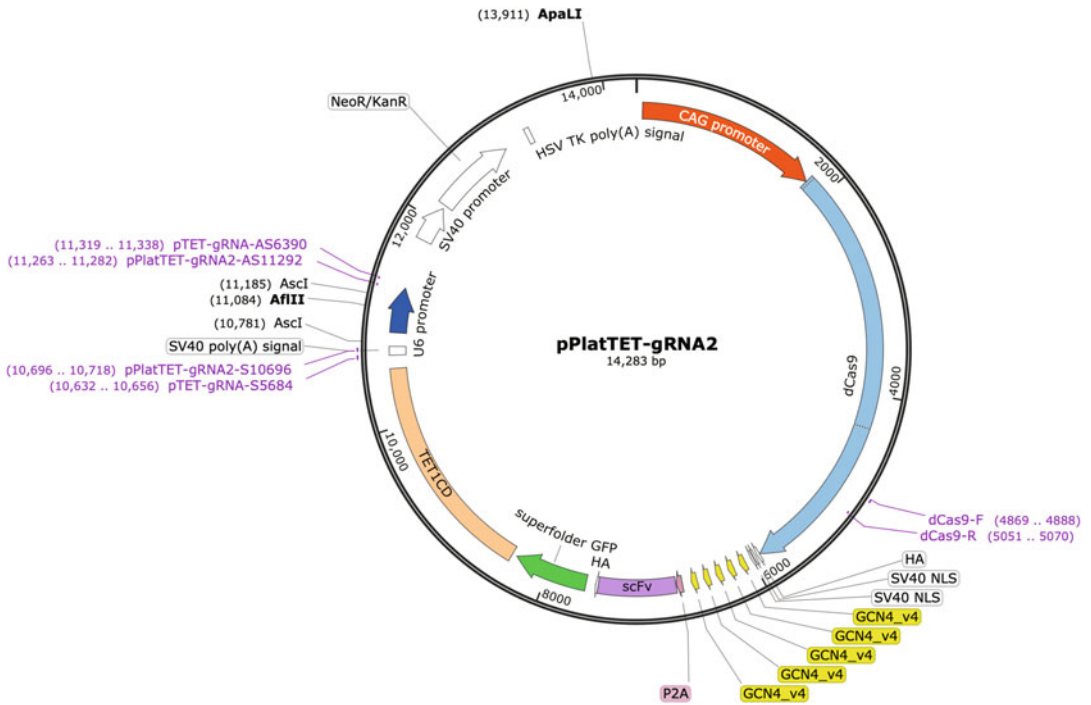
**Fig. 1** Comparison of epigenome editing systems for targeted DNA demethylation

[6]. SunTag-carrying dCas9 recruits the scFv-green fluorescent protein (GFP)-TET1CD fusion protein, which contains the catalytic domain (CD) of the TET1 hydroxylase, to the target locus, leading to targeted DNA demethylation. Based on this system, we have established a method to generate epigenetic disease model mice with targeted DNA demethylation [7]. This chapter focuses on how to design gRNA targets, construct plasmid vectors, and generate epigenetic disease model mice with target-specific DNA demethylation.

## 2 Materials

### 2.1 Vector Construction

1. pPlatTET-gRNA2 (Addgene, #82559) (Fig. 2).
2. gRNA Cloning Vector *Bbs*I (Addgene, #128433).
3. pCR2.1 vector. This vector is generated by TA cloning or self-ligation using a TOPO TA Cloning Kit (Thermo Fisher Scientific). Any vector containing two *Bst*XI sites can be used.
4. Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB).
5. Q5 Hot Start High-Fidelity DNA Polymerase (NEB).
6. TaKaRa Taq™ (TaKaRa).
7. Gibson Assembly Master Mix (NEB).
8. DNA ligation kit: DNA Ligation Kit Ver. 2.1 (TaKaRa).
9. Plasmid miniprep kit.



**Fig. 2** The map of the all-in-one epigenome editing vector (pPlatTET-gRNA2). The positions of the CAG promoter, dCas9, GCN4, scFv, superfolder GFP, and TET1 catalytic domain (TET1CD) are shown. The restriction enzyme sites and primer positions used in this study are also indicated

10. PCR purification kit.
11. Gel extraction kit.
12. DNA purification kit.
13. Media and antibiotics for growing *Escherichia coli* cultures: LB liquid medium and LB agar plates supplemented with kanamycin (20 µg/mL) or ampicillin (50 µg/mL).
14. Calf intestinal alkaline phosphatase (CIAP).
15. Restriction enzymes (*AflIII*, *BbsI*, *BsaI*, *BstXI*, *AscI*, and *ApaLI*).
16. Elution buffer (EB): 10-mM Tris-Cl, pH 8.5 (included in Qiagen kit).
17. Sterile distilled water (DW).
18. Competent cells (*E. coli*).
19. Block incubator or water bath.
20. Microbiological incubator set to 37 °C.
21. Shaker set to 37 °C.
22. Primers:

pTET-gRNA-S5684: 5'-CATAAAATGAATGCAATTGT  
TGTTG-3'

pTET-gRNA-AS6390: 5'-TGGAACAACACTCAACCCT  
A-3'

Universal M13 Forward (-20): 5'-GTAAAACGACGG  
CCAG-3'

Universal M13 Reverse: 5'-CAGGAAACAGCTATGAC-3'

pPlatTET-gRNA2-S10696: 5'-GCAATAGCATCACAAA  
TTTCACA-3'

pPlatTET-gRNA2-AS11292: 5'-CGATTTTCGGCCTATT  
GGTTA-3'

## 2.2 Embryo Manipulation

1. Male and female B6D2F1 mice.
2. M16 embryo culture medium [8].
3. M2 embryo culture medium [8].
4. M2-hyaluronidase medium: M2 medium containing hyaluronidase (300 U/mL).
5. Paraffin oil.
6. Plastic dish: 60 mm for embryo culture and 100 mm for micromanipulation.
7. Micromanipulation system on an inverted microscope.
8. CO<sub>2</sub> incubator.
9. Primers for genotyping:  
dCas9-F: CCAAGAGCGAGCAGGAAATC  
dCas9-R: GGCATAGACAGCACTTTCCG

## 2.3 DNA Methylation Analysis

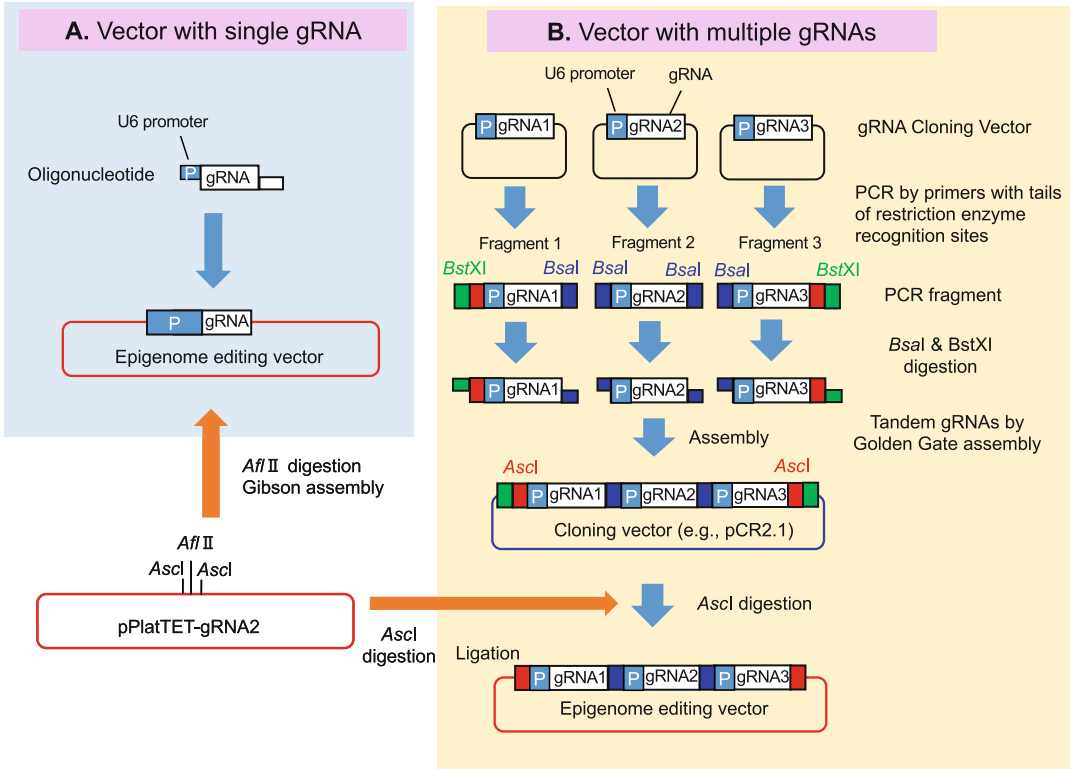
1. DNA bisulfite conversion kit: EpiTect Plus DNA Bisulfite Kit (Qiagen).
2. TA cloning vector: TOPO TA Cloning Kit (Thermo Fisher Scientific).
3. Restriction enzymes (*Bst*UI, *Cla*I, *Mlu*I, and *Taq*I).

---

## 3 Methods

In the present targeted DNA demethylation system, a single gRNA is sufficient for DNA demethylation 200–400 bp upstream and downstream of the target sequence. If you wish to demethylate an area containing a few hundred base pairs, a single gRNA is sufficient (Subheading 3.1). Otherwise, you should select multiple gRNAs to cover the target region (Subheading 3.2).

1. Obtain the sequence of the region to induce targeted DNA demethylation.



**Fig. 3** Overview of the construction of epigenome editing vectors containing a single gRNA (a) and multiple gRNAs (b)

**3.1 Vector Construction for a Single gRNA (Fig. 3a)**

**3.1.1 gRNA Design**

2. Select a candidate target sequence (20 mer) (*see Note 1*) in the region to demethylate using the CRISPR design tools (*see Note 2*). If the 5' end of the selected target sequence is not G, replace it with G; this is defined as **S** (*see Note 3*). The antisense complement of **S** is defined as **AS**. Synthesize the oligonucleotides (60 mer) by adding a 40-mer tail complementary to the vector insertion site to the 5' end of these sequences as follows:

**XgRNA-S:**



**XgRNA-AS:**



The sequence complementary to the vector insertion site is underlined.

### 3.1.2 Preparation of Oligonucleotides

1. Anneal the two oligonucleotides (XgRNA-S and XgRNA-AS). Mix 1  $\mu\text{L}$  of XgRNA-S (5 pmol/ $\mu\text{L}$ ), 1  $\mu\text{L}$  of XgRNA-AS (5 pmol/ $\mu\text{L}$ ), 5  $\mu\text{L}$  of Phusion High-Fidelity PCR Master Mix, and 3  $\mu\text{L}$  of DW.
2. Incubate in a thermal cycler at 98 °C for 30 s, 50 °C for 30 s, and 72 °C for 3 min.
3. Add 20  $\mu\text{L}$  of DW.

### 3.1.3 Preparation of pPlatTET-gRNA2 Vector

1. Digest pPlatTET-gRNA2 (Fig. 2) with the *Afl*III restriction enzyme at 37 °C for 3 h.
2. Incubate at 65 °C for 20 min to inactivate the enzyme.

### 3.1.4 Ligation

1. Mix 100 ng of digested pPlatTET-gRNA2 (prepared in Subheading 3.1.3), 0.5  $\mu\text{L}$  of oligonucleotides (prepared in Subheading 3.1.2), 5  $\mu\text{L}$  of Gibson Assembly Master Mix (2 $\times$ ), and DW up to 10  $\mu\text{L}$ .
2. Incubate at 50 °C for 1 h.

### 3.1.5 Transformation

1. Add 1–2  $\mu\text{L}$  of the ligated vector to 50  $\mu\text{L}$  of competent cells and incubate on ice for 30 min.
2. Incubate the transformant at 42 °C for 30 s using a block incubator or water bath.
3. Immediately cool on ice for 2 min.
4. Plate the transformant onto LB agar containing kanamycin and incubate at 37 °C overnight.
5. Pick a colony and suspend it in PCR mixture containing the primer set pTET-gRNA-S5684 and pTET-gRNA-AS6390.
6. Perform colony PCR at 94 °C for 3 min, followed by 30 cycles of 94 °C for 10 s, 55 °C for 30 s, and 72 °C for 1 min, and then 72 °C for 5 min.
7. Confirm that the target sequence has been inserted by direct sequencing using the primer set employed for colony PCR.
8. Pick the colonies that tested positive by colony PCR from the replica plate. Place them into 3 mL of LB liquid medium containing kanamycin and shake overnight at 37 °C.
9. Extract the plasmid from the culture using a plasmid miniprep kit.
10. Move to Subheading 3.3.

## 3.2 Vector Construction for Multiple gRNAs (Fig. 3b)

### 3.2.1 gRNA Design

1. Obtain the sequence of the region to induce targeted DNA demethylation.
2. Select candidate target sequences (20 mer) (*see Note 1*) at approximately 250-bp intervals in the region to demethylate using the CRISPR design tools (*see Note 2*). If the 5' end of the selected target sequence is not G, replace it with G; this is

defined as **S** (*see Note 3*). The antisense complement of **S** is defined as **AS**. Synthesize the following oligonucleotides for each gRNA:

XgRNA-S: 5'-CACC-**S**-3'

XgRNA-AS: 5'-AAAC-**AS**-3'

### 3.2.2 Preparation of Oligonucleotides

1. Mix 4.5  $\mu\text{L}$  of XgRNA-**S** (100 pmol/ $\mu\text{L}$ ), 4.5  $\mu\text{L}$  of XgRNA-**AS** (100 pmol/ $\mu\text{L}$ ), and 1  $\mu\text{L}$  of TaKaRa Taq™ 10 $\times$  PCR Buffer.
2. Incubate at 95 °C for 5 min in a block bath, 37 °C for 15 min in an air incubator, and room temperature for 15 min to anneal the two oligonucleotides.

### 3.2.3 Preparation of gRNA Cloning Vector *BbsI*

1. Digest gRNA Cloning Vector *BbsI* with the *BbsI* restriction enzyme at 37 °C for 3 h.
2. Purify gRNA Cloning Vector *BbsI* following gel electrophoresis with a gel extraction kit (*see Note 4*). Elute with 30  $\mu\text{L}$  of EB buffer.

### 3.2.4 Ligation

1. Mix 1  $\mu\text{L}$  of the oligonucleotide mixture (prepared in Subheading 3.2.2), 1  $\mu\text{L}$  of gRNA Cloning Vector *BbsI* digested with *BbsI* (prepared in Subheading 3.2.3), and 2  $\mu\text{L}$  of Sol I (DNA Ligation Kit Ver. 2.1).
2. Incubate at 16 °C for 4 h to overnight.

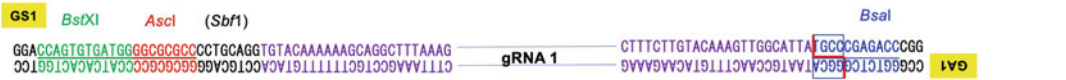
### 3.2.5 Transformation

1. Transform the vector into competent cells, plate the transformant onto LB agar containing kanamycin, and incubate at 37 °C overnight.
2. Pick a few colonies (*see Note 5*) and suspend them in PCR mixture containing the primer set Universal M13 Forward (–20) and Universal M13 Reverse (product size: 658 bp).
3. Perform colony PCR at 94 °C for 3 min, followed by 30 cycles of 94 °C for 10 s, 55 °C for 30 s, and 72 °C for 1 min, and then 72 °C for 5 min.
4. Confirm that the target sequence has been inserted by direct sequencing using the primer set employed for colony PCR.
5. Pick the colonies that tested positive by colony PCR from the replica plate. Place them into 3 mL of LB liquid medium containing kanamycin and shake overnight at 37 °C.
6. Extract the plasmid from the culture using a plasmid miniprep kit.

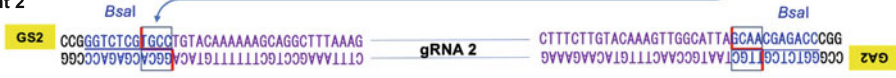
### 3.2.6 Assembly of gRNA Fragments by the Golden Gate Method

An example of assembly of three different gRNA fragments (1, 2, and 3) is provided (Fig. 3b).

**Fragment 1**



**Fragment 2**



**Fragment 3**



Bsal  
 5' -GGTCTCNNNNN-3'  
 3' -CCAGANNNNN-5'  
 BstXI  
 5' -CCANNNNNTGG-3'  
 3' -GGTNNNNNAACC-5'  
 GS sense primer  
 GA antisense primer

GGTCTCNNNNN (*Bsal* site): N is an arbitrary nucleotide.  
 NNNN : Designed to be complementary to neighboring gRNAs. This part becomes an assembly attachment surface during ligation and binds complementarily to neighboring gRNAs.  
 Purple: sequences in the gRNA Cloning Vector *Bbs*I.  
 CCANNNNNTGG (*Bst*XI site): used for ligation to pCR2.1 vector. N is an arbitrary nucleotide.  
 NNNN : Designed to be complementary to the *Bst*XI sequence of the insertion site.  
 GGCGCGCC (*Ascl* site): used for ligation to pPlatTET-gRNA2.

**Fig. 4** Primer design to amplify gRNA fragments for Golden Gate Assembly. This figure shows an example of assembly of three gRNA fragments. PCR is performed to amplify gRNA fragments using primers tailed by *Bsal* or *Ascl* and *Bst*XI recognition sites. gRNA fragments with restriction enzyme sites added to each end are assembled by the Golden Gate method

1. Design primers with tails as shown below to amplify fragments containing each gRNA sequence. Details are shown in Fig. 4.

**Primers for Fragment 1 (gRNA1):**

**GS1**

5'-GGACCAAGTGTGATGGGGCGCGCCCTGCAGGTGTACAAAAAAGCAGGCTTTAAAG-3'

**GA1**

5'-CCGGGTCTCGGGCATAATGCCAACTTTGTACAAGAAAG-3'

**Primers for Fragment 2 (gRNA2):**

**GS2**

5'-CCGGGTCTCGTGCCTGTACAAAAAAGCAGGCTTTAAAG-3'

**GA2**

5'-CCGGGTCTCGTTGCTAATGCCAACTTTGTACAAGAAAG-3'



**Primers for Fragment 3 (gRNA3):****GS3**

5'-CCGGGTCTCGGCAATGTACAAAAAAGCAGGCTTAAAG-3'

**GA3**

5'-CCACCAGTGTTCATGGGGCGCGCCGTTTAAACTAATGCCAACTTTGTACAAGAAAG-3'

Blue, *Bsa*I site; green, *Bst*XI site; red, *Asc*I site; purple, sequences in gRNA Cloning Vector *Bbs*I. The primer sequences including a *Bst*XI site (GS1 and GA3) are used to clone tandem gRNAs into the pCR2.1 vector.

2. Perform PCR to amplify fragments 1, 2, and 3 separately. For example, to amplify fragment 1, prepare a reaction mixture containing 20  $\mu$ L of 5 $\times$  Q5 Reaction Buffer, 2  $\mu$ L of 10-mM dNTPs, 1  $\mu$ L of Q5 Hot Start High-Fidelity DNA Polymerase, and 73  $\mu$ L of DW. Then, add 1  $\mu$ L of GS1 primer (100 pmol/ $\mu$ L), 1  $\mu$ L of GA1 primer (100 pmol/ $\mu$ L), and 2  $\mu$ L of gRNA Cloning Vector *Bbs*I containing the gRNA1 fragment (5 ng/ $\mu$ L) prepared in Subheading 3.2.5 as template DNA.
3. Perform PCR at 98  $^{\circ}$ C for 30 s, followed by 25 cycles of 98  $^{\circ}$ C for 10 s, 60  $^{\circ}$ C for 30 s, 72  $^{\circ}$ C for 30 s, and then 72  $^{\circ}$ C for 2 min.
4. Confirm that a PCR product of about 400 bp is amplified by agarose gel electrophoresis.
5. Purify the remaining PCR products with a PCR purification kit (*see Note 6*).
6. Digest each purified PCR product with *Bsa*I for fragment 2 and with *Bsa*I and *Bst*XI for fragments 1 and 3.
7. Purify with a PCR purification kit and elute with 10  $\mu$ L of EB.
8. Digest the pCR2.1 vector with *Bst*XI and purify it with a gel extraction kit after agarose gel electrophoresis (*see Note 7*).
9. Mix Solution I from DNA Ligation Kit Ver. 2.1 with each of the three PCR products and digested pCR2.1 vector (*see Note 8*).
10. Incubate at 16  $^{\circ}$ C for 4 h to overnight.
11. Transform competent cells using the reaction mixture described above (1–2  $\mu$ L), and spread them over solid medium containing ampicillin to select transformants.
12. Perform colony PCR using the primer set Universal M13 Forward (–20) and Universal M13 Reverse to confirm that the target sequence is present.

13. Pick the colonies that tested positive by colony PCR from the replica plate. Place them into 3 mL of LB liquid medium containing ampicillin and shake overnight at 37 °C.

14. Extract the plasmid from the culture using a plasmid miniprep kit.

### 3.2.7 Preparation of Tandem gRNA Insert

1. Digest the pCR2.1 vector containing tandem gRNAs with *AscI* at 37 °C for 3 h.

2. Extract the digested vector with a gel extraction kit after agarose gel electrophoresis.

### 3.2.8 Preparation of pPlatTET-gRNA2 Vector

1. Digest 3 µg of pPlatTET-gRNA2 (Fig. 2) with *AscI* at 37 °C for 3 h.

2. Perform CIAP treatment in a total volume of 100 µL (*see Note 9*).

3. Excise a band that migrates at 13.9 kb by agarose gel electrophoresis and purify it with a gel extraction kit.

### 3.2.9 Ligation and Transformation

1. Ligate the tandem gRNA fragment into the linearized pPlatTET-gRNA2 vector.

2. Transform competent cells using the reaction mixture, plate the transformants onto LB agar containing kanamycin, and incubate at 37 °C overnight.

3. Perform colony PCR using the primer set pPlatTET-gRNA2-S10696 and pPlatTET-gRNA2-AS11292 to confirm that the target sequence is present (product size: tandem gRNA length + 183 bp).

4. Pick the colonies that tested positive by colony PCR from the replica plate. Place them into 3 mL of LB liquid medium containing kanamycin and shake overnight at 37 °C.

5. Extract the plasmid from the culture using a plasmid miniprep kit.

6. Confirm that the target sequence has been inserted by performing sequencing analysis using the primer set described above (*see Note 10*).

## 3.3 Microinjection of Embryos

1. Linearize the epigenome editing vector containing a single gRNA or multiple gRNAs by *ApaI* digestion.

2. Purify with a DNA purification kit (*see Note 11*).

3. Place several flat droplets of M2 medium (10–15 µL) on the center of the lid of a 100-mm plastic dish, and then place droplets of linearized vector (1 µL, 5–35 ng/µL) (*see Note 12*) to the lower right of these droplets, and cover them with paraffin oil.

4. Move the tip of an injection pipet to a vector drop, and fill the tip with vector by maintaining negative pressure on the syringe for a few minutes.
5. Using an embryo transfer pipet, transfer 20–30 eggs to the droplet of M2 medium in the chamber.
6. Place the holding pipet next to an embryo and affix the embryo to the tip of the pipet by applying a gentle vacuum.
7. Keeping the holding pipet steady, adjust the focus of the microscope to confirm the position of the pronuclei.
8. Carefully insert the injection pipet into the pronuclei and inject the linearized vector.
9. Once the vector buffer is released, quickly remove the injection pipet from the egg.
10. Release the egg from the holding pipet and place the injected egg into a separate area of the droplet of M2 medium.
11. When all the eggs in the droplet have been injected, place them back into a droplet of M16 medium at 37 °C, and transfer new eggs into the droplet of M2 medium in the injection chamber.
12. After all the eggs have been injected, incubate them at 37 °C for an additional 30 min.
13. Separate surviving eggs from any that have lysed (*see Note 13*).
14. Culture eggs in M16 medium at 37 °C under 5% CO<sub>2</sub> in air until embryo transfer or other manipulations (*see Note 14*).
15. Transgenic embryos and pups can be detected by PCR using the primer set dCas9-F and dCas9-R (product size: 202 bp). In addition, expression of epigenome editing factors can be detected by analyzing superfolder GFP expression (*see Note 15*).

### 3.4 DNA Methylation Analysis (Adapted from [9])

Treat genomic DNA with bisulfite using, for example, an EpiTect Plus DNA Bisulfite Kit (Qiagen), according to the manufacturer's instructions.

#### 3.4.1 Bisulfite Modification

#### 3.4.2 Bisulfite PCR Amplification

1. Design bisulfite PCR primers using MethPrimer (<http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi>) (*see Note 16*).
2. Perform bisulfite PCR amplification. For the resulting PCR products, determine the percentage of demethylation at the target site using combined bisulfite restriction analysis (COBRA) [10, 11] or bisulfite sequencing.

### 3.4.3 COBRA

The COBRA technique is based on restriction digestion of PCR products using an enzyme whose recognition sequence is affected by the methylation state of the template DNA. Use of a combination of bisulfite treatment and PCR amplification results in conversion of unmethylated cytosine residues to thymine residues and methylated cytosine residues to cytosine residues. This sequence conversion can lead to the methylation-dependent creation of new restriction enzyme sites or methylation-dependent retention of pre-existing sites. The user must find the restriction enzyme sites within the methylated bisulfite-converted strand that are absent in the unmethylated bisulfite-converted strand or vice versa.

The following restriction enzymes are commonly used in COBRA:

- *Bst*UI: CG/CG
- *Cla*I: AT/CGAT
- *Mlu*I: A/CGCGT
- *Taq*I: T/CGA

1. Digest PCR amplicons with the restriction enzyme and separate the fragments by gel electrophoresis.
2. Calculate the percentage of methylation based on the ratio of cleaved and uncleaved DNA determined by densitometric analysis of gels stained with ethidium bromide.

### 3.4.4 Bisulfite Sequencing

PCR products generated by bisulfite PCR amplification can be cloned and sequenced. Following bisulfite PCR amplification, unmethylated cytosine residues in genomic DNA are converted to thymine residues, and methylated cytosine residues are converted to unmethylated cytosine residues.

1. Ligate PCR products into the TA cloning vector and transform them into competent cells.
2. Perform colony PCR on the following day to detect cloned PCR products. At least 16 clones should be sequenced.
3. Analyze sequences using the methylation analysis tool QUantification tool for Methylation Analysis (QUMA, <http://quma.cdb.riken.jp/top/index.html>).

---

## 4 Notes

1. Ensure that the protospacer adjacent motif (PAM) sequence (NGG) is not included in this 20 mer.
2. We use CRISPRdirect (<http://crispr.dbcls.jp/>) and CRISPOR (<http://crispor.tefor.net/>) as CRISPR design tools.

3. Transcription with T7 RNA polymerase starts at the G in the promoter sequence. Therefore, the first base in the transcript should be G.
4. We generally use a QIAquick Gel Extraction Kit (Qiagen) to purify the digested vector.
5. Sequence errors are rare and therefore it is sufficient to pick three or four colonies.
6. We generally use a MinElute PCR Purification Kit (Qiagen) to purify PCR products.
7. Any vector containing two *Bst*XI sites can be used. *Bst*XI sequences in the text and Fig. 4 are applicable when using the pCR2.1 vector. If you use other vectors, you should change these *Bst*XI sequences to those used in other vectors.
8. We first insert the tandem gRNA fragment into the *Bst*XI site of the pCR2.1 vector for easy handling. In principle, a tandem gRNA fragment can be inserted directly into the *Asc*I site of the pPlatTET-gRNA2 vector.
9. CIAP treatment is performed to prevent self-ligation of the vector. It is not necessary to purify the DNA before performing gel electrophoresis.
10. Tandem gRNAs are inserted into the *Asc*I site of the pPlatTET-gRNA2 vector in the forward or reverse direction.
11. We generally use an AllPrep DNA/RNA Mini Kit (Qiagen).
12. We usually inject linearized vector at a concentration of 35 ng/ $\mu$ L. If the embryonic development rate is extremely low, reduce the concentration of the vector.
13. Surviving eggs have a distinct outline compared with lysed eggs.
14. Vector-injected embryos tend to develop at a slower rate in vitro. It is better to transfer the embryos to oviducts as soon as possible after microinjection. Alternatively, embryos can be cultured until the blastocyst stage to determine the developmental rate in vitro or the epigenome editing efficiency.
15. The fluorescence intensity of transgenic mice into which pPlatTET-gRNA2 has been inserted is lower than that of commonly used EGFP mice.
16. PCR amplification of bisulfite-treated DNA converts uracil residues to thymine residues and methylated cytosine residues to unmethylated cytosine residues. MethPrimer is a program to design bisulfite conversion-based methylation PCR primers. The user only has to input the original source sequence.

## Acknowledgments

We thank Mika Kimura (Gunma University) for critical reading of the manuscript. This work was supported by funding from the Basic Science and Platform Technology Program for Innovative Biological Medicine from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, to T.H., S.M., and I.H.; the Japan Agency for Medical Research and Development (AMED) under Grant Number 21gm6310009h0003 to S.M.; the Platform Project for Supporting Drug Discovery and Life Science Research (Basis for Supporting Innovative Drug Discovery and Life Science Research (BINDS)) from AMED under Grant Number JP21am0101120 to I.H.; the Tokyo Biochemical Research Foundation to I.H.; the Takeda Science Foundation to I.H.; the Practical Research Project for Rare/Intractable Diseases from AMED to I.H.; and the Life Science Institute, Inc. (LSII).

## References

1. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816–821
2. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. *Science* 339:823–826
3. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339:819–823
4. Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159:635–646
5. Morita S, Noguchi H, Horii T, Nakabayashi K, Kimura M, Okamura K, Sakai A, Nakashima H, Hata K, Nakashima K, Hatada I (2016) Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat Biotechnol* 34: 1060–1065
6. Morita S, Horii T, Kimura M, Hatada I (2020) Synergistic upregulation of target genes by TET1 and VP64 in the dCas9-SunTag platform. *Int J Mol Sci* 21:1574
7. Horii T, Morita S, Hino S, Kimura M, Hino Y, Kogo H, Nakao M, Hatada I (2020) Successful generation of epigenetic disease model mice by targeted demethylation of the epigenome. *Genome Biol* 21:77
8. Hogan B, Costantini F, Lacy E (eds) (1986) *Manipulating the mouse embryo*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor
9. Morita S, Horii T, Hatada I (2018) Editing of DNA methylation using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Methods Mol Biol* 1767:419–428
10. Xiong A, Laird PW (1997) COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res* 25:2532–2534
11. Eads CA, Laird PW (2002) Combined Bisulfite Restriction Analysis (COBRA). DNA methylation protocols. Volume 200 of the series *Methods in molecular biology*<sup>TM</sup> 71–86



## In Vivo Tissue-Specific DNA Demethylation in Mouse Liver Through a Hydrodynamic Tail Vein Injection

Koshi Hashimoto and Nozomi Hanzawa

### Abstract

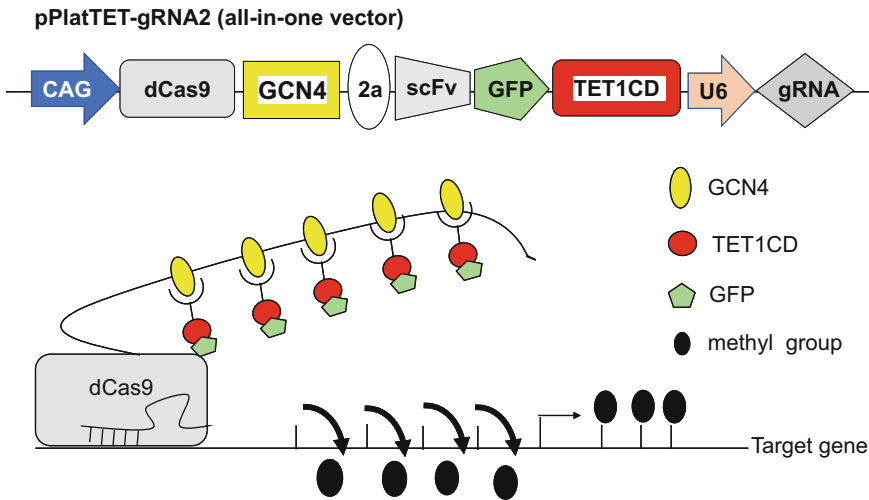
A new technique called the dCas9-SunTag and scFv-TET1CD epigenome editing system has recently been developed to edit the DNA methylation status of specific genes. The transfection of an all-in-one vector containing this system into cells is feasible and induces the DNA demethylation of specific genes; however, due to the large size of the vector, difficulties are associated with its introduction into mice. We herein used a hydrodynamic tail vein injection (HTVi) to introduce the all-in-one vector into mice for in vivo epigenome editing. HTVi needs to be considered for inducing the targeted DNA demethylation of particular genes in the mouse liver.

**Key words** Hydrodynamic tail vein injection (HTVi), DNA demethylation, In vivo epigenome editing, dCas9-SunTag, scFv-TET1CD

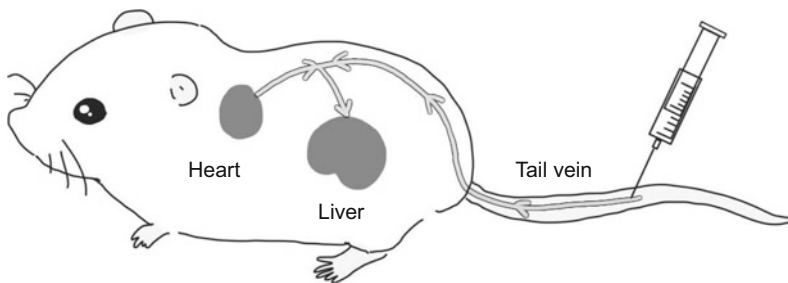
---

### 1 Introduction

In mammalian cells, DNA methylation is a major epigenetic modification that regulates gene expression without altering DNA sequences and, thus, plays a pivotal role in a myriad of physiological and pathological processes, including cell development and differentiation, genome imprinting, and tumorigenesis [1]. To obtain a more detailed understanding of the physiological and functional significance of the DNA methylation status of specific genes, a new technique to edit the DNA methylation status of specific genes using dead Cas9 (dCas9) fused with a SunTag has recently been developed [2, 3]. SunTag consists of a tandem repeat of 5 copies of a 19-amino acid GCN4 peptide separated by a 22-amino acid linker, which recruits several copies of the single-chain variable fragment (scFv) of the anti-GCN4 antibody-fused Ten Eleven Translocation 1 catalytic domain (TET1CD) fusion protein [2, 3], thereby promoting DNA demethylation [2]. This dCas9-SunTag and scFv-TET1CD epigenome editing system, hereafter



**Fig. 1** Structure of the vector components of pPlatTET-gRNA2, all-in-one vector (above). A scheme of target gene DNA demethylation employing dCas9-SunTag and scFv-TET1 catalytic domain (CD) system. dCas9 fused to a peptide repeat can recruit multiple copies of the antibody (scFv)-fused TET1CD, allowing multiple copies of the TET1CD to demethylate the target gene efficiently (below). CAG CAG promoter, U6 U6 promoter, 2A 2A peptide

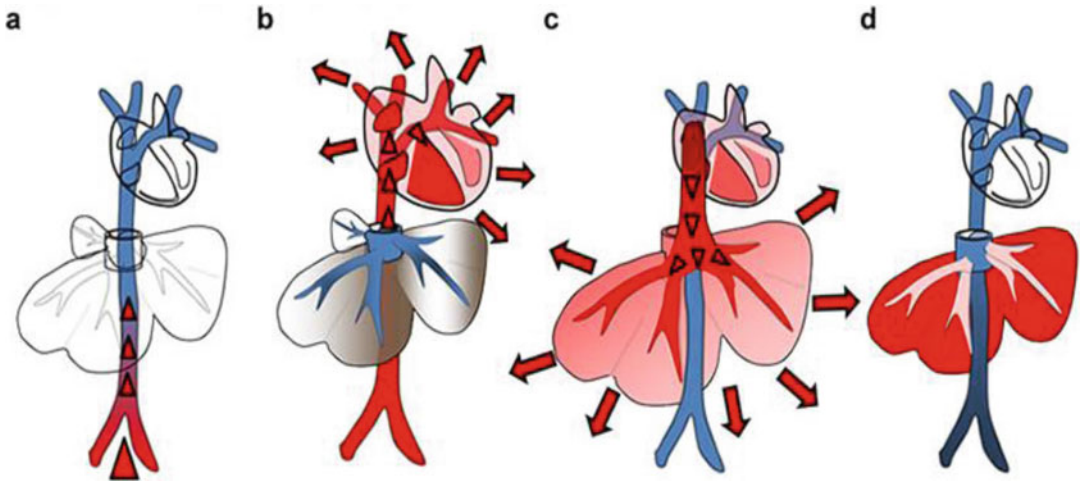


**Fig. 2** Schematic figure of HTVi

referred to as the dCas9-SunTag and scFv-TET1CD system, allows us to induce the DNA demethylation of specific genes (Fig. 1). We transfected the vector, which contains the dCas9-SunTag and scFv-TET1CD system and is referred to as an “all-in-one vector,” into cells and induced the DNA demethylation of specific target genes [4].

However, difficulties are associated with the introduction of the all-in-one vector into mice because it is relatively large at approximately 14 kb and it is challenging to constitute a virus vector. We herein employed a hydrodynamic tail vein injection (HTVi) to introduce the all-in-one vector into mice for in vivo epigenome editing [4–8]. HTVi may be a useful method for inducing the targeted DNA demethylation of particular genes in the liver (Fig. 2).





**Fig. 3** Mechanism underlying HTVi into the mouse liver [12]. (Reprinted with permission from MDPI (Basel, Switzerland))

In HTVi, hydrodynamic pressure caused by the volume and flow of the injection is the driving force for intracellular gene transfer [9, 10]. Various organs show high transgene expression levels, with the highest levels being detected in the liver; around 40% of hepatocytes are transfected by a single injection of less than 50  $\mu\text{g}$  of plasmid DNA with a length of approximately 6 kb [11]. The mechanism underlying gene transfer through HTVi is shown in Fig. 3 [12]. The plasmid solution injected from the tail vein is directly delivered to the heart (Fig. 3a), inducing cardiac congestion and the accumulation of the solution in the inferior vena cava (IVC) and hepatic veins (Fig. 3b). This solution reaches the sinusoids, and hydrodynamic pressure expands the liver and dilates the fenestrae of the endothelium, which opens a hole in the cellular membrane of hepatocytes, thereby allowing the entry of plasmid DNA into the cytoplasm (Fig. 3c) [13]. In mice, the invaginated cellular membrane generally recovers within a few minutes, thereby trapping plasmid DNA within the cytoplasm. The enlarged liver is typically restored within 24 h (Fig. 3d) [14]. Serum hepatic enzyme levels, including aspartate aminotransferase, alanine aminotransferase, and lactate dehydrogenase, transiently increase after the injection and generally return to normal levels within a few days [14].

## 2 Materials

### 2.1 Animals

Eight-week-old male C57BL6J mice (body weight: approximately 20 g). They are allowed free access to water and food.

## 2.2 Plasmids for Targeted DNA Demethylation

1. The pPlatTET-gRNA2 (#82559, Addgene), all-in-one vector (14 kb), containing the dCas9-SunTag (linker length: 22 aa), scFV-sfGFP-TET1CD, and guide (g) RNA expression system [4] (*see Note 1*).
2. pMAX-GFP (Lonza, Basel, Switzerland) (3.5 kb).

## 2.3 Preparation of Vector

1. Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB).
2. Gibson Assembly Master Mix (NEB).
3. Restriction enzyme, *AflIII* (NEB).
4. LB agar with kanamycin: 1.6% agar, 1% tryptone, 0.5% yeast extract, 1% NaCl, 50- $\mu$ g/mL kanamycin (*see Note 2*).
5. LB: 1% tryptone, 0.5% yeast extract, 1% NaCl, 50- $\mu$ g/mL kanamycin (*see Note 2*).
6. Sequencing primer, U6 promoter Fw:5'- GACTATCA TATGCTTACCGT-3'.
7. DNA sequencing kit.
8. Thermal cycler.

## 2.4 Components Used for HTVi

1. The EndoFree Plasmid Maxi Kit (#12362, QIAGEN).
2. TransIT-EE Hydrodynamic Delivery Solution (#MIR5340, Takara).
3. A restrainer to fix the mouse tail.
4. A 45–50 °C water bath .
5. A 2.5-ml syringe.
6. A 27-gauge needle (0.5 inches).
7. A 70% ethanol (EtOH) swab.

## 2.5 Bisulfite Sequencing (BS) Analysis

1. AllPrep DNA/RNA Mini Kit (#80204, QIAGEN).
2. EpiTect Bisulfite Kit (#59104, QIAGEN).
3. pGEM-T easy vectors (#A1360, Promega).

---

## 3 Methods

### 3.1 Vector Construction

Guide (g) RNA design is crucial for efficient DNA demethylation of specific genes; therefore validation of gRNA must be required.

#### 3.1.1 gRNA Design

1. Enter the sequence flanking the region where DNA demethylation is to be induced into a CRISPR/Cas9 target design tool such as CRISPR website from Dr. Feng Zhang's laboratory (<http://crispor.tefor.net>) to select for a highly specific target sequence (20 mer) (*see Note 3*).

- Design and synthesize the oligonucleotides of sense (**S**) and antisense (**AS**) of the specific target sequence (20 mer) with sequences listed below. These sequences include a part of the U6 promoter (underlined) or gRNA tail (*see Note 4*).

5'-TTTCTTGGCTTTATATATCTTGTGGAAAGGACG  
AAACACC-S-3'

5'-GACTAGCCTTATTTTAACTTGCTATTTCTAGCTC  
TAAAAC-AS-3'

### 3.1.2 Preparation of the Double-Stranded Oligonucleotide

Mix Phusion Master Mix (2 × 5 μL) with 1 μL of each oligonucleotide (5 pmol/μL) and 3 μL of dH<sub>2</sub>O. Incubate at 98 °C for 30 s, 50 °C for 30 s, and 72 °C for 3 min, using thermal cycler, and finally add 20 μL of dH<sub>2</sub>O.

### 3.1.3 gRNA Cloning into pPlatTET-gRNA2, All-in-One Vector

- Digest 1-μg pPlatTET-gRNA2 (#82559, Addgene) with *AflIII* at 37 °C for 1 h, and then heat to 65 °C for 20 min to inactivate the enzyme.
- Mix the plasmid digested with *AflIII* (100 ng) as above with 0.5-μL oligonucleotide mixture prepared in Subheading 3.1.2 and 5-μL 2 × Gibson Assembly Master Mix, and fill up to 10 μL with dH<sub>2</sub>O. Incubate at 50 °C for 60 min.
- Perform bacterial transformation using the reaction mixture described above (1–2 μL) into competent *Escherichia coli* such as DH5α, and spread them over LB agar plate with kanamycin to select for successfully transformed clones.
- After overnight incubation of the plate at 37 °C, pick up a few colonies from the plate, and confirm the sequence of some colonies by conventional DNA sequencing using a sequence primer, U6 promoter Fw.
- Culture the validated colony and purify the plasmid.

## 3.2 Hydrodynamic Tail Vein Injection of the All-in-One Vector

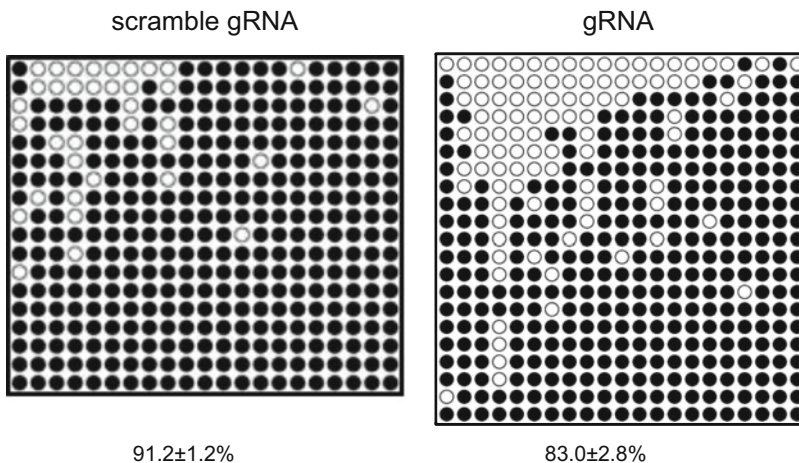
- Prepare the plasmid using the EndoFree Plasmid Maxi Kit (*see Note 5*).
- Prepare a male 8-week-old C57BL6J mouse.
- Dilute 100 μg of the all-in-one vector or pMAX-GFP in TransIT-EE Hydrodynamic Delivery Solution to a volume equivalent to ~10% of the mouse's body weight. Use the solution within 30 min of its preparation (*see Note 6*).
- Prepare a plasmid solution and aspirate 2.1 mL for each injection with a 2.5-mL syringe using a 27G needle (*see Note 7*).
- (*Optional*) Warm the mouse to dilate its blood vessels using a heat pack and incandescent lamp.
- Fix the mouse using a retainer and disinfect the tail with a 70% EtOH swab.
- Warm the tail with lukewarm (45–50 °C) water for 40–50 s.

8. Place the pre-warmed tail in an easy-to-inject position.
9. Disinfect the tail with a 70% EtOH swab.
10. Insert the needle into the tail vein (Fig. 2) (*see Note 8*).
11. Inject the solution within 7 ~ 8 s at a regular speed, which is very important (confirm the time using an alarm tone) (*see Note 9*).
12. A longer injection time may lead to failure, whereas a shorter time may damage the liver.
13. Release the mouse from the retainer and place it into a mouse-raising cage as soon as possible after the injection.
14. The mouse is expected to recover within 1 h.

### 3.3 Bisulfite Sequencing (BS) Analysis

To evaluate DNA demethylation in the mouse liver, bisulfite sequencing (BS) analysis is employed (Fig. 4).

1. Extracted genomic DNA from the mouse liver by an AllPrep DNA/RNA Mini Kit.
2. Use 2  $\mu\text{g}$  of the genomic DNA for bisulfite treatment using an EpiTect Bisulfite Kit.
3. Perform sequential PCR amplification of specific genes, to which DNA demethylation is induced, using specific primers designed with a web tool, MethPrimer (<http://www.urogene.org/methprimer/>) [15]. The reaction profiles are 40 cycles of 98 °C for 10 s, 54 °C for 30 s, and 72 °C for 120 s.



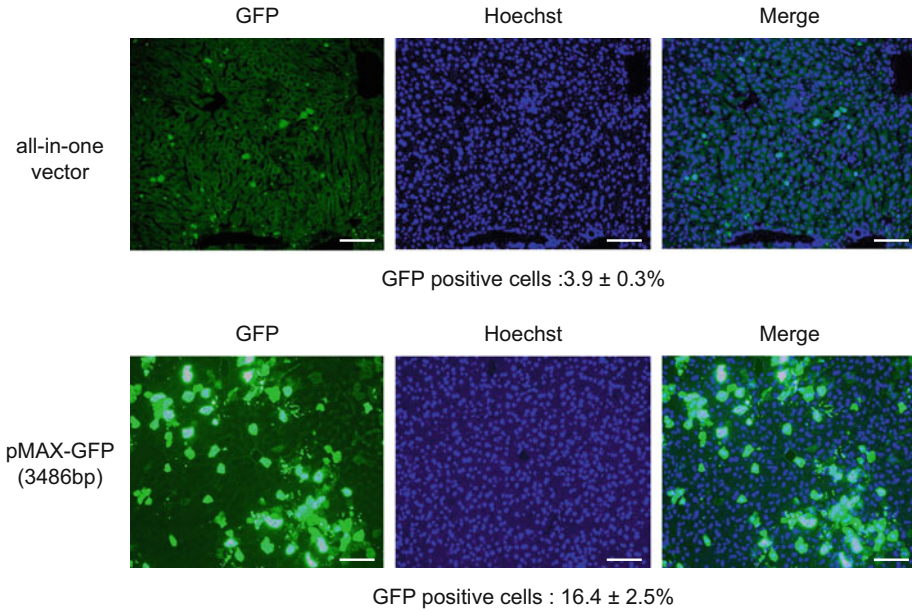
**Fig. 4** The all-in-one vector with scramble gRNA and gRNA into the liver of peroxisome proliferator-activated receptor  $\alpha$ -deficient mice, in which the DNA methylation status of the fibroblast growth factor 21 (*Fgf21*) gene is high, via HTVi. A bisulfite sequencing analysis of the *Fgf21* promoter on day 6 after HTVi ( $n = 5\text{--}9$  per group). Closed and open circles indicate methylated and unmethylated CpGs, respectively. Representative data of four independent experiments are shown. % DNA methylation is indicated below as means  $\pm$  SD [4]

4. Ligate the above amplified fragments into pGEM-T easy vectors (#A1360, Promega). At least 14 clones should be sequenced per reaction.
5. Use a web-based quantification tool for DNA methylation analysis for BS analysis of CpG methylation (<http://quma.cdb.riken.jp/>) [16].

---

## 4 Notes

1. pPlatTET-gRNA2 plasmid, which contains Cas9, GCN4 peptide array (linker length: 22aa), antibody-sfGFP-TET1CD, and gRNA expression system, is available at Addgene (Plasmid #82559). The aa sequence of GCN4 used was EELLSKNYH-LENEVARLKK. The sequence of the linker between each GCN4 peptide unit is GSGSGSGSGSGSGSGSGSGSGSG. All the fusion proteins are expressed under the control of the CAG promoter. The sequence of 2A peptide is GSGATNFSLLK-QAGDVEENPGP (Fig. 1) [2, 4].
2. Kanamycin should be added from a 10-mg/mL stock solution after the autoclaved LB agar media has cooled down.
3. When one gRNA is employed, DNA demethylation can be induced at least 100 bp from target site. Note that the protospacer adjacent motif (PAM) sequence (NGG) should not be included in a highly specific target sequence (20 mer). The U6 promoter absolutely needs a guanine (G) base at the transcription start site; therefore, if the base at the 5' end of the 20-mer sequence (sense strand: S) is not guanine, design to change the base to G (or add a guanine at the 5' end of gRNA sequence). In that case, likewise, change the base to cytosine (C) or add C at the 3' end of gRNA sequence (antisense strand: AS).
4. When all-in-one vector is digested with *AflIII*, U6 promoter is destroyed. However these double-stranded oligos are inserted into vectors; U6 promoter is complemented.
5. The purity of plasmid DNA needs to be high ( $A_{260}/A_{280} > 1.8$ ) and free of endotoxins and proteins.
6. For example, for a mouse weighing 20 g, dilute 100  $\mu$ g of the plasmid vector in 2.1 mL of the TransIT-EE Hydrodynamic Delivery Solution. Consider the residual solution ( $\sim 0.1$  mL) in the needle and syringe after the injection.
7. Avoid air at aspiration.
8. Hold the syringe parallel to the tail vein and insert the needle into the tail vein as close to the tip as possible (Fig. 2). To confirm that the needle is correctly inserted, inject a small amount of the solution. When the needle is correctly inserted,



**Fig. 5** GFP fluorescence and Hoechst staining 2 days after HTVi [4]. GFP-positive cells are indicated as means  $\pm$  SD.  $n = 4$  per group. Scale bars indicate 100  $\mu\text{m}$  (see **Notes 10** and **11**)

the color of the vein will be light. If there is pressure against the plunger of the syringe, the needle is not correctly inserted, and the tail will become discolored and expand. In this case, remove the needle and insert it into another region.

9. Nasal bleeding after the injection indicates a successful injection.
10. Employ pMAX-GFP as a control vector, and the presence of approximately 16 ~ 20% GFP-positive cells suggests that HTVi was successfully performed (Fig. 5).
11. Regarding the all-in-one vector, even a transfection efficiency of 4% may be effective for DNA demethylation induced by HTVi (Fig. 5) [4].

## References

1. Robertson KD, Wolffe AP (2000) DNA methylation in health and disease. *Nat Rev Genet* 1: 11–19
2. Morita S, Horii T, Hatada I (2018) Editing of DNA methylation using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Methods Mol Biol* 1767:419–428
3. Morita S, Noguchi H, Horii T, Nakabayashi K, Kimura M, Okamura K, Sakai A, Nakashima H, Hata K, Nakashima K, Hatada I (2016) Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat Biotechnol* 34: 1060–1065
4. Hanzawa N, Hashimoto K, Yuan X, Kawahori K, Tsujimoto K, Hamaguchi M, Tanaka T, Nagaoka Y, Nishina H, Morita S, Hatada I, Yamada T, Ogawa Y (2020) Targeted DNA demethylation of the Fgf21 promoter by CRISPR/dCas9-mediated epigenome editing. *Sci Rep* 10:5181



5. Hickman MA, Malone RW, Lehmann-Bruinsma K, Sih TR, Knoell D, Szoka FC, Walzem R, Carlson DM, Powell JS (1994) Gene expression following direct injection of DNA into liver. *Hum Gene Ther* 5:1477–1483
6. Rossmanith W, Chabicovsky M, Herkner K, Schulte-Hermann R (2002) Cellular gene dose and kinetics of gene expression in mouse livers transfected by high-volume tail-vein injection of naked DNA. *DNA Cell Biol* 21:847–853
7. Budker VG, Subbotin VM, Budker T, Sebestyen MG, Zhang G, Wolff JA (2006) Mechanism of plasmid delivery by hydrodynamic tail vein injection. II. Morphological studies. *J Gene Med* 8:874–888
8. Miyamura N, Hata S, Itoh T, Tanaka M, Nishio M, Itoh M, Ogawa Y, Terai S, Sakaida I, Suzuki A, Miyajima A, Nishina H (2017) YAP determines the cell fate of injured mouse hepatocytes in vivo. *Nat Commun* 8:16017
9. Liu F, Song Y, Liu D (1999) Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA. *Gene Ther* 7:1258–1266
10. Zhang G, Budker V, Wolff JA (1999) High levels of foreign gene expression in hepatocytes after tail vein injections of naked plasmid DNA. *Hum Gene Ther* 10:1735–1737
11. Zhang G, Song YK, Liu D (2000) Long-term expression of human alpha1-antitrypsin gene in mouse liver achieved by intravenous administration of plasmid DNA using a hydrodynamics-based procedure. *Gene Ther* 15:1344–1349
12. Kamimura K, Yokoo T, Abe H, Kobayashi Y, Ogawa K, Shinagawa Y, Inoue R, Terai S (2015) Image-guided hydrodynamic gene delivery: current status and future directions. *Pharmaceutics* 7:213–223
13. Suda T, Liu D (2007) Hydrodynamic gene delivery: its principles and applications. *Mol Ther* 12:2063–2069
14. Suda T, Gao X, Stolz DB, Liu D (2007) Structural impact of hydrodynamic injection on mouse liver. *Gene Ther* 2:129–137
15. Li LC, Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18:1427–1431
16. Kumaki Y, Oda M, Okano M (2008) QUMA: quantification tool for methylation analysis. *Nucleic Acids Res* 36:W170–W175



## Targeted Manipulation of Histone Modification in Medaka Embryos

Hiroto S. Fukushima, Hiroyuki Takeda, and Ryohei Nakamura

### Abstract

Recent development of targeted manipulation of histone modification enables us to experimentally and directly test the functional relevance of histone modifications accumulated at specific genomic regions. In particular, dCas9 epigenome editing has been widely used for site-specific manipulation of epigenetic modification. Here, we describe how to apply dCas9 epigenome editing in fish (medaka, *Oryzias latipes*) embryos and how to analyze induced changes in histone modification.

**Key words** Epigenome editing, Histone modification, H3K27me3, Fish, Medaka, ChIP, Chromatin, In vivo

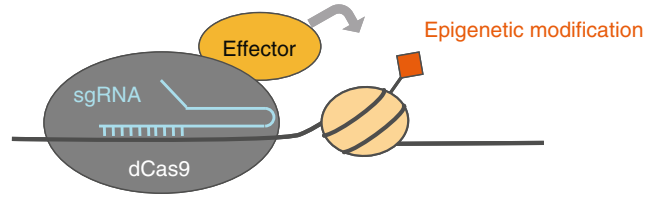
---

### 1 Introduction

Recent development of site-specific manipulation of epigenetic modification enables us to experimentally and directly test the functional relevance of epigenetic modifications accumulated at specific genomic regions [1]. One of the prevalent technologies is to use DNA-targeting proteins such as ZF [2–5], TALE [6, 7], and dCas9 [1] combined with epigenetic effector proteins, which are specifically called “epigenome editing.” In particular, dCas9 is the most popular method for site-specific manipulation of epigenetic modification recently [1].

In dCas9 epigenome editing, fusion proteins of dCas9 and any effector (dCas9 effector) are recruited to targeted genomic regions by single-guide RNAs (sgRNAs); thereby ectopically add or remove epigenetic modifications (Fig. 1) [1]. For instance, dCas9 fused to a histone acetyltransferase p300 (dCas9-p300) can site-specifically accumulate H3K27ac at any promoters or enhancers and activate transcription of the downstream gene [8]. However, most of the





**Fig. 1** Schematics showing epigenome editing. A fusion protein of dCas9 with effector domain is recruited to a targeted genomic region by sgRNAs; thereby site-specifically add or remove epigenetic modifications

previous studies used culture cells, or in vitro, and the number of applications of dCas9 epigenome editing in living organisms, or in vivo, is still limited. Here, we describe the method of dCas9 epigenome editing to site-specifically edit histone modifications in Japanese killifish (medaka, *Oryzias latipes*) embryos.

## 2 Materials

### 2.1 Preparation or Construction of dCas9-Effector Fusions

1. Any dCas9-effector construct such as dCas9-olEzh2 (Addgene, #180216).
2. T7 ARCA mRNA Synthesis Kit (with tailing) (NEB, E2060).
3. RNA purification kit such as RNeasy Mini Kit (Qiagen, 74104).

### 2.2 Designing Targets and Preparation of sgRNA

1. sgRNA oligos (*see* Subheading 3.2 on how to make oligos).
2. pDR274 (Addgene, #42250).
3. BsaI-HFv2 (NEB, R3733).
4. 10× Cut Smart Buffer (included in BsaI-HFv2).
5. DNA purification kit such as Wizard SV Gel and PCR Clean-Up System (Promega, A9281).
6. 10× annealing buffer: 100-mM Tris-HCl, pH 8.0, 500-mM NaCl, 10-mM EDTA in Milli-Q.
7. Ligation high Ver.2 (TOYOBO, LGK-201).
8. Competent cell DH5 alpha.
9. Kanamycin LB plate.
10. Round tube for subculture.
11. Plasmid purification kit such as QIAprep Spin Miniprep Kit (QIAGEN, 27104).
12. 2% agarose gel.
13. Phusion High-Fidelity DNA Polymerase (NEB, M0530).

14. Primers for amplification of template for sgRNA in vitro transcription.

Primer list table:

	<b>Name</b>	<b>Sequence (5'-3')</b>
Forward	IVT sgRNA F	AAAAGCACCGACTCGGTG
Reverse	IVT sgRNA R	GGTCAGGTATGATTTAAATGGTCAGT

15. Dpn I.  
 16. HiScribe T7 Quick High Yield RNA Synthesis Kit (NEB, E2050).  
 17. RNA purification kit such as RNeasy Mini Kit (Qiagen, 74104).

### **2.3 mRNA Injection into Medaka One-Cell Stage Embryo**

1. 0.5% phenol red stock solution.
2. Medaka d-rR strain.

### **2.4 Crosslinking**

1. PBS: place PBS on ice before use.
2. PNPP: 1× PBS containing 20-mM sodium butyrate (Na-butyrate), 1-mM PMSF, and 1× proteinase inhibitor cocktail.

Stock list table:

<b>Name</b>	<b>Catalog #</b>	<b>Stock concentration</b>	<b>Solvent</b>	<b>Store</b>
Sodium butyrate	Sigma, 303410	2 M	Milli-Q	4 °C
PMSF	Roche, 10837091001	100 mM	100% ethanol	-20 °C
Proteinase inhibitor cocktail	Roche, 11873580001	50× (1 tablet in 1-mL Milli-Q)	Milli-Q	-20 °C

3. Formaldehyde solution.
4. 1.25-M glycine in PBS.

### **2.5 Sonication**

1. Lysis buffer: 50-mM Tris-HCl, pH 8.0, 10-mM EDTA, 1% (w/v) SDS in Milli-Q.
2. Lysis-ChIP buffer: lysis buffer containing 20-mM sodium butyrate (Na-butyrate), 1-mM PMSF, and 1× proteinase inhibitor cocktail (stocks are shown in Subheading 2.4, item 2).
3. Covaris microtube: microtube AFA Fiber Snap-Cap 6 × 16 mm (Covaris, 520045).
4. Covaris S220.

5. PCI: phenol/chloroform/isoamyl alcohol, 25:24:1.
6. Either 2.0% agarose gel or Agilent 2100 Bioanalyzer.
7. Qubit 2.0 Fluorometer (Invitrogen).
8. Qubit dsDNA HS Assay Kit (Invitrogen, Q32851).

## 2.6 ChIP

1. Dynabeads Protein A (Invitrogen, 10002D).
2. RIPA buffer: 10-mM Tris-HCl, pH 8.0, 140-mM NaCl, 1-mM EDTA, 0.5% EGTA, 1% Triton X-100, 0.1% (w/v) SDS, and 0.1% sodium deoxycholate in Milli-Q.
3. RIPA-ChIP buffer: RIPA buffer containing 20-mM sodium butyrate (Na-butyrate), 1-mM PMSF, and 1× proteinase inhibitor cocktail (stocks are shown in Subheading 2.4, item 2).
4. Antibodies (e.g., anti-FLAG antibody, Sigma F3165; anti-H3K27me3 antibody, Diagenode C15410069).
5. Magnetic stand.
6. ChIP TE buffer: 10-mM Tris-HCl, pH 8.0, 10-mM EDTA in Milli-Q.
7. 5-M NaCl.
8. Thermomixer.
9. 10-mg/mL RNase A.
10. 20-mg/mL Proteinase K (Sigma, P2308), diluted with Milli-Q.
11. Ethachinmate (NIPPON GENE, 318-01793) as a carrier of ethanol purification.
12. 100% and 70% ethanol.

## 2.7 ChIP-qPCR

1. Primer pairs for ChIP-qPCR.
2. THUNDERBIRD<sup>®</sup> SYBR<sup>®</sup> qPCR Mix (TOYOBO, QPS-201).
3. AriaMX Real-Time PCR.

## 2.8 ChIP-seq and Data Processing

1. KAPA HyperPrep Kit (KAPA Biosystems, KK8504).
2. Illumina adapters.

---

## 3 Methods

Hereafter, we explain the procedure of *in vivo* dCas9 epigenome editing, using manipulation of H3K27me3 by dCas9-olEzh2 (medaka's Ezh2 fused to dCas9) as examples. Ezh2 is a methyltransferase of H3K27, and dCas9-olEzh2 accumulates H3K27me3 at targeted regions specifically. In the following procedure, we

injected sgRNA cocktails with mRNA of either dCas9, dCas9-olEzh2, or dCas9-olEzh2( $\Delta$ SET) (olEzh2 lacking catalytic domain) or without mRNA to medaka embryos (*see Note 1*) and performed ChIP using anti-FLAG and anti-H3K27me3 to detect dCas9 recruitment and successive accumulation of H3K27me3, respectively. Finally, we confirm target-specific enrichment of FLAG and H3K27me3 by ChIP-qPCR and ChIP-seq. See more detailed results in [9] if necessary.

### 3.1 Preparation or Construction of dCas9-Effector Fusions

1. Order dCas9-effector fusion constructs on Addgene if available (*see Note 1*).
2. If not, make dCas9-effector fusion by yourself.
  - (a) dCas9 itself is available on Addgene.
  - (b) Adding any tag (e.g., 3 $\times$  FLAG) to dCas9 is highly recommended.
  - (c) Add T7 promoter sequence for in vitro transcription.
3. In vitro transcribe mRNA using T7 ARCA mRNA synthesis kit with poly(A) tailing.
4. Purify mRNA using RNeasy Mini Kit.

### 3.2 Designing Targets and Preparation of sgRNA

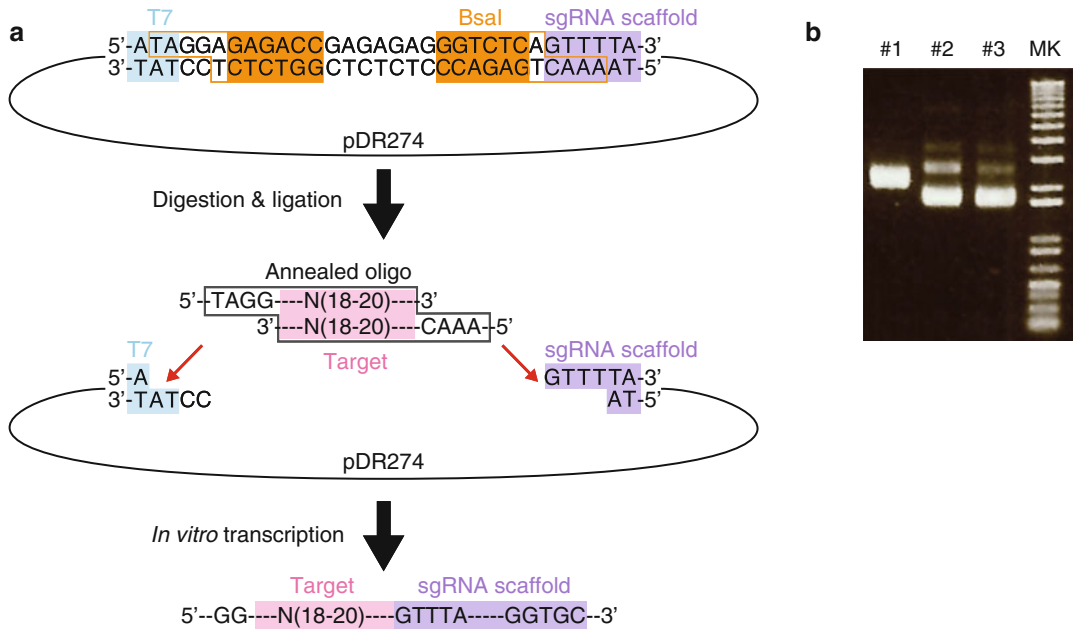
This protocol was slightly modified from the original study [10]. Here, we describe how to in vitro transcribe sgRNA using T7 promoter, but synthesized RNAs can be also used (*see Note 2*).

1. Check any epigenetic information such as chromatin accessibility and DNA methylation around your regions of interest if possible (*see Note 3*).
2. Design sgRNAs using CRISPR target online predictor CCTop [11].
  - (a) Forward: TAGG + N(18–20)
  - (b) Reverse: AAAC + N(18–20)

N(18–20) indicates the 18–20 bases matching the genomic sequences of target.
3. Linearize pDR274 by digesting with BsaI-HFv2 (Fig. 2).
  - (a) Assemble as follows:

pDR274	1 $\mu$ g
10 $\times$ Cut Smart Buffer	5 $\mu$ L
NEB's BsaI-HFv2	1 $\mu$ L
Milli-Q	Up to 50 $\mu$ L

- (b) Incubate at 37  $^{\circ}$ C overnight.
- (c) Purify linearized pDR274 by Wizard SV Gel and PCR Clean-Up System.



**Fig. 2** Construction of sgRNA vector. **(a)** sgRNA vector pDR274 has two BsaI sites between T7 promoter and sgRNA scaffold. After digestion by BsaI, annealed oligos can be complementarily ligated to digested pDR274. Here, N(18–20) indicates the 18–20 bases matching the target genomic sequences. The transcripts *in vitro* synthesized by T7 RNA polymerase include the target sequence and the scaffold of sgRNA. **(b)** Electrophoresis of plasmids picked up from LB plates after transformation of ligation of digested pDR274 with annealed oligos (MK, marker; #1, a plasmid without insertion; #2 & #3, plasmids with insertion)

4. Anneal oligos and ligate to linearized pDR274 (Fig. 2).

(a) Assemble as follows:

10× annealing buffer	1 μL
100-μM forward oligo	1 μL
100-μM reverse oligo	1 μL
Milli-Q	7 μL
Total	10 μL

(b) Incubation at 95 °C for 2 min and gradual cooling down to 25 °C ( $\Delta t/s = -1$  °C).

(c) Assemble as follows:

1/100× diluted products after (a)	1 μL
Linearized pDR274	25 ng
Ligation high ver2	2 μL
Milli-Q	Up to 6 μL

(d) Incubate at 16 °C for 30 min.

5. Transform the products into competent cells and spread on kanamycin LB plate.
6. Incubate at 37 °C for 16 h.
7. Pick colonies; subculture; purify plasmids by QIAprep Spin Miniprep Kit.
8. Digest plasmid with BsaI-HFv2, check digestion by agarose gel electrophoresis, and select the undigested ones (*see Note 4*) (Fig. 2).
9. Amplify templates for in vitro transcription from inserted plasmid by PCR using Phusion DNA polymerase with primers (IVT sgRNA F and IVT sgRNA R) and parameters as follows:
  - (a) Assemble as follows on ice:

5× Phusion HF buffer	10 μL
2.5-mM dNTP	4 μL
10-μM IVT sgRNA F	5 μL
10-μM IVT sgRNA R	5 μL
Plasmid	1 ng
Phusion high-Fidelity DNA polymerase	0.5 μL
Milli-Q	Up to 50 μL

- (b) PCR synthesis.

Initial denaturation	98 °C	10 s	
Denaturation	98 °C	5 s	× 35 cycles
Annealing and extension	63 °C	5 s	
Hold	4 °C	∞	

- (c) Add 2 μL of Dpn I and incubate at 37 °C for 2 h to degenerate template plasmids.
  - (d) Purify PCR products by Wizard SV Gel and PCR Clean-Up System.
10. In vitro transcribe sgRNAs using PCR-synthesized template and HiScribe T7 Quick High Yield RNA Synthesis Kit following the supplied protocol.
  11. Purify sgRNAs using RNeasy Mini Kit.

### 3.3 mRNA Injection into Medaka One-Cell Stage Embryo

1. Mate a male and a female, collect embryos 20 min after mating, and keep them on ice.
2. Inject dCas9-editor mRNA along with sgRNA cocktail (contents are found below) into one-cell stage medaka embryo on ice (*see Note 5*).

- (a) 350-nM dCas9-effector mRNA
  - (b) 120-ng/ $\mu$ L each sgRNA
  - (c) 0.025% phenol red
3. Dechorionate medaka embryos.
  4. Incubate dechorionated embryos at 28 °C with mild shaking until appropriate stages.

### 3.4 Crosslinking

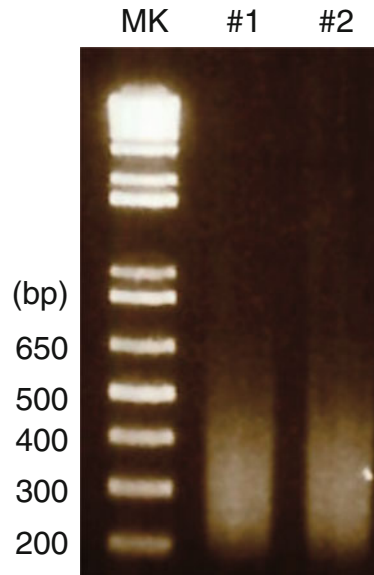
1. Transfer embryos into 500  $\mu$ L of ice-cold PBS in flat 1.5-mL or 2-mL tube; homogenize embryos by gently pipetting up and down using P1000 pipet (*see Note 6*).
2. Centrifuge at 4 °C, 500 g for 10 min; discard supernatant.
3. Add PNPP up to 700  $\mu$ L; add 20  $\mu$ L of formaldehyde solution to crosslink chromatin; vortex for 5 s; place tubes at R.T. for 8 min (*see Note 7*).
4. Add 80  $\mu$ L of 1.25-M glycine/PBS to quench crosslinking, vortex for 5 s, and place tubes on ice for 5 min.
5. Centrifuge at 4 °C, 500 g for 10 min; discard supernatant.
6. Add 500  $\mu$ L of ice-cold PNPP, gently vortex for 5 s to dissolve pellets, and centrifuge at 4 °C, 500 g for 10 min, and discard supernatant.
7. Add 500  $\mu$ L of ice-cold PNPP (without vortex), centrifuge at 4 °C, 500 g for 5 min, and discard supernatant.
8. Store dry pellet at  $-80$  °C.

### 3.5 Sonication

1. Resuspend the dry pellet in 130  $\mu$ L of lysis–ChIP buffer; transfer 130  $\mu$ L into Covaris microtube (*see Note 8*).
2. Sonicate chromatin using Covaris S220 with the following parameters (*see Note 9*).

Peak incident power	105 watts
Duty factor	4.0%
Cycles/burst	200
Duration	750 s

3. Transfer sonicated samples in 1.5-mL tube; centrifuge at 4 °C, 12,000 g for 10 min; transfer supernatant in new tubes; store sonicated samples at  $-80$  °C.
4. Dilute sonicated samples with ice-cold RIPA–ChIP buffer to  $\sim 1/5\times$  concentration.
5. Purify DNA from 10 to 20  $\mu$ L of products after **step 4** using PCI, and check the distribution of DNA by agarose gel electrophoresis or bioanalyzer to confirm successive chromatin shearing (ideally within 200–500 bp) (*see Note 10*) (Fig. 3).



**Fig. 3** Electrophoresis of fragmented chromatin after DNA purification (MK, marker; #1 & #2, DNA purified from chromatin in medaka blastula embryos)

6. Quantify the concentration of DNA in **step 5** using Qubit, and estimate the original concentration of sonicated chromatin stock stored at  $-80^{\circ}\text{C}$ .

### 3.6 ChIP

1. Pipet 30  $\mu\text{L}$  of Dynabeads Protein A to 1.5-mL tube; wash once with  $2.5\times$  volume and twice with  $2\times$  volume of ice-cold RIPA buffer (use magnetic stand for replacement of buffers).
2. Discard supernatant; add 300  $\mu\text{L}$  of ice-cold RIPA buffer ( $10\times$  volume).
3. Add 3  $\mu\text{g}$  of antibody (*see Note 11*).
4. Rotate the tubes containing antibody and beads overnight at  $4^{\circ}\text{C}$ .
5. On the next day, prepare diluted chromatin for ChIP and input (*see Note 12*).
  - (a) We usually dilute chromatin as follows:

Sonicated chromatin (the concentration was estimated in Subheading 3.5)	200 ng
Ice-cold RIPA-ChIP buffer	Up to 500 $\mu\text{L}$ per IP sample

- (b) Keep 100  $\mu\text{L}$  of diluted chromatin for input.
- (c) For example, if you are performing two immunoprecipitations (e.g., anti-FLAG and anti-H3K27me3),  $\geq 1100\text{-}\mu\text{L}$  ( $500\text{ }\mu\text{L} \times 2 + 100\text{ }\mu\text{L} \times 1$ ) diluted chromatin is required.



6. Place the tubes with antibody–beads complex on magnetic stand and remove supernatant. Add 500  $\mu\text{L}$  of prepared chromatin dilution.
7. Rotate the tubes overnight at 4  $^{\circ}\text{C}$ .
8. Keep the chromatin dilution for input at 4  $^{\circ}\text{C}$ .
9. On the next day, wash the beads captured chromatin with 500  $\mu\text{L}$  of ice-cold RIPA buffer on magnet for three times (*see Note 13*).
10. Wash immunoprecipitated samples once with 500  $\mu\text{L}$  of ice-cold ChIP TE buffer on magnet.
11. Discard supernatant of immunoprecipitated tubes; add 188- $\mu\text{L}$  lysis buffer and 12  $\mu\text{L}$  of 5-M NaCl (final 300-mM NaCl, total 200  $\mu\text{L}$ ).
12. Add 88  $\mu\text{L}$  of lysis buffer and 12  $\mu\text{L}$  of 5-M NaCl to 100  $\mu\text{L}$  of input (from **step 8**) (final 300-mM NaCl, total 200  $\mu\text{L}$ ).
13. Shake the tubes overnight using thermomixer (65  $^{\circ}\text{C}$ , 800 rpm) for reverse crosslinking.
14. On the next day, place the tubes on magnetic stand; transfer all supernatant ( $\sim 200$   $\mu\text{L}$ ) into new tubes; add 200  $\mu\text{L}$  of ChIP-TE buffer.
15. Add 8  $\mu\text{L}$  of 10-mg/mL RNase A; incubate at 37  $^{\circ}\text{C}$  for 2 h.
16. Add 4  $\mu\text{L}$  of 20-mg/mL Proteinase K; incubate at 55  $^{\circ}\text{C}$  for 2 h.
17. Add 400  $\mu\text{L}$  of PCI; vortex for 20 s; centrifuge at R.T., 14,000 rpm for 10 min; transfer the upper (aqueous) phase ( $\sim 400$   $\mu\text{L}$ ) into new tube.
18. Add 1  $\mu\text{L}$  of ethachinmate (or other carrier) and 16  $\mu\text{L}$  of 5-M NaCl; vortex well.
19. Add 1 mL of 100% ethanol; vortex well; store tubes at  $-80$   $^{\circ}\text{C}$  overnight (*see Note 14*).
20. Centrifuge at 4  $^{\circ}\text{C}$ , 15000 rpm for 25 min; discard supernatant.
21. Add 1 mL of 70% ethanol; invert tubes for two to three times.
22. Centrifuge at 4  $^{\circ}\text{C}$ , 15000 rpm for 5 min; discard supernatant; air-dry for 2–3 min; dissolve pellet with 20–40  $\mu\text{L}$  of Milli-Q.
23. Quantify the concentration using Qubit.
24. Test the enrichment of histone modifications by ChIP-qPCR and/or ChIP-seq as follows:

### 3.7 ChIP-qPCR

We recommend testing site-specific accumulation of epigenetic modification by ChIP-qPCR first. In our case, we first confirmed by ChIP-qPCR that only dCas9-olEzh2 injection increased H3K27me3 site-specifically while dCas9 or dCas9-olEzh2( $\Delta\text{SET}$ ) did not and subsequently conducted ChIP-seq [9].

1. Prepare primer pairs for the target regions and for the control regions.
  - (a) For example, highly active promoters and intergenic regions can be used for positive and negative control regions, respectively, for H3K27ac ChIP.
2. Prepare input dilution (e.g., 1×, 1/4×, 1/16×, and 1/64×).
3. Assemble qPCR mix on ice according to manufacturer's protocol as follows:

10- $\mu$ M forward primer	0.3 $\mu$ L
10- $\mu$ M reverse primer	0.3 $\mu$ L
2× Thunderbird SYBR qPCR Mix	5 $\mu$ L
Immunoprecipitated DNA or input dilution	1 $\mu$ L
Milli-Q	3.4 $\mu$ L
Total	10 $\mu$ L/well on qPCR plate

4. Perform qPCR using AriaMX Real-Time PCR as shown in the following:

95 °C	1 min	
95 °C	25 s	40× cycles
60 °C	55 s	
95 °C	1 min	
55 °C	30 s	
95 °C	30 s	

5. Compare IP/input.

### 3.8 ChIP-seq and Data Processing

After confirmation by ChIP-qPCR, genome-wide changes can be detected by ChIP-seq as follows. In our case, we confirmed both site-specific binding of dCas9 (anti-FLAG) in all injection and site-specific accumulation of H3K27me3 only in dCas9-olEzh2 injection [9].

1. Prepare libraries using KAPA HyperPrep Kit according to manufacturer's protocol.
  - (a) We usually use 1 ng of immunoprecipitated or input DNA as input of library preparation.
  - (b) In our experiences, 18× cycles are best for PCR amplification process.
2. Conduct next-generation sequencing.
3. Data processing.

- (a) Trimming of low-quality reads and adapter-derived sequences by Trimmomatic [12].
- (b) Alignment of trimmed reads to medaka genome using BWA [13].
- (c) Generates SAM files (bwa samse).
- (d) SAM to BAM conversion and removal of low-quality alignment (samtools view -bS -q 20) [14].
- (e) Remove PCR duplicates (samtools rmdup).
- (f) Bam to bed conversion (bedtools bamtobed) [15].
- (g) MACS2 peak calling [16].
- (h) Check distribution of modifications on IGV or UCSC genome browser.

---

## 4 Notes

1. Side effects described as follows should be considered. Since these side effects would depend on the size of constructs, expression of dCas9 would not be an ideal control. Therefore, we highly recommend using dCas9 fused to catalytically inactivated effector domain as a negative control.
  - (a) dCas9 binding itself can inhibit proper transcription probably by interfering transcriptional elongation, RNA polymerase binding, or transcription factor binding [9, 17].
  - (b) dCas9 binding can increase chromatin accessibility in originally inaccessible loci [18, 19].
2. We introduced in vitro transcribed sgRNA (fusion of crRNA and tracrRNA) from T7 promoter in this protocol. However, the sgRNA in vitro transcribed in the protocol includes GG bases at the 5' end since T7 promoter needs GG just adjacent to the promoter (Fig. 2a), and these extra bases ("GG" upstream of "N(18–20)" targets sequence in Fig. 2a) potentially mismatch the genomic sequences upstream of the targets. Indeed, it was reported that such extra and mismatched "GG" bases at the 5' end of sgRNA correlate with decrease in efficiency of CRISPR-Cas9 mutagenesis in zebrafish [20]. Instead, synthesized crRNA and tracrRNA are now commercially available. Actually, a previous paper showed higher efficiency of mutagenesis in zebrafish in which crRNA and tracrRNA were used instead of sgRNA [21].
3. Some histone modifications could be antagonized by other epigenetic modifications: DNA methylation antagonizes H3K27me3 and H3K4me2 [22], H3K27me3 and H3K27ac antagonize each other [23], and H3K36me3 antagonizes

H3K27me3 [24]. Indeed, we found that DNA methylation at target regions inhibited H3K27me3 accumulation by dCas9-olEzh2 [9]. Furthermore, closed chromatin prevents sgRNA/Cas9 complex binding [25, 26]. Therefore, the target regions should be carefully chosen.

4. sgRNA vector pDR274 has two BsaI sites between T7 promoter and sgRNA scaffold. Once annealed oligos are inserted into linearized pDR274, BsaI recognition site will be eliminated from vectors, and oligo sequences will be inserted. Therefore, only non-inserted ones will show single band, while oligo sequence-inserted plasmids show more than two bands (super-coiled, open circular, and so on) by electrophoresis (Fig. 2).
5. Injection of multiple sgRNA at each target increases the efficiency of epigenome editing, compared to injection of a single sgRNA [17, 27, 28]. Therefore, sgRNA cocktail should be injected along with dCas9-effector mRNAs. In addition, it is also important to titrate the expression level of a effector domain to control the ratio of on-target and off-target efficiency [29, 30].
6. In our experience, the number of embryos in a single tube during crosslinking should be controlled for high reproducibility of ChIP experiments. For example, we usually put 100–150 medaka embryos at stage 11 (the late blastula stage, ~4000 cells/embryo) per tube.
7. Duration of crosslinking strongly influences the efficiency of chromatin fragmentation. Thus, the crosslinking procedure should be done exactly on time to achieve reproducible results.
8. Since lysis buffer for ChIP contains SDS, do not place tubes on ice for a long time. Avoid bubbles in Covaris microtube when transferring chromatin solution into the tubes.
9. When you sonicate samples for the first time, the parameter should be optimized. In most cases, titrating the duration time of sonication is sufficient, and check the size distribution of sonicated chromatin by electrophoresis or bioanalyzer.
10. Early embryos are highly rich in maternally derived RNA, and this causes the overestimation of chromatin concentration even if you measure DNA concentration by Qubit, in our experience. Therefore, we strongly recommend performing RNase A treatment before starting purification of sonicated chromatin.
11. The amount of antibody should be carefully determined. Generally, excess amount of antibody increases the background signal, while too little antibody decreases the true-positive signal. To maximize the signal to noise ratio, the amount of antibody should be titrated if necessary.

12. We always estimate the concentration of sonicated chromatin as described above and use the same amount of chromatin during ChIP procedure to make results highly reproducible.
13. Drying beads or tube negatively affects specificity. Do not make tubes dry. Replace buffers immediately during repeated wash.
14. Since the amount of DNA after IP is very limited, we recommend to perform ethanol precipitation at  $-80^{\circ}\text{C}$  overnight to maximize the efficiency.

---

## Acknowledgments

This work was supported by JSPS Grant-in-Aid for JSPS Research Fellow Grant Number JP18J21761 to H.S.F., by Japan Agency for Medical Research and Development (AMED) under Grant Number JP18gm1110007h0001 to H.T., by Japan Society for the Promotion of Science (JSPS) grant number JP21K06013, and by Grant-in-Aid for Scientific Research on Innovative Areas grant number JP21H00245 to R.N.

## References

1. Nakamura M, Gao Y, Dominguez AA, Qi LS (2021) CRISPR technologies for precise epigenome editing. *Nat Cell Biol* 23:11–22. <https://doi.org/10.1038/s41556-020-00620-7>
2. Heller EA, Cates HM, Peña CJ et al (2014) Locus-specific epigenetic remodeling controls addiction- and depression-related behaviors. *Nat Neurosci* 17:1720–1727. <https://doi.org/10.1038/nn.3871>
3. Aleyasin H, Flanigan ME, Golden SA et al (2018) Cell-type-specific role of  $\Delta\text{FosB}$  in nucleus accumbens in modulating intermale aggression. *J Neurosci* 38:5913–5924. <https://doi.org/10.1523/JNEUROSCI.0296-18.2018>
4. Hamilton PJ, Burek DJ, Lombroso SI et al (2018) Cell-type-specific epigenetic editing at the Fosb gene controls susceptibility to social defeat stress. *Neuropsychopharmacology* 43:272–284. <https://doi.org/10.1038/npp.2017.88>
5. Bustos FJ, Ampuero E, Jury N et al (2017) Epigenetic editing of the Dlg4/PSD95 gene improves cognition in aged and Alzheimer's disease mice. *Brain* 140:3252–3268. <https://doi.org/10.1093/brain/awx272>
6. Mendenhall EM, Williamson KE, Reyon D et al (2013) Locus-specific editing of histone modifications at endogenous enhancers. *Nat Biotechnol* 31:1133–1136. <https://doi.org/10.1038/nbt.2701>
7. Konermann S, Brigham MD, Trevino AE et al (2013) Optical control of mammalian endogenous transcription and epigenetic states. *Nature* 500:472–476. <https://doi.org/10.1038/nature12466>
8. Hilton IB, D'Ippolito AM, Vockley CM et al (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 33:510–517. <https://doi.org/10.1038/nbt.3199>
9. Fukushima HS, Takeda H, Nakamura R (2019) Targeted in vivo epigenome editing of H3K27me3. *Epigenetics Chromatin* 12:17. <https://doi.org/10.1186/s13072-019-0263-z>
10. Hwang WY, Fu Y, Reyon D et al (2013) Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* 31:227–229. <https://doi.org/10.1038/nbt.2501>
11. Stemmer M, Thumberger T, Del Sol KM et al (2015) CCTop: an intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS One* 10:e0124633. <https://doi.org/10.1371/journal.pone.0124633>

12. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
13. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698>
14. Li H, Handsaker B, Wysoker A et al (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
15. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842. <https://doi.org/10.1093/bioinformatics/btq033>
16. Zhang Y, Liu T, Meyer CA et al (2008) Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* 9:R137. <https://doi.org/10.1186/gb-2008-9-9-r137>
17. Qi LS, Larson MH, Gilbert LA et al (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152:1173–1183. <https://doi.org/10.1016/j.cell.2013.02.022>
18. Barkal AA, Srinivasan S, Hashimoto T et al (2016) Cas9 functionally opens chromatin. *PLoS One* 11:e0152683. <https://doi.org/10.1371/journal.pone.0152683>
19. Polstein LR, Perez-Pinera P, Kocak DD et al (2015) Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. *Genome Res* 25:1158–1169. <https://doi.org/10.1101/gr.179044.114>
20. Varshney GK, Pei W, LaFave MC et al (2015) High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res* 25:1030–1042. <https://doi.org/10.1101/gr.186379.114>
21. Hoshijima K, Jurynek MJ, Klatt Shaw D et al (2019) Highly efficient CRISPR-Cas9-based methods for generating deletion mutations and F0 embryos that lack gene function in zebrafish. *Dev Cell* 51:645–657.e4. <https://doi.org/10.1016/j.devcel.2019.10.004>
22. Nakamura R, Tsukahara T, Qu W et al (2014) Large hypomethylated domains serve as strong repressive machinery for key developmental genes in vertebrates. *Development* 141:2568–2580. <https://doi.org/10.1242/dev.108548>
23. Tie F, Banerjee R, Stratton CA et al (2009) CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. *Development* 136:3131–3141. <https://doi.org/10.1242/dev.037127>
24. Yuan W, Xu M, Huang C et al (2011) H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J Biol Chem* 286:7983–7989. <https://doi.org/10.1074/jbc.M110.194027>
25. Wu X, Scott DA, Kriz AJ et al (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* 32:670–676. <https://doi.org/10.1038/nbt.2889>
26. Singh R, Kuscus C, Quinlan A et al (2015) Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Res* 43:e118–e118. <https://doi.org/10.1093/nar/gkv575>
27. Maeder ML, Linder SJ, Cascio VM et al (2013) CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* 10:977–979. <https://doi.org/10.1038/nmeth.2598>
28. Perez-Pinera P, Kocak DD, Vockley CM et al (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* 10:973–976. <https://doi.org/10.1038/nmeth.2600>
29. Lin L, Liu Y, Xu F et al (2018) Genome-wide determination of on-target and off-target characteristics for RNA-guided DNA methylation by dCas9 methyltransferases. *Gigascience* 7. <https://doi.org/10.1093/gigascience/gy011>
30. Pflueger C, Tan D, Swain T et al (2018) A modular dCas9-SunTag DNMT3A epigenome editing system overcomes pervasive off-target activity of direct fusion dCas9-DNMT3A constructs. *Genome Res* 28:1193–1206. <https://doi.org/10.1101/gr.233049.117>

# INDEX

## A

- Abscisic acid (ABA)..... 199, 205, 208  
 Accessible chromatin..... 65–80, 93,  
 94, 96, 98  
 Assay for transposase accessible chromatin sequencing  
 (ATAC-seq)..... 66, 67, 75, 77–79, 93  
 Assay of transposase accessible chromatin with  
 visualization (ATAC-seq) ..... 93–101

## B

- Biomarker detection ..... 4  
 Bisulfite conversion ..... 6–8, 11, 12,  
 14, 18, 39, 42, 44, 49, 181, 251, 257

## C

- Chromatin ..... 3, 55–59, 61, 65, 66,  
 77, 79, 83, 84, 93–101, 124, 130, 134, 161, 162,  
 168, 170–172, 177, 181, 197, 245, 247,  
 286–288, 291, 292  
 Chromatin accessibility ..... 65, 93–101,  
 123, 283, 290  
 Chromatin contact ..... 123–144  
 Chromatin immunoprecipitation (ChIP)..... 55–57,  
 59, 62, 139, 168, 283, 286–289, 291, 292  
 Chromatin immunoprecipitation sequencing  
 (ChIP-seq) ..... 55–63, 78, 90,  
 124–128, 130–133, 138–140, 162, 168, 282,  
 283, 288, 289  
 Clustered regularly interspaced short palindromic repeat  
 (CRISPR)..... 179, 198,  
 211, 212, 215, 223, 229, 255, 259, 260, 266,  
 272, 283  
 Coiled-coil peptide..... 180, 198  
 CpG methyltransferase..... 244, 247  
 CpG site..... 6–8, 10–14, 17–19,  
 39, 47, 177, 181, 182  
 CRISPR RNA (crRNA) ..... 229–234,  
 236–239, 290  
 CRISPR–Cas9 ..... 290  
 CRISPR–Cpf1 (Cas12a) ..... 229  
 CRISPR/dCas9 ..... 198

## D

- dCas9 ..... 179–181, 186, 189,  
 190, 201, 204, 206, 211, 215, 244, 255–257,  
 269, 270, 279, 280, 282, 283, 288–290  
 dCas9–SunTag..... 190, 191, 269, 270, 272  
 Demethylation..... 190, 244, 255, 265  
 DNA demethylation ..... 191, 255–257, 260,  
 269, 270, 272, 274–276  
 DNA methylation ..... 3, 4, 6, 7, 11,  
 12, 17, 21, 39, 40, 103, 106, 178, 180–183, 186,  
 187, 189, 243–245, 247, 249–252, 255, 269,  
 274, 275, 283, 290, 291  
 DNMT3A ..... 180, 181  
 Droplet digital PCR (ddPCR) ..... 212, 214,  
 221, 222, 224  
 Dual crosslink ..... 57, 59

## E

- Enhancer..... 44, 65, 89, 93,  
 100, 103–106, 123–125, 128–130, 134, 139,  
 142, 279  
 Epigenetic editing ..... 189  
 Epigenetics ..... 3, 39, 55, 83, 123,  
 161, 189, 197, 255, 256, 269, 279, 280, 283,  
 288, 290  
 Epigenome editing..... 178, 179, 181,  
 189, 244, 247, 255–257, 259, 265, 267, 269,  
 279, 280, 282, 291

## F

- Fish..... 280, 290  
 Fluorescent microscopy ..... 96

## G

- Gene activation..... 190, 212, 230,  
 231, 236–238  
 Gene regulation..... 79, 124, 139, 140  
 Genome editing ..... 189, 211,  
 229–231, 235  
 Genomic PCR ..... 235, 236  
 Germ cells ..... 70, 147, 244

**H**

- H3K27me3 .....62, 90, 282, 283, 288–291  
 Hi-C ..... 63, 124, 126, 134–142  
 Histone ..... 56, 59, 65, 161, 162,  
 168, 170, 172, 177, 189, 245, 247, 249, 250, 279  
 Histone modification ..... 3, 55–63, 103,  
 123, 255, 280, 288, 290  
 Hydrodynamic tail vein injection (HTVi) .....238,  
 270, 271, 274, 276

**I**

- ImageJ .....109, 113, 114, 246, 249  
 Immunoprecipitation .....55–63, 83, 170, 287  
 In vivo ..... 83, 236, 238,  
 239, 243, 280, 282  
 In vivo epigenome editing .....270  
 In vivo gene activation .....231, 238

**L**

- Live-cell imaging .....109, 247, 251, 253  
 Live imaging .....106, 107  
 Long Interspersed Element-1 (LINE-1, L1) ..... 147–158

**M**

- Medaka ..... 280–283, 285–287, 290, 291  
 Meiosis .....124, 130  
 Methylome .....21–23, 34, 40, 243  
 Mouse ..... 11, 19, 34, 39, 66–67,  
 70–71, 76, 79, 83–91, 95, 99, 124–127,  
 133–135, 161–172, 232, 236–240, 243–253,  
 269–276  
 MspI ..... 39, 40, 42, 43

**N**

- Next-generation sequencing (NGS) .....v, 40, 55,  
 58, 60, 83, 126, 131, 150, 153, 154, 162, 181,  
 205, 243, 289  
 Non-homologous end joining (NHEJ) .....229–231,  
 234, 235  
 Nuclear decondensation ..... 163–164  
 Nucleoplasmin ..... 162–170

**O**

- Oocytes ..... 83–85, 87,  
 88, 90, 125, 247  
 Optogenetics ..... 230, 231, 235

**P**

- Pericentromere .....243, 244, 247, 249, 250  
 Post-bisulfite adaptor tagging (PBAT) .....22, 23  
 Postmortem brain ..... 148, 149, 152  
 Preimplantation embryos ..... 83–91, 244, 247

- Primer design .....11, 14, 262  
 Promoter-enhancer interaction .....137, 141–144  
 Pyrosequencing technology ..... 4–6  
 Python ..... 109, 117–119,  
 121, 127, 130, 135

**R**

- Real-time PCR .....31, 69, 192,  
 200, 201, 212, 214, 220–221, 224, 236, 282, 289  
 Reduced representation bisulfite sequencing  
 (RRBS) .....39–50  
 Reporter assay .....212–214, 216  
 RNA polymerase II (RNAPII) ..... 103, 104, 106

**S**

- ScFv-TET1CD .....269, 270  
 Single-cell nuclei ..... 147–158  
 Single-nucleotide polymorphisms (SNPs) ..... 7, 148,  
 150, 153, 156, 157  
 Single-stranded DNA (ssDNA) ligation .....21–36  
 Solubilization ..... 161–172  
 Sperm .....90, 125, 161–163,  
 165, 168–172, 247  
 Spermatogenesis ..... 66, 71, 79, 124,  
 125, 129, 130, 132  
 STREAMING-tag .....106–109, 120  
 SunTag ..... 180, 190, 198,  
 211, 255, 269  
 Super-enhancer (SE) ..... 75, 124,  
 128–134, 138, 140–143, 154, 158

**T**

- TACS ligation .....23, 29, 33  
 T7 endonuclease I (T7E1) ..... 231, 235  
 TET .....255  
 TET1 .....255  
 Tn5 transposase ..... 66, 93–96, 100  
 Trackpy .....109, 111, 117, 121  
 Transcription activator-like effector (TALE) .....179,  
 206, 211, 244, 279  
 Transcriptional activation ..... 190, 191,  
 198, 199, 215  
 Transcriptional bursting ..... 104–106  
 Transcriptional regulatory factor .....103  
 Transcription factor-binding ..... 77, 290

**V**

- VP64 .....190, 191, 194, 211, 215

**W**

- Whole genome amplification .....148  
 Whole-genome bisulfite sequencing (WGBS) ..... 21,  
 22, 24, 39, 47, 50