Methods in
Molecular Biology 2458

Springer Protocols

Julia Horsfield
Judith Marsman *Editors*

Chromatin

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.

Chromatin

Methods and Protocols

Edited by

Julia Horsfield

Department of Pathology
Dunedin School of Medicine, University of Otago, Dunedin, New Zealand

Judith Marsman

Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, The Netherlands

Humana Press

Editors
Julia Horsfield
Department of Pathology
Dunedin School of Medicine
University of Otago
Dunedin, New Zealand

Judith Marsman
Cardiology, Division Heart & Lungs
University Medical Center Utrecht
Utrecht, The Netherlands

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-0716-2139-4 ISBN 978-1-0716-2140-0 (eBook)
https://doi.org/10.1007/978-1-0716-2140-0

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

The study of chromatin has emerged as an important technological wave in molecular biology over the last 15 years. Chromatin research has become crucial to the understanding of how modifications of DNA and its associated proteins affect the transcriptional output of the genome. This book contains cutting-edge gold standard techniques that are best practice for the study of chromatin biology today. Themes and chapters cover well-established methods for the analysis of DNA methylation, DNA-associated proteins and their modifications, and methods used to study chromatin accessibility and three-dimensional structure. Where deep sequencing-based tools are used, the methods include data analysis pipelines that can be used by researchers with basic bioinformatics skills.

Dunedin, New Zealand — *Julia Horsfield*
Utrecht, The Netherlands — *Judith Marsman*

Contents

Contributors

Suzan Al Momani • *Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand; Maurice Wilkins Centre for Molecular Biodiscovery, Auckland, New Zealand*

Rhys S. Allan • *The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia; Department of Medical Biology, The University of Melbourne, Parkville, VIC, Australia*

Allegra Angeloni • *Genomics and Epigenetics Division, Garvan Institute of Medical Research, Sydney, NSW, Australia; School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia*

Folkert W. Asselbergs • *Department of Cardiology, Division Heart & Lungs, UMC Utrecht, University of Utrecht, Utrecht, The Netherlands; Health Data Research UK and Institute of Health Informatics, University College London, London, UK; Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK*

Rakesh Banerjee • *Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand*

Basharat Bhat • *Department of Surgical Sciences, Dunedin School of Medicine, University of Otago, Dunedin School of Medicine, Dunedin, New Zealand*

Ozren Bogdanovic • *Genomics and Epigenetics Division, Garvan Institute of Medical Research, Sydney, NSW, Australia; School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW, Australia*

Paloma Cejas • *Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA; Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA; Translational Oncology Laboratory, Hospital La Paz Institute for Health Research (IdiPAZ) and CIBERONC, La Paz University Hospital, Madrid, Spain*

Aniruddha Chatterjee • *Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand; Maurice Wilkins Centre for Molecular Biodiscovery, Auckland, New Zealand*

Robert C. Day • *Department of Biochemistry, University of Otago, Dunedin, New Zealand*

Caroline Delandre • *Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia*

Sarah D. Diermeier • *Department of Biochemistry, University of Otago, Dunedin, New Zealand*

Michael R. Eccles • *Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand; Maurice Wilkins Centre for Molecular Biodiscovery, Auckland, New Zealand*

James Ferguson • *Genomics and Epigenetics Division, Garvan Institute of Medical Research, Sydney, NSW, Australia; St Vincent's Clinical School, University of New South Wales, Sydney, NSW, Australia*

Gregory Gimenez • *Dunedin School of Medicine, University of Otago, Dunedin, New Zealand*

Macarena Moronta Gines • *Department of Cell Biology, Erasmus MC, Rotterdam, The Netherlands*

William J. Greenleaf • *Department of Genetics, Stanford University, Stanford, CA, USA; Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA, USA; Department of Applied Physics, Stanford University, Stanford, CA, USA; Chan Zuckerberg Biohub, San Francisco, CA, USA*

Anders S. Hansen • *Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA*

Magdalena Harakalova • *Department of Cardiology, Division Heart & Lungs, UMC Utrecht, University of Utrecht, Utrecht, The Netherlands; Regenerative Medicine Utrecht (RMU), UMC Utrecht, University of Utrecht, Utrecht, The Netherlands*

Asmita Jha • *Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA*

Timothy M. Johanson • *The Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia; Department of Medical Biology, The University of Melbourne, Parkville, VIC, Australia*

Gregory T. Jones • *Department of Surgical Sciences, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand*

Michael-C. Keogh • *EpiCypher Inc., Durham, NC, USA*

Anja S. Knaupp • *Department of Anatomy and Developmental Biology, Monash University, Clayton, VIC, Australia; Development and Stem Cells Program, Monash Biomedicine Discovery Institute, Clayton, VIC, Australia; Australian Regenerative Medicine Institute, Monash University, Clayton, VIC, Australia*

Nils Krietenstein • *The Novo Nordisk Center for Protein Research (CPR), Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark*

Anshul Kundaje • *Department of Genetics, Stanford University, Stanford, CA, USA; Department of Computer Science, Stanford University, Stanford, CA, USA*

Henry W. Long • *Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA; Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA*

Stefano G. Manzo • *Oncode Institute and Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, The Netherlands*

Georgi K. Marinov • *Department of Genetics, Stanford University, Stanford, CA, USA*

Owen J. Marshall • *Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia*

Judith Marsman • *Department of Cardiology, Division Heart & Lungs, University Medical Centre Utrecht, Utrecht, The Netherlands*

Matthew R. Marunde • *EpiCypher Inc., Durham, NC, USA*

Michal Mokry • *Department of Cardiology, Division Heart & Lungs, UMC Utrecht, University of Utrecht, Utrecht, The Netherlands; Laboratory of Clinical Chemistry and Hematology, UMC Utrecht, Utrecht, The Netherlands*

Jiayi Pei • *Department of Cardiology, Division Heart & Lungs, UMC Utrecht, University of Utrecht, Utrecht, The Netherlands; Regenerative Medicine Utrecht (RMU), UMC Utrecht, University of Utrecht, Utrecht, The Netherlands*

Jose M. Polo • *Department of Anatomy and Developmental Biology, Monash University, Clayton, VIC, Australia; Development and Stem Cells Program, Monash Biomedicine Discovery Institute, Clayton, VIC, Australia; Australian Regenerative Medicine Institute, Monash University, Clayton, VIC, Australia; Adelaide Centre for Epigenetics and The South Australian immunoGENomics Cancer Institute, The University of Adelaide, Adelaide, SA, Australia*

Irina K. Popova • *EpiCypher Inc., Durham, NC, USA*
Tamara A. Potapova • *Stowers Institute for Medical Research, Kansas City, MO, USA*
Oliver J. Rando • *Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA*
Euan J. Rodger • *Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand; Maurice Wilkins Centre for Molecular Biodiscovery, Auckland, New Zealand*
Debina Sarkar • *Department of Biochemistry, University of Otago, Dunedin, New Zealand*
Ralf B. Schittenhelm • *Monash Proteomics and Metabolomics Facility, Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia*
Zohar Shipony • *Department of Genetics, Stanford University, Stanford, CA, USA*
Jim Smith • *Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand*
Peter A. Stockwell • *Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand*
Maria Elize van Breugel • *Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, The Netherlands*
Noortje A. M. van den Dungen • *Laboratory of Clinical Chemistry and Hematology, UMC Utrecht, Utrecht, The Netherlands*
Fred van Leeuwen • *Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, The Netherlands; Department of Medical Biology, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands*
Tom van Schaik • *Oncode Institute and Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, The Netherlands*
Bas van Steensel • *Oncode Institute and Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, The Netherlands; Department of Cell Biology, Erasmus University Medical Center, Rotterdam, The Netherlands*
Robert J. Weeks • *Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand*
Ellen N. Weinzapfel • *EpiCypher Inc., Durham, NC, USA*
Kerstin S. Wendt • *Department of Cell Biology, Erasmus MC, Rotterdam, The Netherlands*
Zulin Yu • *Stowers Institute for Medical Research, Kansas City, MO, USA*

Part I

DNA Methylation

Chapter 1

Generating Sequencing-Based DNA Methylation Maps from Low DNA Input Samples

Suzan Al Momani, Euan J. Rodger, Peter A. Stockwell, Michael R. Eccles, and Aniruddha Chatterjee

Abstract

Reduced representation bisulfite sequencing (RRBS) is a technique used for assessing genome-wide DNA methylation patterns in eukaryotes. RRBS was introduced to focus on CpG-rich regions that are likely to be of most interest for epigenetic regulation, such as gene promoters and enhancer sequence elements (Meissner et al., Nature 454:766–770, 2008). This "reduced representation" lowers the cost of sequencing and also gives increased depth of coverage, facilitating the resolution of more subtle changes in methylation levels. Here, we describe a modified RRBS sequencing (RRBS-seq) library preparation. Our protocol is optimized for generating single base–resolution libraries when low input DNA is a concern (10–100 ng). Our protocol includes steps to optimize library preparation, such as using deparaffinization solution (when formalin-fixed material is used), and a replacement of gel size-selection with sample purification beads. The described protocol can be accomplished in 3 days and has been successfully applied to tissues or cells from different organisms, including formalin-fixed tissues, to yield robust and reproducible results.

Key words Epigenetics, DNA methylation, Reduced representation bisulfite sequencing, Multiplexed, Methylome, FFPE, CpG island, DMAP

1 Introduction

Considered to be the most stable epigenetic mechanism, DNA methylation refers to the addition of a methyl group to cytosine residues to form 5-methylcytosine (5mC), generally in the context of a CpG dinucleotide [1]. Interestingly, human and other mammalian genomes have a lower proportion of CpG dinucleotides than expected. It is possible that this is a protective response to the inherent mutagenic potential of 5mC, which is predisposed to cytosine-to-thymine transition mutations via spontaneous or

Suzan Al Momani and Euan J. Rodger are joint first author.

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_1,

enzymatic deamination. This apparent disadvantage, however, is offset by the extensive regulatory potential offered by DNA methylation.

DNA methylation is crucial for normal mammalian development. The significance of DNA methylation in the epigenetic regulation of gene expression is well-documented, yet methylation also plays a vital role within a variety of other contexts, including X-inactivation, the maintenance of genomic stability, and imprinting [2]. Throughout this wide range of attributed functions, DNA methylation confers an element of added plasticity and dynamism to genetic regulation, whilst retaining the capacity for stable propagation of epigenetic marks during cell replication [3]. Failure to propagate epigenetic information precisely results in deviations from the normal pattern of gene expression, which in turn can result in a failure of developmental programs and, potentially, cancer [3, 4].

DNA methylation analysis has become more tractable and cost-efficient, and thanks to next generation sequencing (NGS), it is possible to screen the DNA methylome to obtain comprehensive information regarding epigenetic modifications associated with phenotypes of interest. However, whole-genome bisulfite sequencing (WGBS, BS-seq) remains cost-prohibitive and requires intensive computational analysis. Currently, restriction enzyme–based reduced representation bisulfite sequencing (RRBS) and its modified protocols are widely used to study methylation status on a large scale and at a single-base resolution by enriching for CpGs in the CpG-rich regions by MspI digestion [5, 6].

In RRBS, the DNA is digested with a restriction enzyme (frequently MspI) and enrichment occurs by size selection of the MspI fragments corresponding to 40–220 bp. The MspI enzyme cuts at 5′-C↓CGG-3′ irrespective of cytosine methylation status, enriching for CpG-rich promoter regions and thus achieving a high coverage of CpG-rich regions while hugely reducing sequencing read requirements [7]. DNA fragments from the digestion can be in a wide range of lengths; however, only the fragments ranging from 40 to 220 bp are selected for sequencing [8, 9]. After PCR amplification of bisulfite-converted DNA libraries, these are normalized and pooled at the same molarity before sequencing, resulting in equal representation of each library. Apart from numerous human studies, RRBS has also been used to produce DNA methylation maps in several organisms including zebrafish [10–12], mice [13, 14], pigs [15, 16], and rats [17]. RRBS was shown to be a highly reproducible method for genome-scale methylation profile [18–24].

In this chapter, we detail the RRBS protocol, with a particular focus on low input and/or degraded DNA samples, such as those from FFPE [25]. Initially, several micrograms of DNA were required to perform genome-wide DNA methylation analysis, but

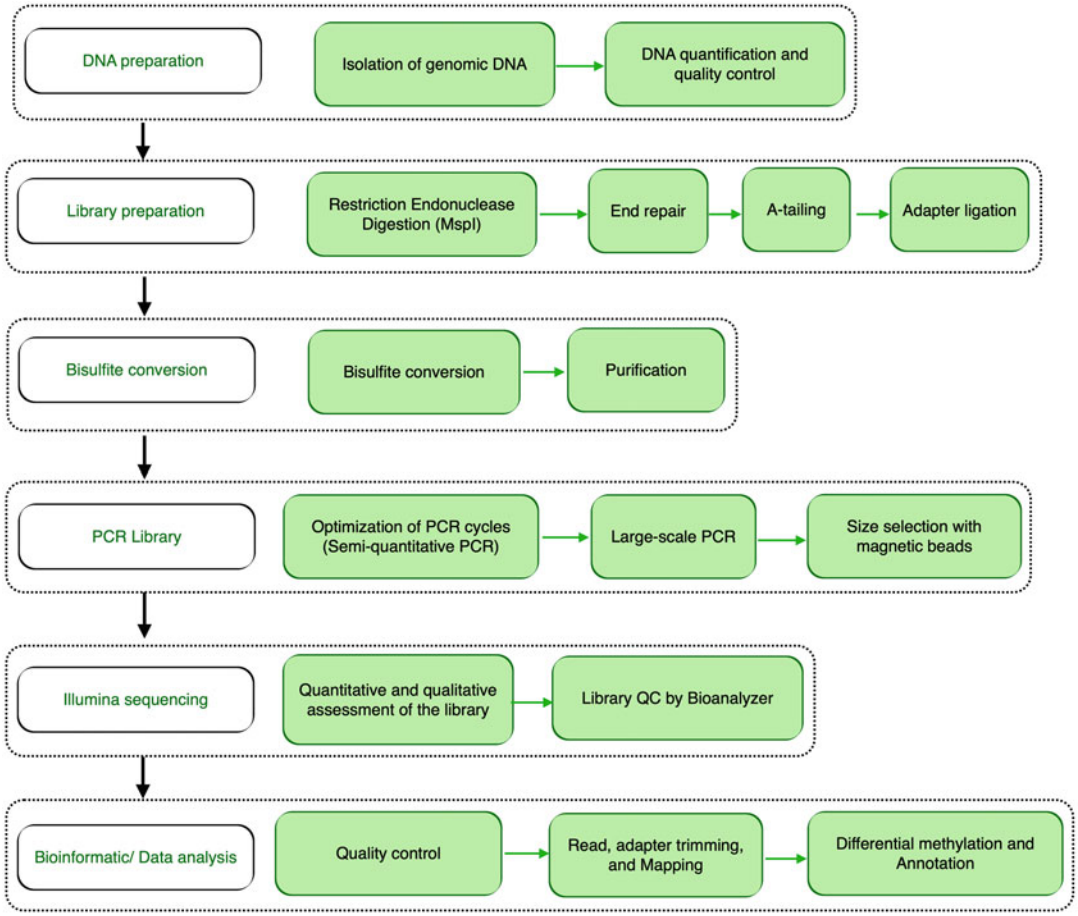


Fig. 1 Workflow of the reduced representation bisulfite sequencing protocol with magnetic bead fragment size selection

the replacement of electrophoresis steps and gel extraction for purification by magnetic beads cleanup has enabled construction of libraries suitable for sequencing from ~10 ng of input material. Moreover, size selection using sample purification beads is less labor-intensive and more precise [26]. The workflow of the library preparation protocol is detailed below and follows the classical library preparation protocol, in which methylated adaptors are ligated to the fragmented DNA prior to bisulfite conversion (Fig. 1). For adequate coverage of the human genome, we multiplex up to eight DNA libraries to be sequenced in a single lane of a flow cell. These can be distinguished bioinformatically after sequencing, using the six-base index sequence in each adapter.

With the protocol described in this chapter, we routinely obtain 200 M reads per HiSeq 2500 lane, of which 60–70% can be mapped to a bisulfite-treated human genome [27], resulting in coverage comparable to that of the original RRBS protocol starting with 2 μg

of DNA [28]. A conversion efficiency of ~99% is routinely achieved, which for mammalian genomes is estimated by evaluating methylation outside the CpG context. We regularly use the protocol described here to obtain high quality DNA methylation data from samples with low amounts of DNA.

2 Materials

2.1 Experimental Workflow

1. Commercial DNA/RNA decontaminating solution such as DNAZap.
2. Commercial Plasmid DNA extraction kit (*see* **Note 1**).
3. Commercial PCR purification kits (1× standard, 1× low elution volume).
4. Centrifuge (*see* **Note 2**).
5. DNA Clean and Concentrator kit.
6. Deparaffinization Solution (*see* **Note 3**).
7. Nanophotometer.
8. 1.5 mL low-bind DNA microcentrifuge tubes.
9. 0.2 mL PCR tubes.
10. Absolute ethanol, analytical grade.
11. TE buffer: 10 mM Tris–HCL, 1 mM EDTA, pH 8.0.
12. High sensitivity DNA detection instrument such as a Qubit fluorometer (Life Technologies) and associated reagents.
13. MspI restriction endonuclease [20 U/μL] with compatible restriction buffer.
14. Milli-Q purified water.
15. PCR thermocycler.
16. Dry thermoblock with shaking function.
17. Vortex.
18. TruSeq Nano DNA sample preparation kit (Illumina cat #20015964) containing: End Repair Buffer Mix (ERP2), Resuspension Buffer (RSB), Ligation Buffer Mix (Ligation Mix 2), A-Tailing Mix (ATL), Stop Ligation Buffer (STL), PCR Primer Cocktail (PPC), Enhanced PCR Mix (EPM), and Sample Purification Beads (SPB, similar to Beckman Coulter AMPure XP or SRPI).
19. Index methylated adapters: TruSeq adapters sets A and/or B.
20. 96-well magnetic plate.
21. DNA gel materials including: DNA ladder, agarose, TAE buffer 0.5× with ethidium bromide, XC loading dye 6×, trays, combs, running apparatus, and UV gel imaging system.

22. Sodium bisulfite treatment kit, such as EZ DNA Direct Methylation Kit (Zymo Research), or equivalent.
23. KAPA HiFi HotStart ReadyMix: containing all reaction components except primers and DNA template.
24. 2100 Bioanalyzer and high sensitivity DNA kit (Agilent Technologies), or equivalent.
25. Illumina HiSeq2500 Sequencer and corresponding reagents.

2.2 Data Analysis

1. A Linux/Unix multicore machine with at least 16 GB of RAM and 5 cores (see Bismark user guide).
2. FastQC: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [29].
3. Cutadapt for read trimming. Illumina adapter removal: http://code.google.com/p/cutadapt/ [30] or Trim Galore: www.bioinformatics.babraham.ac.uk/projects/trim_galore/ [31].
4. Bismark: http://www.bioinformatics.babraham.ac.uk/projects/bismark/ [32].
5. Bowtie2 (Alignment software): http://bowtie-bio.sourceforge.net/bowtie2/index.shtml.
6. Samtools: http://samtools.sourceforge.net/ [33].
7. DMAP: from https://github.com/peterstockwell/DMAP [34].
8. Methylkit: from https://bioconductor.org/packages/release/bioc/html/methylKit.html [35].

3 Methods

The specific protocol for library construction will depend on, for example, starting material available and DNA input and quality (*see* **Note 3**). Below, we give general directions for the library construction. While initially several micrograms of DNA were required to perform RRBS-Seq, the replacement of electrophoretic steps and gel extraction for purification by magnetic beads has enabled creating libraries suitable for sequencing from ~100 ng of input material. For this protocol, we substitute toxic xylene with safer deparaffinization buffer without compromising DNA quality. We also replaced gel size-selection by magnetic bead-based clean up. Compared to gel size selection, bead-based size selection of DNA fragments is more efficient, easier to handle and to standardize [36]. We established a two-step bead-based size selection cleanup. While in the first cleanup, larger DNA fragments are removed, the second cleanup discards PCR primers and adapters before sequencing. The magnetic beads work better where sufficient starting material is available (low adapter dimers). We use diluted adapters (1:4) to avoid generating excess adapter dimer products.

3.1 DNA Purification

1. In case of FFPE tissues, deparaffinization is achieved with xylene or deparaffinization buffer according to the manufacturer's recommendation (*see* **Notes 3** and **4**).
2. Genomic DNA (gDNA) can be isolated from fresh, frozen, or FFPE tissue using any standard extraction protocol. Use the QIAamp DNA mini kit, or equivalent, according to the manufacturer's instructions for specific cell or tissue type, including steps to remove protein and RNA contamination (*see* **Notes 5** and **6**).
3. Quantify the concentration of DNA using a Qubit fluorometer and estimate DNA quality using a Nanophotometer, according to the manufacturer's instructions. Ideally, DNA should be of high quality as indicated by a 260/280 ratio of 1.8–1.9. If it is less than 1.8 it is generally carbohydrate contamination. If the value is above 2.0, it may be protein contamination.
4. (Optional) If a gDNA sample is of poor quality, it can be cleaned up using a Genomic DNA Clean and Concentrator kit in order to remove residual contaminants.
5. Dilute gDNA to between 10 and 50 ng/μL using TE buffer. Standardizing all DNA samples to one concentration will simplify subsequent steps.
6. Store gDNA in sealed tubes at 4 °C until required.

3.2 Restriction Endonuclease Digestion (MspI)

MspI cuts at 5′-C↓CGG-3′ irrespective of cytosine methylation status, which generates a defined and predictable number of DNA fragments which contain at least two CpG loci, one at each terminus, regardless of fragment length or CpG methylation status.

1. Set up the following 40 μL reaction for digestion of gDNA:
   (a) 10–500 ng of gDNA.
   (b) 8 μL of 20/U μL of MspI.
   (c) 4 μL of 10× Restriction Enzyme Buffer.
   (d) Milli-Q H_2O up to 40 μL.
2. Mix well by pipetting, and briefly spin down all liquid in the tubes in a mini centrifuge. The total volume of each reaction should be 40 μL. If preparing multiple digests, you can create a master mix containing 110% of all the components excluding the DNA sample in a 1.5 mL microcentrifuge tube, and then combine the mix with each DNA sample in separate reaction tube (assuming DNA concentration was adjusted to the same concentration as suggested in Subheading 3.1, **step 5**).
3. Incubate overnight (~16 h) at 37 °C in a thermocycler with the heated lid turned off (*see* **Note 7**).

4. After digestion, purify the DNA using a PCR purification kit following the manufacturer's protocol, eluting in 60 μL TE buffer.

3.3 End Repair

Enzymatic digestion of gDNA yields fragments with overhangs at each end. In this step, an end repair reaction will remove the sticky (over-hanging) ends and fill in the gaps, thereby generating blunt ends for A-tailing and adapter ligation. These A-tails will facilitate the ligation of methylated adapters in a subsequent step.

1. Thaw the End Repair Buffer (ERP2) of the TruSeq DNA sample preparation kit.
2. In a PCR tube, add 40 μL of End Repair mix 2 (ERP2) to the digested gDNA sample (from Subheading 3.2, **step 4**). Mix by pipetting, cap and briefly spin down the tubes. The total volume of each reaction should be 100 μL.
3. Incubate for 30 min at 30 °C in a heat block or a PCR thermocycler with the heated lid turned off.
4. Remove the tubes from the thermal cycler, spin briefly.
5. Purify the product following end repair using a small elution volume PCR purification kit (such as Qiagen MinElute) according to the manufacturer's protocol, eluting in 18 μL kit elution buffer.

3.4 Adenylation of 3′ Ends

To prevent the blunt end fragments from annealing to each other, an adenine residue is attached to the 3′ end of each blunt ended fragment. The Illumina adapters have a single "T" base overhang at the 3′ end. As such, addition of "A" base to the 3′ end of each fragment ensures that the DNA fragments will bind uniquely to the adapters and not to each other.

1. Add 12.5 μL TruSeq A-Tailing mix (ATL) to the end-repaired gDNA (from Subheading 3.3, **step 5**) and mix thoroughly.
2. Incubate in a thermocycler as follows.
   (a) 30 min at 37 °C.
   (b) 5 min at 70 °C.
   (c) 5 min at 4 °C.
3. Immediately proceed to the next step.

3.5 Methylated Adapter Ligation

Indexed, methylated adapters are used to maintain integrity during bisulfite sequencing. These adapters include universal sequences that enable PCR amplification of adapter-ligated DNA fragments. In addition, each of the different adapters contains a unique six-base index sequence (or barcode). In this step, a different adapter is ligated to A-tailed DNA fragments from each of the samples that will later be pooled and run in a single sequencing

lane. The unique index sequences will tag each of the different samples in a pool, allowing them to be distinguished bioinformatically after sequencing.

1. Briefly spin the PCR tubes containing the mix to ensure all liquid is at the bottom of the tubes.
2. To carry out the ligation reaction, set up the following reaction in the same PCR tube.
   (a) gDNA from Subheading 3.4, **step 3**.
   (b) 2.5 μL resuspension buffer (RSB).
   (c) 2.5 μL TruSeq Adapter Index (choosing a single barcoded adaptor (15 μM) from the 24 available adapters, *see* **Notes 8–10**).
   (d) 2.5 μL TruSeq DNA Ligase mix 2 (LIG2), total reaction volume should be 37.5 μL
3. Incubate for 10 min at 30 °C in a thermocycler.
4. Immediately after incubation, add 5 μL of Stop Ligase mix (STL) and mix.
5. Use a low volume elution PCR purification kit according to the manufacturer's instructions, eluting in 18 μL TE buffer.
6. Store at 4 °C until ready to perform bisulfite conversion.

3.6 Bisulfite Conversion

Bisulfite treatment allows discrimination of methylated and unmethylated cytosines by selectively converting unmethylated cytosines to uracils while leaving 5-methylcytosine intact.

1. Prepare the DNA methylation CT conversion reagent according to the manufacturer's instructions, which may vary between different kits. Mix thoroughly until fully dissolved.
2. Incubate for 15 min at 37 °C in a thermocycler.
3. Add 130 μL of the prepared CT conversion reagent to the sample and mix. Incubate at 64 °C in a thermocycler for 3.5 h (*see* **Note 11**).
4. After incubation, place the tubes on ice for 10 min and proceed with the DNA methylation kit according to the manufacturer's instruction, eluting in 20 μL TE buffer. Do not leave the desulfonation buffer step longer than the time recommended in the instructions (*see* **Note 12**).

3.7 Determine the Optimal Amplification Cycles by Semiquantitative Amplification

Prior to sequencing, the bisulfite converted libraries should be amplified using KAPA HiFi HotStart. However, to minimize the PCR amplification bias, the libraries should be amplified using the lowest number of cycles that results in robust amplification, as excessive amplification could increase the prevalence of point mutations and increase amplification of short fragments [37]. Successful

amplification is determined by visualizing the products under the UV gel imager. Then, the remainder of the pool is amplified using the minimum number of cycles, as determined from the small-scale tests (*see* **Notes 13** and **14**).

1. In a PCR tube, add the following and mix to make a total volume of 25 μL and mix well.
   (a) 6.5 μL of Milli-Q Water.
   (b) 12.5 μL of KAPA HiFi HotStart Uracil+ ReadyMix.
   (c) 3 μL of TruSeq PCR Primer Cocktail.
   (d) 3 μL of DNA sample (from Subheading 3.6, **step 4**).
2. Divide the above mix into 2 PCR tubes (12 μL each) and run using separate PCR thermocyclers adjusting the number of cycles to 15, and 20 using the following reaction conditions.
   (a) 95 °C 4 min.
   (b) 98 °C 30 s.
   (c) 65 °C 30 s.
   (d) 72 °C 45 s.
   (e) Return to **step 2** for $n - 1$ cycles.
   (f) 72 °C 7 min.
   (g) 4 °C hold.
3. Check the results by running and imaging the PCR products on a 2% agarose gel using the UV gel imager.
4. Assess the gel and determine the optimal cycle number carefully for large-scale PCR amplification of libraries (*see* **Note 15**).

3.8 Large-Scale PCR Amplification

This step enriches DNA fragments that are ligated to the methylated adapters at both ends.

1. In a PCR tube, set up PCR reaction mixture on ice and add the following (25 μL total volume).
   (a) 6.5 μL of Milli-Q water.
   (b) 12.5 μL of KAPA HiFi HotStart Uracil+ ReadyMix.
   (c) 3 μL of TruSeq PCR Primer Cocktail.
   (d) 3 μL of DNA sample (from Subheading 3.6, **step 4**).
2. Set up PCR program on a PCR thermal cycler: 95 °C for 4 min followed by determined cycles for each sample of 95 °C for 30 s, 65 °C for 30 s, 72 °C for 45 s, with final extension of 72 °C for 7 min.
3. Use a low volume elution PCR purification kit according to the manufacturer's instructions to purify large-scale amplified libraries, eluting in 18 μL TE buffer.

3.9 Size Selection with Beads

This step removes larger sized DNA fragments from the libraries in addition to removing any unligated adapters and residual dNTPs from the sample libraries.

1. Briefly centrifuge the tubes containing the DNA libraries to ensure the liquid is at the bottom of the tube.
2. Resuspend the Sample Purification Beads (SPB) by vortexing.
3. Add the first bead selection volume of 15 μL (0.6×) resuspended magnetic beads to each library. These beads will bind to unwanted large DNA fragments.
4. Mix well by pipetting up and down at least ten times. The solution should become homogeneous.
5. Incubate at room temperature for 5 min.
6. Place the tubes into the DynaMag-96 side magnet.
7. Once the beads have separated from the supernatant, and the solution becomes clear (about 5–10 min), carefully transfer the supernatant containing the desired smaller DNA fragments and into a new tube. Avoid collecting any beads with the supernatant.
8. Discard the beads containing unwanted larger DNA fragments.
9. Add the second bead selection volume of 10 μL (0.4× of original volume) resuspended SPB to each DNA library. These beads will bind to the desired DNA fragments.
10. Mix well by pipetting up and down at least ten times to create a homogeneous mixture.
11. Incubate at room temperature for 5 min.
12. Spin the tubes briefly and place the tubes onto the DynaMag-96 side magnet.
13. Once the beads have separated from the supernatant and the solution becomes clear (about 5–10 min), remove and discard the supernatant containing unwanted DNA fragments in addition to unligated adapters and residual dNTPs. Avoid disturbing the beads that now contain the desired DNA fragments .
14. While the tubes are in the magnetic rack, add 200 μL of 70% ethanol to wash the beads (*see* **Note 16**) that are bound to desired DNA fragments. Incubate for at least 30 s at room temperature, then carefully remove and discard the supernatant without disturbing the beads.
15. Repeat the wash step one more time for a total of two washes. Air-dry the beads while the tube is open on the magnetic plate for 10–15 min, until all liquid has evaporated (*see* **Note 17**).
16. Remove the tubes from the magnetic plate, and then immediately add 20–30 μL of TE Buffer to elute the DNA from the beads, mixing by pipetting up and down.

17. Place the tube into the magnet plate for 5–10 min or until the liquid appears clear in order to separate the beads from the liquid supernatant, which now contains the eluted DNA library. Once the liquid becomes clear, carefully transfer the liquid supernatant to a fresh tube. Avoid disturbing the beads. Approximately 3 μL of liquid will be retained by the beads.
18. Store at 4 °C for up to 1 week or at −20 °C until needed.

3.10 Quantitative and Qualitative Assessment of the Library

1. Quantify 1 μL of the final DNA library (from Subheading 3.9, **step 18**) using the Qubit fluorometer (or equivalent) according to the manufacturer's instructions.
2. Prior to sequencing, perform quality control using the 2100 Bioanalyzer using the high sensitivity DNA kit (1 μL required) according to the manufacturer's instructions.
3. Profiles should display a peak at approximately 300 bp (200 and 400 bp) corresponding to 150–325 base pair inserts (*see* **Notes 18** and **19**).
4. Determine the average size (bp) of each library, in conjunction with the Qubit reading, this will be used to determine the molarity (*see* **Note 20**).
5. DNA libraries can be stored at −80 °C until ready to pool (see Subheading 3.11).

3.11 Cluster Generation for Multiplexed RRBS

In consultation with a sequencing provider, calculate the volume of each library that should be added such that an equal quantity of each library will be present in the final pool, making a 10 nM in a total volume of 30 μL (top up with nuclease-free water if necessary). Ten microliters from the pooled libraries is then denatured with NaOH and diluted to a final concentration of 8 pM. One hundred and twenty microliters of the diluted sample is used for the cluster generation on the Illumina cBot machine.

3.12 Sequencing and Methylation Calls

This protocol is optimized for sequencing libraries on the HiSeq 2000/2500 platforms (Illumina). We recommend using 100 bp single-read (SE) sequencing kits. Using paired-end sequencing (PE) is also an option, but keep in mind that PE ends that overlap in the middle yield redundant methylation information for the same fragment, and therefore do not yield twice the amount of methylation data as predicted [38].

3.13 Assessing Data Quality and Alignment

Most sequencing providers will perform MiSeq QC prior to running on the HiSeq platform. Both the MiSeq QC and HiSeq run metrics and data files will be shared to the user on the Illumina BaseSpace platform (https://basespace.illumina.com/).

1. To evaluate the quality of the sequence data, we use the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which gives good summary information of per base

sequence quality, per base sequence content, adaptor content and other metrics. Other tools which can be used for quality evaluation include SolexaQA [39], FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), PRINSEQ (http://prinseq.sourceforge.net) [40], and MethyQA [41].

2. To trim sequence reads based on Phred score quality, remove filled-in CpG bases from the 3′ end and remove adaptor sequences from the reads, we use our in-house cleanadaptors tool [28]. The following command will trim adaptors from reads (adaptor sequences contained in contam.fa file), trim three bases back from the 3′ ends of matching reads to remove the C residue inserted during library preparation and leave no reads shorter than 4 bp:

```
cleanadaptors -I contam.fa -t 3 -x 4 -z -F HB4TFADXX-1551-
01_L001_R1.fastq.gz -Z > HB4TFADXX-1551-01_R1_adtrimmed.fastq
```

A similar operation can be performed using tools like FASTX toolkit, Trim Galore (www.bioinformatics.babraham.ac.uk/projects/trim_galore/) or Cutadapt (http://code.google.com/p/cutadapt/).

3. We use Bismark [32] to align RRBS reads as it is rapid and relatively unbiased [28]. Bismark provides information for both CpG and non-CpG (CHG, CHH) methylation. There are several other powerful bisulfite alignment programs which can also be used for similar purpose such as BSMAP [42], BS Seeker [43], RRBSMAP [44], BatMeth [45], PASS-bis [46], and SAAP-RRBS [47]. The following command will align the sequence reads to a Bismark-prepared reference genome (hs_ref_GRCh37) allowing for one mismatch in the seed (i.e., in the first 28 bp of the sequence reads):

```
bismark -n 1 hs_ref_GRCh37 HB4TFADXX-1551-01_R1_adtrimmed.
fastq
```

Bismark generates a BAM output file and an alignment report file, which gives summary data including mapping efficiency and proportions of context-specific cytosine methylation. For libraries generated from fresh-frozen tissues and cells, we routinely achieve 60–70% alignment efficiency and ~50% for FFPE tissues. As methylation in mammalian somatic tissues is exclusively in a CpG context, the proportion of cytosines "methylated" in a non-CpG context gives an indirect measure of bisulfite conversion efficiency (ideally should be <1%).

To extract the methylation call for every single C analyzed from the BAM output file generated from Bismark, the

following command can be used to generate a context-dependent (--comprehensive option) methylation bedGraph output file with a minimum of ten reads per methylation call:

```
bismark_methylation_extractor --bedGraph -comprehensive --cutoff 10 HB4TFADXX-1551-01_R1_adtrimmed.bam
```

For each CpG, the bedGraph file gives the following:

```
<chromosome> <start position> <end position> <methylation percentage>
```

3.14 Tools for Detecting Differential Methylation from RRBS Data

There are many different available tools for detecting differential methylation, each with different strengths and weaknesses. We have developed a differential methylation analysis package (DMAP) that we routinely use to generate reference DNA methylomes and identify differentially methylated regions across multiple samples from RRBS and WGBS data [34]. DMAP directly works with BAM or SAM files and contains a suite of statistical tools for differential analysis of genomic regions/fragments. The source code, documentation and test data are freely available from https://github/peterstockwell/DMAP [34]. The following command can be used to perform an ANOVA/F-test (−B) of methylation values from 40 to 220 bp MspI fragments with at least 2 CpGs (−F) and a minimum coverage of 10 (−t) between samples grouped by "R" or "S."

```
diffmeth -F 2 -t 10 -g Ref_Genome -z -B 40,220 \
-R HB4TFADXX-1551-01_R1_adtrimmed.bam \
-R HB4TFADXX-1551-02_R1_adtrimmed.bam \
-S HB4TFADXX-1551-03_R1_adtrimmed.bam \
-S HB4TFADXX-1551-04_R1_adtrimmed.bam \
> diffmeth_output.txt
```

DMAP also provides information and distances from nearest genes, CpG features. Using the diffmeth output file, the following command can be used to annotate each fragment with information on genomic feature data (−Q), the closest CpG island (−U, −R) and the closest protein-coding gene (−B).

```
identgeneloc -Q -U -R -B "protein_coding" -p Ref_Genome -s ".dat" \
-r diffmeth_output.txt > diffmeth_output.txt
```

4 Notes

1. Between 10 and 100 ng of input DNA allows for obtaining high-diversity libraries with comprehensive genomic coverage. However, below this amount the duplicate rates increase significantly and extensive sequencing is required for sufficient coverage. An input of 500 ng is usually required to generated decent libraries from paraffin-embedded (FFPE) samples.
2. Carry out all centrifugation steps at room temperature (15–25 °C).
3. Prior to DNA purification, the paraffin in an FFPE sample needs to be removed to allow exposure of the sample to proteinase K. We found that Qiagen deparaffinization solution is a good nontoxic alternative for xylene. The DNA concentration and purity were found to be comparable with xylene.
4. Use only filter pipette tips throughout the whole procedure of RRBS library generation. Moreover, to control for possible DNA contamination, use *DNAZap* (a nucleic acid decontamination solution that effectively degrades nucleic acids). *DNAZap* can be used to safely decontaminate any surface, including pipettors, gloves, microcentrifuge tubes, tube racks, and work benches. To avoid contamination of any equipment and solutions used for library construction, pre-PCR steps, subsequent steps, and post-PCR steps should be performed at a different working place.
5. Two deparaffinization solutions were compared based on xylene and xylene-free deparaffinization methods. Obtained results showed that both methods are similarly effective. However, deparaffinization with the utilization of deparaffinization solution was more efficient as the paraffin was easily dissolved and removed with fewer cycles, and no paraffin residues were observed on the surface of the tubes.
6. The presence of cellular proteins, particularly proteins that bind DNA, leads to problems with enzymatic reactions as well as sodium bisulfite treatment. In addition, RNA contamination can lead to inaccurate quantification of the DNA. We use QIAamp protease solution or QIAamp proteinase K (Qiagen) for complete protein digestion. While both proteases are used for protein digestion, subtle differences between both enzymes should be considered (e.g., Qiagen Protease is not compatible with Buffer ATL solution). RNA contamination can be removed by adding 2 μL of RNase A (10 mg/mL). Although TruSeq protocol recommends 1 μg input DNA, in our experience, 100 ng of DNA input is optimal to generate a high-quality library for sequencing while conserving the sample.

7. When using a thermocycler for certain enzymatic reactions, it should be noted that the temperature of the lid should be ~85 °C, instead of default settings of 105 °C. This prevents overheating and degradation of the samples. However, for reactions that require enzymatic activities, it is critical that the heated lid is turned off in order to avoid raising the temperature and prematurely deactivating the enzymatic activity.
8. The HiSeq2500 allows for a large number samples to be pooled. Different adapters and/or batches of adapters may perform differently. Sample libraries must be normalized at the same DNA concentration. Normalized libraries are then pooled in equal volumes before sequencing, this will allocate the reads fairly equally among the libraries.
9. To reduce the occurrence of adapter dimers, we use a lower concentration of adapters (diluted 1:4 in nuclease-free water) than recommended by the original protocol.
10. There are slight differences in the ligation efficiencies of the 24 different indexed, methylated TruSeq adapters from Illumina. Therefore, we recommend using the pooling guidelines in the TruSeq DNA sample preparation guide and the Illumina Experiment Manager software to determine adaptor combinations suitable for multiplexing.
11. wThe harsh chemical treatment during sodium bisulfite treatment can potentially damage DNA during a prolonged exposure, leading to a drastic reduction of library integrity. Therefore, it is best to minimize the length of the desulfonation step [48].
12. CT conversion solution can be stored at −20 °C up to 1 month, when needed, heat the bisulfite solution to 37 °C and vortex until all precipitates are dissolved again.
13. An optimal number of cycles is defined by the visualization of bands after semiquantitative amplification under each of the different reaction conditions (cycle numbers). Excessive amplification of the libraries increases the amplification of short fragments leading to skewed CpG coverage after sequencing [37]. At this stage, a wide range of adapter-ligated DNA fragments is obtained. The desired size range of fragments will be selected later to enrich for DNA regions with dense CpG loci using two subsequent rounds of SPB bead cleanups.
14. PCR with an optimal cycle number will generate a reasonable amount of 160–340 bp products without nonspecific amplification (a 125 bp adaptor-adaptor dimer band is often present, but this is removed during the second gel size selection).
15. NuSieve GTG agarose (low melting temperature agarose) gives fine resolution of small DNA fragments (10–1000 bp). Below 3% the NuSieve gel becomes fragile and hence requires more delicate handing. Because this type of agarose melts at 65 °C,

maintain a low voltage (6–7 V/cm of gel) to prevent the gel from overheating and/or melting.

16. Note that 70–80% ethanol is hygroscopic and should be prepared fresh to achieve optimal results.
17. Do not overdry beads as this will led to significant library loss.
18. Take the Bioanalyzer kit reagents out of the fridge 30 min before proceeding and leave to warm up to room temperature.
19. Bioanalyzer results might indicate PCR artifacts if the sequencing libraries display DNA molecules about twice the expected size (called "PCR-bubbles"). These PCR artifacts result from over-amplification of the libraries, which cause the PCR products to anneal to each other after depletion of available primers. The resulting annealing products are double/single-stranded; thus, they migrate slower on agarose gels as well as on Bioanalyzer assays. In most cases, PCR bubbles cannot be removed by SPRI bead size selections or Blue Pippin size selections. If necessary, the PCR bubbles can be eliminated by reamplifying of a tenfold diluted mixed-template PCR product with a low cycle number (a so-called reconditioning PCR) [49] (Fig. 2). However, to avoid unnecessary complexity and PCR bias, it would be best to optimize the library preparation protocol for a lower number of PCR cycles beforehand.

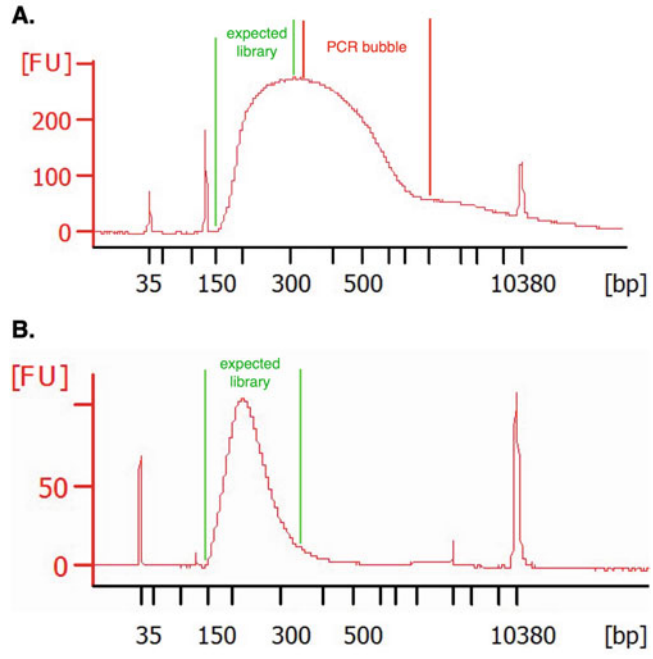


Fig. 2 Representative Agilent Bioanalyzer trace of the RRBS libraries before and after reconditioning PCR. (**a**) This shows a library with a "PCR bubble." (**b**) This shows a library with a uniform band and good peak size after performing reconditioning PCR

20. Use the following formula to determine library molarity:

$$\text{nM} = \text{concentration (ng/μL)}/(\text{average fragment length (bp)} \times 0.00065)$$

Elevated amounts of adapter dimers (which are observed as a peak ~125 bp) have been found to negatively impact the sequencing of the library. To remove excess adapter dimers, a further one-step cleanup of these libraries is recommended, using bead selection (1.8× of original volume), following similar steps to Subheading 3.9.

Acknowledgments

We gratefully acknowledge the help and support of the Otago Genomics and Facility (OGF), Dunedin during the development of this method. This work was supported by the Health Research Council (HRC) of New Zealand, the Maurice Wilkins Centre for Molecular Biodiscovery, and the New Zealand Institute for Cancer Research Trust. Aniruddha Chatterjee would like to thank the Rutherford Discovery Fellowship from the Royal Society of New Zealand and University of Otago for funding.

References

1. Conerly M, Grady WM (2010) Insights into the role of DNA methylation in disease through the use of mouse models. Dis Model Mech 3(5–6):290–297
2. Hackett JA, Surani MA (2013) DNA methylation dynamics during the mammalian life cycle. Philos Trans R Soc Lond Ser B Biol Sci 368(1609):20110328
3. Sarkies P, Sale JE (2012) Cellular epigenetic stability and cancer. Trends Genet 28(3): 118–127
4. Flavahan WA, Gaskell E, Bernstein BE (2017) Epigenetic plasticity and the hallmarks of cancer. Science 357(6348):eaal2380
5. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454(7205):766–770
6. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res 33(18):5868–5877
7. Lee YK, Jin S, Duan S, Lim YC, Ng DP, Lin XM, Yeo GS, Ding C (2014) Improved reduced representation bisulfite sequencing for epigenomic profiling of clinical samples. Biol Proced Online 16(1):1
8. Gu H, Smith ZD, Bock C, Boyle P, Gnirke A, Meissner A (2011) Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. Nat Protoc 6(4):468–481
9. Guo H, Zhu P, Wu X, Li X, Wen L, Tang F (2013) Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res 23(12):2126–2135
10. Chatterjee A, Ozaki Y, Stockwell PA, Horsfield JA, Morison IM, Nakagawa S (2013) Mapping the zebrafish brain methylome using reduced representation bisulfite sequencing. Epigenetics 8(9):979–989
11. Chatterjee A, Lagisz M, Rodger EJ, Zhen L, Stockwell PA, Duncan EJ, Horsfield JA, Jeyakani J, Mathavan S, Ozaki Y (2016) Sex differences in DNA methylation and expression in zebrafish brain: a test of an extended 'male sex drive' hypothesis. Gene 590(2):307–316
12. Chatterjee A, Stockwell PA, Horsfield JA, Morison IM, Nakagawa S (2014) Base-resolution DNA methylation landscape of zebrafish brain and liver. Genomics Data 2:342–344

13. Zhang C, Hoshida Y, Sadler KC (2016) Comparative epigenomic profiling of the DNA methylome in mouse and zebrafish uncovers high interspecies divergence. Front Genet 7: 110
14. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454(7205): 766–770
15. Choi M, Lee J, Le MT, Nguyen DT, Park S, Soundrarajan N, Schachtschneider KM, Kim J, Park J-K, Kim J-H (2015) Genome-wide analysis of DNA methylation in pigs using reduced representation bisulfite sequencing. DNA Res 22(5):343–355
16. Yuan X-L, Zhang Z, Pan R-Y, Gao N, Deng X, Li B, Zhang H, Sangild PT, Li J-Q (2017) Performances of different fragment sizes for reduced representation bisulfite sequencing in pigs. Biol Proced Online 19(1):1–8
17. Hartung T, Zhang L, Kanwar R, Khrebtukova I, Reinhardt M, Wang C, Therneau TM, Banck MS, Schroth GP, Beutler AS (2012) Diametrically opposite methylome-transcriptome relationships in high- and low-CpG promoter genes in postmitotic neural rat tissue. Epigenetics 7(5):421–428
18. Bock C, Kiskinis E, Verstappen G, Gu H, Boulting G, Smith ZD, Ziller M, Croft GF, Amoroso MW, Oakley DH (2011) Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell 144(3):439–452
19. Chatterjee A, Stockwell PA, Ahn A, Rodger EJ, Leichter AL, Eccles MR (2017) Genome-wide methylation sequencing of paired primary and metastatic cell lines identifies common DNA methylation changes and a role for EBF3 as a candidate epigenetic driver of melanoma metastasis. Oncotarget 8(4):6085
20. Chatterjee A, Rodger EJ, Stockwell PA, Le Mée G, Morison IM (2017) Generating multiple base-resolution DNA methylomes using reduced representation bisulfite sequencing. In: Oral biology. Springer, Berlin, pp 279–298
21. Chatterjee A, Macaulay EC, Ahn A, Ludgate JL, Stockwell PA, Weeks RJ, Parry MF, Foster TJ, Knarston IM, Eccles MR (2017) Comparative assessment of DNA methylation patterns between reduced representation bisulfite sequencing and Sequenom EpiTyper methylation analysis. Epigenomics 9(6):823–832
22. Rodger EJ, Chatterjee A, Stockwell PA, Eccles MR (2019) Characterisation of DNA methylation changes in EBF3 and TBC1D16 associated with tumour progression and metastasis in multiple cancer types. Clin Epigenetics 11(1):1–11
23. Bock C, Tomazou EM, Brinkman AB, Müller F, Simmer F, Gu H, Jäger N, Gnirke A, Stunnenberg HG, Meissner A (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. Nat Biotechnol 28(10):1106–1114
24. Baranzini SE, Mudge J, Van Velkinburgh JC, Khankhanian P, Khrebtukova I, Miller NA, Zhang L, Farmer AD, Bell CJ, Kim RW (2010) Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. Nature 464(7293): 1351–1356
25. Ludgate JL, Wright J, Stockwell PA, Morison IM, Eccles MR, Chatterjee A (2017) A streamlined method for analysing genome-wide DNA methylation patterns from low amounts of FFPE DNA. BMC Med Genet 10(1):1–10
26. Quail MA, Swerdlow H, Turner DJ (2009) Improved protocols for the illumina genome analyzer sequencing system. Curr Protoc Hum Genet Chapter 18:Unit 18.12
27. Chatterjee A, Rodger EJ, Stockwell PA, Weeks RJ, Morison IM (2012) Technical considerations for reduced representation bisulfite sequencing with multiplexed libraries. J Biomed Biotechnol 2012:741542
28. Chatterjee A, Stockwell PA, Rodger EJ, Morison IM (2012) Comparison of alignment software for genome-wide bisulphite sequence data. Nucleic Acids Res 40(10):e79
29. Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Babraham bioinformatics. Babraham Institute, Cambridge
30. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17(1):10–12
31. Babraham Bioinformatics (2017) Trim galore. Babraham Institute, Cambridge
32. Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27(11): 1571–1572
33. 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(16): 2078–2079
34. Stockwell PA, Chatterjee A, Rodger EJ, Morison IM (2014) DMAP: differential methylation analysis package for RRBS and WGBS data. Bioinformatics 30(13):1814–1822

35. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome Biol 13(10):R87
36. Buchbender A, Mutter H, Sutandy FXR, Körtel N, Hänel H, Busch A, Ebersberger S, König J (2020) Improved library preparation with the new iCLIP2 protocol. Methods 178: 33–48
37. Smith ZD, Gu H, Bock C, Gnirke A, Meissner A (2009) High-throughput bisulfite sequencing in mammalian genomes. Methods 48(3): 226–232
38. Babraham Bioinformatics (2016) FastQC: a quality tool for high throughput sequence data. Babraham Institute, Cambridge
39. Cox MP, Peterson DA, Biggs PJ (2010) SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. BMC Bioinformatics 11:485
40. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27(6):863–864
41. Sun S, Noviski A, Yu X (2013) MethyQA: a pipeline for bisulfite-treated methylation sequencing quality assessment. BMC Bioinformatics 14(1):259
42. Xi Y, Li W (2009) BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics 10:232
43. Chen PY, Cokus SJ, Pellegrini M (2010) BS Seeker: precise mapping for bisulfite sequencing. BMC Bioinformatics 11:203
44. Xi Y, Bock C, Muller F, Sun D, Meissner A, Li W (2012) RRBSMAP: a fast, accurate and user-friendly alignment tool for reduced representation bisulfite sequencing. Bioinformatics 28(3):430–432
45. Lim JQ, Tennakoon C, Li G, Wong E, Ruan Y, Wei CL, Sung WK (2012) BatMeth: improved mapper for bisulfite sequencing reads on DNA methylation. Genome Biol 13(10):R82
46. Campagna D, Telatin A, Forcato C, Vitulo N, Valle G (2013) PASS-bis: a bisulfite aligner suitable for whole methylome analysis of Illumina and SOLiD reads. Bioinformatics 29(2): 268–270
47. Ziller MJ, Gu H, Muller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A, Meissner A (2013) Charting a dynamic DNA methylation landscape of the human genome. Nature 500(7463):477–481
48. Munson K, Clark J, Lamparska-Kupsik K, Smith SS (2007) Recovery of bisulfite-converted genomic sequences in the methylation-sensitive QPCR. Nucleic Acids Res 35(9):2893–2903
49. Thompson JR, Marcelino LA, Polz MF (2002) Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. Nucleic Acids Res 30(9):2083–2088

Chapter 2

Data Analysis of DNA Methylation Epigenome-Wide Association Studies (EWAS): A Guide to the Principles of Best Practice

Basharat Bhat and Gregory T. Jones

Abstract

Array-based EWAS have become an increasingly popular technique to identify population epigenetic effects, particularly in humans. With the arrival of nonhuman species arrays, such as the mouse, this is likely to become an even more widely used technology. This chapter provides the less experienced researcher a guide to the analysis of data from the most widely used platform, the Illumina Infinium Methylation assay. This includes an overview of quality filtering, data normalization, analysis options, and techniques to improve the interpretation of results.

Key words DNA methylation, Epigenome-wide association studies, Methylation quantitative trait loci (meQTL)

1 Introduction

Epigenome-wide association studies (EWAS) have become increasingly popular in exploring and understanding the interaction between an individual's genetic background and their environment, particularly in the context of understanding how this interplay affects human health [1]. Epigenetic modifications do not change the DNA sequence but can nevertheless cause various changes in gene expression and cellular functions. EWAS quantify genome-wide patterns of epigenetic marks to identify associations between epigenetic variation and phenotypic traits. There are many types of epigenetic alterations, including histone modifications and noncoding RNAs, however the most widely investigated in genome-wide studies are DNA methylation (DNAm). While there are multiple forms of DNAm, in mammalian species the most predominant involve the enzymatic transfer of a methyl group from *S*-adenosylmethionine to the 5′-position of the cytosine

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_2,

within cytosine–guanine dinucleotides (CpG sites). Consequently, CpG sites methylation has been the focus of most population epigenetics studies.

Current methods for genome-wide DNA methylation studies include the use of microarrays or bisulfite-genomic DNA sequencing. The most recent iteration of Illumina Infinium Methylation assays for use in humans measures the DNA methylation level of over 850,000 specific sites. This array is commonly referred to as the Illumina EPIC array, with its predecessor being the 450k array. It is important to note that there are technical differences between these two assays [2]; nevertheless, most of the analysis principles discussed in this chapter are valid for both dataset types. While these Illumina Methylation Assays do not provide complete genome coverage, as would be the case with whole genome bisulfite sequencing, they do provide researchers with a highly reproducible and cost-effective means of performing a genome-wide screen for changes in DNAm within human samples.

The subsequent interrogation of such DNA methylation datasets, using appropriate statistical tools, allows researchers to study the epigenetic mechanisms underlying complex biological processes. However, these analyses need to be carefully designed and conducted. While there are similarities to both genome-wide association studies (GWAS) and gene expression (mRNA) studies, EWAS have their own distinct differences that need to be considered. This chapter aims to provide a guideline for performing EWAS analysis by outlining the current options for data normalization, covariant adjustment and data visualization.

2 EWAS Design Considerations

2.1 Study Type

An EWAS experiment typically involves a correlation between a quantified level of CpG site methylation (β-values, ranging from 0 to 1) and the phenotype of interest. There are no particular limitations on the types of epigenome-wide studies that can be performed, with designs including:

- Unrelated singleton cases and controls in a retrospective design.
- Case–control studies from prospective cohorts.
- Family-based or twin studies.
- Time course experiments.
- Continuous variable correlation model designs.

While these designs essentially mirror that of other types of genome-wide association studies (such as genetic variant or gene expression studies), the researcher should be aware that in many ways EWAS are more complex, particular due to the potential interacting effects of environmental exposures [3]. Nevertheless,

many EWAS follow the fairly typical observational case–control study design widely used in molecular epidemiology. Case–control studies are particularly suitable when a well-defined quantitative measure is not available to identify the severity or extend of a disease. However, if the phenotype of interest has a well-established quantitative measure (e.g., body mass index (BMI) [4], blood pressure [1], and cholesterol [5]) a continuous comparison can be utilized. Indeed, continuous traits are often preferable as they can improve the ability to detect an epigenetic effect when compared with categorical phenotypes.

2.2 Phenotype/Trait Selection

As discussed above, EWAS can either investigate a specific categorical condition or disease (e.g., current smoking [6] and T2DM [7–9]) or a correlation with a continuous phenotype (e.g., birthweight [10], age [11], BMI [4], HDL [5], and blood pressure [12]). If a continuous phenotype is to examined care must be taken to ensure that appropriate (parametric or nonparametric) statistical tests are applied depending on variable distribution (normal or nonnormal respectively).

Selection of cases and controls The starting point of most case–control studies is the identification of a case phenotype. This designation should be conducted in such a way as to limit bias. The more specific the phenotype classification, for example clinically verified versus participant self-declared, will likely impact on discovery statistical power. Similar consideration should be given to control phenotype-free status.

Several other factors must be taken into consideration when designing an EWAS or analyzing resulting data, including the choice of sample type, data normalization methods, how to appropriately adjust for confounding factors and correction for multiple testing.

2.3 Sample Type and Quality

The choice of the appropriate biological samples, storage, and processing can all influence the DNA methylation patterns observed in an EWAS. DNA methylation is generally considered to be a relatively chemically stable marker and appear to be robust and well-preserved, even after long term storage [13]. Commonly used sources of DNA include whole blood (WB), purified cells, fresh-frozen tissue, and formalin-fixed, paraffin-embedded (FFPE) tissues.

Despite all these sample types being able to successfully generate DNA methylation profiles using Illumina bead-chip assays, a direct comparison of samples of different collection types is not recommended (e.g., fresh frozen versus FFPE [14]) due to the possibility of introducing sample type bias.

2.4 Cellular Heterogeneity

Cell specific DNAm patterns are well recognized [15, 16], and consequently there has been concern about the effect of source tissue cellular heterogeneity on the resulting differential methylation patterns observed in an EWAS. These concerns are wide ranging and include how best to identify global (concordant) versus tissue specific (discordant) differential methylation patterns and the cryptic effects of cellular heterogeneity when comparing between samples. This was exemplified in an analysis comparing paired WB and adipose tissue samples, in which both tissue specific (discordant) and tissue nonspecific (concordant) methylation patterns were observed [15].

Despite this issue, and because of its convenience of accessibility, WB has become the most widely used source tissue in contemporary EWAS; however, this must be recognized as a heterogeneous tissue when analyzing differential methylation patterns. Moreover, it cannot be assumed that blood will represent a suitable surrogate or sentinel tissue for all phenotypes of interest. These issues notwithstanding, there are well established DNA methylation-based techniques that can impute cellular composition for a given sample and help identify if these cellular components are potentially confounding the observed associations. In blood, the Houseman method can be used to impute specific white cell counts [17, 18], while the Horvath method includes additional blood cell related coefficients that have been suggested as potential confounders of methylation-correlated aging [19].

2.5 Sample Size and Independent Validation

Regardless of the type of EWAS being conducted, as with genome wide association studies, the large number of statistical comparisons performed in EWAS results in sample size related issues impacting on detection power. It is, however, reassuring that robust top hit associations, such as cg19693031 (*TXNIP*) with type 2 diabetes [7–9], and cg05575921 (*AHRR*) with smoking [6], appear to be consistently evident within independent EWAS, most of which have relatively small (less than 1000) sample sizes. This is in striking contrast to the many thousands of participants typically required to see such reliable results in GWAS. It must be noted that, while top hit associations for common traits often appear relatively easy to detect, even in modest sized cohorts, other associations with weaker effect sizes will likely not be observed in small or modest sizes cohorts. Smoking exposure is a good example of this, wherein differential methylation of CpG sites associated with genes such as *AHRR*, *F2RL3*, *ALPPL2*, and *IER3* can be readily detected, even when examining only a few dozen smokers versus nonsmokers, there are at least 60 sites which appear to be reproducible associations in larger scale studies [6].

Determining the appropriate sample size to detect differential methylation associations will depend on a range of factors, not the least being the effect size of the trait being investigated [20], but

also the influence of other confounding variables within the cohort. Careful consideration of sample size, with reference to estimated power [20], should be a central factor in the design of any EWAS. Software tools, such as pwrEWAS [21], are publicly available to aid with this aspect of EWAS design and planning.

Of equal importance to discovery phase statistical power is the ability to validate associations within independent cohorts. Independent validation not only helps exclude false positive associations but also provides the opportunity for meta-analysis to identify true positive associations of weaker effect, as clearly demonstrated in a recent type 2 diabetes EWAS meta-analysis [9]. This also highlights another advantage of utilizing a common analysis platform (Illumina methylation arrays) because it makes it relatively easy to combine studies in a meta-analysis [6, 9], or to perform simpler confirmatory look-ups of both the magnitude and direction of effect of discovery based associations.

3 Data Analysis

3.1 Data Preprocessing

Once intensity data (IDAT) files are obtained these can be interrogated in the Illumina Genome Studio software package, following the manufacturer's recommendations, to check for poorly performing samples. A range of metrics for assessing output data quality are available (including completeness of sample bisulfite conversion and probe hybridization, green and red channel signal intensities and the number of CpG sites detected). In addition, negative controls probe preference should be assessed, prior to using these for background subtraction.

In addition, an often useful secondary QC check is to take the post–normalization output β-values, convert them to M-values [22], and generate a sample principal components plot using all variables (CpG sites). This technique will identify outlier samples (Fig. 1), for whom careful reintegration of their quality metrices should be conducted. Such samples (blinded to phenotype to limit possible selection bias) may be considered for subsequent analysis exclusion.

3.1.1 Normalization

A particularly important aspect of the analysis workflow is the selection of the normalization method that will be applied. There are a number of options that have been reported in the literature and that are publicly available within various EWAS analysis packages.

1. *Raw (no normalization):* Converts the Red/Green channel intensities for an Illumina methylation array into methylation signals, without using any normalization. Although this method has some specific applications it is not recommended for most EWAS.

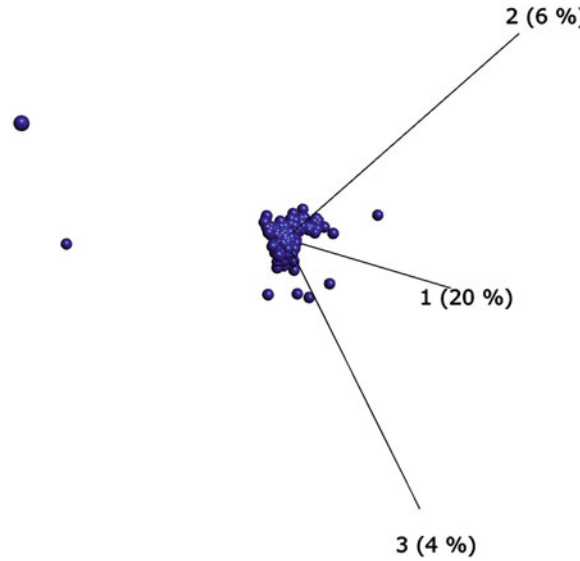


Fig. 1 Using sample principal component analysis to identify outlier samples of likely poor quality. This plot is of the first three principal components of all 850,000 CpG sites in a Type 2 Diabetes EWAS (1039 samples). No post–array scanning filtering, to exclude samples with poor QC metrics, was applied to emphasize the impact of passing such samples through to subsequent analysis. The two samples that stand out as extreme outliers, on the left-hand side of the plot, had very low post–array scanning signal intensities (and would normally be automatically excluded on this basis). Slightly more problematic are the five samples at the fringe of the central cluster which had more marginal QC values (these had red and/or green array channel signal intensities that were just below the product manufacturers recommended QC threshold intensities). We would recommend that all seven of these samples be excluded from subsequent analysis. In this case, this would be of little consequence as they represent less than 1% of the total sample size. Reassuringly, our experience is that poor quality samples (for a variety of reasons including highly degraded DNA, poor array hybridization, and incorrect bisulfite conversion) are easy to identify using a combination of the Genome Studio software's quality control metrics and subsequent quality assurance rechecking with PCA

2. *Quantile Normalization*: This function implements stratified quantile normalization preprocessing for Illumina methylation microarrays. Quantile normalization is a global adjustment method that assumes the statistical distribution of each sample is the same. Normalization is achieved by forcing the observed distributions to be the same and the average distribution, obtained by taking the average of each quantile across samples, is used as the reference. This method has worked very well in practice, however it is important to note that when the assumptions are not met, global changes in distribution that may be of biological interest may be lost and false positive associations may be induced.

3. *Illumina (Genome Studio) Normalization*: These functions implement preprocessing for Illumina methylation microarrays as used in Genome Studio, the standard normalization method provided by Illumina. This schema also includes both control and background normalization.
4. *SWAN Normalization*: Subset-quantile Within Array Normalization (SWAN) is a within array normalization method for the Illumina Infinium platform. The SWAN method has two parts. First, an average quantile distribution is determined using a subset of probes defined to be biologically similar based on CpG content. The subset for each probe type, from each channel (methylated or unmethylated), is sorted by increasing intensity. The value of each of the $3N$ pairs of observations is then assigned to be the mean intensity of the two probe types for that row or "quantile." This is the standard quantile procedure. The second step is to then adjust the intensities of the remaining probes, of which there are many more Infinium II than I, by interpolation onto the distribution of the subset probes. This is done for each probe type separately using linear interpolation between the subset probes to define the new intensities.
5. *Noob Normalization*: Noob (normal-exponential out-of-band) is a background correction method with dye-bias normalization for Illumina Infinium methylation arrays. Briefly, Noob performs within-array normalization correcting for background fluorescence and dye bias. It fits a normal-exponential convolution model to estimate the true signal conditional on the observed intensities. The background-corrected intensities are normalized by the variation in average intensity of the red and green channels via a multiplicative scale factor computed using the average intensities of the positive control probes.
6. *FUNNORM Normalization*: Funnorm is a between-sample (functional) normalization method that attempts to remove unwanted variation by adjusting for covariates estimated from a control probe matrix. Briefly, 42 summary measures are estimated from the combined 848 control probes and type I "out-of-band" intensities, with the first $m = 2$ principal components of the summarized measures chosen as covariates for intensity adjustment. Adjustment is performed separately in methylated and unmethylated intensities, and in type I and II probes. For probes mapped to X and Y chromosomes, males and females are processed separately, with ordinary quantile normalization used for probes on the Y chromosome because of the small number of probes ($N = 416$). By default, the functional normalization is applied after Noob.

7. *BMIQ Normalization*: BMIQ is a mixture model–based normalization method designed to correct the type II probe bias and make the methylation distribution of type II features comparable to the distribution of type I features. BMIQ fits a three-state (unmethylated, 50% methylated and fully methylated) beta mixture model for the type I and type II probes separately, with probes assigned to the state with maximum probability. Beta values for the type II features are normalized by state to the distributions of the same estimated in type I features. Like Noob, it is a within-array method.
8. *PBC Normalization*: PBC rescale the Infinium II data on the basis of the Infinium I data. While this technique should be viewed as an approximation method, it significantly improves the quality of Infinium II data.

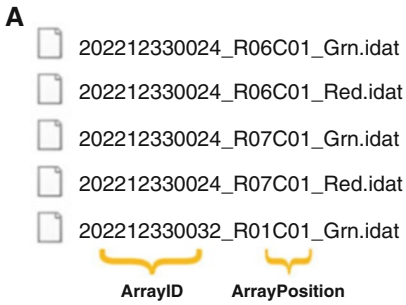
At the time of writing this guide, the authors personal view is that BMIQ Normalization is the most optimal approach for most EWAS. Knowing which normalization method has been applied to a given dataset is very important. While numerous DNA methylation datasets are publicly available via sites such as the NCBI's Gene Expression Omnibus (GEO), these are typically in the form of *ß*-value matrices. It is particularly important to note that different normalization methods may produce subtle differences in the resulting *ß*-values. It is, therefore, not recommended to directly compare CpG site *ß*-values derived using different normalization methods. Where possible, original signal intensity (IDAT) files should be reanalyzed using the same preprocessing and normalization schema for all cohorts being compared.

3.1.2 QC and Data Filtering

Quality control (QC) and filtering of microarray data are important preprocessing steps to produce accurate results. The work-flow usually proceeds as follows: (a) removing unexpressed and nonspecific probes, (b) Filtering population-based SNPs and nonautosomal sites, (c) Data imputation for correction of NA values on beta matrix file, (d) fixing outliers.

3.2 EWAS Analysis Using ChAMP; the Chip Analysis Methylation Pipeline

The following is a brief description of the steps implemented by a widely used software package, the *Ch*ip *A*nalysis *M*ethylation *P*ipeline (ChAMP), for the assessment of Infinium Methylation Assay datasets. The analysis begins with a quality control assessment, where checks for any potential sample issues, array failures or major batch effects are conducted. Subsequent normalization steps attempt to reduce technical variability and batch effects through both intra- and intersample normalization procedures. Quality control filtering and normalization procedures are essential to reducing bias and ensuring that the subsequent analyses provide reliable and more reproducible results.



B

Gender	Disease_state	Sentrix_ID	Sentrix_Position	Pool_ID	Sample_Plate	Sample_Well
M	Yes	202212330024	R06C01	6	1	1
M	Yes	202212330024	R07C01	7	1	1
F	No	202212330032	R01C01	1	1	1
F	No	202212330034	R03C01	3	1	1

Fig. 2 Overview of various commonly used files in the ChAMP analysis pipeline

The Chip Analysis Methylation pipeline ChAMP [23] version 2.20, is a freely available open-source package that implements quality control and statistical analysis procedures. It can be downloaded and installed from https://bioconductor.org/. The ChAMP package is designed for the analysis of either Illumina EPIC or 450k datasets. The package accepts input either as IDAT files or a beta-valued matrix (though we recommend using original IDAT files whenever possible). Each sample will have two IDAT files, consisting of red and green channel intensity data. IDAT file names should be formatted so as to indicate the Illumina array (Sentrix) identifier and the sample (row and column) position on that array (Fig. 2a). This allows for the linking of sample annotation files (Fig. 2b) to corresponding IDAT files.

3.2.1 Data Loading

Data analysis begins with loading data to R environment. ChAMP provides *champ.import* function to get data from IDAT files (Fig. 3a). This function accepts a minimum of two input parameters *arraytype* and *directory*. *Arraytype* specifies the microarray type "EPIC" or "450K." *Directory* specifies the location of IDAT files (Fig. 3a), this folder should contain phenotype file in .CSV format containing phenotype of interest and confounding factors (Fig. 3b).

3.2.2 Data Filtering and QC

The ChAMP package provides *champ.filter* function for data filtering. Data filtering in ChAMP involve the following.

- Removing probes with detection *P*-value (default <0.01).
- Remove probes with <3 beads in at least 5% of samples per probe.
- Remove nonspecific probes and unexpressed probes.
- By default, ChAMP filters out polymorphic and nonautosomal sites.

In addition to the data filtering capability, ChAMP package provides *champ.QC* function for data QC. Usually, three plots would be generated by *champ.QC* function.

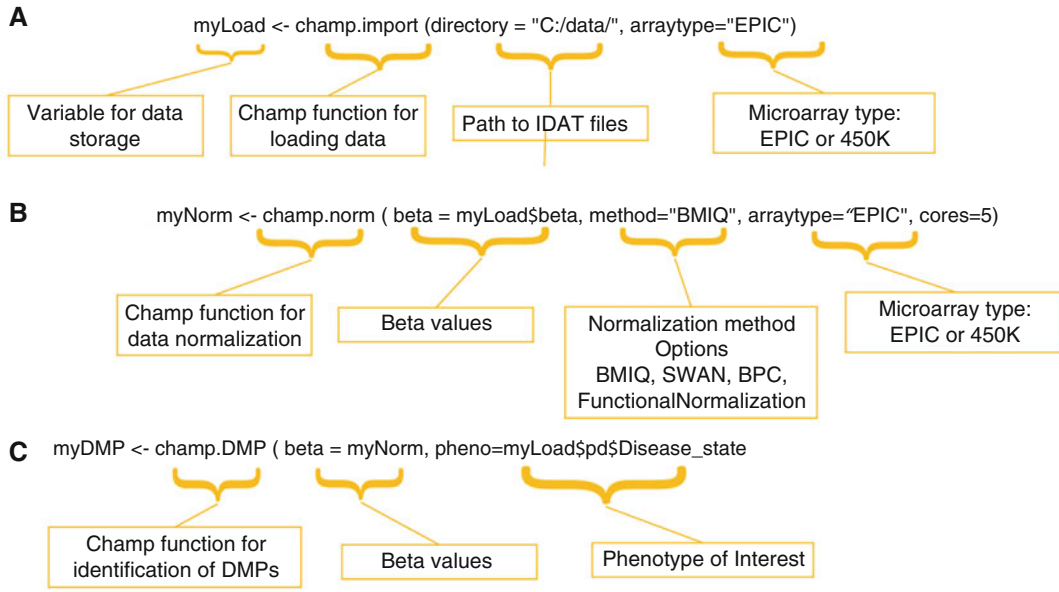


Fig. 3 Structure of the various commands used in the ChAMP pipeline

- *Multidimensional Scaling (MDS) Plot*: This allows visualization of the similarity of samples based on the top 1000 most variable probes amongst all samples.
- *Density Plot*: Beta value distribution for each sample.
- *Dendogram*: The clustering plot for each sample.

3.2.3 Data Normalization

In *champ.norm* function ChAMP package provide four data normalization methods BMIQ, SWAN, PBC, and FunctionalNormalization (Fig. 3b).

3.2.4 Differential Methylation Probes

ChAMP provides *champ.DMP* function for the identification of differentially expressed methylated sites for categorical and continuous variables (Fig. 3c). This function implements the Bioconductor limma package to calculate the *P*-value for differential methylation using a linear model. Of note, linear models assume that the test data is symmetric about the mean, for nonsymmetric (non–normally distributed) datasets it is recommended to utilize RANK regression for large datasets or log-transformation in small datasets.

3.3 Independence of Association/Adjusting for Confounding Factors

By definition, DNA methylation can be affected by each individual's genetic background and environment. Studies on whole blood derived DNA have robustly identified numerous CpG sites associated with a diverse range of chronic disease risk factors such as aging [11], smoking exposure [6], type 2 diabetes [7, 9], dyslipidemia [5], and BMI [4, 24]. Many of these associations show a

Table 1
Iterative modelling of confounding variables

T2DM EWAS, cg06500161_AGBG1	*P*-value	Fold change
Unadjusted	7.2×10^{-15}	1.0144
Age, sex	1.2×10^{-14}	1.0143
Blood cell composition (BCC; Houseman)	1.9×10^{-18}	1.0156
Age, sex, BCC	3.8×10^{-18}	1.0155
Body mass index (BMI)	7.9×10^{-11}	1.0123
HDL-C	2.4×10^{-11}	1.0126
LDL-C	4.1×10^{-10}	1.0120
Triglycerides	1.5×10^{-11}	1.0124
HDL-C, LDL-C, triglycerides,	9.7×10^{-7}	1.0094
HDL-C, LDL-C, triglycerides, BMI	1.4×10^{-5}	1.0085
HDL-C, LDL-C, triglycerides, BMI, BCC	4.8×10^{-7}	1.0092

To determine potential confounding effects on a top hit association, iterative statistical regression modelling should be applied, beginning with the unadjusted association and sequentially adding potential confounders. In the above example, age and sex appear to have little effect on the cg06500161 (*AGBG1*) association with diabetes. Moreover, this association does not appear to be significantly confounded by blood cell composition. However, notice that addition of BMI and dyslipidemia substantially reduced the strength of this association, observations which are consistent with the known epidemiological overlap between diabetes, obesity, and dyslipidemia. Avoiding overadjustment in multiple regression modeling of EWAS should be a key feature in the prospective analysis plan

degree of overlap, for example cg06500161 (*ABCG1*) is strongly associated with type 2 diabetes [9], dyslipidemia (both HDL-cholesterol and triglycerides) [5], and obesity [4]. In the case of the association between cg06500161 (*ABCG1*) and T2DM, inclusion of BMI and dyslipidemia (Table 1) could obscure these markers' association. We therefore suggest that substantial caution should be applied to the selection of variables in an adjusted risk model, as true associations may be lost in overly inclusive multivariate models. An iterative approach, sequentially adding components to determine effects on associations, should always be applied (Table 1). There is no universal dogma for which variables to include as EWAS confounders in multivariate regression, though it has become common to see models that apply a "default" correction of age, sex and (blood) cell composition. While there is some logic to this, particularly if the case/control groups are not well matched for age and sex, again some caution is suggested, as such corrections may mask real phenotype-of-interest associated biological effects. For example, with regard to blood cell composition, correcting for cellular fractions (Houseman method) [18] or more functional cellular markers (Horvath Method) [19], may obscure disease specific differences such as altered white cell counts or activation states.

Other key factors that should be considered are cryptic relatedness and differences in age and sex between participants. Related individuals may produce false-positive associations, particular at CpG sites that are methylation quantitative trait loci (meQTL, see Subheading 4.3). A large number of CpG sites have a strong association with age and/or sex [11, 25] and these factors should therefore be matched, as much as possible, in population EWAS design. Introduction of potential bias through case/control inclusion and exclusion criteria should also be carefully considered, particularly with regard to cryptic confounding due to "design constructed" differences in environmental exposures.

3.4 Singular Value Decomposition (SVD) Analysis and Adjusting for Confounding Factors

Determining the independence of DNAm associations from other potentially confounding variables remains one of the greatest challenges in the analysis of population EWAS. One method that can be applied to this problem is the use of Singular Value Decomposition (SVD). SVD can "decompose" a complex association matrix into a weighed ordered sum of separable matrices. By applying SVD to EWAS datasets it is possible to identify independent statistical associations between sets of related markers (CpG sites) and potentially confounding variables. As shown in Fig. 4, this method not only identifies the (PCA) related sets of differentially methylated CpG sites associated with the trait of interest, but also indicates potential confounders of these associations. Often these can be useful in suggesting the environmental exposures or biological components which are driving the observed epigenetic effect.

This includes understanding the relationship between differentially methylated CpG sites and specific cell types. In the example shown in Fig. 4, the methylation changes associated with the case (abdominal aortic aneurysm) group appear to be strongly related to chronological age and CD8 T-cell markers, both of which are known contributory factors in aneurysm pathogenesis. Similarly, this approach has been used to demonstrate that smoking predominantly alters myeloid DNA methylation patterns, thereby improving our understanding of how smoking disproportionally affects immune-cells and influences smoking related disease [23].

4 Interpreting the Data Output

Accurate information extraction and interpretation from a population EWAS depends heavily on careful bioinformatics analysis and is aided by appropriate selection of data visualization tools.

4.1 Data Visualization

Manhattan plots, in conjunction with Q-Q plots, are widely used to summarize EWAS associations (Fig. 5). It is important to note that the appearance of these plots will have distinct differences from that typically observed in a genetic association study (GWAS). Because

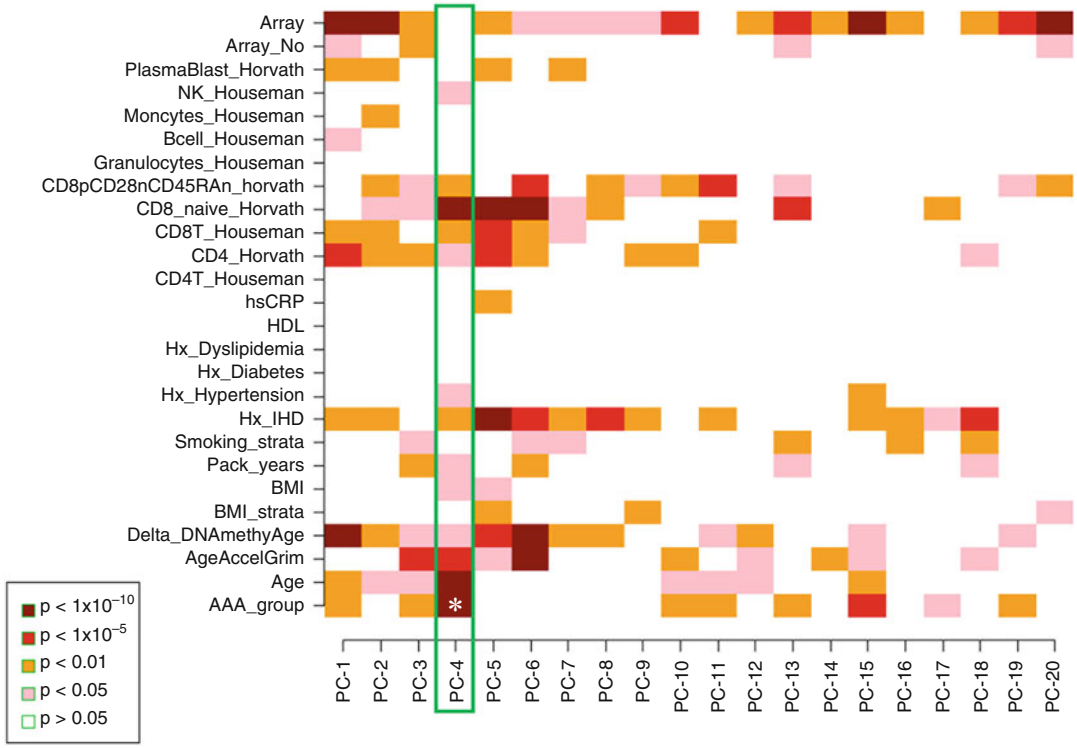


Fig. 4 Singular Value Decomposition (SVD) analysis to identify confounding variables. In this example from a case control EWAS for abdominal aortic aneurysm (AAA), phenotypic variables are listed in rows, while various CpG site principal components are in columns. Note that for the phenotype of interest (AAA) the most significant association is with markers within principal component 4 (PC4, asterisk), and that this association is potentially confounded by age and methylation predicted (Horvath Method [19]) CD8 naïve T cell levels, but not strongly by other common cardiovascular disease risk factors such as smoking, hypertension or dyslipidemia

of SNP linkage disequilibrium, a GWAS will normally have multiple significant markers within a locus. Although multiple significantly differentially methylated CpG sites can certainly be observed in a specific gene region, it is also common to see singular marker associations (Fig. 5), which are shown to be consistently reproducible in multiple independent cohorts. Similarly, interpretation of Q-Q plots is different between EWAS and GWAS. In a GWAS substantial lambda inflation (genome wide deviation of the test statistic from the null-hypothesis) typically suggests a systemic bias, often caused by differences in the (ancestral) population substructure of cases versus controls. In an EWAS, it is common to see much greater lambda inflation values. While it is possible for genetic population substructure to influence some of the methylation signal through meQTLs, this is typically very modest. Most of the high lambda inflation effect observed in many EWAS is due to environmental exposure differences which can be influenced by study design, particularly case control inclusion exclusion criteria

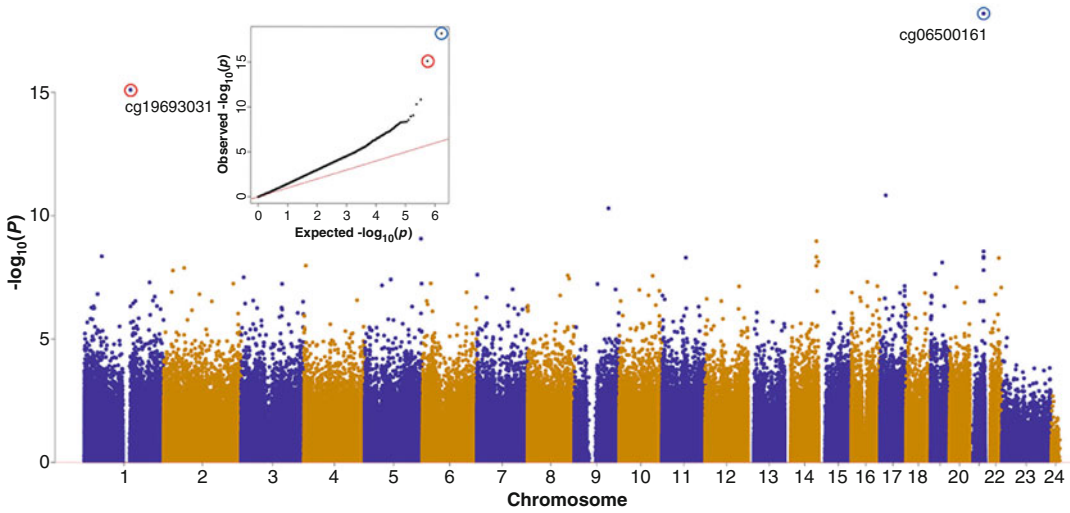


Fig. 5 Type 2 Diabetes EWAS Manhattan and QQ plot (adjusted for age, sex, blood cell composition)

and intergroup age and sex matching. Substantial differences can also be observed due to differences in cellular composition. The magnitude of the effect for each confounding component will vary depending on the phenotype being examined and the underlying nature of the study design. For example, in the case–control EWAS shown in Fig. 6, the relatively global effects that potential confounders, such as age, can have on test-statistics is clearly evident.

Another useful way of visualizing summary data is the volcano plot (Fig. 7). Unlike the Manhattan plot it partitions differentially methylated sites in hypo and hypermethylated clusters and also separates markers on effect size (such as fold change difference between groups).

4.2 Interpreting Top Hit Associations

EWAS results are often reported in a table listing the CpG site, Chr: position, Gene, effect coefficient, and *P*-value. The gene ascribed as being associated with a particular CpG site clearly becomes central to subsequent bioinformatic/pathway analysis, and it is therefore essential that this is as accurate a designation as possible. The data analyst should be aware that gene annotation to each CpG site is typically based on physical proximity (as known at the time of annotation list generation). It is always important to double check this with the most up to data information, this should be done in conjunction with an examination of the potential tissue specific functionality states (enhancer, promoter, insulator, etc.) using appropriate tracks (e.g., ENCODE/Broad ChromHMM) within tools such as the UCSC Genome Browser (https://genome.ucsc.edu). Once CpG site gene associations have been verified as accurately as possible, these can then be carried forward into biological and disease pathway analyses.

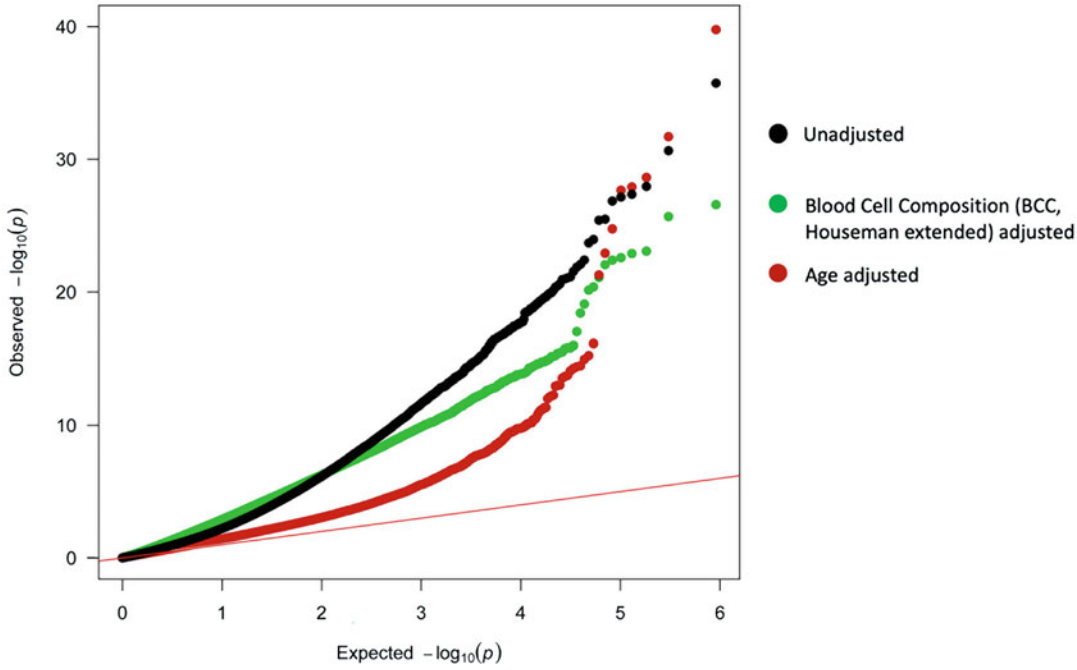


Fig. 6 An EWAS QQ plot for abdominal aortic aneurysm (AAA). This plot shows the observed versus expected *P*-values for an Illumina 450K Human Methylation array dataset comparing males study participants with either AAA ($n = 473$, mean age 76 years) or AAA-free controls ($n = 488$, mean age 69 years). When age is included as a confounding variable the global effect of the case–control age mismatch is clearly evident. In contrast, adjusting for blood cell composition appears to only significantly influence a small subset of methylation markers in this particular analysis

While simple lists of differentially methylated CpG sites are useful in identifying top hit associations, they do not inform the relationship between these markers. The use of variable PCA can identify discrete clusters of correlated CpG sites that are associated with the phenotype of interest. As was discussed for SVD, such analysis can be useful in identifying potentially overlapping biological/environmental effects, for example if a novel marker is strongly correlated with a set of CpG sites which have been previously linked to a specific environmental exposure (such as smoking or obesity), this may suggest, at least part of, the underlying biological drivers for the marker in question.

4.3 Identification of Methylation Quantitative Trait Loci (meQTLs)

It is well recognized that some components of differential methylation patterns are particularly prone to be influenced by the presence of genetic variants. Most concern centers around polymorphisms which are either located directly within CpG sites or alter bead assay probe binding. Lists of CpG sites which are potentially prone to this problem are publicly available and can be used to cross-check potential associations [2, 26].

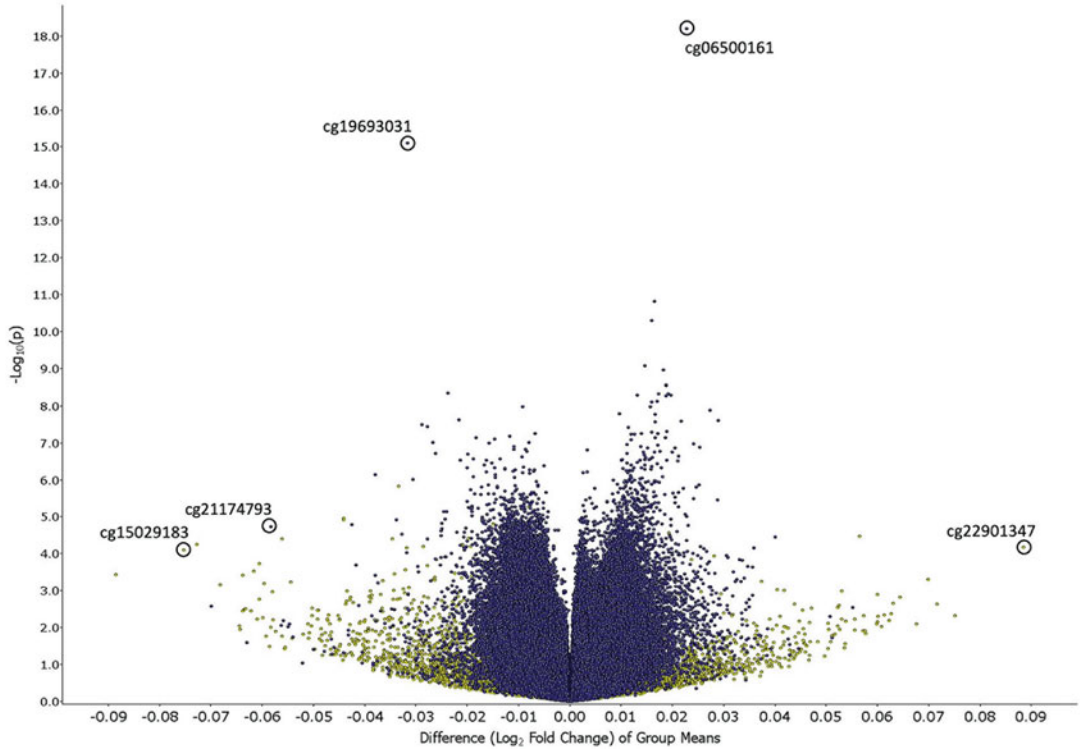


Fig. 7 Type 2 Diabetes EWAS volcano plot (adjusted for age, sex, blood cell composition). CpG sites which are hypomethylated in diabetics have negative fold changes (x = axis). Those sites prospectively annotated as being associated with a common SNP (MAF > 0.03) are labelled yellow. Notice that these form the majority of sites with high fold changes but relatively modest (non–genome-wide significant) *P*-values. The *ß*-values for two such sites (cg15029183 and cg22901347) are shown in Fig. 9, while the equivalent plot for the top hit diabetes association (cg06500161) is shown in Fig. 8

One strategy that should be discouraged is the prefiltering of these sites. If this is done for all potentially SNP variant probes and CpG sites it can result in a very large loss of possibly interesting data. For example, a recent paper which performed two EWAS for circulating Lp(a) levels showed that the strongest differential methylation association was in fact tagging a large effect size meQTL [27].

MeQTL are usually readily apparent within (categorical phenotype) EWAS volcano plots, where most have a high fold change in *ß*-value between groups but relatively modest *P*-values (Fig. 7). While it is easy to flag sites within such plots, using publicly available annotation lists, this will not always guarantee identification of every polymorphic site. The *ß*-values for all associations of interest should always be individually visualized. For example, in the type 2 diabetes EWAS example (Figs. 5 and 7), flagging previously annotated common CpG site polymorphisms suggests that the top hits are not polymorphic, and this is confirmed when the *ß*-

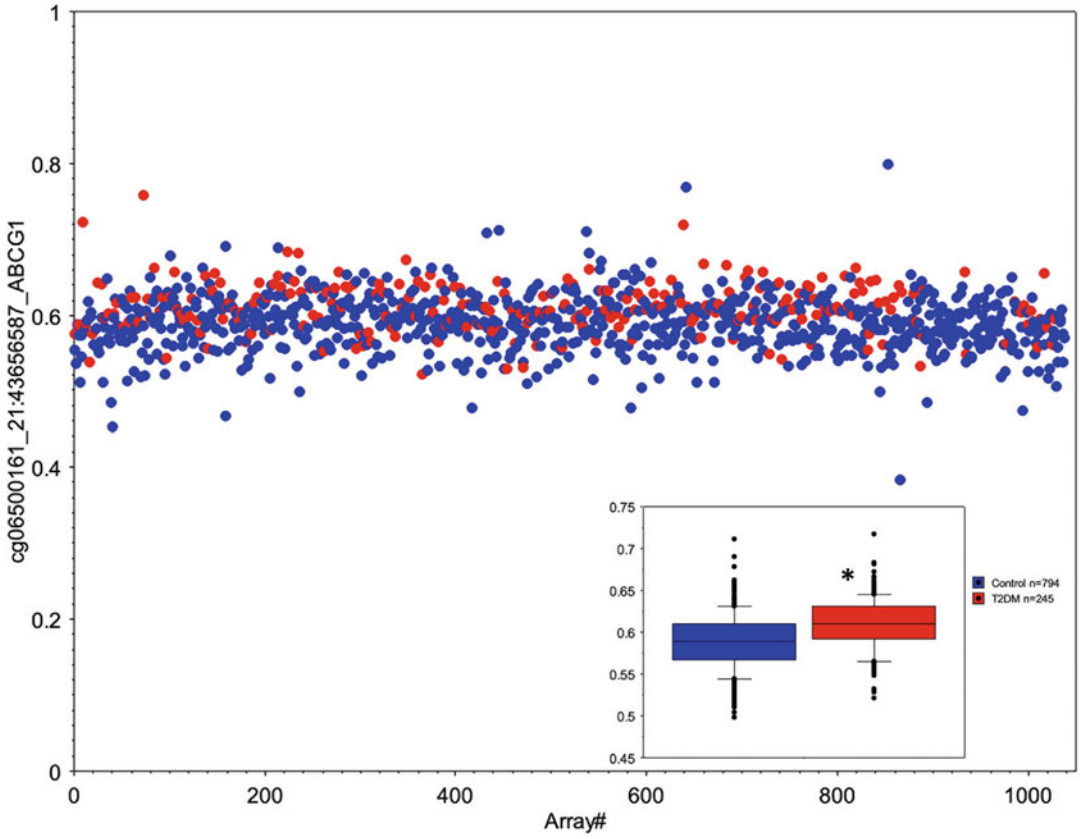


Fig. 8 Type 2 Diabetes EWAS, ß-value plots for top hit CpG site (cg06500161). The scatter plot shows the unadjusted ß-values ($y =$ axis) arranged in order of sample arraying (x-axis). Diabetes (red) and nondiabetics (blue) are labelled separately. While diabetics were on average only approximately 2% more hypermethylated at this site than nondiabetics, due to the relatively low variance of ß-values, this difference was highly statistically significant (*unadjusted P-value $= 5.2 \times 10^{-17}$, Mann–Whitney U-test)

values for these sites are visualized (Fig. 8). While the mean ß-values are significantly different between groups the variances have a relatively normal distribution (as suggested in the associated box plot). Three examples of sites with high fold ß-value change between groups but relatively modest P-values (cg22901347, cg15029183, cg21175793) are also highlighted. Two sites (cg22901347, cg15029183) were preannotated as being commonly polymorphic in Europeans (the ethnicity of the participants in this EWAS), but one (cg21175793) was not (Figs. 9 and 10 respectively). All three sites were confirmed as polymorphic, based on ß-value patterns, with the initially "cryptic" site (cg21175793, Fig. 10) being shown to be a likely error in the publicly available annotation reference file. Notice that the ß-values for polymorphic CpG sites form distinct groups which conform to genotype frequencies (adhering to the proportions predicted by the Hardy–Weinberg equilibrium). Although it can be advantageous to have

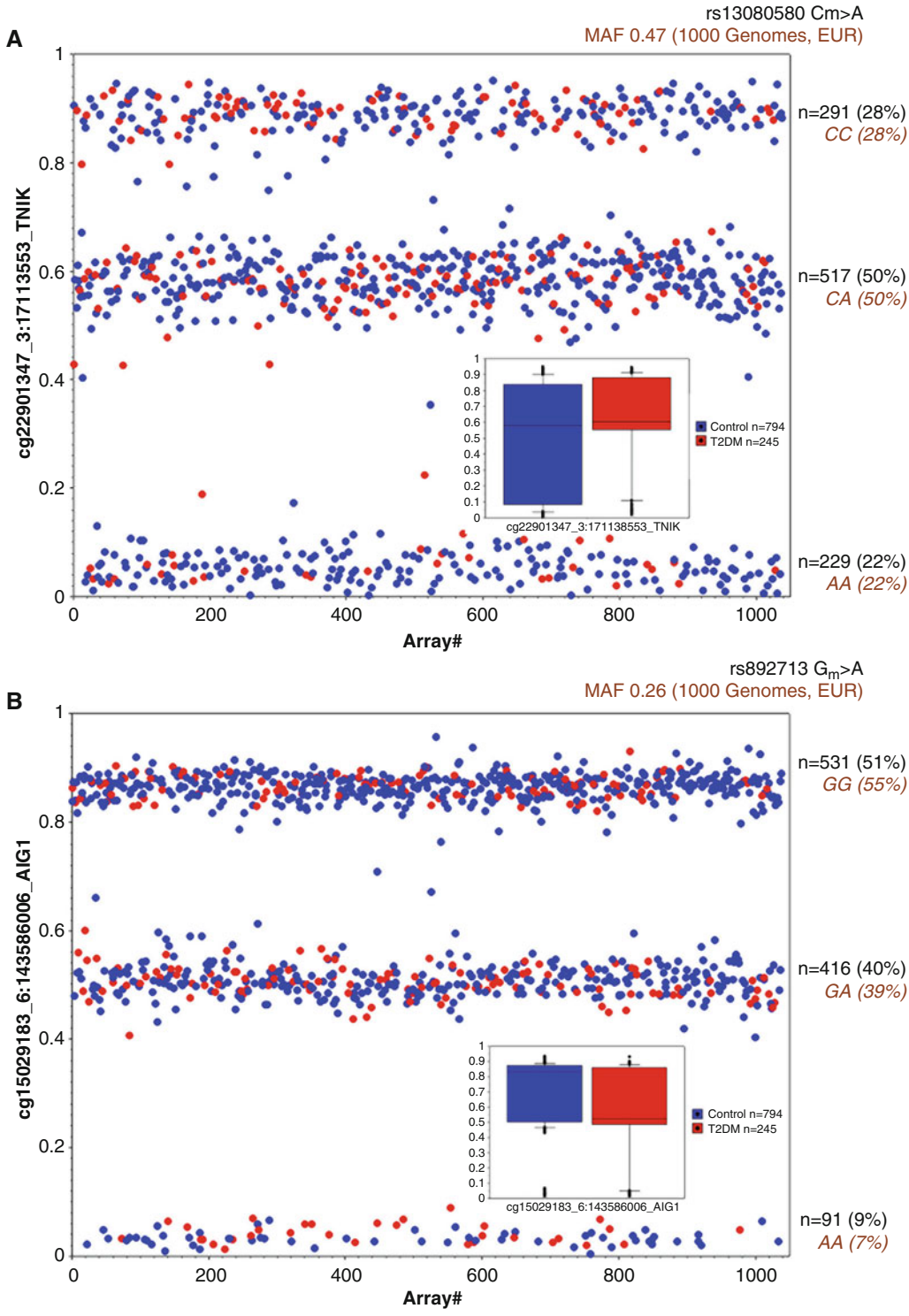


Fig. 9 Type 2 Diabetes EWAS, *ß*-value plots for two sites (**a**) cg22901347 and (**b**) cg15029183, that appear to be methylation quantitative trait loci (meQTLs). These scatter plots show the unadjusted *ß*-values ($y =$ axis)

matching sample genotype data available, these examples show that, at least for relatively common SNPs, the ß-values in and of itself can accurately identify meQTL/SNP associations. For rare SNPs, depending on the EWAS sample size, it may be that only a

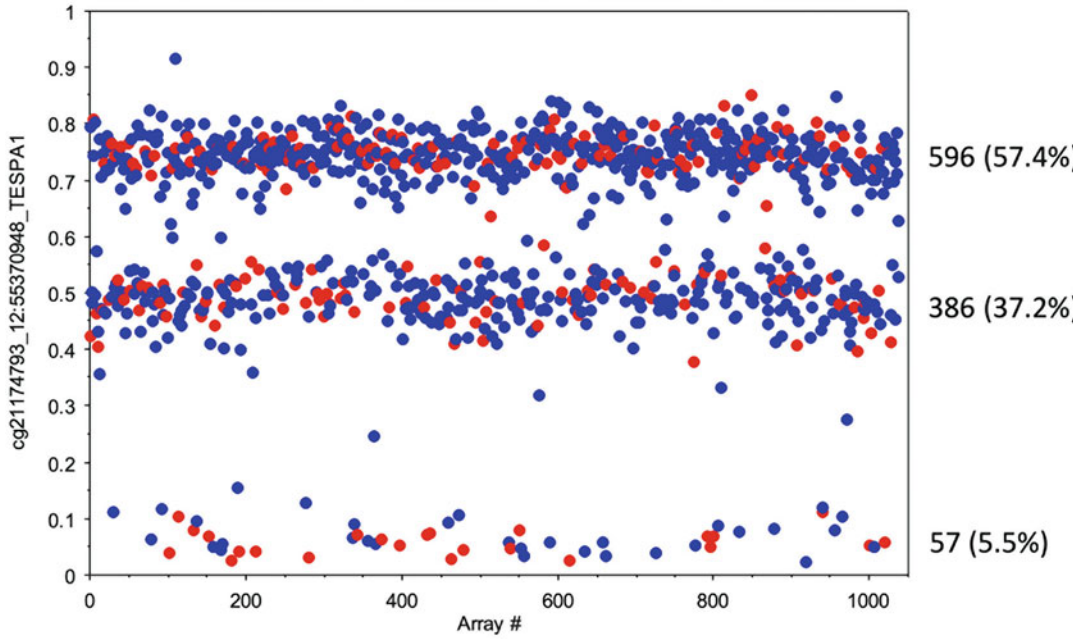


Fig. 10 A "cryptic" CpG site with a high fold ß-value change but modest *P*-value. This site was not flagged as a potential common meQTL (Fig. 7) as, although it was annotated being associated with a SNP (rs7314940 $C_m > T$), the Illumina probe site annotation file denoted the corresponding MAF as only 0.0024. The observed methylation ß-values pattern predicts an MAF of 0.23, with the observed proportions of high, medium and low values being in corresponding Hardy–Weinberg equilibrium. A subsequent dbSNP lookup reported an MAF of 0.24 in CEU, suggesting an error in the annotation file. Similar to other meQTL sites, the observed difference in cg21174793 ß-values between diabetics and controls, appears likely to be driven by a greater proportion of diabetics having a genotype which cannot be methylated (TT 9.9% diabetics versus 4.2% controls, χ^2 *P*-value 0.0007)

Fig. 9 (continued) arranged in order a sample arraying (*x*-axis). Diabetes (red) and nondiabetics (blue) are labelled separately. While both of these sites had large fold-change intergroup differences (see Fig. 7), these were of modest statistical significance (cg22901347 unadjusted *P*-value = 0.001, cg15029183 unadjusted *P*-value = 0.0002, Mann–Whitney *U*-test). (**a**) The cg22901347 site was is known to be associated with a SNP (rs13080580, minor allele frequency 0.47 in Europeans) for which the more common allele can be methylated (C_m) while the minor allele (A) cannot. Notice the predicted 1000 Genomes allele frequencies match the proportion of samples that have high (CC genotype), intermediate (CA) or low (AA) methylation patterns. (**b**) The cg15029183 site is known to be associated with a SNP (rs892713, MAF 0.26 in Europeans) for which the major allele can be methylated (G_m), while the minor allele (A) cannot. Again, notice that the predicted allele frequencies match the proportion of samples that have high (GG genotype), intermediate (GA) or low (AA) methylation patterns

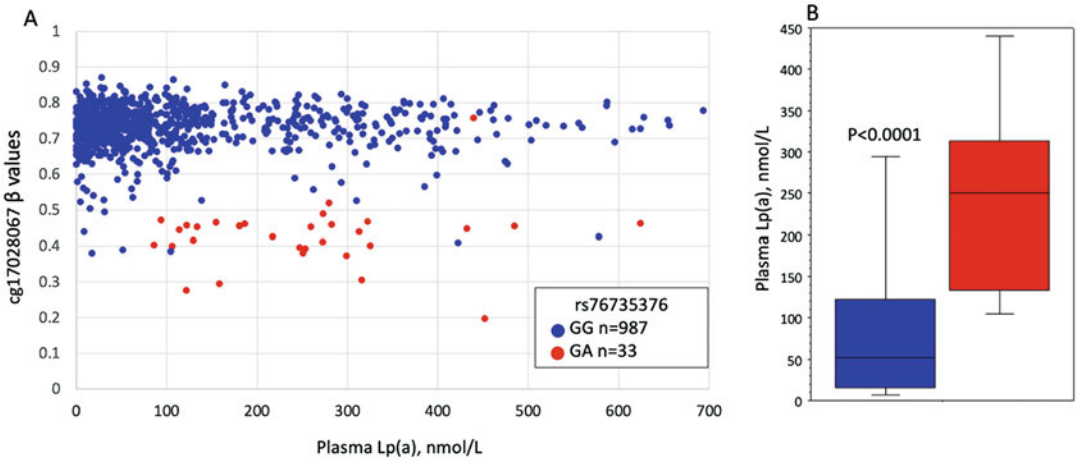


Fig. 11 The relationship between plasma Lp(a), cg1702867 methylation and rs76735376 genotype. In this EWAS for plasma Lp(a), the top association was cg1702867 [27]. (**a**) When cg1702867 β values (*Y*-axis) were plotted against Lp(a) (*X*-axis) a group with lower methylation values was apparent. Subsequent genotyping showed that the vast majority of these samples were heterozygotes for a CpG SNP (rs76835376 $G_m > A$, MAF 0.017). Notice that those samples with a GA genotype (red) had *ß*-values approximately half that of the GG genotypes. (**b**) This observation is biologically significant as rs76735376 heterozygotes (3.3% of participants) had significantly higher plasma Lp(a) levels than major allele homozygotes ($P < 0.0001$, Mann–Whitney *U*-test)

modest number of "heterozygotes" are present, but as shown in the Lp(a) example (Fig. 11) [27] such associations can still represent potentially biologically important effects. In brief, the identification of meQTL in EWAS has the potential to highlight mechanistically novel genetic associations [27] and their potential presence should be incorporated into all population EWAS analysis plans.

In summary, the design and analysis of contemporary EWAS has distinct differences from that of other genome-wide association studies. This chapter has summarized some of the key study design, data processing and statistical analysis steps than should be considered, particularly for studies using the Illumina Methylation Array technology.

Acknowledgments

The authors were supported in this work by grants from the Health Research Council of New Zealand (14/155, 17/402), The Healthier Lives, National Science Challenge and Genomics Aotearoa (a New Zealand Ministry of Business, Innovation and Employment funded research platform).

References

1. Rakyan VK, Down TA, Balding DJ, Beck S (2011) Epigenome-wide association studies for common human diseases. Nat Rev Genet 12(8):529–541. https://doi.org/10.1038/nrg3000
2. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Djik S, Muhlhausler B, Stirzaker C, Clark SJ (2016) Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biol 17(1):208. https://doi.org/10.1186/s13059-016-1066-1
3. Chadwick LH, Sawa A, Yang IV, Baccarelli A, Breakefield XO, Deng H-W, Dolinoy DC, Fallin MD, Holland NT, Houseman EA, Lomvardas S, Rao M, Satterlee JS, Tyson FL, Vijayanand P, Greally JM (2015) New insights and updated guidelines for epigenome-wide association studies. Neuroepigenetics 1:14–19. https://doi.org/10.1016/j.nepig.2014.10.004
4. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, Tsai PC, Ried JS, Zhang W, Yang Y, Tan S, Fiorito G, Franke L, Guarrera S, Kasela S, Kriebel J, Richmond RC, Adamo M, Afzal U, Ala-Korpela M, Albetti B, Ammerpohl O, Apperley JF, Beekman M, Bertazzi PA, Black SL, Blancher C, Bonder MJ, Brosch M, Carstensen-Kirberg M, de Craen AJ, de Lusignan S, Dehghan A, Elkalaawy M, Fischer K, Franco OH, Gaunt TR, Hampe J, Hashemi M, Isaacs A, Jenkinson A, Jha S, Kato N, Krogh V, Laffan M, Meisinger C, Meitinger T, Mok ZY, Motta V, Ng HK, Nikolakopoulou Z, Nteliopoulos G, Panico S, Pervjakova N, Prokisch H, Rathmann W, Roden M, Rota F, Rozario MA, Sandling JK, Schafmayer C, Schramm K, Siebert R, Slagboom PE, Soininen P, Stolk L, Strauch K, Tai ES, Tarantini L, Thorand B, Tigchelaar EF, Tumino R, Uitterlinden AG, van Duijn C, van Meurs JB, Vineis P, Wickremasinghe AR, Wijmenga C, Yang TP, Yuan W, Zhernakova A, Batterham RL, Smith GD, Deloukas P, Heijmans BT, Herder C, Hofman A, Lindgren CM, Milani L, van der Harst P, Peters A, Illig T, Relton CL, Waldenberger M, Jarvelin MR, Bollati V, Soong R, Spector TD, Scott J, McCarthy MI, Elliott P, Bell JT, Matullo G, Gieger C, Kooner JS, Grallert H, Chambers JC (2017) Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. Nature 541(7635):81–86. https://doi.org/10.1038/nature20784
5. Braun KVE, Dhana K, de Vries PS, Voortman T, van Meurs JBJ, Uitterlinden AG, BIOS consortium, Hofman A, Hu FB, Franco OH, Dehghan A (2017) Epigenome-wide association study (EWAS) on lipids: the Rotterdam Study. Clin Epigenetics 9:15. https://doi.org/10.1186/s13148-016-0304-4
6. Gao X, Jia M, Zhang Y, Breitling LP, Brenner H (2015) DNA methylation changes of whole blood cells in response to active smoking exposure in adults: a systematic review of DNA methylation studies. Clin Epigenetics 7:113. https://doi.org/10.1186/s13148-015-0148-3
7. Soriano-Tarraga C, Jimenez-Conde J, Giralt-Steinhauer E, Mola-Caminal M, Vivanco-Hidalgo RM, Ois A, Rodriguez-Campello A, Cuadrado-Godia E, Sayols-Baixeras S, Elosua R, Roquer J, GENESTROKE Consortium (2016) Epigenome-wide association study identifies TXNIP gene associated with type 2 diabetes mellitus and sustained hyperglycemia. Hum Mol Genet 25(3):609–619. https://doi.org/10.1093/hmg/ddv493
8. Florath I, Butterbach K, Heiss J, Bewerunge-Hudler M, Zhang Y, Schottker B, Brenner H (2016) Type 2 diabetes and leucocyte DNA methylation: an epigenome-wide association study in over 1,500 older adults. Diabetologia 59(1):130–138. https://doi.org/10.1007/s00125-015-3773-7
9. Juvinao-Quintero DL, Marioni RE, Ochoa-Rosales C, Russ TC, Deary IJ, van Meurs JBJ, Voortman T, Hivert MF, Sharp GC, Relton CL, Elliott HR (2021) DNA methylation of blood cells is associated with prevalent type 2 diabetes in a meta-analysis of four European cohorts. Clin Epigenetics 13(1):40. https://doi.org/10.1186/s13148-021-01027-3
10. Kupers LK, Monnereau C, Sharp GC, Yousefi P, Salas LA, Ghantous A, Page CM, Reese SE, Wilcox AJ, Czamara D, Starling AP, Novoloaca A, Lent S, Roy R, Hoyo C, Breton CV, Allard C, Just AC, Bakulski KM, Holloway JW, Everson TM, Xu CJ, Huang RC, van der Plaat DA, Wielscher M, Merid SK, Ullemar V, Rezwan FI, Lahti J, van Dongen J, Langie SAS, Richardson TG, Magnus MC, Nohr EA, Xu Z, Duijts L, Zhao S, Zhang W, Plusquin M, DeMeo DL, Solomon O, Heimovaara JH, Jima DD, Gao L, Bustamante M, Perron P, Wright RO, Hertz-Picciotto I, Zhang H, Karagas MR, Gehring U, Marsit CJ, Beilin LJ,

Vonk JM, Jarvelin MR, Bergstrom A, Ortqvist AK, Ewart S, Villa PM, Moore SE, Willemsen G, Standaert ARL, Haberg SE, Sorensen TIA, Taylor JA, Raikkonen K, Yang IV, Kechris K, Nawrot TS, Silver MJ, Gong YY, Richiardi L, Kogevinas M, Litonjua AA, Eskenazi B, Huen K, Mbarek H, Maguire RL, Dwyer T, Vrijheid M, Bouchard L, Baccarelli AA, Croen LA, Karmaus W, Anderson D, de Vries M, Sebert S, Kere J, Karlsson R, Arshad SH, Hamalainen E, Routledge MN, Boomsma DI, Feinberg AP, Newschaffer CJ, Govarts E, Moisse M, Fallin MD, Melen E, Prentice AM, Kajantie E, Almqvist C, Oken E, Dabelea D, Boezen HM, Melton PE, Wright RJ, Koppelman GH, Trevisi L, Hivert MF, Sunyer J, Munthe-Kaas MC, Murphy SK, Corpeleijn E, Wiemels J, Holland N, Herceg Z, Binder EB, Davey Smith G, Jaddoe VWV, Lie RT, Nystad W, London SJ, Lawlor DA, Relton CL, Snieder H, Felix JF (2019) Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. Nat Commun 10(1):1893. https://doi.org/10.1038/s41467-019-09671-3

11. Garagnani P, Bacalini MG, Pirazzini C, Gori D, Giuliani C, Mari D, Di Blasio AM, Gentilini D, Vitale G, Collino S, Rezzi S, Castellani G, Capri M, Salvioli S, Franceschi C (2012) Methylation of ELOVL2 gene as a new epigenetic marker of age. Aging Cell 11(6):1132–1134. https://doi.org/10.1111/acel.12005
12. Richard MA, Huan T, Ligthart S, Gondalia R, Jhun MA, Brody JA, Irvin MR, Marioni R, Shen J, Tsai PC, Montasser ME, Jia Y, Syme C, Salfati EL, Boerwinkle E, Guan W, Mosley TH Jr, Bressler J, Morrison AC, Liu C, Mendelson MM, Uitterlinden AG, van Meurs JB, BIOS Consortium, Franco OH, Zhang G, Li Y, Stewart JD, Bis JC, Psaty BM, Chen YI, SLR K, Zhao W, Turner ST, Absher D, Aslibekyan S, Starr JM, AF MR, Hou L, Just AC, Schwartz JD, Vokonas PS, Menni C, Spector TD, Shuldiner A, Damcott CM, Rotter JI, Palmas W, Liu Y, Paus T, Horvath S, O'Connell JR, Guo X, Pausova Z, Assimes TL, Sotoodehnia N, Smith JA, Arnett DK, Deary IJ, Baccarelli AA, Bell JT, Whitsel E, Dehghan A, Levy D, Fornage M (2017) DNA methylation analysis identifies loci for blood pressure regulation. Am J Hum Genet 101(6):888–902. https://doi.org/10.1016/j.ajhg.2017.09.028
13. Li Y, Pan X, Roberts ML, Liu P, Kotchen TA, Cowley AW Jr, Mattson DL, Liu Y, Liang M, Kidambi S (2018) Stability of global methylation profiles of whole blood and extracted DNA under different storage durations and conditions. Epigenomics 10(6):797–811. https://doi.org/10.2217/epi-2018-0025
14. Moran S, Vizoso M, Martinez-Cardus A, Gomez A, Matias-Guiu X, Chiavenna SM, Fernandez AG, Esteller M (2014) Validation of DNA methylation profiling in formalin-fixed paraffin-embedded samples using the Infinium HumanMethylation450 Microarray. Epigenetics 9(6):829–833. https://doi.org/10.4161/epi.28790
15. Huang YT, Chu S, Loucks EB, Lin CL, Eaton CB, Buka SL, Kelsey KT (2016) Epigenome-wide profiling of DNA methylation in paired samples of adipose tissue and blood. Epigenetics 11(3):227–236. https://doi.org/10.1080/15592294.2016.1146853
16. Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H, Yamamoto H, Hidalgo M, Tan AC, Galm O, Ferrer I, Sanchez-Cespedes M, Villanueva A, Carmona J, Sanchez-Mut JV, Berdasco M, Moreno V, Capella G, Monk D, Ballestar E, Ropero S, Martinez R, Sanchez-Carbayo M, Prosper F, Agirre X, Fraga MF, Grana O, Perez-Jurado L, Mora J, Puig S, Prat J, Badimon L, Puca AA, Meltzer SJ, Lengauer T, Bridgewater J, Bock C, Esteller M (2012) A DNA methylation fingerprint of 1628 human samples. Genome Res 22(2):407–419. https://doi.org/10.1101/gr.119867.110
17. Koestler DC, Christensen B, Karagas MR, Marsit CJ, Langevin SM, Kelsey KT, Wiencke JK, Houseman EA (2013) Blood-based profiles of DNA methylation predict the underlying distribution of cell types: a validation analysis. Epigenetics 8(8):816–826. https://doi.org/10.4161/epi.25430
18. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics 13:86. https://doi.org/10.1186/1471-2105-13-86
19. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome Biol 14(10):R115. https://doi.org/10.1186/gb-2013-14-10-r115
20. Tsai PC, Bell JT (2015) Power and sample size estimation for epigenome-wide association scans to detect differential DNA methylation. Int J Epidemiol 44(4):1429–1441. https://doi.org/10.1093/ije/dyv041
21. Graw S, Henn R, Thompson JA, Koestler DC (2019) pwrEWAS: a user-friendly tool for comprehensive power estimation for epigenome wide association studies (EWAS). BMC

Bioinformatics 20(1):218. https://doi.org/10.1186/s12859-019-2804-7

22. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM (2010) Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 11:587. https://doi.org/10.1186/1471-2105-11-587
23. You C, Wu S, Zheng SC, Zhu T, Jing H, Flagg K, Wang G, Jin L, Wang S, Teschendorff AE (2020) A cell-type deconvolution meta-analysis of whole blood EWAS reveals lineage-specific smoking-associated DNA methylation changes. Nat Commun 11(1):4779. https://doi.org/10.1038/s41467-020-18618-y
24. Sayols-Baixeras S, Subirana I, Fernandez-Sanles A, Senti M, Lluis-Ganella C, Marrugat J, Elosua R (2017) DNA methylation and obesity traits: an epigenome-wide association study. The REGICOR study. Epigenetics 12(10):909–916. https://doi.org/10.1080/15592294.2017.1363951
25. McCartney DL, Zhang F, Hillary RF, Zhang Q, Stevenson AJ, Walker RM, Bermingham ML, Boutin T, Morris SW, Campbell A, Murray AD, Whalley HC, Porteous DJ, Hayward C, Evans KL, Chandra T, Deary IJ, McIntosh AM, Yang J, Visscher PM, McRae AF, Marioni RE (2019) An epigenome-wide association study of sex-specific chronological ageing. Genome Med 12(1):1. https://doi.org/10.1186/s13073-019-0693-z
26. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8(2):203–209. https://doi.org/10.4161/epi.23470
27. Jones GT, Marsman J, Bhat B, Phillips VL, Chatterjee A, Rodger EJ, Williams MJA, van Rij AM, McCormick SPA (2020) DNA methylation profiling identifies a high effect genetic variant for lipoprotein(a) levels. Epigenetics 15(9):949–958. https://doi.org/10.1080/15592294.2020.1739797

Chapter 3

Next-Generation Bisulfite Sequencing for Targeted DNA Methylation Analysis

Jim Smith, Robert C. Day, and Robert J. Weeks

Abstract

Bisulfite sequencing is the "gold-standard" technique for DNA methylation analysis. By combining bisulfite sequencing with high-throughput, next-generation sequencing technology, we can document methylation from many thousands of individual reads (equivalent to alleles or "cells"), for multiple target regions and from many samples simultaneously. Here, we describe a next-generation bisulfite-sequencing assay for targeted DNA methylation analysis which offers scope for the simultaneous interrogation of multiple genomic loci across numerous samples.

Key words DNA methylation, Bisulfite conversion, Next-generation sequencing, Analysis

1 Introduction

DNA methylation is an important epigenetic regulator of gene expression and genome stability and is essential for many processes, including embryogenesis, cellular differentiation, and cellular identity [1]. Indeed, improper control of DNA methylation has been implicated in many diseases, including cancer [2]. In mammals, DNA methylation occurs predominantly at cytosine residues of CpG dinucleotides. Of relevance to gene expression, it has been known since the pioneering work of Bird and others, that many gene promoters contain CpG-dense regions, termed CpG islands, and that these are frequently methylated in transcriptionally silenced genes [3–5]. In the context of malignancy, pathognomonic patterns of aberrant DNA methylation have since been widely characterized. Typically, genome-wide hypomethylation, paired with promoter-specific hypermethylation of key tumor-suppressor genes, is associated with the genomic instability and altered expression of crucial genes which drive oncogenesis [2]. Our understanding of DNA methylation profiles in tumorigenesis is central to deciphering the underlying pathobiology and informing treatment

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_3,

avenues for respective malignancies. Hence, accurate analysis of local methylation patterns is an essential component in the investigation of mammalian diseases such as cancer.

Many methods and protocols have been developed to detect and quantify methylation of DNA [6]. Some, for example restriction digestion with methylation-sensitive restriction enzymes, can interrogate only one or a few CpG sites at a time [7]. However, others such as Sequenom™ Epityper [8] are able to simultaneously interrogate multiple sites. We, along with many other groups, have combined one such common technique, bisulfite sequencing [9], with high-throughput, next-generation sequencing technology.

Bisulfite sequencing combines bisulfite treatment of DNA with PCR amplification of specific regions and DNA sequencing. Bisulfite conversion of DNA will convert unmethylated cytosines to uracil residues, while methylated cytosines (5′methylcytosine) are resistant to conversion. After PCR amplification and sequencing, unmethylated cytosines will appear as thymidine and methylated cytosines as cytosine. Thus, bisulfite sequencing can generate nucleotide-level methylation data for an entire region of interest, that is, for many CpG sites.

The Illumina MiSeq system is a next-generation sequencing platform which allows for high-throughput analysis of multiple genes within a single run. The emergence of such massive parallel sequencing platforms has led to significant advancements in the depth of coverage and overall sequencing efficiency achievable in the modern era [10–12]. The MiSeq platform, in conjunction with bisulfite conversion of genomic DNA, allows for efficient and detailed analysis of the DNA methylome.

Targeted sequencing on the MiSeq platform requires the amplification of the specific genomic region of interest using target-specific primers that have been "tagged" with Illumina-specific sequences. These Illumina sequences are required for capturing sequencing libraries onto the MiSeq flow cell. Each selected amplicon must be limited to under 500 bp in length, as the MiSeq platform paired-end sequencing can generate two reads of up to 300 bp, one from each of the forward and reverse ends of the amplicon. Paired reads can be joined bioinformatically post-sequencing to cover the entire amplimer [13]. Additionally, optimal bisulfite-converted amplicon length is between 250 and 400 bp, as larger amplimers may be difficult to amplify due to the fragmentation of DNA during bisulfite treatment. As bisulfite sequencing is performed here, primer sequences flanking the target region are also designed to avoid overlap with CpG dinucleotides [14].

Furthermore, sequencing technology allows for the simultaneous interrogation of multiple different genomic locations (e.g., tumor suppressor gene promoters) across many samples. After first round PCR amplification with bisulfite-specific primers for each

target region, amplicons for each sample can be combined into a "sub-library" and amplified with Illumina index primers. These second-round primers add unique index sequences to each "sub-library" product, as well as sequences necessary for binding to the sequencing flow cell and sequencing primer. Using this two-step PCR amplification protocol, we are able to bisulfite sequence multiple promoter targets for multiple samples at great depth, from each MiSeq run. For example, in a recent study of 6 gene promoters, amplified from 62 tumor samples (PPFE sections), we obtained 321,879 reads at an average read-depth of 865 reads per gene per tumor.

In this chapter, we document one such bisulfite sequencing assay developed to interrogate methylation in the promoter region of the *TES* gene. This protocol can be easily modified to target any locus of interest within the limits of successful primer design.

2 Materials

Prepare and store all materials at room temperature unless otherwise stated.

2.1 Bisulfite-Specific Primer Design

1. DNA sequence for target region.
2. Access to the MethPrimer web tool [15].

2.2 Genomic DNA Purification and Bisulfite Conversion

1. QIAamp DNA Blood Mini Kit (QIAGEN).
2. EZ Methylation-Gold or EZ Methylation-Direct bisulfite conversion kit (Zymo Research).
3. Nanophotometer or similar UV spectrophotometer.

2.3 Bisulfite-Specific PCR Amplification.

1. 2× KAPA HiFi HotStart Uracil+ ReadyMix.
2. Bisulfite-specific primers.
3. Illumina index primers (Illumina TruSeqHT, i5 and i7 primers).
4. Nuclease-free H_2O.
5. Thermal cycler.

2.4 Agarose Gel Electrophoresis

1. Molecular grade agarose.
2. 1 kb Plus DNA Plus Ladder.
3. 0.5× TAE with ethidium bromide (10 μg/mL).
4. Xylene cyanol loading dye: 15% Ficoll/1 mM EDTA (pH 8.0) (w/v), 0.025% xylene cyanol.
5. Agarose gel electrophoresis system.
6. UV gel doc system.

2.5 PCR Clean-Up

1. Agencourt AMPure XP magnetic beads.
2. Magnetic plate.
3. Nuclease-free H_2O.
4. TE buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0.
5. LoBind microcentrifuge tubes (Eppendorf®).

2.6 PCR Product Quantitation

1. Qubit dsDNA HS assay.
2. Qubit assay tubes.
3. Qubit Fluorometer.

2.7 Sequencing Library Quality Assessment

1. High-Sensitivity DNA chips (Agilent).
2. BioAnalyzer 2100 (Agilent).

2.8 High-Throughput Sequencing

1. MiSeq V2 Nano reagents kit (Illumina).
2. PhiX Control V3 library (Illumina).
3. Illumina MiSeq next-generation sequencer.

2.9 Downstream Analysis of High-Throughput Sequencing Data

1. Computer capable of running UNIX or Linux programs (e.g., Apple Mac OS computer).

3 Methods

This protocol provides details for analysis of DNA methylation using targeted, high-throughput bisulfite sequencing. As this protocol involves multiple PCR amplification steps, extreme care needs to be taken to ensure that PCR contamination does not contribute to sample reads. Prepare multiple "clean" aliquots of bisulfite-specific primers, nuclease-free H_2O, Illumina index primers and 2× KAPA Uracil+ mix and make use of "no-template controls" to monitor contamination. To minimize the risk of contamination, use "clean" laboratory practices including frequent changes of disposable gloves. PCRs should be set-up in PCR hoods, which have been UV-irradiated, using sterile and UV-irradiated nuclease-free low-binding plasticware, aerosol barrier pipette tips and fresh aliquots of 2× KAPA HiFi HotStart Uracil+ ReadyMix and primers. To illustrate each step, we will use the CpG island located in the region upstream of the transcriptional start site of the *TES* gene promoter.

3.1 Bisulfite-Specific PCR Primer Design

1. Firstly, obtain the DNA sequence for your region of interest. We recommend using the UCSC Genome Browser (http://genome.ucsc.edu/) or the Ensemble web site (https://grch37.ensembl.org/index.html) for the latest genome builds.

2. Secondly, bisulfite-specific primers are designed to amplify the target region using the MethPrimer online tool [15] (*see* **Note 1**). For our example, we selected and copied 2 kb of genomic sequence (UCSC Genome Browser; hg38; chr7:116209548–116211547) of the *TES* promoter region immediately upstream of the transcriptional start site into the MethPrimer design window (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi). Parameters, such as defining regions of interest, can be modified as necessary. For the *TES* promoter, the "Product Size" parameters were modified (*Min Size: 300; Opt: 350; Max: 400*) (Fig. 1) (*see* **Note 2**). After submission, bisulfite-specific primer sequences specific for the *TES* promoter region are displayed (Table 1) (*see* **Note 3**).
3. Once bisulfite-specific primers have been designed, Illumina-specific "tag" sequences are appended to the 5′ ends of the forward and reverse primers, respectively, before oligonucleotide synthesis (as shown in Table 2) (*see* **Note 4**).

3.2 DNA Purification and Bisulfite Conversion

1. Purify genomic DNA from tissue or cell (>1 × 10^6 cells) samples using the QIAamp DNA blood mini kit, as per manufacturer's instructions, and quantify using a Nanophotometer.
2. Bisulfite convert up to 500 ng (or maximum of 20 μL) of purified DNA using the EZ DNA Methylation-Gold kit, according to manufacturer's instructions (*see* **Note 5**).
3. Bisulfite-treated DNA is eluted with 10 μL of Elution Buffer (*see* **Note 6**).

3.3 First Round Touchdown PCR

1. Prepare a master mix containing the following components, volumes shown are per sample/reaction (*see* **Note 7**):
   (a) 2.8 μL nuclease-free H_2O.
   (b) 5.0 μL of 2× KAPA Uracil+ PCR mix.
   (c) 0.1 μL of [10 μM] Forward primer.
   (d) 0.1 μL of [10 μM] Reverse primer.
   (e) 2.0 μL of bisulfite-converted DNA.
2. Amplify using the following "Touchdown" cycle conditions in a thermocycler (*see* **Note 8**):
   (a) 95 °C, 4 min.
   (b) 94 °C, 30 s.
   (c) 60 °C, for 30 s.
   (d) 72 °C, for 30 s.
   (e) Repeat **steps b-d** 19×, reducing "**step c**" temperature by 0.5 °C per cycle.
   (f) 94 °C, 30 s.

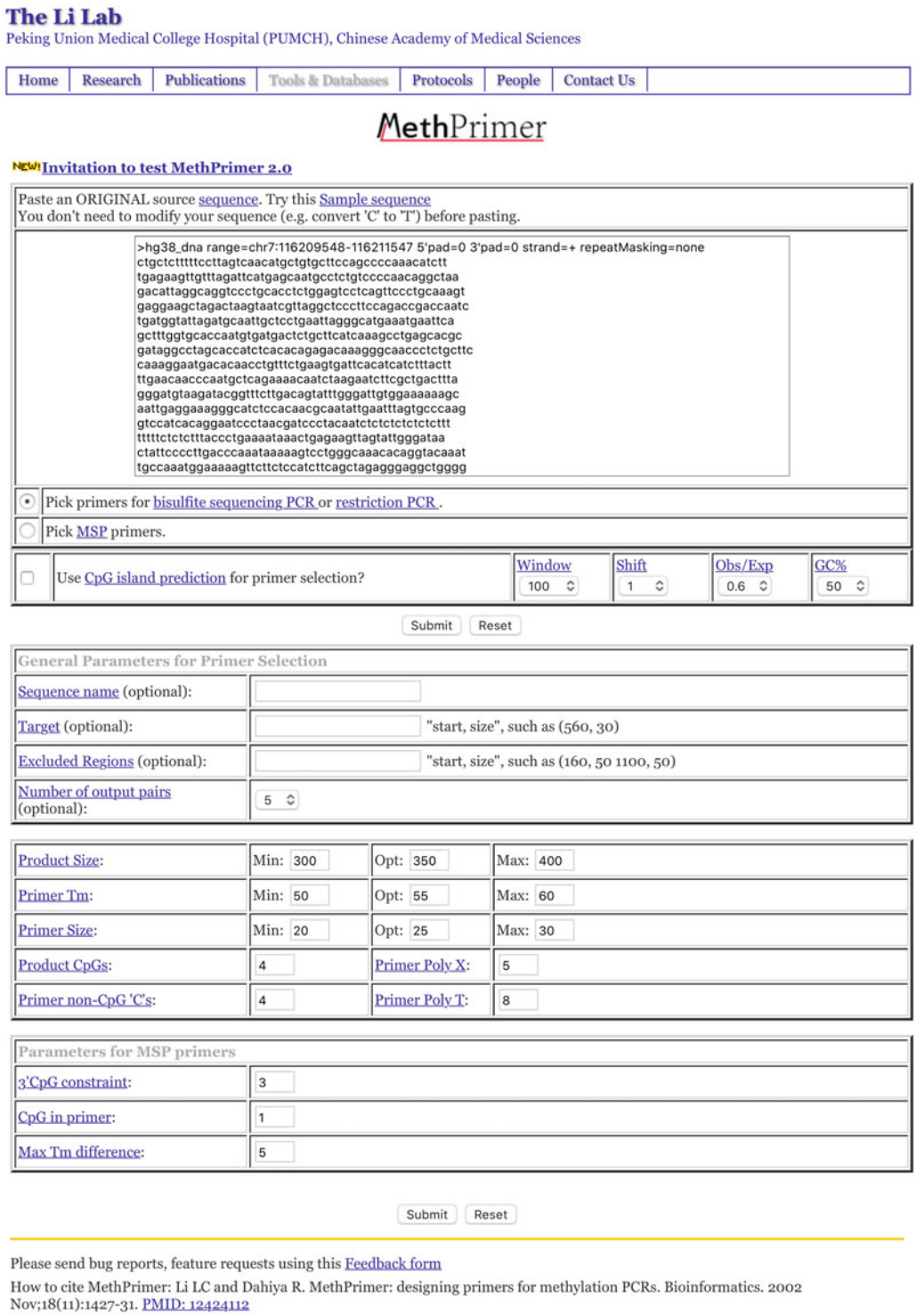


Fig. 1 Image showing the MethPrimer interface with the *TES* promoter sequence and General Parameters and options

Table 1
MethPrimer output showing primer details

Primer	Start	Size	Tm	GC%	C's	Sequence
Left primer	809	27	57.70	62.96	8	TTAGGGTTATTGAGTTTGTTTAGTAGG
Right primer	1179	25	54.76	40.00	4	CTTTATTTTCCAAATCCATATTAAC

Product size: 371 Tm: 67.3 CpGs in product: 48

Table 2
Table showing the MethPrimer primer sequences and Illumina-specific tags (in bold) and the increased product size

MiSeq primers	Sequence	Size	Product size
TES_MiSeq_For	**ACGACGCTCTTCCGATCT** TTAGGGTTATTGAGTTTGTTTAGTAGG	45	407
TES_MiSeq_Rev	**CGTGTGCTCTTCCGATCT** CTTTATTTTCCAAATCCATATTAAC	43	

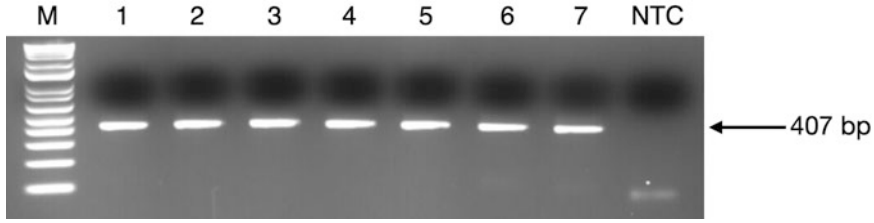


Fig. 2 Example gel showing successful amplification of *TES* from first round PCR. 2% agarose gel (containing ethidium bromide) electrophoresis image showing successful amplification of *TES* promoter region from seven human bisulfite-treated DNA samples and one "no template control" (NTC). (M – 1 kb DNA Plus Ladder)

(g) 50 °C, 30 s.
(h) 72 °C, 30 s.
(i) Repeat **steps f** and **g**, 30×.
(j) 72 °C, 5 min.
(k) Cool to 4 °C.

3. Following PCR amplification, remove 3.0 μL from each sample PCR tube and add to 3.0 μL of xylene cyanol loading dye in a clean tube. Carefully load the sample/xylene cyanol mix into wells alongside 1 kb DNA ladder and electrophorese through 2% (w/v) agarose gel (0.5× TAE containing ethidium bromide; 100 V, 30 min) to confirm that a single product of the expected size is present for each sample. For our example, one single band of 407 bp corresponding to the *TES* product is shown (*see* Fig. 2).

3.4 Bead Clean-Up of PCR Products (See Note 9)

1. For multiple targets, combine the PCR products for each sample together (*see* **Note 10**). If necessary add nuclease-free H_2O to each sample for a minimum volume of 40 μL.
2. Add an equal volume of room temperature AMPure XP beads and mix by pipetting. Incubate for a minimum of 5 min at room temperature.
3. Collect beads on the side of the tube using a neodymium magnet plate.
4. Remove supernatant and dispose, taking care not to disturb the pellet.
5. Wash beads with 100 μL of freshly prepared 80% ethanol. Collect beads using the magnet as before, remove ethanol and repeat.
6. Air-dry bead pellet at room temperature until "cracking" of bead pellet is evident.
7. Resuspend bead pellet in 20 μL of TE. Incubate for a minimum of 5 min at room temperature.
8. Return to the magnet and collect supernatant containing purified DNA into a clean LoBind tube (*see* **Note 11**).

3.5 Qubit Concentration and Dilution

1. Quantify bead-purified first round PCR samples using a Qubit Fluorometer and 1× High-Sensitivity dsDNA assay reagents, according to the manufacturer's instructions.
2. Dilute each bead-purified PCR sample to 1 ng/μL with TE (pH 8.0).

3.6 Second Round PCR

The purpose of the second round PCR amplification stage, is to "end-label" purified PCR or pooled PCR sample with a unique combination of forward and reverse index primers [16] (*see* **Note 12**).

1. For each sample, set up the second round PCR, as below:
   (a) 2.0 μL of purified first round PCR product (1 ng/μL).
   (b) 0.25 μL [10 μM] i7 index primer.
   (c) 0.25 μL [10 μM] i5 index primer.
   (d) 5.0 μL 2× KAPA HiFi Uracil+ mix.
2. Amplify samples with the following program:
   (a) 95 °C; 4 min.
   (b) 95 °C; 30 s.
   (c) 55 °C; 30 s.
   (d) 72 °C; 30 s.
   (e) Repeat **steps b-d**, 9×.
   (f) Cool to 4 °C.

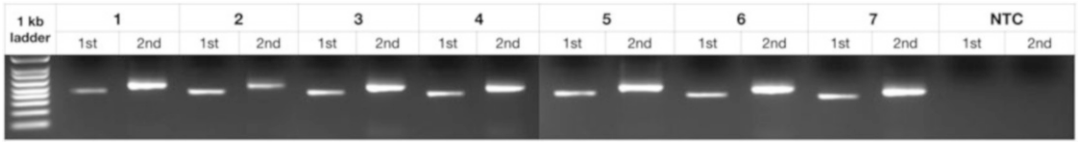


Fig. 3 Second round PCR amplification of purified first round *TES* PCR products. Composite gel electrophoresis image of first and second round PCR products, showing increased sized of second round products

3. Following PCR amplification, remove 2.0 μL from each sample PCR tube and add to 3.0 μL of xylene cyanol loading dye in a clean tube. Carefully load sample/xylene cyanol mix into wells alongside an aliquot of purified first round product and 1 kb Plus DNA ladder, electrophorese through 2% agarose gel (0.5× TAE containing ethidium bromide; 100 V, 30 min) to confirm that the expected size product is present for each sample. One single band of approximately 100 bp larger than the first round PCR product should be observed (*see* Fig. 3) (*see* **Note 13**).

3.7 Library Preparation and Bead Clean-Up

1. After gel visualization of second round products, combine approximately equal amounts of PCR samples together (as determined by intensity of ethidium bromide staining) into a "library" tube (*see* **Note 14**).
2. Purify the combined library, containing second round PCR product from all samples, using a 1:1 (vol:vol) ratio with AMPure XP magnetic beads (as per Subheading 3.4) and quantify by Qubit fluorimeter (*see* Subheading 3.5).

3.8 Library Quality Check

1. Dilute the bead-purified library to 1 ng/μL with nuclease-free water.
2. Analyse 1 μL of the diluted library using a DNA High Sensitivity DNA chip and an Agilent BioAnalyzer 2100 system, as per manufacturer's instructions (Fig. 4).
3. Dilute the combined library to 4 nM. This calculation is based on the average size of the combined amplicon library trace taken from the BioAnalyzer and the quantification by Qubit (*see* **Note 15**).

3.9 MiSeq Sequencing

1. The MiSeq V2 reagent cartridge and the tube of HS buffer should be taken out from the −20 °C freezer at least 1 h prior to diluting and denaturing the libraries. The HS buffer can stand at room temperature on your bench. Place the cartridge into a container containing tap water such that the bottom section of the cartridge is submerged to the level marked on the white plastic bottom section of the cartridge (*see* **Note 16**).

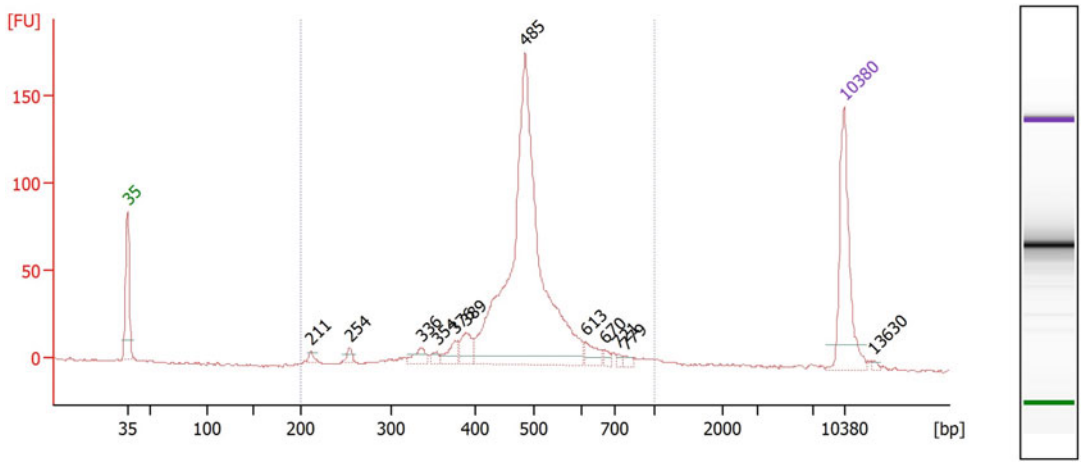


Fig. 4 BioAnalyzer 2100 electropherogram of the bead-purified *TES* library, showing one peak of 485 bp length

2. Add 4.0 μL of nuclease-free water and 1.0 μL of 1N NaOH to a 1.5 mL LoBind Eppendorf tube (total = 5 μL of 0.2N NaOH). To this, add 5 μL of the 4 nM combined library and mix by pipetting. Incubate at room temperature for 5 min to denature the library, before adding 990 μL of HS buffer (total = 1 mL of 20 pM denatured library). This can be stored in a −20 °C freezer.
3. **Step 2** is repeated for the PhiX Control V3 library (*see* **Note 17**).
4. Combine 30 μL of the denatured PhiX library, 270 μL of the denatured library and 300 μL of HS buffer in a 1.5 mL LoBind Eppendorf tube to make a final volume of 600 μL, and pulse on a vortex mixer. The final preparation is now at a concentration of 8 pM (*see* **Note 18**).
5. Remove the MiSeq cartridge from the water bath and dry well. Puncture the foil, covering the position marked "Sample" using a sterile 1 mL pipette tip. Using a new 1 mL pipette tip, transfer 600 μL of denatured library/PhiX mix into position 18.
6. Follow the manufacturer's instructions and visual guides to load the cartridge and flow cell and begin the run (*see* **Note 19**).

3.10 Sequencing Analysis

After paired-end sequencing, the MiSeq sequencer will demultiplex the samples according to the Illumina index adaptors used and will output two "fastq.gz" sequence files for each sample. These two files have R1 (i.e., forward reads) and R2 (i.e., reverse reads) appended to their sample names. To limit the size of these files, they are output in compressed .gz format. Processing and analysis

of the sequence files can be performed with basic bioinformatic knowledge and we have included a brief summary of our workflow as an example (*see* **Note 20**).

1. Make directory with sample name:

```
$ mkdir <directory>
```

2. Copy both R1 and R2 files into directory:

```
 $ cp -f <filenameR1> <directory>
$ cp -f <filenameR2> <directory>
```

3. Use the PEAR end-joiner program to join R1 and R2 sequences together (*see* **Note 21**) [13]:

```
$ pear -f <filenameR1> -r <filenameR2> -o <filename>
```

4. Remove low quality reads (default: Phred score < 30) from the "filename.assembled.fastq" file with the *trim_galore* program (https://github.com/FelixKrueger/TrimGalore):

```
$ mkdir <Trimmed>
$ trim_galore <filename.assembled.fastq> --output_dir <./
Trimmed>
```

5. For multiple amplimers, it is necessary to separate reads from each "filename.assembled.fq" file into "filename.amplimer_-name.fasta" files. This can be achieved using the *fastq_to_fasta* and the *fastx_barcode_splitter* programs (available as components of the *fastx_toolkit* (http://hannonlab.cshl.edu/fastx_toolkit/download.html)). These programs will compare each sequence read in the "filename.assembled.fq" file to the amplimer primer sequences and will copy each read into its correct "amplimer.fasta" file (*see* **Note 22**):

```
$ fastq_to_fasta -i <filename.assembled.fq | fastx_barcode_s-
plitter.pl --bcfile <amplimers.txt> --bol ./Trimmed/${file-
name.amplimer_name.fasta}
```

6. To determine methylation status and to generate methylation scores, each "filename.amplimer_name.fasta" file has to be aligned to its respective amplimer sequence, for which we use

the BiQ_Analyzer_HT program [17]. This JAVA program can be run in a GUI window, which we recommend for novice users, or in command line UNIX for more experienced users. Simply, fasta sequencing files are loaded into the interface along with the corresponding unconverted, genomic region and analysis is performed (*see* **Note 23**). After alignment, BiQ_Analyzer_HT will generate "heatmap" and "pearl_necklace" images and a Results Table for all amplimers and samples. The Results Table can be imported into MS Excel or R to calculate mean methylation across the whole amplimer or to calculate mean methylation for each CpG site.

4 Notes

1. MethPrimer website: http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi [15].
2. For two reasons, it is important that the proposed amplimer is less than 500 bp: firstly, bisulfite treatments involve harsh conditions and lead to fragmentation of DNA limiting the efficient amplification of larger amplimers; and secondly, the maximum sequencing read length obtainable from the MiSeq Nano v2 sequencing kit is 500 nucleotides or 250 bp paired-ends.
3. As bisulfite-specific PCR primers are designed to not contain CpG sites, it is not uncommon for MethPrimer to initially fail to design primers, particularly for highly CpG-rich regions. Careful adjustment of the design parameters may allow primers to be designed by the tool, however, if MethPrimer is still unable to design primers, then we have had success with designing primers manually.
4. These Illumina tag sequences are required during the second round PCR for MiSeq library preparation. The Illumina specific tag sequences increase the expected first round PCR product by 36 bp.
5. For smaller samples or fewer cells, we recommend bisulfite conversion using the EZ DNA Methylation-Direct kit (Zymo Research, cat.no. D5020).
6. Eluted, bisulfite-converted DNA (10 μL of elution buffer) can be stored for up to 1 month at −20 °C.
7. Volumes shown are for one reaction only. As per normal PCR setup, prepare enough master mix for bisulfite DNA samples and for a "no-template control" and aliquot 8.0 μL of master mix into PCR tubes, before adding 2.0 μL of each bisulfite-treated DNA sample (or water for the "no-template control").
8. Using this Touchdown PCR protocol, we have had good success with amplification of multiple, specific PCR products, without prior primer-annealing temperature optimizations.

9. Bead purifications are performed to remove PCR primers and salts. Routinely to remove short primer sequences from longer PCR products, we perform bead purifications using PCR reaction to bead ratio of 1:1 (vol:vol).
10. For this example, we have one target amplification (*TES* promoter) per sample and each sample PCR product is bead-purified separately. For multiple targets amplified per sample, the PCR products should be combined into their sample tube prior to bead purification, for example, for four amplimers and eight samples: after first round PCR amplification, the four amplimers per sample should be combined into their sample tubes (eight tubes) before bead purification.
11. Care should be taken to minimize carryover of beads into the sample tube, however the presence of beads in the sample is not detrimental to downstream steps.
12. Illumina index primers (TruSeq HT equivalent) were purchased from IDT (Table 3).
13. One single PCR product larger than the first-round product is required to proceed. The presence of "un-end-labeled" first round product after second round PCR will result in overestimation of the final library concentration and underloading of the sequencing flow cell and a reduction in output.
14. Bead purification and quantitation of individual samples prior to mixing is time-consuming and unnecessary. We achieve good read numbers after combining individual samples based on relative band intensities from the gel electrophoresis and bead purification of this combined library mix.
15. The concentration of the combined library can be calculated with the following:

$$\text{concentration [nM]} = \text{Qubit conc (ng/}\mu\text{L)}/(\text{average length (bp)} \times 0.00065)$$

Table 3
Index primers used for amplification of PCR products. Oligonucleotide sequences shown in 5′ to 3′ direction. Underlined X's denote six base-pair index positions. Bold denotes bases that overlap step one PCR products

Primer name	Primer sequence
i5_Fwd	AATGATACGGCGACCACCGAGATCTACAC<u>XXXXXX</u>ACACTCTTTCCCTAC **ACGACGCTCTTCCGATCT**
i7_Rev	CAAGCAGAAGACGGCATACGAGAT<u>XXXXXX</u>GTGACTGGAGTTCAGA **CGTGTGCTCTTCCGATC**

For our typical combined libraries, with an average fragment length of 400 bp, the 4 nM combined library preparation equates to a concentration of 1.04 ng/μL. The concentration of the diluted combined library should be confirmed by Qubit measurement.

16. For early starts, we recommend placing the reagent cartridge at 4 °C overnight to thaw.
17. Bisulfite sequencing libraries have a relatively low sequence complexity, due to the conversion of unmethylated bases during bisulfite treatment. Illumina sequencers rely on a good representation of all four bases during initialization and so adding a high complexity control library is highly recommended. For low complexity libraries Illumina recommends at least 5% of the molecules in the mix are PhiX, but we prefer 10% PhiX in our final combined library.
18. While metrics will vary by individual MiSeq machine, loading 8 pM final library/PhiX mix on to our local machine results in approximately 700–800 cluster density, 7–10% aligned reads to PhiX and 800,000–1,000,000 reads pass filter.
19. Most of this process is shown to the user by the MiSeq software. The sequencing setup software includes visual guides to identify the locations for reagent loading.
20. We have included this section to illustrate our workflow for analyzing bisulfite sequencing results, but many tools are available. For novice users, we would recommend gaining experience with processing of bisulfite sequencing data through using the online GALAXY interface (https://usegalaxy.org/) [18]. For more experienced users, we suggest that they develop a UNIX BASH terminal script to enable reliable and efficient processing and analysis of their data locally.
21. The PEAR [13] program is able to join overlapping, paired-end reads. PEAR will output four files with added suffixes ".assembled.fastq", ".discarded.fastq", "unassembled.forward.fastq", and "unassembled.reverse.fastq".
22. The *fastx_barcode_splitter* program systematically compares the beginning or end of each sequence read with the expected amplimer sequence. To do this, you must prepare a text file containing the amplimer name and the first 10 bases of their primer sequences, either forward or reverse, in the following tabbed format:

```
Amplimer_name1<TAB>xxxxxxxxxx
Amplimer_name2<TAB>xxxxxxxxxx
```

This tabbed text file can be named "amplimers.txt" and specified with the "--bcfile" option and the "--bol" option or the "--eol" option can be used to specify forward or reverse primer sequences, respectively. Output files will be called "filename.Amplimer1.fasta", "filename.Amplimer2.fasta", and so on.

23. Initially, the "Results Table" should be examined, as reads with low "Conversion rate" and "Sequence identity" should be removed from the analysis. We suggest removing reads with "Conversion rate" <0.95 and "Sequence identity" <0.95. This can be achieved in the "Settings" tab of the GUI window or with the "-minconv 0.95 -minsi 0.95" option in the command line.

Acknowledgments

We would like to acknowledge Professor Ian Morison, as the "driving force" behind the development of this high-throughput, next-generation bisulfite sequencing protocol and analysis workflow. While this protocol is the work of many, contributions from Jackie Ludgate, Luke Bridgman, Dr. Suzan Almomani, and Dr. Issam Mayyas were instrumental in developing the bisulfite sequencing protocol and the analysis workflow.

References

1. Smith J, Sen S, Weeks RJ, Eccles MR, Chatterjee A (2020) Promoter DNA hypermethylation and paradoxical gene activation. Trends Cancer 6(5):392–406. https://doi.org/10.1016/j.trecan.2020.02.007
2. Kulis M, Esteller M (2010) DNA methylation and cancer. Adv Genet 70:27–56. https://doi.org/10.1016/B978-0-12-380866-0.60002-2
3. Bird A, Tate P, Nan X, Campoy J, Meehan R, Cross S, Tweedie S, Charlton J, Macleod D (1995) Studies of DNA methylation in animals. J Cell Sci Suppl 19:37–39. https://doi.org/10.1242/jcs.1995.supplement_19.5
4. 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
5. Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet 13(7):484–492. https://doi.org/10.1038/nrg3230
6. Beck S, Rakyan VK (2008) The methylome: approaches for global DNA methylation profiling. Trends Genet 24(5):231–237. https://doi.org/10.1016/j.tig.2008.01.006
7. Weeks RJ, Morison IM (2006) Detailed methylation analysis of CpG islands on human chromosome region 9p21. Genes Chromosomes Cancer 45(4):357–364. https://doi.org/10.1002/gcc.20297
8. Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ (2007) Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res 35(18):e119. https://doi.org/10.1093/nar/gkm662
9. Grunau C, Clark SJ, Rosenthal A (2001) Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. Nucleic Acids Res 29(13):E65. https://doi.org/10.1093/nar/29.13.e65
10. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM,

Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6(8):1621–1624. https://doi.org/10.1038/ismej.2012.8

11. King JL, LaRue BL, Novroski NM, Stoljarova M, Seo SB, Zeng X, Warshauer DH, Davis CP, Parson W, Sajantila A, Budowle B (2014) High-quality and high-throughput massively parallel sequencing of the human mitochondrial genome using the Illumina MiSeq. Forensic Sci Int Genet 12:128–135. https://doi.org/10.1016/j.fsigen.2014.06.001
12. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13:341. https://doi.org/10.1186/1471-2164-13-341
13. Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30(5):614–620. https://doi.org/10.1093/bioinformatics/btt593
14. Masser DR, Stanford DR, Freeman WM (2015) Targeted DNA methylation analysis by next-generation sequencing. J Vis Exp (96): 52488. https://doi.org/10.3791/52488
15. Li LC, Dahiya R (2002) MethPrimer: designing primers for methylation PCRs. Bioinformatics 18(11):1427–1431. https://doi.org/10.1093/bioinformatics/18.11.1427
16. Hakkaart C, Ellison-Loschmann L, Day R, Sporle A, Koea J, Harawira P, Cheng S, Gray M, Whaanga T, Pearce N, Guilford P (2019) Germline CDH1 mutations are a significant contributor to the high frequency of early-onset diffuse gastric cancer cases in New Zealand Maori. Familial Cancer 18(1): 83–90. https://doi.org/10.1007/s10689-018-0080-8
17. Lutsik P, Feuerbach L, Arand J, Lengauer T, Walter J, Bock C (2011) BiQ Analyzer HT: locus-specific analysis of DNA methylation by high-throughput bisulfite sequencing. Nucleic Acids Res 39(Web Server issue):W551–W556. https://doi.org/10.1093/nar/gkr312
18. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Gruning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46(W1): W537–W544. https://doi.org/10.1093/nar/gky379

Chapter 4

Editing of DNA Methylation Patterns Using CRISPR-Based Tools

Jim Smith, Rakesh Banerjee, Robert J. Weeks, and Aniruddha Chatterjee

Abstract

DNA methylation is an epigenetic modification with an established role in both normal cellular function and mammalian disease. Despite well-characterized associations between aberrant DNA methylation changes and gene expression, evidence for a causal relationship in this context has been difficult to obtain. Early techniques for interrogating the role of DNA methylation in the regulation of gene transcription lack specificity and, where more specific techniques such and ZNFs and TALEs have been developed, they are limited by their extensive cost and labor requirements. However, the recent advent of CRISPR-based technologies has revolutionized our potential for site-specific epigenomic editing. Here, we provide a detailed protocol for the design, construction, and utilization of a transient, CRISPR-based DNA methylation-editing system in mammalian cells.

Key words CRISPR, dCas9, DNA, Methylation, Editing, Transfection

1 Introduction

DNA methylation (5-methylcytosine, 5mC) is a stable, heritable epigenetic modification which has an established role in numerous biological contexts, including gene expression regulation, cellular differentiation, and genomic imprinting [1, 2]. Furthermore, aberrant changes in DNA methylation are associated with a range of disease phenotypes, including neurodevelopmental disease [3], diabetes [4], and cancer development [5, 6], progression [7, 8] and inter-individual variation [9, 10]. For many decades now, aberrant DNA methylation patterns have been associated with altered transcriptional programs and dysregulated gene expression, particularly within the context of tumorigenesis [11]. Despite the identification of characteristic DNA methylation anomalies within key tumor-associated genes, causality between these methylation changes and transcriptional alterations has been difficult to establish [12]. An inability to establish true causality in this context is largely due to

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_4,

the limitations of early investigative technologies, in particular, the lack of specificity offered by global DNA methylation-modifying agents such as decitabine or nitric oxide [12–14]. Decitabine (5-aza-2′-deoxycytidine) is a demethylating agent which inhibits DNA methyltransferase (DNMT) enzymes to prevent both de novo methylation and the maintenance of methylation patterns during replication [13]. In contrast, nitric oxide is an inflammatory mediator which can induce DNA methylation through the upregulation of DNMT activity [14]. Both of these agents are nonspecific, acting globally to induce methylation changes across the entire DNA methylome and, therefore, are unable to provide causal evidence between locus-specific methylation changes and gene expression. Hence, more modern methylation-editing technologies were developed with the aim of producing controlled, site-specific DNA methylation changes with minimal off-target impacts [12]. Initial methods, including zinc finger proteins (ZNFs) and transcription activator-like effector proteins (TALEs), utilized modular DNA-binding proteins with binding domains engineered to target a specific nucleotide sequence. However, these techniques are labor-intensive and expensive, requiring complete reengineering of ZNFs and TALEs proteins for each target locus [15, 16]. More recently, the advent of clustered regularly interspaced short palindromic repeats (CRISPR)-based technologies has revolutionized our capacity for epigenomic editing, offering the potential for locus-specific manipulation of DNA methylation, alongside minimal off-target impacts. Further, CRISPR-based systems have shown to be effective across multiple contexts, are much less labor-intensive to engineer, and are easily modifiable to target a range of unique genomic loci [12, 17]. Here, we provide a detailed protocol for the design, construction, and utilization of a CRISPR-based DNA methylation-editing system (Fig. 1). Our described method is optimized for locus-specific, transient manipulation of DNA methylation within mammalian cells [18]. Although we describe this protocol for human cell lines, a similar method could be applied to other organisms such as mice [17] and zebrafish [19, 20].

2 Materials

Prepare and store all reagents at room temperature unless otherwise specified.

2.1 gRNA Design

1. Access to the online Benchling platform (http://benchling.com).

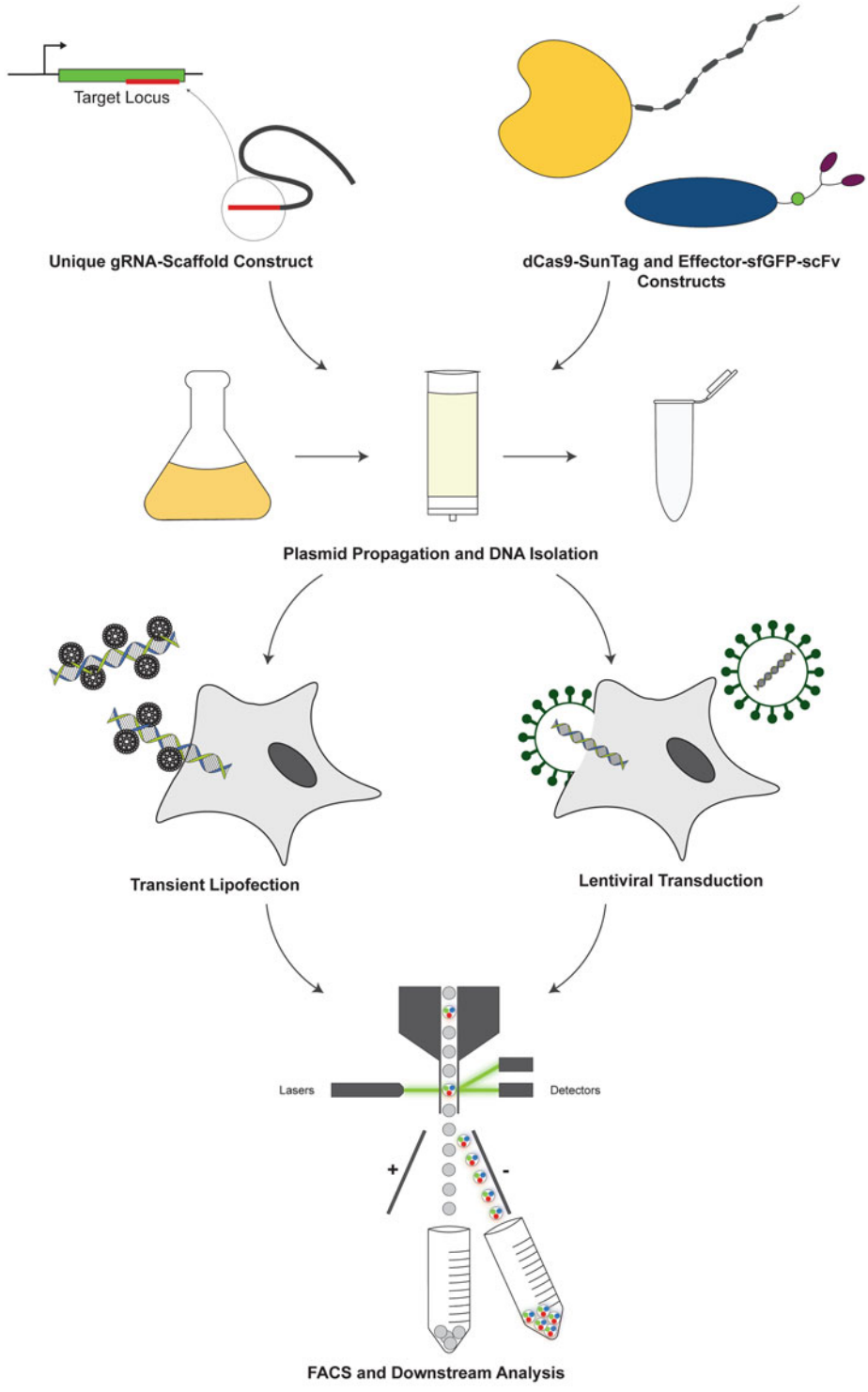


Fig. 1 Overview of the DNA methylation editing process. Three major components are required for targeted DNA methylation editing using this system: a CRISPR-dCas9-SunTag DNA binding protein (yellow); a unique gRNA molecule designed to the target locus; and an effector protein construct composed of a methylation-editing enzyme (blue) fused to a short chain variable fragment (scFv; purple) domain, which facilitates binding

2.2 gRNA and Other Plasmid Construct Preparation

1. Commercially available plasmids for construction (*see* **Note 1**).
2. Antibiotic for plasmid selection (*see* **Note 2**).
3. Lysogeny broth (LB) and LB agar plates (*see* **Note 3**).
4. Incubator capable of incubation at 37 °C and oscillation at 200 rpm for up to mid-scale bacterial cultures (200–500 mL).
5. Miniprep kit for plasmid DNA isolation (*see* **Note 4**).
6. Restriction endonucleases RsrII, SpeI, BsmBI and appropriate buffer(s) (*see* **Note 5**).
7. Thermal cycler for annealing, PCR, and restriction digest procedures.
8. T4 DNA ligase. Store at −20 °C. Use on ice when required.
9. Competent *E. coli* cells for cloning (*see* **Note 6**).
10. Maxi-prep DNA isolation kit (*see* **Note 7**).
11. Nanophotometer for measuring DNA concentration.

2.3 Cell Culture and Transfection

1. Appropriate cell line for transfection (*see* **Note 8**).
2. Cell culture medium for chosen cell line(s).
3. Cell culture facilities and incubator (*see* **Note 9**).
4. Cell culture plasticware (*see* **Note 10**).
5. 0.05% trypsin–EDTA (ethylenediaminetetraacetic acid). Store at 4 °C. Prewarm to approximately 37 °C prior to use.
6. 1× Dulbecco's phosphate buffered saline (DPBS).
7. Opti-MEM serum-free medium or appropriate equivalent. Store at 4 °C. Prewarm to approximately 37 °C prior to use.
8. Lipofectamine 3000 reagent and P3000 reagent. Store at 4 °C.
9. Hemocytometer for cell counting.

2.4 FACS

1. Ice for keeping cells pre- and post-FACS.
2. Sterile autoMACS buffer: 1× DPBS, 1% FCS, 2 mM EDTA (ethylenediaminetetraacetic acid). Store at 4 °C (*see* **Note 11**).
3. Transfected cells (72 h posttransfection).
4. Cell culture medium for chosen cell line(s).
5. Cell culture facilities and incubator (*see* **Note 9**).
6. Cell culture plasticware (*see* **Note 10**).

Fig. 1 (continued) to the SunTag protein scaffold. Each of these constructs is propagated in competent *E. coli* and plasmid DNA is isolated. Isolated DNA from each construct is transferred into target cell lines; either via lentiviral packaging and transduction which promotes stable integration and expression, or transiently, via lipofection (described here). Transfected cells can then be sorted via FACS to isolate a population of triple-positive cells containing all three components of the methylation editing system

7. 0.05% trypsin–EDTA. Store at 4 °C.
8. 1× Dulbecco's phosphate buffered saline (DPBS).
9. FACS system capable of sorting cells with multiple fluorescent markers, simultaneously (*see* **Note 12**).

3 Methods

Carry out all procedures at room temperature, unless otherwise specified. Procedures in Subheadings 3.3 and 3.4 should be performed in appropriate sterile cell culture facilities. gRNA design is performed using the Benchling online platform, as per described methods [21]. Transfection of the CRISPR-methylation system is performed via reverse lipofection using the Lipofectamine 3000 transfection system, with minor variation from the manufacturer's protocol.

3.1 gRNA Design and Preparation

1. Design unique gRNA sequences for a target locus using the cloud based Benchling platform. gRNA design for targeted DNA methylation editing requires multiple specialist considerations (*see* **Note 13**).
2. Once designed, gRNA sequences should be ordered as pairs of oligonucleotides with the following sequences: forward: 5′-CACC**G**$(N)_{20}$-3′, reverse: 3′-**C**$(N')_{20}$CAAA-5′ (*see* **Note 14**).
3. To begin gRNA construct preparation, first digest the pLKO5.sgRNA.EFS.tRFP657 vector using BsmBI restriction endonuclease, overnight at 55 °C (*see* **Note 15**).
4. Then dephosphorylate the free ends of the digested vector by adding 2 μL of shrimp alkaline phosphatase (rSAP) directly to the digested product and incubating at 37 °C for 30 min, followed by 65 °C for 5 min to heat-inactivate the rSAP enzyme. Purify the digested, dephosphorylated vector (*see* **Note 16**).
5. Anneal respective forward and reverse gRNA oligonucleotides to form a unique oligonucleotide duplex (*see* **Note 17**).
6. Dilute the oligonucleotide duplex 1:500, then ligate into the digested pLKO5.sgRNA.EFS.tRFP657 vector at room temperature overnight using T4 DNA Ligase (*see* **Note 18**).
7. Transform ligated constructs into competent *E. coli* and propagate on LB agar with 100 μg/mL ampicillin selection via incubation overnight at 37 °C.
8. Select a single, isolated colony for further propagation in LB liquid medium. Culture overnight via incubation at 37 °C with 100 μg/mL ampicillin selection (*see* **Note 19**). Proceed to DNA isolation as per Subheading 3.3.

3.2 CRISPR-dCas9-SunTag and Effector Construct Preparation

1. For the CRISPR-dCas9-SunTag construct, we use the commercially available pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS plasmid (*see* **Note 1**). Culture the *E. coli* strain containing this plasmid first on LB agar with 100 μg/mL ampicillin selection overnight at 37 °C, then select a single, isolated colony for liquid culture in LB, again cultured with 100 μg/mL ampicillin selection overnight at 37 °C, oscillating at 200 rpm.
2. For preparation of each respective effector construct, perform restriction cloning with RsrII and SpeI restriction endonucleases, as per the protocol previously described by Huang et al. [22] (*see* **Note 20**).
3. Once prepared, grow plasmid-containing bacterial cultures in 350–500 mL of LB medium, containing an appropriate antibiotic, for 12–18 h at 37 °C, oscillating at 200 rpm. Begin this step the day before plasmid isolation.
4. Isolate plasmid DNA as per the instructions of the maxi-prep kit you are using.
5. Collect the eluted DNA a 1.5 mL Eppendorf tube. Measure the concentration and purity using a nanophotometer and store at −20 °C.

3.3 CRISPR-Methylation Editing System Transfection

1. Culture the respective cell line(s) to be transfected such that cells will be >85% confluent on the day of transfection (*see* **Note 21**).
2. Prior to transfection, prepare DNA combinations to be transfected by combining CRISPR-dCas9-SunTag, effector construct and gRNA plasmid DNA, respectively, in a 1:1:1 ratio, to a total amount of 1500 ng per combination (i.e., 500 ng of each plasmid) (*see* **Note 22**). Also have an empty tube prepared with no DNA to serve as a negative control sample.
3. To begin the transfection process, add 5 μL of P3000 reagent combined with 125 μL of serum-free medium, to each respective DNA combination, as well as the negative control sample (*see* **Note 23**).
4. Next, add 8 μL of Lipofectamine 3000 reagent diluted in a further 125 μL of serum-free medium, to each respective tube.
5. Mix by pipetting, then incubate at room temperature for 15–20 min to allow DNA-lipid complexes to form.
6. Whilst the DNA-lipid complexes are incubating, prepare and count cells as appropriate for the respective cell line(s) used (*see* **Note 24**).
7. Once incubation is complete, add each respective DNA-lipid complex to an individual well of a 6-well cell culture plate.

8. Resuspend cells in cell culture medium to a total concentration of 5.0×10^5 cells/mL and add 1 mL of cell suspension to each respective DNA-lipid complex-containing well.
9. Add a further 1–2 mL of cell culture medium to each well, bringing the total volume to 2–3 mL, then incubate under appropriate cell culture conditions.
10. After the optimal exposure period (*see* **Note 25**), remove transfection reagents and wash cells with 1× DPBS. Replace with 2–3 mL of culture medium and incubate under appropriate conditions until 72-h posttransfection.

3.4 FACS for Edited Cell Collection

Perform all steps in this section within a sterile cell culture hood. Perform preparation steps immediately before FACS to maximize cell health.

1. At 72 h posttransfection, trypsinize cells and wash with 1× DPBS.
2. Stain cells with LIVE/DEAD Fixable Near-IR Dead Cell Stain Kit for 20–25 min in 15 mL Falcon tubes (*see* **Note 26**).
3. Centrifuge cells and remove supernatant before resuspending cells in 250 μL of sterile autoMACS buffer (*see* **Note 27**).
4. Prepare and label a 15 mL Falcon collection tube for each respective sample, containing 250 μL of 1× DPBS (*see* **Note 28**).
5. Once prepared, keep cells on ice pre- and post-FACS.
6. Perform FACS using an appropriate sorting apparatus (*see* **Note 29**), collecting each sample into its respective collection tube.
7. Once sorted, cells should be centrifuged, the supernatant removed, and then stored at −80 °C until required for analysis.

4 Notes

1. Commercially available plasmids required for system construction are summarized in Table 1; all are available from Addgene and catalogue numbers are as listed. It should be noted that the required plasmids will vary dependent on the goal of the editing system (e.g., for methylation versus demethylation purposes).
2. For the suggested plasmids detailed above, we use ampicillin at a final concentration of 100 μg/mL.
3. Prepare fresh, containing the appropriate antibiotic for plasmid selection.

Table 1
Commercially available plasmids for CRISPR-methylation editing system construction

Plasmid name	Purpose	Catalogue #
pHRdSV40-dCas9-10xGCN4_v4-P2A-BFP	CRISPR-dCas9-SunTag	60,903
pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS	Effector protein vector	60,904
Fuw-dCas9-TET1CD	TET1 effector for demethylation	84,475
Fuw-dCas9-DNMT3A	DNMT3A effector for methylation	84,476
pLKO5.sgRNA.EFS.tRFP657	gRNA scaffold vector	57,824

4. We recommend using the Zyppy Plasmid Miniprep Kit (Zymo Research) or similar.
5. We use restriction endonucleases RsrII, SpeI, and BsmBI with CutSmart™, CutSmart™, and NEBuffer™ 3.1 buffers (New England BioLabs), respectively.
6. We use One Shot TOP10 Chemically Competent *E. coli* (Invitrogen) for cloning with very good results.
7. We have had the best results with respect to DNA yield and purity when using the GenCatch™ Plasmid Plus DNA Maxiprep columns (Epoch Life Science Inc).
8. Our transfection method has been optimized specifically for the human melanoma cell lines NZM40, WM115, and WM266-4. Any deviation from these cell lines would benefit from reoptimization of transfection conditions, including DNA amount, Lipofectamine 3000 concentration, and transfection reagent exposure period.
9. Cell culture and transfection procedures should be performed in a class II biological safety cabinet to maintain sterility. Incubator(s) should have capacity to maintain appropriate cell culture conditions for the chosen cell line(s).
10. Plasticware would generally include 1.5 mL Eppendorf (or other brand microfuge) tubes, 15 mL Falcon tubes and 75 and 175 cm^2 cell culture flasks.
11. Combine sterile 1× DPBS, 1% FCS, and 2 mM EDTA. autoMACS buffer is used to keep cells happier during FACS to maximise survival.
12. The FACS system used must have appropriate lasers for sorting the respective fluorescent markers of each CRISPR-methylation editing construct. We use the BD FACSAria Fusion (BD Biosciences) flow cytometer. An experienced operator is also valuable to streamline to sorting process.
13. In order to maximize CRISPR-methylation editing efficacy, each gRNA should be selected with the following considerations in mind: (1) Binding of the dCas9 module to the target

locus will mechanically obstruct 20–30 bp directly overlying the gRNA target sequence; (2) Current evidence suggests that dCas9-SunTag based methylation editing systems are effective up to 1 kb in distance from the PAM site of the target locus [18, 23].

14. gRNA oligonucleotides are designed such that the 20 nt target sequence ($(N)_{20}$) and reverse complement sequence ($(N')_{20}$) also have a 5′-CACC-3′ or 3′-CAAA-5′ "overhang" sequence, respectively, for "in-frame" restriction cloning into the gRNA scaffold vector. Further, a 5′ guanine (bold) is added to the target sequence (corresponding cytosine added to the reverse sequence) to act as a U6 promoter transcriptional initiation site in the final construct. Where possible, oligonucleotides should be ordered prephosphorylated at the 5′ ends to remove the requirement for this during construction.
15. Restriction digest reaction components are detailed in Table 2.
16. We use rSAP purchased from New England BioLabs Incorporated (NEB). Incubation at 65 °C for 5 min is required to inactivate the rSAP. We recommend the use of the DNA Clean and Concentrator-5 Kit (Zymo Research) for purification of the digested, dephosphorylated vector.
17. Reaction components and reaction conditions for annealing gRNA oligonucleotides are detailed in Tables 3 and 4, respectively.
18. Reaction components for ligation are listed in Table 5.
19. For DNA isolation, we recommend using 350–500 mL of cultured plasmid in LB liquid medium. Prior to bulk culture, confirmation of the correctly inserted gRNA target sequence can be performed via capillary Sanger sequencing, using a primer designed to the adjacent U6 promoter: 5′-TTTGCTG TACTTTCTATAGTG-3′.
20. Briefly, the protocol described by Huang et al. [22] involves (1) amplification of the respective effector protein target sequence from the parent plasmid using primers designed with flanking SpeI and RsrII restriction sites for in-frame cloning, (2) SpeI and RsrII restriction digestion of the pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS vector, and (3) cloning of the effector protein sequence into the digested vector.
21. As timing will vary by cell line, we recommend being familiar with each cell line prior to planning a transfection.
22. Prepare each respective combination in a 1.5 mL Eppendorf tube.
23. Mix combinations by pipetting, ensuring no cross-contamination of the samples. Negative control samples comprise transfection reagents only.

Table 2
BsmBI restriction digest of the pLKO5-sgRNA-EFS-tRFP657 vector

Reagent	Volume (μL)
NEBuffer™ 3.1 (or alternative buffer)	10.0
BsmBI enzyme	2.0
H_2O	68.0
Plasmid DNA	25.0
Total	**105.0**

Table 3
Reaction conditions for gRNA oligonucleotide annealing

Reagent	Volume (μL)
Sense oligonucleotide (100 μM)	1.0
Antisense oligonucleotide (100 μM)	1.0
T4 DNA ligase buffer	1.0
H_2O	7.0
Total	**10.0**

Table 4
Thermal cycling protocol for gRNA oligonucleotide annealing

Protocol		
Step 1	95 °C	2:30 min
Step 2	−1.0 °C per cycle	10 s
Step 3	GOTO Step 2	× 72
Step 4	22 °C	Infinite hold

Table 5
Reagents for ligation of gRNA sequences into the pLKO5-sgRNA-EFS-tRFP657 vector

Reagent	Volume (μL)
10× T4 DNA ligase reaction buffer	0.5
T4 DNA ligase	0.5
H_2O	1.0
Vector DNA	2.0
gRNA duplex (1:500 dilution)	1.0
Total	**5.0**

24. Cell preparation should include washing with 1× DPBS and resuspension in fresh culture medium. Adherent cell lines will require trypsinization or cell scraping as appropriate.
25. Optimal exposure period will vary between cell lines and may require dedicated optimization for best results. For human melanoma cell lines NZM40, WM115, and WM266-4, we found the optimal exposure period to transfection reagents to be 12-h, 6-h, and 6-h, respectively.
26. For optimal FACS with our melanoma cell lines, we adapted the manufacturer's protocol to use stain at a concentration of 1 μL per 6–8 mL of 1× DPS.
27. Centrifuge cells at an appropriate speed, we use 350 rcf at 4 °C for 5 min for human melanoma cells.
28. Collection tube size may vary dependent on the FACS system used.
29. The FACS system used must have capacity (i.e., appropriate lasers) to sort each of the three respective fluorophores simultaneously. We use the BD FACSAria (BD Biosciences) system. An experienced operator is also desirable to ensure accurate sorting.

Acknowledgments

The research team and the related methylation editing work is funded by a Marsden grant (grant number 11465801PNE), a Rutherford Discovery Fellowship to Chatterjee (11495101PNE), Maurice & Phyllis Paykel Trust (11459701PNE) and funding support from the Otago Medical School, Dunedin Campus. We acknowledge the support of Mike Eccles, Ian Morison, Robert Day, Gregory Gimenez, and Peter Stockwell in our work on DNA methylation editing.

References

1. Smith J, Sen S, Weeks RJ, Eccles MR, Chatterjee A (2020) Promoter DNA hypermethylation and paradoxical gene activation. Trends Cancer 6(5):392–406
2. Chatterjee A, Eccles MR (2015) DNA methylation and epigenomics: new technologies and emerging concepts. Genome Biol 16:103
3. Hwang J-Y, Aromolaran KA, Zukin RS (2017) The emerging field of epigenetics in neurodegeneration and neuroprotection. Nat Rev Neurosci 18(6):347–361. https://doi.org/10.1038/nrn.2017.46
4. Mutize T, Mkandla Z, Nkambule BB (2018) Global and gene-specific DNA methylation in adult type 2 diabetic individuals: a protocol for a systematic review. Syst Rev 7(1):1–5
5. Hattori N, Ushijima T (2014) Compendium of aberrant DNA methylation and histone modifications in cancer. Biochem Biophys Res Commun 455(1–2):3–9. https://doi.org/10.1016/j.bbrc.2014.08.140
6. Takeshima H, Yamada H, Ushijima T (2019) Cancer epigenetics. Elsevier, Amsterdam, pp 65–76. https://doi.org/10.1016/b978-0-12-811785-9.00005-3
7. Chatterjee A, Stockwell PA, Ahn A, Rodger EJ, Leichter AL, Eccles MR (2017) Genome-wide methylation sequencing of paired primary and

metastatic cell lines identifies common DNA methylation changes and a role for EBF3 as a candidate epigenetic driver of melanoma metastasis. Oncotarget 8(4):6085

8. Chatterjee A, Rodger EJ, Ahn A, Stockwell PA, Parry M, Motwani J, Gallagher SJ, Shklovskaya E, Tiffen J, Eccles MR, Hersey P (2018) Marked global DNA hypomethylation is associated with constitutive PD-L1 expression in melanoma. iScience 4:312–325. https://doi.org/10.1016/j.isci.2018.05.021
9. Chatterjee A, Stockwell PA, Rodger EJ, Duncan EJ, Parry MF, Weeks RJ, Morison IM (2015) Genome-wide DNA methylation map of human neutrophils reveals widespread inter-individual epigenetic variation. Sci Rep 5(1): 17328. https://doi.org/10.1038/srep17328
10. Chatterjee A, Stockwell PA, Rodger EJ, Morison IM (2016) Genome-scale DNA methylome and transcriptome profiling of human neutrophils. Sci Data 3(1):1–9
11. Chatterjee A, Rodger EJ, Eccles MR (2018) Epigenetic drivers of tumourigenesis and cancer metastasis. Semin Cancer Biol 51:149–159. https://doi.org/10.1016/j.semcancer.2017.08.004
12. Urbano A, Smith J, Weeks RJ, Chatterjee A (2019) Gene-specific targeting of DNA methylation in the mammalian genome. Cancers 11(10):1515
13. Derissen EJB, Beijnen JH, Schellens JHM (2013) Concise drug review: azacitidine and decitabine. Oncologist 18(5):619–624
14. Takeshima H, Niwa T, Yamashita S, Takamura-Enya T, Iida N, Wakabayashi M, Nanjo S, Abe M, Sugiyama T, Kim Y-J, Ushijima T (2020) TET repression and increased DNMT activity synergistically induce aberrant DNA methylation. J Clin Investig 130(10): 5370–5379. https://doi.org/10.1172/jci124070
15. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11(9):636
16. Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol 14(1):49
17. Nakamura M, Gao Y, Dominguez AA, Qi LS (2021) CRISPR technologies for precise epigenome editing. Nat Cell Biol 23(1):11–22
18. Smith J, Banerjee R, Waly R, Urbano A, Gimenez G, Day R, Eccles MR, Weeks RJ, Chatterjee A (2021) Locus-specific DNA methylation editing in melanoma cell lines using a CRISPR-based system. Cancers 13(21):5433. https://doi.org/10.3390/cancers13215433
19. Chatterjee A, Ozaki Y, Stockwell PA, Horsfield JA, Morison IM, Nakagawa S (2013) Mapping the zebrafish brain methylome using reduced representation bisulfite sequencing. Epigenetics 8(9):979–989
20. Chatterjee A, Lagisz M, Rodger EJ, Zhen L, Stockwell PA, Duncan EJ, Horsfield JA, Jeyakani J, Mathavan S, Ozaki Y (2016) Sex differences in DNA methylation and expression in zebrafish brain: a test of an extended 'male sex drive' hypothesis. Gene 590(2):307–316
21. Pellegrini R (2016) How to synthesize your gRNAs for CRISPR. Available via https://www.benchling.com/2016/02/23/how-to-synthesize-your-grnas-for-crispr/ Accessed on 16 May 2020
22. Huang Y-H, Su J, Lei Y, Brunetti L, Gundry MC, Zhang X, Jeong M, Li W, Goodell MA (2017) DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A. Genome Biol 18(1):176
23. Stepper P, Kungulovski G, Jurkowska RZ, Chandra T, Krueger F, Reinhardt R, Reik W, Jeltsch A, Jurkowski TP (2016) Efficient targeted DNA methylation with chimeric dCas9–Dnmt3a–Dnmt3L methyltransferase. Nucleic Acids Res 45(4):1703–1713

Chapter 5

Nanopore Sequencing and Data Analysis for Base-Resolution Genome-Wide 5-Methylcytosine Profiling

Allegra Angeloni, James Ferguson, and Ozren Bogdanovic

Abstract

Whole-genome bisulfite sequencing (WGBS) is currently the gold standard for DNA methylation (5-methylcytosine, 5mC) profiling; however, the destructive nature of sodium bisulfite results in DNA fragmentation and subsequent biases in sequencing data. Such issues have led to the development of bisulfite-free methods for 5mC detection. Nanopore sequencing is a long read nondestructive approach that directly analyzes DNA and RNA fragments in real time. Recently, computational tools have been developed that enable base-resolution detection of 5mC from Oxford Nanopore sequencing data. In this chapter, we provide a detailed protocol for preparation, sequencing, read assembly, and analysis of genome-wide 5mC using Nanopore sequencing technologies.

Key words 5-Methylcytosine, Nanopore, Long read sequencing, Cytosine–guanine

1 Introduction

DNA methylation (5-methylcytosine, 5mC) is a chemical modification of the DNA molecule occurring almost exclusively at CpG dyads in metazoans [1, 2]. 5mC is associated with biological processes involving long-term gene silencing such as genomic imprinting, X-chromosome inactivation and silencing of repetitive elements [3, 4]. The development of base-resolution 5mC profiling techniques has enabled quantitative analysis of locus-specific alterations in the dynamic 5mC landscape, such as changes that occur during cell differentiation and disease progression [5–7]. Currently, the gold standard for 5mC profiling is whole-genome bisulfite sequencing (WGBS). In WGBS, sodium bisulfite treatment converts nonmethylated cytosines to uracil that is subsequently PCR

Data Availability: Nanopore and WGBS data were deposited at the Gene Expression Omnibus under the record: GSE179673.

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_5,

amplified and sequenced as thymine, while 5mC is sequenced as cytosine. Overall, WGBS is a widely used and accurate technique that enables quantitative base-resolution detection of 5mC, with conversion efficiencies greater than 99% [8]. However, while WGBS has long been a benchmark for 5mC analysis, bisulfite-induced fragmentation of DNA and decreased sequence complexity are associated with WGBS data [9, 10].

Such challenges have led to the development of bisulfite-free sequencing methods for base-resolution detection of 5mC, such as long read Oxford Nanopore sequencing technologies [11, 12]. Nanopore sequencing functions through characterization of sequence-dependent changes in ionic current as a single-stranded oligonucleotide travels through a protein nanopore. Methylated bases generate interpretable signatures that enable 5mC sequencing without the requirement to convert the DNA sequence itself. It has been reported that Nanopore technologies can sequence DNA fragments in excess of 100s to 1000s of kilobase pairs (kbp) [13, 14]. Therefore, a major advantage of Nanopore is that long reads can be assembled with less ambiguity, increasing genome coverage at regions that are not well covered by short reads, such as repetitive elements.

Here, we provide a detailed protocol for the preparation and sequencing of genomic DNA using Nanopore and methods for base-calling, read assembly, and 5mC detection as well as example data analyses. This chapter describes genomic DNA extraction, preparation, and sequencing from zebrafish embryos at 24 h post-fertilization (hpf). Zebrafish are a valuable model organism for DNA methylation studies. Similar to the mammalian DNA methylome, zebrafish exhibit a fully methylated genome, DNA methyltransferase enzymes and orthologues of human Ten-Eleven Translocation (TET) proteins involved in actively removing 5mC [15, 16]. However, the protocols outlined in this chapter can be extended to the methylation profile of any tissue or cell line.

2 Materials

2.1 Zebrafish Embryos

1. 25–50 fresh or frozen whole zebrafish embryos (24 hpf).

2.2 Phenol–Chloroform Genomic DNA Extraction Followed by Ethanol Precipitation

1. Homogenization buffer: 20 mM Tris–HCl pH 8.0, 100 mM NaCl, 15 mM EDTA, 1% SDS, 0.5 mg/mL Proteinase K.
2. RNase A (4 mg/mL).
3. Phenol–chloroform–isoamyl alcohol (25:24:1).
4. 3 M sodium acetate (pH 5.2).
5. Ethanol (>99%). Place in −20 °C freezer for approximately an hour prior to experiment to keep ice-cold.

6. Linear acrylamide (5 mg/mL).
7. Nuclease-free water.
8. Freshly prepared 70% ethanol: add 3 mL nuclease-free water to 7 mL ethanol (>99%) the day of experiment. Place in −20 °C freezer for approximately an hour prior to experiment to keep ice-cold.

2.3 Genomic DNA Quality Control

1. Qubit dsDNA BR Assay Kit (Invitrogen) **or** Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) or similar high sensitivity DNA assay kit.
2. TapeStation Genomic DNA ScreenTape Analysis Reagents (Agilent) and ScreenTape (Agilent) or similar system.

2.3.1 Size Selection (Optional)

1. Short Read Eliminator XS Kit (Circulomics).

2.4 Library Preparation and Sequencing

1. Ligation Sequencing Kit (Oxford Nanopore Technologies, SQK-LSK109 for multiplexing samples, SQK-LSK110 for processing singleplex samples).
2. AMPure XP beads.
3. NEBNext FFPE Repair Mix (New England Biolabs).
4. NEBNext Ultra II End repair/dA-tailing Module (New England Biolabs).
5. NEBNext Quick Ligation Module (New England Biolabs).
6. Nuclease-free water.
7. Freshly prepared 70% ethanol: add 3 mL nuclease-free water to 7 mL ethanol (>99%) the day of experiment.
8. If multiplexing: Native Barcoding Expansion 1–12 (Oxford Nanopore Technologies, EXP-NBD104) and/or Native Barcoding Expansion 13–24 (Oxford Nanopore Technologies, EXP-NBD114) (*see* **Note 1**).

2.5 Base-Calling, Read Alignment and Analysis

1. Oxford Nanopore Technologies guppy base-caller (available from Oxford Nanopore Technologies).
2. *Danio rerio* UCSC danRer10 genome (https://support.illumina.com/sequencing/sequencing_software/igenome.html?langsel=/us/).
3. *Enterobacteriophage lambda* NCBI 1993-04-28 (https://support.illumina.com/sequencing/sequencing_software/igenome.html?langsel=/us/).
4. minimap2 (https://github.com/lh3/minimap2).
5. samtools 1.12 (http://www.htslib.org/).
6. Nanopolish 0.13.2 (https://github.com/jts/nanopolish).

7. calculate_methylation_frequency.py script (https://github.com/jts/nanopolish/blob/master/scripts/calculate_methylation_frequency.py).
8. trimmomatic 0.39 (https://github.com/igordot/trimmomatic).
9. WALT v1.0 (https://github.com/smithlabcode/walt).
10. Picard 2.3.0 (https://github.com/broadinstitute/picard).
11. MethylDackel 0.5.0 (https://github.com/dpryan79/MethylDackel).
12. bedtools 2.30.0 (https://github.com/arq5x/bedtools2).
13. kentUtils 302.1 (https://github.com/ENCODE-DCC/kentUtils).
14. deepTools 3.5.0 (https://github.com/deeptools/deepTools) (requires python 2.7 or 3.x).

3 Methods

Unless otherwise specified, carry out all procedures at room temperature.

3.1 Genomic DNA Extraction

3.1.1 Phenol–Chloroform Genomic DNA Extraction Followed by Ethanol Precipitation

1. Homogenize zebrafish embryos in 200 μL homogenization buffer and incubate sample at 55 °C for no more than 2 h. Invert tube several times to ensure embryos are homogenized following incubation.
2. To remove RNA contamination, incubate sample at 95 °C for 10 min. Add 1 μL RNase A per 100 μL homogenized genomic DNA and incubate at room temperature for 30 min.
3. Perform two phenol–chloroform extractions on samples. Add phenol–chloroform–isoamyl alcohol to sample in 1:1 ratio (approximately 200 μL). Invert tube several times to mix solutions. Spin for 5 min at 13,000 rpm (16,000 rcf).
4. Remove approximately 180 μL from the aqueous upper layer of the mixed solutions and transfer to a new eppendorf tube, being cautious not to remove any solution from the phenol–chloroform–isoamyl alcohol phase. Repeat phenol–chloroform extraction on the separated aqueous layer.
5. Add 1/10 volume 3 M sodium acetate (pH 5.2), 2.5 volumes ice-cold absolute ethanol and 2 μL linear acrylamide to genomic DNA. Precipitate at −80 °C for 30 min to 1 h or at −20 °C for at least 1 h. Extended precipitation periods (such as overnight) may increase recovery.
6. Following precipitation, spin samples at full speed for 5 min. Remove supernatant and wash pellet in 500 μL cold 70% ethanol. Spin samples at full speed for 5 min. Remove supernatant.

7. Resuspend pellet in nuclease-free water or elution buffer to desired volume. Store at −20 °C.

3.1.2 Alternative Purification of Genomic DNA

Alternatively genomic DNA can be extracted from tissue of interest using a genomic DNA extraction kit according to manufacturer's instructions (*see* **Note 2**). Examples include the Monarch Genomic DNA Purification Kit (New England Biolabs, T3010S), the Nanobind CBB Big DNA Kit (Circulomics, SKU NB-900-001-01), or the DNeasy Blood & Tissue Kit (QIAGEN, 69504).

3.2 Quality Control

1. Use a sensitive fluorometric assay such as Qubit dsDNA BR Assay Kit or Quant-iT PicoGreen dsDNA Assay Kit according to manufacturer's instructions to accurately quantify the amount of genomic DNA present in your sample (*see* **Note 3**). At least 1 μg of gDNA should be present in your sample. Use NanoDrop 2000 Spectrophotometer to determine purity of the gDNA sample. The following parameters should be met when sequencing a sample.

gDNA input amount	260/280 ratio	260/230 ratio
>1 μg	~1.8	2.0–2.2

2. Run gDNA on a TapeStation using Genomic DNA ScreenTape Analysis reagents to determine the size distribution of the DNA fragments. If DNA fragments shorter than 1 kb are present, proceed to **step 3** (*see* **Note 4**). If DNA fragments are larger than at least 1 kb, proceed to Library Preparation and Sequencing.
3. DNA fragments shorter than 1 kb can be removed using the Short Read Eliminator XS Kit according to manufacturer's instructions. This protocol performs near complete depletion of fragments less than 5 kb and progressive depletion of fragments less than 10 kb. The recovery efficiency is approximately 50–90% of input DNA. To ensure sample meets all requirements for size selection, repeat all quality control steps detailed in **Quality Control** following size selection.

3.3 Library Preparation and Sequencing

1. Library preparation and sequencing is performed using the Ligation Sequencing Kit according to manufacturer's instructions. Third party consumables required for this kit are listed in Subheading 2. Alternatively, library preparation and sequencing can be performed at a facility that offers Nanopore sequencing services. This kit is compatible with the following Nanopore sequencing devices.
   (a) Flongle.
   (b) MinION Mk1B.

Table 1
Summary of throughput specifications for each sequencing platform. Data from: https://nanoporetech.com/products/comparison

	Flongle	MinION Mk1B	MinION Mk1C	GridION Mk1	PromethION P24/P48
Approximate yield (dependent on sample and preparation methods)	1–2 GB	10–50 GB	10–50 GB	10–50 GB	100–300 GB
Run time	1 min–16 h	1 min–72 h	1 min–72 h	1 min–72 h	1 min–72 h
Multiplexing	Yes	Yes	Yes	Yes	Yes
Number of flow cells	1	1	1	5	24/48
Number of channels per flow cell	126	512	512	512	2675

(c) MinION Mk1C.
(d) GridION Mk1.
(e) PromethION P24/P48.

The sequencing device used is largely dependent on the genome of interest, multiplexing requirements and the desired sequencing throughput. For this chapter, 250 ng of the zebrafish DNA library was loaded onto a PromethION flow cell (pore version 9.4.1, flow cell version FLO-PRO002) and run on a PromethION sequencing device for 64 h. Table 1 may be used as a guide for selecting a sequencing device fit for the purpose of the study.

3.4 Read Alignment and Methylation Calling

In this chapter, Nanopore sequencing and whole-genome bisulfite sequencing (WGBS) data files for whole zebrafish embryos 24 hpf are compared, however the scripts can be edited for different input files. For example, these scripts can be used for comparing two replicates of Nanopore sequencing data. Nanopore sequencing raw data and methylation frequency files used in this tutorial are available for download from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5426094. WGBS raw data and methylation frequency files used in this tutorial are available for download from: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM5426095.

3.4.1 Data Download

3.4.2 Nanopore Base-Calling and Methylation Profiling

Oxford Nanopore reads are stored in a format called fast5. fast5 files contain the stream of current measurements taken during sequencing which are converted into nucleobases using base-calling software. Reads can either be base-called live during sequencing or following the sequencing run.

1. Download zebrafish reference genome (*Danio rerio* UCSC danRer10) sequence in fasta format:

```
###--- download and untar
wget http://igenomes.illumina.com.s3-website-us-east-1.amazo-
naws.com/Danio_rerio/UCSC/danRer10/Danio_rerio_UCSC_danRer10.
tar.gz
tar -xzvf Danio_rerio_UCSC_danRer10.tar.gz
# genome fasta found in Danio_rerio/UCSC/danRer10/Sequence/
WholeGenomeFasta/
```

2. Perform base-calling on fast5 reads using Guppy Basecaller (*see* **Notes 5** and **6**):

```
FAST5_DIR="path-to-fast5/"
FASTQ_DIR="path-to-fastq-output/"
mkdir -p ${FASTQ_DIR}

guppy_basecaller -x "cuda:0" \
-r \
-i ${FAST5_DIR} \
-c dna_r9.4.1_450bps_hac_prom.cfg \
-s ${FASTQ_DIR}
```

3. Merge pass fastq files into a single fastq:

```
for fastq in ${FASTQ_DIR}*pass*/*.fastq; do
cat $fastq; done >> ${FASTQ_DIR}pass_merged.fastq
```

4. Align fastq reads to reference genome and output in sam format [17]:

```
ALIGNMENT_DIR="path-to-alignment-output/"
mkdir -p ${ALIGNMENT_DIR}
minimap2 -x map-ont \
-a \-t16 \
--secondary=no \
genome.fa \
${FASTQ_DIR}pass_merged.fastq > ${ALIGNMENT_DIR}alignment.sam
```

5. Sort and index sam file [18]:

```
cd ${ALIGNMENT_DIR}
samtools sort alignment.sam > alignment.bam
samtools index alignment.bam
```

6. Index fast5 files for methylation calling [11]:

```
nanopolish index -d ${FAST5_DIR} \
-s ${FASTQ_DIR}sequencing_summary.txt \
${FASTQ_DIR}pass_merged.fastq
```

7. Call methylation from reads and calculate methylation frequency:

```
GENOME_FASTA_DIR="path-to-genome-fasta/"
METH_DIR="path-to-methylation-output/"
mkdir -p ${METH_DIR}
nanopolish call-methylation \
-t 8 \
-r ${FASTQ_DIR}pass_merged.fastq \
-b ${ALIGNMENT_DIR}alignment.bam \
-g ${GENOME_FASTA_DIR}genome.fa > ${METH_DIR}methylation_calls.tsv

SCRIPT_DIR="path-to-script/"
python3 ${SCRIPT_DIR}calculate_methylation_frequency.py\
${METH_DIR}methylation_calls.tsv > ${METH_DIR}nanopore_methyla-
tion_freq.tsv
```

8. Generate bigwig file for visualization of methylation frequency [19]:

```
###--- generate tsv file containing chromosome sizes
GENOME_FASTA_DIR="path-to-genome-fasta/"
samtools faidx ${GENOME_FASTA_DIR}genome.fa
cut -f1,2 ${GENOME_FASTA_DIR}genome.fa.fai > ${GENOME_FASTA_
DIR}chrom.sizes

###--- convert methylation_freq.tsv to bedGraph
cd ${METH_DIR}
tail -n +2 nanopore_methylation_freq.tsv |
awk '{print $1"\t"$2"\t"$3+2"\t"$7}' |
sort -k1,1 -k2,2n > methylation_freq.bedGraph

###--- generate bigwig
bedGraphToBigWig methylation_freq.bedGraph \
${GENOME_FASTA_DIR}chrom.sizes \
methylation_freq.bw
```

The resulting bigwig file can be visualised in a genome browser such as the Integrative Genomics Viewer (https://software.broadinstitute.org/software/igv/) or the UCSC Genome Browser (https://genome.ucsc.edu/).

3.4.3 WGBS Data Assembly and Methylation Calling

1. Download lambda genome and concatenate with zebrafish reference genome (*see* **Note 7**).

```
###--- download and untar
wget http://igenomes.illumina.com.s3-website-us-east-
1.amazonaws.com/Enterobacteriophage_lambda/NCBI/1993-04-
28/Enterobacteriophage_lambda_NCBI_1993-04-28.tar.gz
tar -xzvf Enterobacteriophage_lambda_NCBI_1993-04-28.tar.gz
# genome fasta found in Enterobacteriophage_lambda/NCBI/1993-
04-28/Sequence/WholeGenomeFasta

###--- rename lambda genome fasta file and add header
LAMBDA_DIR="Enterobacteriophage_lambda/NCBI/1993-04-28/Se-
quence/WholeGenomeFasta/"
cd ${LAMBDA_DIR}
mv genome.fa lambda.fa
sed "1s/.*/>chrLambda/" lambda.fa > /tmp/out
mv /tmp/out lambda.fa

###--- concatenate reference and lambda genome
cp lambda.fa ${GENOME_FASTA_DIR}
cd ${GENOME_FASTA_DIR}
cat genome.fa lambda.fa > genome_lambda.fa
```

2. Perform adapter clipping and read quality trimming using Trimmomatic [20]:

```
DIR="path-to-trimmomatic-0.36/"
FASTQ_DIR="path-to-fastq/"
cd ${FASTQ_DIR}
java -jar ${DIR}trimmomatic-0.36.jar PE \
-threads 24 \
-trimlog trimlog.txt \
WGBS_r1.fastq.gz WGBS_r2.fastq.gz \
r1_paired.fastq r1_unpaired.fastq \
r2_paired.fastq r2_unpaired.fastq \
ILLUMINACLIP:${ADAPTERS_DIR}NEXTflex.fa:2:30:10 \
SLIDINGWINDOW:5:20 \
LEADING:5 \
TRAILING:5 \
MINLEN:50
```

3. Index reference genome using WALT [21]:

```
GENOME_FASTA_DIR="path-to-genome-fasta/"
WALT_INDEX_DIR="path-to-WALTIndex/"
mkdir -p ${WALT_INDEX_DIR}
makedb -c ${GENOME_FASTA_DIR}"genome_lambda.fa" \
-o ${WALT_INDEX_DIR}"genome_lambda.dbindex"
```

4. Align fastq reads to reference genome:

```
ALIGNMENT_DIR="path-to-alignment-output/"
mkdir -p ${ALIGNMENT_DIR}
nice walt -i ${WALT_INDEX_DIR}"genome_lambda.dbindex" \
-1 r1_paired.fastq -2 r2_paired.fastq \
-m 10 \
-o ${ALIGNMENT_DIR}alignment.sam \
-t 24 \
-N 10000000 \
-L 2000
```

5. Convert sam files to bam format:

```
cd ${ALIGNMENT_DIR}
samtools view -Sb alignment.sam > alignment.bam
samtools sort -o alignment_sorted.bam \
alignment.bam;
    samtools index alignment_sorted.bam
```

6. Remove duplicate reads:

```
DIR="path-to-picard-tools-2.3.0/"

java -jar ${DIR}picard.jar MarkDuplicates \
        REMOVE_DUPLICATES=true \
   I=alignment_sorted.bam \
   O=alignment_dedup.bam \
   M=alignment_dedup.txt
```

7. Sort and index deduplicated bam file:

```
samtools sort -o alignment_dedup_sorted.bam \
alignment_dedup.bam;
 samtools index alignment_dedup_sorted.bam
```

8. Call methylation with MethylDackel:

```
###--- mbias
METH_DIR="path-to-methylation-output/"
mkdir -p ${METH_DIR}
MethylDackel mbias ${GENOME_FASTA_DIR}"genome_lambda.fa"
\alignment_dedup_sorted.bam \
${METH_DIR}alignment_dedup_sorted_mbias.svg

###--- call CpG methylation
MethylDackel extract ${GENOME_FASTA_DIR}"genome_lambda.fa" \
alignment_dedup_sorted.bam \
-o ${METH_DIR}wgbs_MethylDackel \
--mergeContext \
--minOppositeDepth 5 \
--maxVariantFrac 0.5 \
--OT 10,0,10,130 --OB 10,140,30,145

###--- output non-conversion rate to txt file
cd ${METH_DIR}
grep chrLambda wgbs_MethylDackel_CpG.bedGraph |
awk '{ sum += $4; n++ } END { if (n > 0) print sum / n; }' >
chrLambda_nonconversion.txt
```

9. Generate bigwig file for visualization of methylation frequency:

```
###--- generate bedGraph containing 5mC levels only
awk '{print $1"\t"$2"\t"$3"\t"$4/100}' wgbs_MethylDackel_CpG.
bedGraph |
sort -k1,1 -k2,2n >wgbs_methylation_freq.bedGraph

###--- generate bigwigGENOME_FASTA_DIR="path-to-chrom-sizes/
"bedGraphToBigWig wgbs_methylation_freq.bedGraph \
${GENOME_FASTA_DIR}chrom.sizes \
wgbs_methylation_freq.bw
```

3.5 Example Downstream Analyses

3.5.1 Bedgraph File Structure

For ease, the bedGraph files containing methylation data (`nanopore_methylation_freq.tsv` and `wgbs_MethylDackel_CpG.bedGraph` files for Nanopore and WGBS data respectively) should be manipulated to have a consistent structure. The bedGraph file structure used for downstream analyses in this chapter is shown below, where column 4 is the number of reads indicating methylation, column 5 is the total number of reads and column 6 is the fraction of all reads indicating methylation at a given CpG

site (between 0.0 and 1.0). This can be generated using the following commands (*see* **Note 8**).

```
chr start end mCG_reads total_reads percent_meth
chr1 81 83 4 4 1.000
chr1 92 94 5 5 1.000
chr1 149 151 4 5 0.800
chr1 234 236 2 3 0.667
chr1 317 319 6 6 1.000
chr1 357 359 3 5 0.600
chr1 374 376 4 5 0.800
chr1 609 611 1 2 0.500
chr1 716 718 6 6 1.000

# Nanopore
###--- rearrange and sort columns by genomic coordinate
awk 'FNR > 1 {print $1"\t"$2"\t"$3+2"\t"$6"\t"$5"\t"$7}' na-
nopore_methylation_freq.tsv |
sort -k1,1 -k2,2n > nanopore_5mC.bedGraph

###--- add header
echo -e "chr\tstart\tend\tmCG_reads\ttotal_reads\tpercent_-
meth" |
cat - nanopore_5mC.bedGraph > /tmp/out
mv /tmp/out nanopore_5mC.bedGraph

# WGBS
###--- rearrange and sort columns by genomic coordinate
awk 'FNR > 1 {print $1"\t"$2"\t"$3"\t"$5"\t"$6+$5"\t"$4/100}'
wgbs_MethylDackel_CpG.bedGraph |
grep -v chrLambda | sort -k1,1 -k2,2n > wgbs_5mC.bedGraph

###--- add header
echo -e "chr\tstart\tend\tmCG_reads\ttotal_reads\tpercent_-
meth" |
cat - wgbs_5mC.bedGraph > /tmp/out
mv /tmp/out wgbs_5mC.bedGraph
```

The bedGraph files containing methylation data used in downstream analysis for WGBS and Nanopore sequencing are referred to as `wgbs_5mC.bedGraph` and `nanopore_5mC.bedGraph` respectively.

3.5.2 CpG Count

1. Count number of CpG sites identified by at least one read and save value to output file (Fig. 1).

```
wc -l nanopore_5mC.bedGraph > nanopore_5mC_CpG.count.txt
wc -l wgbs_5mC.bedGraph > wgbs_5mC_CpG.count.txt
```

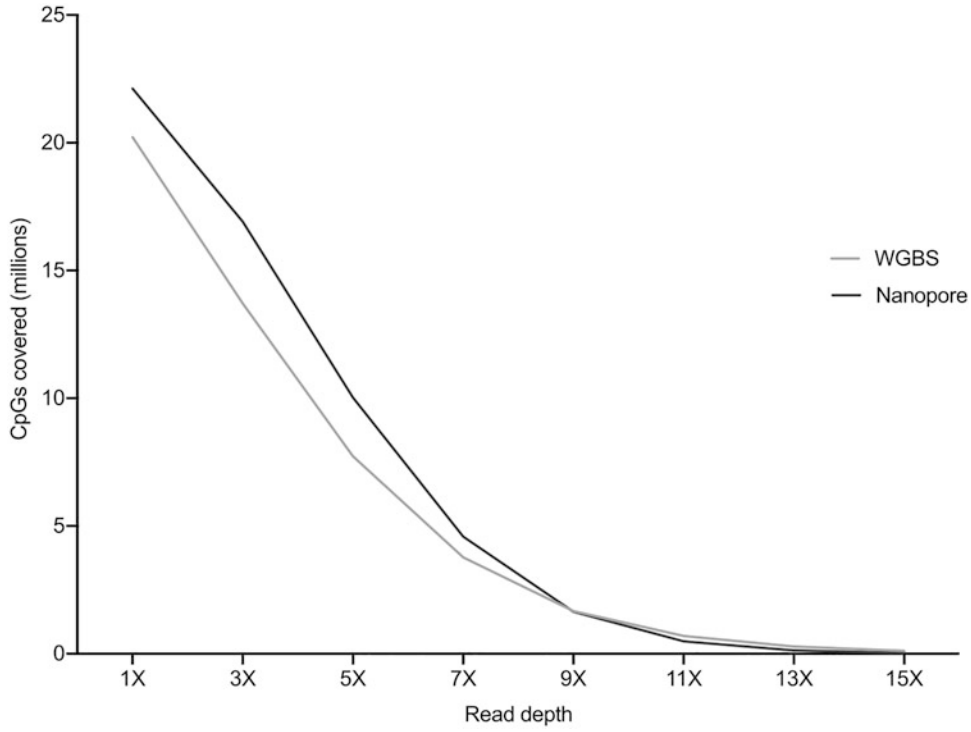


Fig. 1 CpG sites covered by WGBS and Nanopore sequencing at increasing read depths

2. Identify the number of CpG sites identified by at least 3 reads and save value to output file (Fig. 1).

```
awk '$5 >= 3' nanopore_5mC.bedGraph |
wc -l > nanopore_5mC_CpG_3X.count.txt
awk '$5 >= 3' wgbs_5mC.bedGraph |
wc -l > wgbs_5mC_CpG_3X.count.txt
```

3. Count the number of CpG sites covered by both sequencing techniques by at least one read (Fig. 2).

```
bedtools intersect -u -a nanopore_5mC.bedGraph -b wgbs_5mC.
bedGraph |
wc -l > nanopore_wgbs_5mC_CpG.overlap.txt
```

3.5.3 Genomic Tracks Coverage

Read depth for each sequencing technique is plotted over CpG islands (CGI) tracks downloaded from the *Danio rerio* UCSC danRer10 genome.

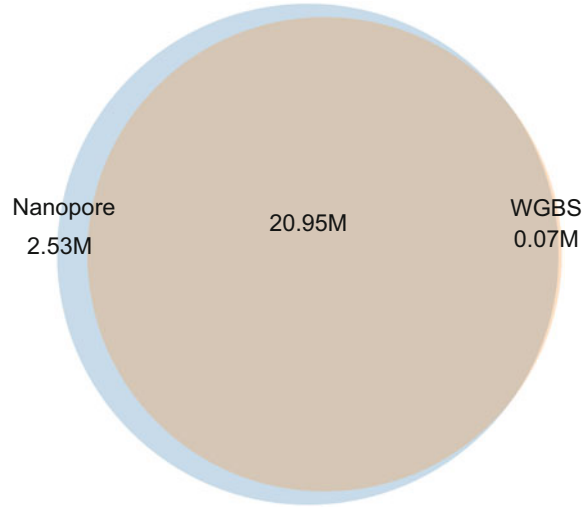


Fig. 2 CpG sites identified commonly and uniquely by Nanopore and WGBS. Venn diagrams drawn to scale, values rounded to 2 decimal places. M = millions of CpG sites. All CpG sites covered by at least one read

1. Generate bigwig file containing read depth information for each sample.

```
###--- specify path to folder containing chrom.sizes file for
genome of interest
GENOME_FASTA_DIR="path-to-genome-chrom.sizes/"
# Nanopore
###--- generate read depth file from alignment
samtools depth nanopore_alignment.bam > nanopore_read_depth
awk 'OFS="\t" {print $1,$2-1,$2,$3}' nanopore_read_depth |
sort -k1,1 -k2,2n > nanopore_read_depth.bed

###--- generate bigwig
bedGraphToBigWig nanopore_read_depth.bed \
${GENOME_FASTA_DIR}chrom.sizes \
nanopore_read_depth.bw#

WGBS
###--- generate read depth file from alignment
samtools depth wgbs_alignment.bam > wgbs_read_depth
awk 'OFS="\t" {print $1,$2-1,$2,$3}' wgbs_read_depth |
sort -k1,1 -k2,2n > wgbs_read_depth.bed

###--- generate bigwig
bedGraphToBigWig wgbs_read_depth.bed \
${GENOME_FASTA_DIR}chrom.sizes \
wgbs_read_depth.bw
```

2. Plot read depth over CGI tracks in a heatmap using deepTools (Fig. 3) [22].

```
###--- generate matrix of read depth at CGIs
computeMatrix scale-regions \
-R UCSC_danRer10_CGI_tracks.bed \
-S nanopore_read_depth.bw wgbs_read_depth.bw \
--afterRegionStartLength 1000 \
--beforeRegionStartLength 1000 \
--binSize 50 \
--smartLabels \
--missingDataAsZero \
-out CGI_read_depth.mat.gz

###--- plot matrix as heatmap
plotHeatmap -m CGI_read_depth.mat.gz \
--yAxisLabel "Read Depth" \
--xAxisLabel "" \
--yMin 0 \
--yMax 30 \
--startLabel Start \
--endLabel End \
--regionsLabel CGI \
--colorMap YlOrRd \
-out CGI_read_depth.pdf
```

3.5.4 Methylation Analysis

1. Generate stacked bar plot of global 5mC levels (performed in R) (Fig. 4).

```
###--- load libraries and set the working directory
library(data.table)
library(tidyr)
library(dplyr)
library(ggplot2)
library(RColorBrewer)
setwd("path-to-directory/")

###--- load data into R
nanopore <- fread("nanopore_5mC.bedGraph")
wgbs <- fread("wgbs_5mC.bedGraph")

###--- generate table of global 5mC levels for input into
ggplot
##-- nanopore
input_nanopore <- nanopore %>%
select(percent_meth) %>%
rename(Nanopore = percent_meth) %>%
gather()
```

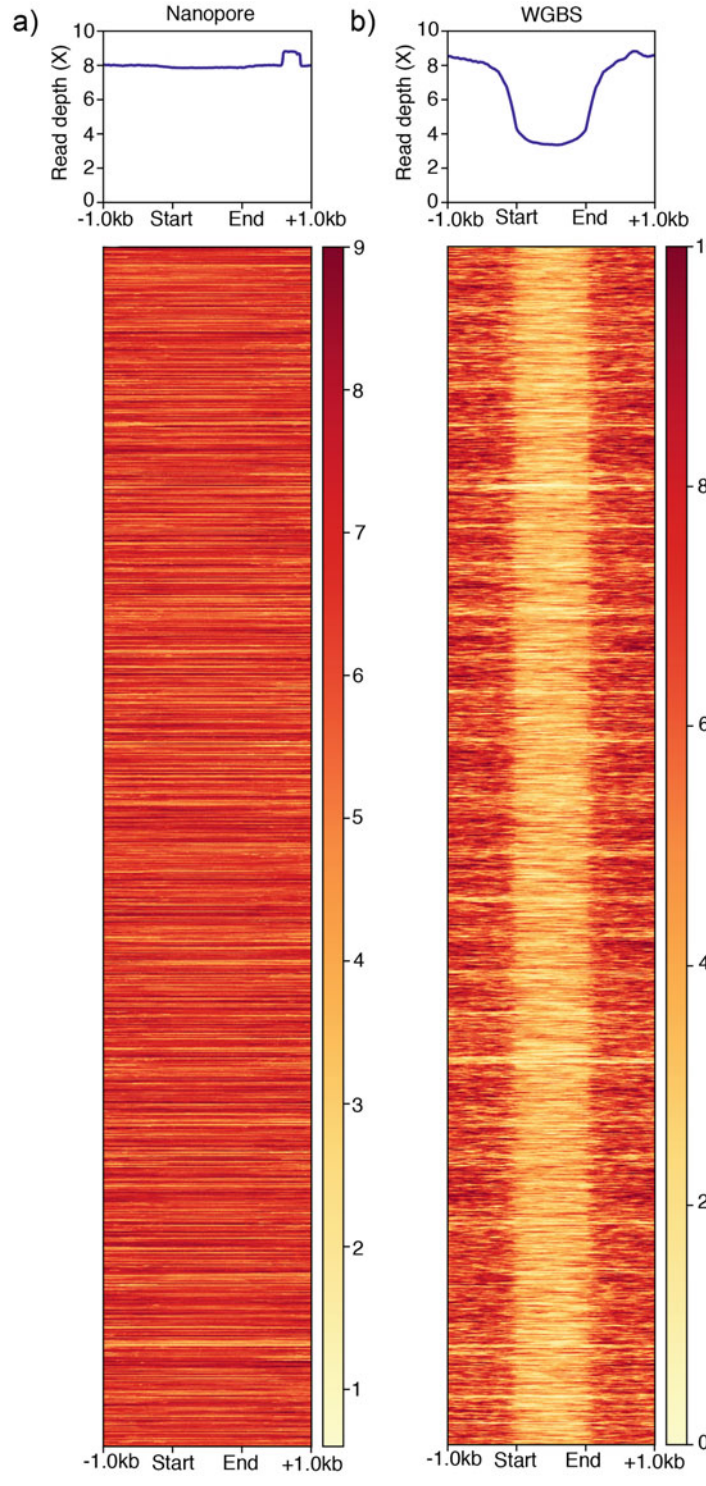


Fig. 3 Sodium bisulfite treatment during WGBS fragments nonmethylated CpG sites. Coverage plot (top) and heatmap (bottom) of mean read depth calculated over CGIs and 1.0 kb upstream and downstream in (**a**) Nanopore and (**b**) WGBS data

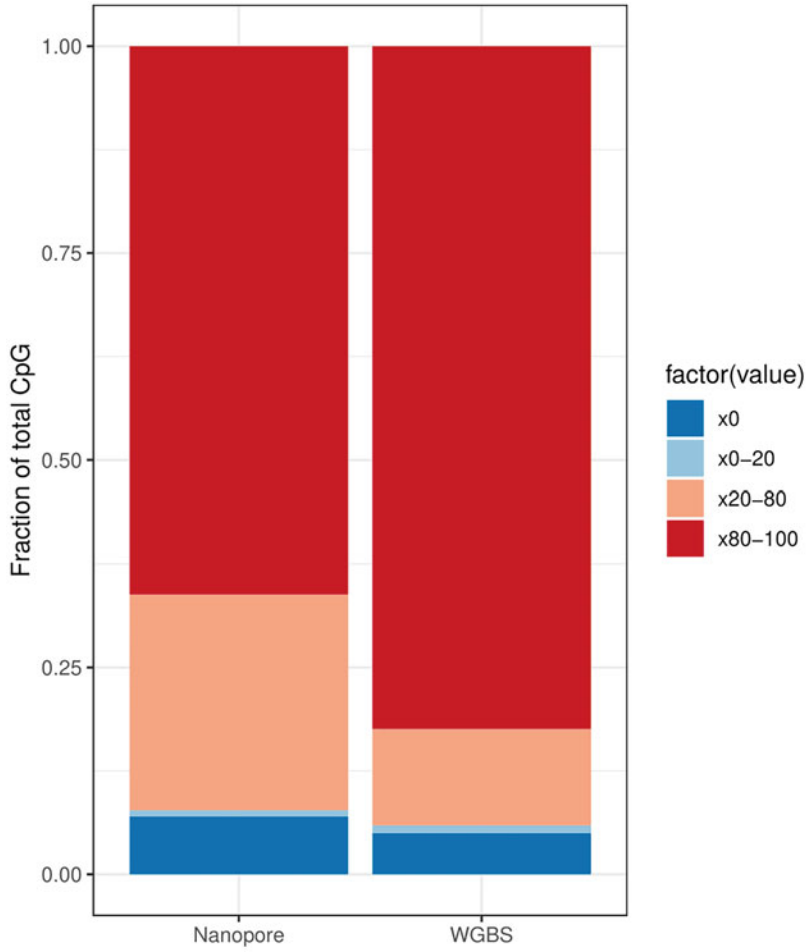


Fig. 4 Stacked bar plot of the proportion of CpG sites in genome with low (<20%), intermediate (20–80%) and high (>80%) mCG/CG levels. Number of CpGs is displayed as a fraction between 0.00 and 1.00, with 0.00 being no CpG sites and 1.00 being all CpG sites identified by each sequencing technique

```
input_nanopore_grouped <- input_nanopore %>%
group_by(key,  value=cut(input_nanopore$value,  breaks=c
(0, 0.01, 0.2, 0.8, 1.0), include.lowest = TRUE)) %>%
summarise(n=(n()/nrow(input_nanopore)*100))
vec <- c("x0", "x0-20", "x20-80", "x80-100")
input_nanopore_grouped$value <- vec

##-- wgbs
input_wgbs <- wgbs %>%
select(percent_meth) %>%
rename(WGBS = percent_meth) %>%
gather()
```

```
input_wgbs_grouped <- input_wgbs %>%
group_by(key, value=cut(input_wgbs$value, breaks=c(0, 0.01,
0.2, 0.8, 1.0), include.lowest = TRUE)) %>%
summarise(n=(n()/nrow(input_wgbs)*100))
vec <- c("x0", "x0-20", "x20-80", "x80-100")
input_wgbs_grouped$value <- vec

###--- merge tables
mCG <- rbind(input_nanopore_grouped, input_wgbs_grouped)

###--- plot table as stacked bar plot, save as PDF
pdf("5mC_stacked_barplot.pdf", width=5, height=6)
print(ggplot(mCG, aes(x = key, y = n, fill = factor(value))) +
geom_bar(stat="identity", position = position_fill(reverse =
TRUE)) +
scale_fill_brewer(palette = "RdBu", direction=-1) +
xlab("") +
ylab("Fraction of total CpG") +
theme_bw())
dev.off()
```

4 Notes

1. DNA methylation information is lost during PCR. This kit must be selected if multiplexing as it uses a PCR-free protocol.
2. Some genomic DNA extraction kits include numerous vortexing steps in their protocol. This may lead to increased DNA shearing and consequently shorter read lengths.
3. It is imperative that genomic DNA is accurately quantified. Ensure a sensitive fluorometric technique is used to quantify genomic DNA. Do not use Nanodrop for this purpose.
4. Short fragments may result in lower sequencing throughput due to inefficient pore utilization. However, size selection will result in loss of DNA. The amount of DNA recovered is dependent on technique used for size selection and should first be tested using a nonprecious DNA sample. Proceed with size selection only if fragments shorter than 1 kb are present and if genomic DNA sample is in excess.
5. As of guppy version 4.5.2+ fastq files will be automatically split into pass/fail folders based on quality, where the "High Accuracy" model (HAC) will use average read quality of Q9 or higher for pass, and the "fast" model will use Q7 or higher for pass.

6. The flow cell and kit used must be specified during base-calling. User can specify the flow cell and kit using the parameters `--flowcell <flowcell name> --kit <kit name>` or by specifying the configuration file associated with each flow cell/kit combination using the parameter `-c <config file>`. To view the list of possible flow cell/kit combinations and associated configuration files, type the following command: `guppy_basecaller --print_workflows`.
7. Unmethylated DNA from Enterobacteria phage λ is added during library preparation as a spike-in control to determine the conversion rate of sodium bisulfite. The nonconversion rate is calculated in Subheading 3.4.3. Lambda reads are removed in the first step of Subheading 3.5.
8. Methylation data in this chapter is 0-based for compatibility with UCSC tools.

References

1. Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328: 916–919
2. Schübeler D (2015) Function and information content of DNA methylation. Nature 517: 321–326
3. Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14:204–220
4. Greenberg MVC, Bourch'is D (2019) The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol 20:590–607
5. Lister R, Pelizzola M, Dowen RH et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322
6. Stadler MB, Murr R, Burger L et al (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480: 490–495
7. Neri F, Rapelli S, Krepelova A et al (2017) Intragenic DNA methylation prevents spurious transcription initiation. Nature 543:72–77
8. Urich MA, Nery JR, Lister R et al (2015) MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. Nat Protoc 10:475–483
9. Ji L, Sasaki T, Sun X et al (2014) Methylated DNA is over-represented in whole-genome bisulfite sequencing data. Front Genet 5:341
10. Olova N, Krueger F, Andrews S et al (2018) Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. Genome Biol 19:33
11. Simpson JT, Workman RE, Zuzarte PC et al (2017) Detecting DNA cytosine methylation using nanopore sequencing. Nat Methods 14: 407–410
12. Rand AC, Jain M, Eizenga JM et al (2017) Mapping DNA methylation with high-throughput nanopore sequencing. Nat Methods 14:411–413
13. Jain M, Koren S, Miga KH et al (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 36:338–345
14. Amarasinghe SL, Su S, Dong X et al (2020) Opportunities and challenges in long-read sequencing data analysis. Genome Biol 21:30
15. Feng S, Cokus SJ, Zhang X et al (2010) Conservation and divergence of methylation patterning in plants and animals. Proc Natl Acad Sci U S A 107:8689–8694
16. Bogdanović O, Smits AH, de la Calle ME et al (2016) Active DNA demethylation at enhancers during the vertebrate phylotypic period. Nat Genet 48:417–426
17. Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34: 3094–3100

18. Li H, Handsaker B, Wysoker A et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079
19. Kent WJ, Zweig AS, Barber G et al (2010) BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 26: 2204–2207
20. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120
21. Chen H, Smith AD, Chen T (2016) WALT: fast and accurate read mapping for bisulfite sequencing. Bioinformatics 32:3507–3509
22. Ramírez F, Ryan DP, Grüning B et al (2016) deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44:W160–W165

Part II

Protein-DNA Interactions

Chapter 6

Chromatin Immunoprecipitation Sequencing (ChIP-seq) Protocol for Small Amounts of Frozen Biobanked Cardiac Tissue

Jiayi Pei, Noortje A. M. van den Dungen, Folkert W. Asselbergs, Michal Mokry, and Magdalena Harakalova

Abstract

Chromatin immunoprecipitation and sequencing (ChIP-seq) is a well-established method to study the epigenetic profile at the genome-wide scale, including histone modifications and DNA–protein interactions. It provides valuable insights to better understand disease mechanisms. Here we present an optimized ChIP-seq protocol suitable for human cardiac tissues, especially the frozen biobanked small biopsy samples.

Key words Chromatin immunoprecipitation, Sequencing, Small biopsy, Cardiac tissues, Antibody, Promoters, Enhancers

1 Introduction

Chromatin immunoprecipitation (ChIP) is commonly used to study protein–DNA interactions and it was traditionally coupled with quantitative real-time polymerase chain reaction (PCR) to study the enriched DNA regions, which had several limitations, such as long amplicon, primer-related bias, and poor primer integrity [1]. ChIP interrogated with microarrays or ChIP-on-chip was developed to address PCR-related issues, however, it has other drawbacks, including the available microarray probes on the chip and the normalization of the array data [2]. ChIP followed by high-throughput sequencing (ChIP-seq) is adapted from ChIP-on-chip and provides better coverage, fewer artifacts, and higher resolution, which subsequently improve the quality of the data [3]. Nowadays, ChIP-seq is commonly used to identify the epigenetic signature at the genome-wide scale, which is a hallmark of the disease [4–6]. Antibodies targeting different histone marks, such as H3 lysine 27 acetylation (H3K27ac), H3 lysine 4 trimethylation

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_6,

(H3K4me3), and H3 lysine 36 trimethylation (H3K36me3), are used in ChIP-seq to capture specific DNA regions (i.e., enhancers and promoters) [7]. Valuable information can be obtained, including motif enrichment analysis to study transcription factors that are likely to bind to those regions, gene expression prediction, and gene-enhancer interaction analysis [8]. Thanks to rapidly developed single-cell assays, in recent years single-cell ChIP-seq is highly used to obtain the cell-specific epigenetic profiles from the bulk samples [9–11].

Thus far, ChIP-seq has successfully revealed the pathological mechanisms and molecular factors underlying the diseases. For example, the genome-wide DNA replication profile in cultured human colon cancer cells was profiled by ChIP-seq [12]. Key transcription factors in human tumor cells obtained by needle biopsy were also identified using ChIP-seq coupled with the H3K27ac antibody [13]. ChIP-seq coupled with H4K16ac antibody identified normal ageing dependent chromatin modification using postmortem human brain samples from cognitively normal younger individuals, cognitively normal elder individuals, and individuals with Alzheimer's disease [14]. A study used ChIP-seq and showed transcriptional enhancers involved in heart development and function using fresh human fetal and adult hearts [15]. Another study used ChIP-seq to further reveal affected key enhancers and promoters and their enriched biological functions in failing human hearts with dilated cardiomyopathy when compared to the controls [16]. We also employed ChIP-seq coupled with the H3K27ac antibody and revealed the epigenomic reorganization in remodelled human hearts due to severe aortic stenosis when compared to the controls [17]. We further demonstrated that histone acetylation profiles correlated with the transcriptome profile per sample-wise and identified a set of promising transcription factors as potential key upstream regulators during the myocardial remodeling. Besides the bulk level, Churko and colleagues identified a set of DNA regions and nearby genes that interacted with transcription factors of interest using ChIP-seq in human induced pluripotent stem cell-derived cardiomyocytes [18]. Another study further isolated cardiomyocyte nuclei from diseased and nonfailing hearts and obtained their epigenetic profiles using ChIP-seq coupled with antibodies targeting H3K27ac, H3K9ac, H3K4me3, H3K36me3, H3K9me3, H3K27me3, and H3K4me1, respectively [19].

Piling studies are now integrating ChIP-seq data with other data, such as whole-genome data and transcriptome data, to identify novel and key regulatory factors that drive the pathological mechanism of the disease [20, 21]. Nevertheless, several challenges of incorporating ChIP-seq to study heart failure remain and are well summarized in our review [22]. Thus, it is important to have an established ChIP-seq protocol that ensures and/or improves the quantity and quality of the captured DNA fragments from cardiac

tissues in the first place. Here, we present a well-optimized ChIP-seq protocol from the manufacturer's instructions for small human cardiac tissues, including the handling of cardiac samples and the visualization of amplified DNA fragments on a gel.

2 Materials

Prepare all solutions using ultrapure water and analytical grade reagents. Store all reagents at room temperature, unless otherwise indicated.

2.1 Chemicals and Reagents

1. 37% formaldehyde.
2. 1.25 M glycine (store at 4 °C).
3. Cold phosphate buffered saline (PBS).
4. Dynabeads® (store at 4 °C).
5. NEXTFLEX® Cleanup Beads 2.0 (store at 4 °C).
6. Reverse Cross-linking Buffer (store at 4 °C).
7. 20 mg/mL Proteinase K (store at 4 °C).
8. MAGnify™ IP Buffer 1 (store at 4 °C).
9. MAGnify™ IP Buffer 2 (store at 4 °C).
10. MAGnify™ DNA Wash Buffer (store at 4 °C).
11. MAGnify™ DNA Elution Buffer (store at 4 °C).
12. MAGnify™ DNA Purification Buffer are stored at 4 °C (store at 4 °C).
13. MAGnify™ Protease Inhibitors (200×) (store at 4 °C).
14. MAGnify™ Dilution Buffer (store at 4 °C).
15. MAGnify™ Lysis Buffer (store at 4 °C).
16. NEXTFLEX® End Repair & Adenylation Buffer Mix 2.0 (store at −20 °C).
17. NEXTFLEX® End Repair & Adenylation Enzyme Mix 2.0 (store at −20 °C).
18. NEXTFLEX® Ligase Buffer Mix 2.0 (store at −20 °C).
19. NEXTFLEX® Ligase Enzyme 2.0 (store at −20 °C).
20. NEXTFLEX® PCR Master Mix 2.0 (store at −20 °C).
21. NEXTFLEX® Primer Mix 2.0 (store at −20 °C).
22. Resuspension Buffer.
23. Nuclease-free water.
24. 100% ethanol.
25. 80% ethanol.
26. Zymo ChIP DNA Clean & Concentrator Kit.

2.2 Working Buffers

1. Lysis Buffer: 74.625 μL of Lysis Buffer stock added with 0.375 μL of 200× Protease Inhibitors per sample.
2. Dilution Buffer: 199 μL of Dilution Buffer stock added with 1 μL of 200× Protease Inhibitors per sample.
3. Diluted Dynabeads® Buffer: 50 μL of Dilution Buffer added with 10 μL of resuspended Dynabeads® per sample.
4. Reverse Cross-linking Solution: 53 μL of Reverse Cross-linking Buffer added with 1 μL of Proteinase K per sample.
5. Purification Buffer: add 5 volumes of DNA Binding Buffer to 1 volume of sample.

2.3 Equipment

1. Covaris System or other methods for shearing chromatin.
2. Adhesive polymerase chain reaction (PCR) Plate Seal.
3. Magnetic holder.
4. Thermocycler.
5. Nuclease-free barrier pipette tips.
6. 1.5 mL LoBind Eppendorf tubes.

3 Methods

Use 1.5 mL LoBind Eppendorf tubes during the whole procedure from Subheadings 3.1 to 3.9. Use sterile filter tips during the entire procedure and make sure to always change tips for each sample to avoid possible contamination. Make sure to first vortex shortly and spin down every single tube used in the protocol, including tubes with samples and chemicals.

3.1 Preparing Tissue and Cross-Linking the Chromatin

1. Set the cryostat to −20 °C and cut the frozen cardiac tissue at the thickness of 10 μm (Fig. 1). Collect tissues (approximately 10 slides at the size of 2 cm × 2 cm × 10 μm) per sample in a sterile 1.5 mL Eppendorf tube and store on dry ice (*see* **Note 1**).
2. In the fume hood, add 21 μL of 37% formaldehyde for every 25 mg of tissue at room temperature.
3. Swirl the tube gently to mix and incubate for exactly 10 minutes (min) at room temperature. Swirl the tube gently every 2 min during incubation.
4. Add 80 μL of 1.25 M glycine for every 25 mg of tissue for a final concentration of 0.125 M.
5. Swirl the tube gently to mix evenly and incubate for 5 min at room temperature, swirl gently every 2 min during incubation.
6. In a cold centrifuge at 4 °C, spin the tubes at 600 × *g* for 10 min. A pellet will form against the sidewall of the tube.

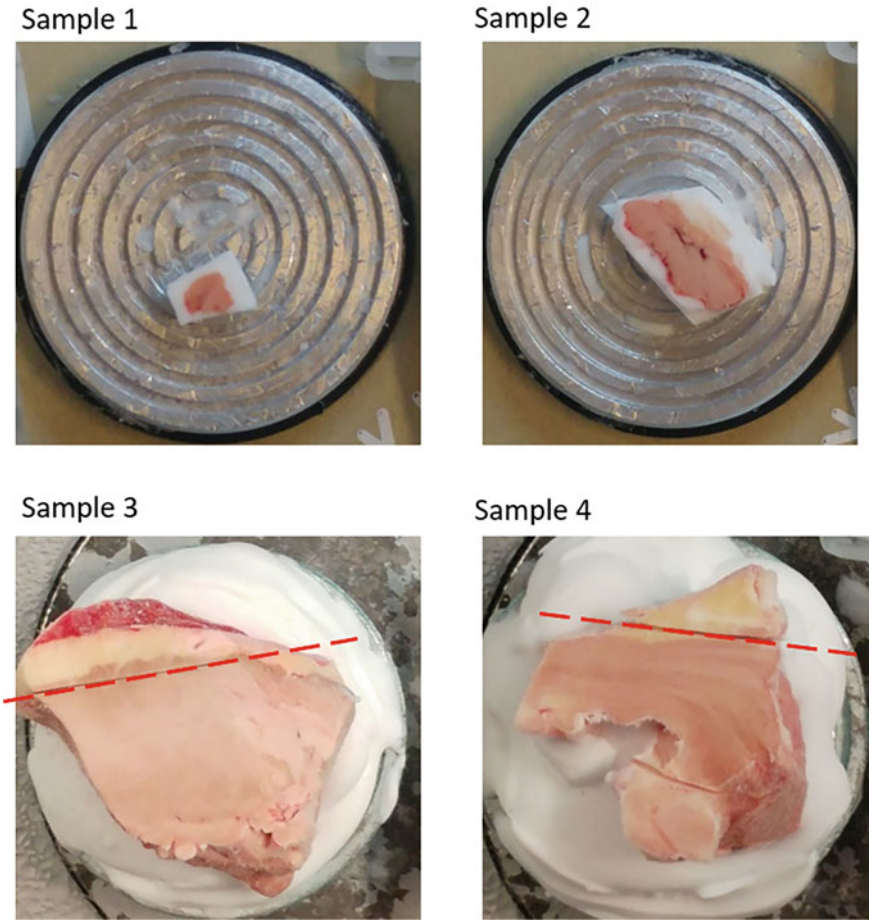


Fig. 1 Examples of snap-frozen human cardiac samples. Samples were sliced at a thickness of 10 μm. However, the number of sectioned slides ranged from 10 to more than 50, depending on the sample size. At least 80 slides of sample 1, around 50 slides of sample 2, and around 15 slides of samples 3 and 4 are estimated to provide an adequate amount of cells for the following experiment. Since cell compositions change during the disease progression, it is advised to use a sterile scalpel and remove the fibrotic and/or fatty region of the cardiac tissue, which subsequently maximizes the chromatin signals derived from cardiomyocytes. Red dashed lines in samples 3 and 4 indicate the possible cutting line by the scalpel

7. Transfer the tubes to ice and keep them on ice for all subsequent steps.
8. Remove and discard the supernatant from each tube, leaving around 30 μL behind to be sure not to disturb the pellet.
9. Add 500 μL of cold PBS to each tube and resuspend the sample by flicking it with your fingers.
10. Spin at 600 × *g* for 10 min at 4 °C.
11. Aspirate the PBS and resuspend the sample once more in 500 μL of cold PBS.

12. Spin at 600 × *g* for 10 min at 4 °C.
13. Aspirate the PBS as much as possible without disturbing the cell pellet.

3.2 Lysing the Cells

1. Prepare 75 μL of Lysis Buffer containing Protease Inhibitors per sample (*see* **Note 2**). For example, (74.625 μL of Lysis Buffer stock + 0.375 μL of 200× concentrated Protease Inhibitors) × (number of samples + 1). Vortex the mixture well to resuspend.
2. Add 75 μL of working Lysis Buffer to each cell pellet from Subheading 3.1.
3. Resuspend by mild pulses on the vortex mixer.
4. Incubate the tubes on ice for at least 5 min.
5. Proceed to Subheading 3.3 or snap-freeze the samples in liquid nitrogen and store at −80 °C until use.

3.3 Shearing the Chromatin

1. Switch on the Covaris 2.5 h before using it (*see* **Note 3**) and set the settings using the S-Series SonoLab Single Software (v2.5) as the following: Duty Cycle: 5%; Cycles: 10; Intensity: 2; Temperature (bath): 4 °C; Cycles per Burst: 200; Power Mode: Frequency Sweeping; Cycle Time: 60 s; Degassing Mode: Continuous (*see* **Note 4**).
2. Take the lid off a Covaris® MicroTube with AFA fiber and transfer 75 μL of the cell lysate including any remaining tissue from Subheading 3.2 into the MicroTube.
3. Insert the tube into a Covaris®-2 series M220 Machine Holder for one 6 mm tube.
4. Sonicate the sample using the settings shown above. The liquid inside the tube should look milky afterward (Fig. 2).
5. Spin the tube briefly (around 5 s) to bring the samples to the bottom of the tube and transfer the freshly sonicated chromatin into a new, sterile 1.5 mL microcentrifuge tube.
6. Centrifuge the tubes at 20,000 × *g* at 4 °C for 5 min.
7. Carefully collect the supernatant that contains the chromatin and transfer it into a new sterile 1.5 mL tube (*see* **Note 5**). Discard the original tubes containing the debris and keep the new tubes on ice until further use.
8. Proceed to Subheading 3.4 or snap-freeze the samples in liquid nitrogen and store at −80 °C until use.

3.4 Diluting the Chromatin

1. Prepare 200 μL of working Dilution Buffer containing protease inhibitors per sample (*see* **Note 2**). For example, (199 μL of dilution buffer stock + 1 μL of 200× concentrated protease inhibitors) × (number of samples + 1).

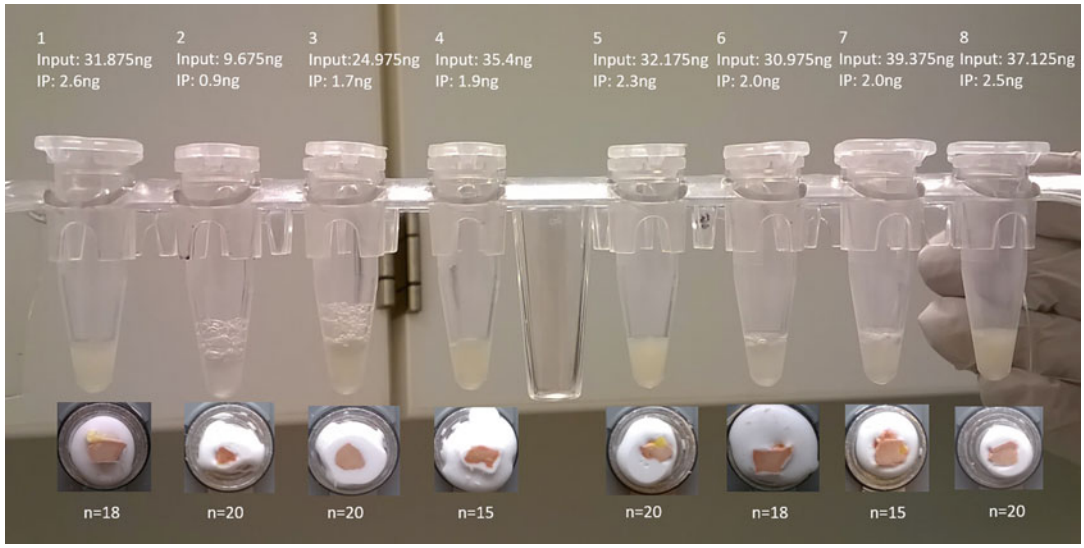


Fig. 2 Examples of sonicated chromatin samples from eight hearts. The DNA concentrations of the input control without coupling to the antibody–Dynabeads® complex and the chromatin immunoprecipitated sample from the same heart were measured by Qubit and shown in the upper panel. The size of each heart and the included number of slides per heart are shown in the lower panel

2. Dilute the sheared chromatin in the cold working buffer by adding 180 μL of the cold working buffer with 20 μL of the sheared chromatin sample to reach the final dilution volume of 200 μL per sample.
3. Of the diluted samples, 100 μL is used for the actual immunoprecipitation reaction and the remaining 100 μL may be used as the input control (required volume = 10 μL).

3.5 Coupling the Antibody to Dynabeads®

1. Briefly spin down the tube containing Dynabeads® and resuspend them by gently pipetting up-and-down several times (*see* **Note 6**) to make sure there are no bead sediments in the bottom of the tube.
2. For each antibody that is going to be used, prepare a separate master mix of cold working dilution buffer with Dynabeads® at the ratio of 5:1. The final volume of the master mix per sample per antibody is 60 μL. For example, (50 μL of the working dilution buffer + 10 μL resuspended Dynabeads®) × (number of samples + 1).
3. Place the tube(s) in the magnetic holder and wait until the beads form a tight pellet.
4. Remove the liquid from the tubes on the magnetic holder (*see* **Notes 7–10**).
5. Add 1 μL of an antibody of interest (1 μg/μL) per sample, such as antibodies targeting Histone acetyl K27 or RNA Polymerase

II. If there are problems with the chromatin immunoprecipitations, the amount of antibody might be increased up to 10 μg.

6. Close the tubes and flick gently to resuspend the beads.
7. Rotate the tubes end-over-end at 4 °C until use in **step 5** of Subheading 3.6.

3.6 Binding Chromatin to the Beads

1. Divide the antibody–Dynabeads® mixture (51 μL per tube) over an appropriate amount of 1.5 mL tubes.
2. Spin the tubes briefly to bring all liquid down to the bottom of the tube and place the tubes in the magnet holder.
3. Let stand for at least 30 s or until the beads form a tight pellet.
4. Remove and discard the liquid from the tubes without disturbing the bead pellet.
5. Remove the tubes from the magnet holder and immediately add 100 μL of diluted chromatin samples from **step 7** in Subheading 3.5 to each tube containing the antibody–Dynabeads® complex.
6. Close the tubes and flick gently to resuspend the beads.
7. Rotate the tubes end-over-end at 4 °C for 2 h.

3.7 Washing the Bound Chromatin

1. Spin the tubes briefly to bring any of the liquid trapped in the cap to the bottom of the tube and place the tubes back to the magnetic holder.
2. Let stand for at least 30 s or until the beads form a tight pellet.
3. Remove and discard the liquid from the tubes without disturbing the bead pellet.
4. Remove the tubes from the magnet holder and add 100 μL of IP Buffer 1 to each tube (*see* **Note 11**).
5. Close the tubes and flick gently to resuspend the beads.
6. Wash the beads by taking the tubes out of the magnetic holder, turn them around and place them back in the holder a couple of times.
7. Repeat **steps 1–6** two more times and always use new 1.5 mL tubes after each round of washing.
8. Spin the tubes briefly to bring any of the liquid trapped in the cap to the bottom of the tube and place the tubes back in the magnetic holder.
9. Let stand for at least 30 s or until the beads form a tight pellet.
10. Remove and discard the liquid from the tubes without disturbing the bead pellet.
11. Remove the tubes from the magnet holder and add 100 μL of IP Buffer 2 to each tube.

12. Close the tubes and flick gently to resuspend the beads.
13. Put the tubes on ice for 5 min and gently flick them a couple of times.
14. Repeat **steps 8–13** one more time.

3.8 Reversing the Cross-Linking

1. For each IP sample, prepare 54 μL of reverse cross-linking buffer containing proteinase K (*see* **Note 12**). For example, (53 μL of reverse cross-linking buffer stock + 1 μL of proteinase K) × (number of samples + 1). Vortex briefly to mix well and keep the working buffer at room temperature until use.
2. For each input control sample from **step 3** in Subheading 3.4, which was not coupled to the antibody–Dynabeads® complex, add 43 μL of reverse cross-linking buffer containing proteinase K to 10 μL of the input control sample.
3. Place the tubes from Subheading 3.7, **step 14** in the magnetic holder and wait at least 30 s or until a pellet forms.
4. Remove and discard the liquid from the tubes without disturbing the bead pellet.
5. Remove the tubes from the magnetic holder and add 54 μL of reverse cross-linking buffer containing proteinase K to each tube. Vortex lightly to fully resuspend the beads.
6. Incubate the IP sample tubes and the input control tubes at 55 °C for 15 min in a water bath or other heating source of choice.
7. Spin the tubes briefly.
8. Place the IP sample tubes in the magnetic holder and wait at least 30 s for a pellet to form.
9. Carefully transfer the liquid to a new, sterile 1.5 mL tube. Do not discard the liquid, it contains the IP sample.
10. Spin the IP sample tubes and input control tubes briefly, then incubate them at 65 °C for 15 min in a water bath or other heating source of choice.
11. Cool down the tubes on ice for 5 min (*see* **Note 13**).

3.9 Purifying the DNA

1. Purify the samples with the ChIP DNA Clean & Concentrator Kit by adding 5 volumes of DNA binding buffer to each sample. For example, 250 μL of DNA binding buffer + 50 μL of the sample.
2. Transfer the mixture of sample and binding buffer to a Zymo Spin column on top of a collection tube.
3. Centrifuge at 10,000 × *g* for 30 s and discard the flow-through.
4. Add 200 μL of washing buffer to the column.

5. Centrifuge at 10,000 × *g* for 30 s and discard the flow-through.
6. Repeat **steps 4** and **5**.
7. Add 20 μL of elution buffer directly to the column matrix and transfer the column to a new, sterile 1.5 mL tube.
8. Centrifuge at 10,000 × *g* for 30 s.
9. Discard the column, label the tubes properly and store them at −20 °C until use.

3.10 End-Pair and Adenylation of the Purified DNA Fragments

1. Measure DNA concentrations in each sample from Subheading 3.9 and start by taking 1–2 μL per sample (*see* **Note 14**).
2. In nuclease-free 96-well PCR strip tubes, add 7.5 μL of NEXTflex™ End-Repair & Adenylation Buffer Mix, 1.5 μL of NEXTflex™ End-Repair & Adenylation Enzyme Mix, 10 ng of purified DNA samples (*see* **Note 15**), and nuclease-free water to reach the final volume of 25 μL.
3. Incubate the samples on a thermocycler using the following program: 22 °C for 20 min, 72 °C for 20 min, and an infinite 4 °C hold.

3.11 Adaptor Ligation

1. Thaw NEXTflex™ Ligase Enzyme Mix to room temperature and vortex for 5–10 s (*see* **Note 16**).
2. For 25 μL of each end-repaired DNA sample from Subheading 3.10, add 33.75 μL of NEXTflex™ Ligase Enzyme Mix (*see* **Notes 17** and **18**) and 1.5 μL of diluted NEXTflex™ DNA Barcode (*see* **Note 19**).
3. Mix thoroughly by pipetting (*see* **Note 20**).
4. Apply adhesive PCR plate seal and incubate at room temperature for 15 min or on a thermocycler at 22 °C for 15 min (*see* **Note 21**).

3.12 Cleanup

1. In case strip tubes are used, transfer the samples from Subheading 3.11 to 1.5 mL tubes.
2. Add 40 μL of AMPure XP Beads to each sample and mix thoroughly until homogenized.
3. Incubate samples at room temperature for 5 min.
4. Place the tubes on the magnetic holder at room temperature for 5 min.
5. Remove and discard the supernatant.
6. Add 200 μL of 80% ethanol to each bead pellet and incubate at room temperature for 30 s.
7. Carefully remove ethanol without touching the bead pellet.

8. Remove the tubes from the magnetic holder and let dry at room temperature till seeing the cracks on the pellet (*see* **Note 22**).
9. Add 50 μL of Resuspension Buffer to the dried beads and mix thoroughly until homogenized.
10. Incubate samples at room temperature for 5 min.
11. Place the tubes on the magnetic holder at room temperature for 5 min.
12. Collect the clear supernatant and transfer them to new tubes.
13. Repeat **steps 2–8**.
14. Add 22 μL of Resuspension Buffer to the dried beads and mix thoroughly until homogenized.
15. Incubate samples at room temperature for 5 min.
16. Place the tubes on the magnetic holder at room temperature for 5 min.
17. Transfer 20 μL of clear sample to a new tube.

3.13 PCR Amplification

1. Add 51 μL of nuclease-free water, 25 μL NEXTflex™ PCR Master Mix, and 4 μL of NEXTflex™ Primer Mix to 20 μL of the ligated DNA sample from Subheading 3.12 to reach the final volume of 100 μL.
2. Apply adhesive PCR plate seal and incubate the samples on a thermocycler using the following program: 98 °C for 2 min, 13 cycles of (98 °C for 30 s, 65 °C for 30 s, 72 °C for 60 s), 72 °C for 4 min, and an infinite 4 °C hold (*see* **Note 23**).
3. Take 4 μL of the sample and mix it with 1 μL of dye and load the mixture to the gel and check the amplified products (Fig. 3).
4. Add 80 μL of AMPure XP beads to each sample (beads–sample = 8:10), mix thoroughly until homogenized.
5. Place the tubes on the magnetic holder and incubate at room temperature for 5 min.
6. Remove and discard the supernatant without disturbing the bead pellets.
7. Add 200 μL of 80% ethanol to each pellet and incubate at room temperature for 30 s.
8. Carefully remove the ethanol without disturbing the bead pellets.
9. Remove the tubes from the magnetic holder and let dry at room temperature for 5 min or till seeing the cracks on the pellet.
10. Add 21 μL of Resuspension Buffer to dried pellets and mix thoroughly until homogenized.

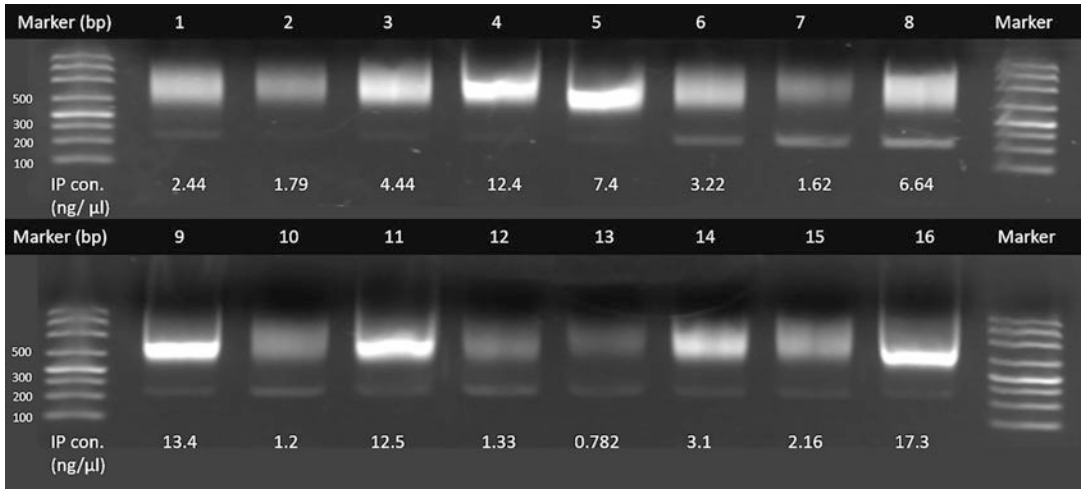


Fig. 3 Examples of checking amplified DNA products of 16 samples after 13 cycles of polymerase chain reaction cycles. The sample concentration per chromatin immunoprecipitated (IP) sample is shown below each band

11. Place the tubes on the magnetic holder and incubate at room temperature for 5 min.
12. Transfer 20 µL of clear sample to a new tube.
13. Optional: check the final product on a gel.
14. Measure the concentrations of the final products using Qubit or other suitable methods (*see* **Note 14**).

4 Notes

1. Additional slides from the same cardiac sample could be collected for other omics experiments, such as RNA sequencing and proteomics, which will provide extra layers of information next to ChIP-seq.
2. When calculating the volume of needed chemicals, that is, for the preparation of the master mixes, always count 1 extra sample due to the pipetting error and avoid introducing batch effects of preparing additional chemicals.
3. Covaris machine for shearing the chromatin takes around 2.5 h to function, so make sure to switch on the machine beforehand to fit the experiment schedule.
4. Make sure the degas button is on.
5. Label the tubes properly to distinguish collected samples throughout the protocol clearly, that is, sheared chromatin samples in Subheading 3.3, chromatin input controls in Subheading 3.4, and purified DNA in Subheading 3.9.

6. Coupling the antibodies to the Dynabeads takes about 1.5 h, so make sure the chromatin isolation is ready before preparing the antibody–Dynabeads mixture.
7. Do not vortex and freeze the Dynabeads®, this will damage the beads.
8. Avoid touching the beads with the pipette tip when removing the liquid from the Dynabeads®, because this will disturb the bead pellet.
9. Do not allow the beads to dry out and make sure to resuspend the beads within 1 min after removing the liquid from them.
10. It is acceptable to remove the tubes from the magnetic holder during the washing steps in Subheading 3.5.
11. During the washing steps in Subheading 3.7, keep the magnets, tubes, and buffers on ice to keep cold.
12. A considerable amount of fibrotic tissue is present in the cardiac samples, especially in diseased hearts. Therefore, it is important to incubate samples with proteinase K to digest muscle fibers [23, 24].
13. It is important to proceed from **steps 6** to **11** in Subheading 3.8 without breaks to make sure the reverse cross-linking reaction occurs without sacrificing the sample qualities.
14. We recommend measuring DNA concentration using the Qubit Kit rather than Nanodrop for higher accuracy. For input samples or samples with relatively more starting tissues than the others, it is advised to take 1 μL of sample and diluted in nuclease-free water at a 1:1 ratio for concentration measurements. For the others, it is advised to start with 2 μL of each sample for concentration measurements.
15. The working range of purified DNA may range from 1 ng to 1 μg, but it needs to be the same among samples.
16. Do not spin down the NEXTflex™ Ligase Enzyme Mix to avoid separating the components of the mix and affect the performance.
17. Depending on the input amount and the starting adapter concentration, dilute NEXTflex™ Adaptor with nuclease-free water.
18. It is recommended to add 1.25 μM Adaptor to 1–10 ng of DNA samples, 3 μM Adaptor is desired for 100 ng of DNA samples, and 25 μM Adaptor is desired for 250–1000 ng of DNA samples.
19. Each sample needs a specific barcode. Do not add the same barcode to different DNA samples in the same experiment.
20. The NEXTflex™ Ligase Enzyme Mix is very viscous. Mixing the DNA samples, NEXTflex™ Ligase Enzyme Mix, and

NEXTflex™ DNA Barcode thoroughly is critical to obtain optimal results. It is suggested to pipette the mixture up and down 15 times.

21. Set the lid temperature at 37 °C on the thermocycler.
22. It takes 1–3 min to dry and see the cracks appearing on the bead pellet.
23. In the PCR program, start with 13 cycles and check the amplified product on the gel. If the band is too weak, extra PCR cycles could be added. However, it is not advised to add more than 10 cycles, which may introduce PCR-biased effects to the samples.

References

1. Mukhopadhyay A, Deplancke B, Walhout AJM, Tissenbaum HA (2008) Chromatin immunoprecipitation (ChIP) coupled to detection by quantitative real-time PCR to study transcription factor binding to DNA in Caenorhabditis elegans. Nat Protoc 3:698
2. Pellegrini M, Ferrari R (2012) Epigenetic analysis: ChIP-chip and ChIP-seq. In: Next generation microarray bioinformatics. Humana Press, Totowa, NJ, pp 377–387
3. Park PJ (2009) ChIP-Seq: advantages and challenges of a maturing technology. Nat Rev Genet 10:669
4. Bonifer C, Cockerill PN (2011) Chromatin mechanisms regulating gene expression in health and disease. Adv Exp Med Biol 711: 12–25
5. Kundu TK, Dasgupta D (2007) Chromatin and disease. Springer Science & Business Media, New York, NY
6. Landt SG, Marinov GK, Kundaje A et al (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 22:1813
7. Busby M, Xue C, Li C et al (2016) Systematic comparison of monoclonal versus polyclonal antibodies for mapping histone modifications by ChIP-seq. Epigenetics Chromatin 9:49. https://doi.org/10.1186/s13072-016-0100-6
8. Nakato R, Sakata T (2020) Methods for ChIP-seq analysis: a practical workflow and advanced applications. Methods 187:44. https://doi.org/10.1016/j.ymeth.2020.03.005
9. Rotem A, Ram O, Shoresh N et al (2015) Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. Nat Biotechnol 33:1165–1172
10. Grosselin K, Durand A, Marsolier J et al (2019) High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. Nat Genet 51:1060–1066
11. Behjati Ardakani F, Kattler K, Heinen T et al (2020) Prediction of single-cell gene expression for transcription factor analysis. GigaScience 9:giaa113. https://doi.org/10.1093/gigascience/giaa113
12. Kobunai T, Matsuoka K, Takechi T (2019) ChIP-seq analysis to explore DNA replication profile in trifluridine-treated human colorectal cancer cells in vitro. Anticancer Res 39:3565–3570
13. Singh AA, Schuurman K, Nevedomskaya E et al (2019) Optimized ChIP-seq method facilitates transcription factor profiling in human tumors. Life Sci Alliance 2:e201800115. https://doi.org/10.26508/lsa.201800115
14. Nativio R, Donahue G, Berson A et al (2018) Dysregulation of the epigenetic landscape of normal aging in Alzheimer's disease. Nat Neurosci 21:497–505
15. May D, Blow MJ, Kaplan T et al (2012) Large-scale discovery of enhancers from human heart tissue. Nat Genet 44:89
16. Liu C-F, Abnousi A, Bazeley P et al (2020) Global analysis of histone modifications and long-range chromatin interactions revealed the differential cistrome changes and novel transcriptional players in human dilated cardiomyopathy. J Mol Cell Cardiol 145:30–42
17. Pei J, Harakalova M, Treibel TA et al (2020) H3K27ac acetylome signatures reveal the epigenomic reorganization in remodeled non-failing human hearts. Clin Epigenetics 12:1–18

18. Churko JM, Garg P, Treutlein B et al (2018) Defining human cardiac transcription factor hierarchies using integrated single-cell heterogeneity analysis. Nat Commun 9:1–14
19. Gilsbach R, Schwaderer M, Preissl S et al (2018) Distinct epigenetic programs regulate cardiac myocyte development and disease in the human heart in vivo. Nat Commun 9: 391. https://doi.org/10.1038/s41467-017-02762-z
20. Gusev FE, Reshetov DA, Mitchell AC et al (2019) Epigenetic-genetic chromatin footprinting identifies novel and subject-specific genes active in prefrontal cortex neurons. FASEB J 33:8161–8173
21. Cherry TJ, Yang MG, Harmin DA et al (2020) Mapping the cis-regulatory architecture of the human retina reveals noncoding genetic variation in disease. Proc Natl Acad Sci U S A 117: 9001. https://doi.org/10.1073/pnas.1922501117
22. Harakalova M, Asselbergs FW (2018) Systems analysis of dilated cardiomyopathy in the next generation sequencing era. Wiley Interdiscip Rev Syst Biol Med 10:e1419
23. Kikuchi A, Naruse A, Sawamura T, Nonaka K (2020) A high-temperature, pre-incubation step before proteinase K treatment notably improves recovery of genomic DNA in formalin-fixed, paraffin-embedded tissue samples. Clin Lab 66. https://doi.org/10.7754/Clin.Lab.2020.200432
24. Ford KL, Anwar M, Heys R et al (2019) Optimisation of laboratory methods for whole transcriptomic RNA analyses in human left ventricular biopsies and blood samples of clinical relevance. PLoS One 14:e0213685

Chapter 7

A Robust Protocol for Investigating the Cohesin Complex by ChIP-Sequencing

Macarena Moronta Gines and Kerstin S. Wendt

Abstract

The investigation of cohesin binding sites throughout different mammalian genomes by ChIP-sequencing has been fundamental to discover how cohesin and CTCF collaborate to form chromatin loops and to gain insight in the intricate regulation of cohesin. Here we describe a detailed ChIP protocol that has been successfully used for different cohesin subunits and cohesin regulators in various cell lines.

Key words Cohesin, ChIP, Chromatin immunoprecipitation

1 Introduction

The cohesin complex has crucial functions for the 3D-organization of chromatin, transcriptional regulation, DNA damage repair and cell division (for a general review *see* [1, 2]). Malfunctions of the complex or its regulators have been linked to developmental syndromes and cancer. Interrogating the localization and occupancy of cohesin binding sites in the genome has been fundamental for our current understanding of the complex and how it shapes the 3D organization of mammalian genomes.

The most effective tool for investigating genomic binding sites of cohesin is ChIP-sequencing. Cohesin binding sites have been mapped with very different ChIP-sequencing protocols, for example, native or SDS-containing protocols, in a wide range of cell types in different species. Here, we describe a general protocol for performing cohesin ChIP-sequencing, with focus on the ChIP part of the protocol since preparation of the sequencing libraries, next-generation sequencing and the downstream analysis depend on the available sequencing platform. If suitable, we point to alternatives

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_7,

Table 1
List of commercial antibodies used to ChIP the cohesin complex in mammalian cells

Subunit	Species	Antibody	Publication
RAD21	HeLa, HCT116, MCF7, HEPG2, mESCs, CH12, adult mouse hepatocytes	Abcam #ab992	Wendt et al. [3], Schwarzer et al. [4], Rao et al. [5], Pugacheva et al. [6]
RAD21	mESCs	Abcam #ab154769	Hansen et al. [7]
SMC1	HCT116	Bethyl #A300-055A	Rao et al. [5]
SMC3	adult mouse hepatocytes	Abcam #ab9263	Schwarzer et al. [4]
SMC3	Hela	Bethyl #A300-060A	Wutz et al. [8]
SA1/STAG1	HSCs	Bethyl #A302-579A	Viny et al. [9]
SA2/STAG2	HSCs	Bethyl #A302-580A	Viny et al. [9]
WAPL	MEF	Peters laboratory ID A960[a]	Busslinger et al. [10]

[a]Although not commercially available, we include this antibody since there is no commercial antibody described

in the protocol that we have tested in the past. In addition, we provide an overview about the different commercial antibodies that have been used (Table 1).

Cohesin is a very abundent protein and only a fraction of the complexes are actually chromatin-bound [11]. We therefore cannot expect to quantitatively precipitate all cohesin complexes in ChIP experiments. It might be plausible to reduce the unbound pool, as done in the ChIP-seq protocols published by other groups [12].

The protocol we describe here was designed to investigate changes of cohesin site occupancy in knockdown experiments for cohesin regulators [3, 13]. To correctly monitor reduced cohesin on chromatin, for example when cohesin-loading factors are depleted but the overall abundance of cohesin is not changed, we decided against depleting the soluble cohesin pool from the nuclei. This protocol has been used to perform ChIP for different subunits and regulators of the cohesin comples (SMC3, SMC1A, STAG1, STAG2, and NIPBL) in different cell lines (HeLa, HB2, HCT116, HEK293T, LCLs, and cultured human skin fibroblasts) (Fig. 1).

2 Materials and Equipment

Prepare all solutions using ultrapure water (MilliQ) and analytical grade reagents. Protease inhibitors and PMSF need to be added to the lysis buffer and wash buffers immediately before use.

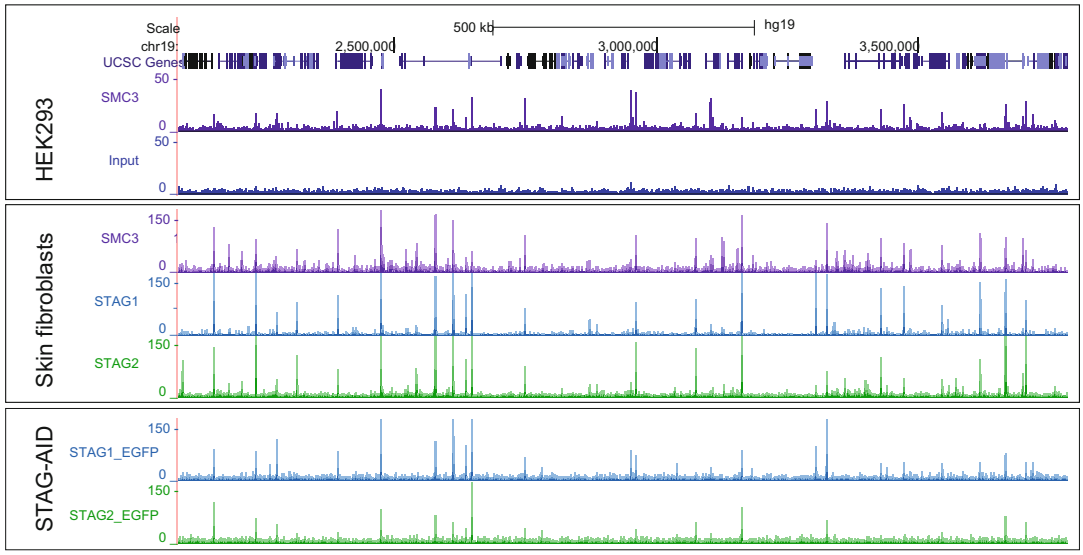


Fig. 1 ChIP-sequencing of cohesin subunits form different human cell lines. ChIP-seq for SMC3 and the respective input are shown for HEK293 cells [14], ChIP-seq for SMC3, STAG1 and STAG2 are shown for human skin fibroblasts (unpublished data). The ChIP-seq with anti-EGFP from STAG1-AID and STAG2-AID HCT116 OsTir1 cells was performed with a protocol reducing the unbound pool [12, 15]

2.1 Cross-Linking and Cell Harvesting

1. Cross-linking Mix: 11% formaldehyde (from 37% stock solution), 100 mM NaCl, 0.5 mM EGTA, 50 mM HEPES (pH 8.0). This stock can be stored at 4 °C and used for 1–2 weeks.
2. Quenching solution: 2.5 M Glycine in water.
3. PBS: Phosphate buffered saline (PBS) (137 mM NaCl, 2.7 mM KCl, 10 mM Na_4HPO_4, 1.76 mM KH_2PO_4).
4. Protease inhibitor cocktail: dissolve one tablet protease inhibitor mix (for e.g., Roche cat # 11697498001) in 1 mL water to obtain a 50× stock. This stock can be stored at −20 °C for a few weeks.
5. PMSF protease inhibitor for washing buffers: dissolve PMSF (for e.g., Merck, cat # 10837091001) in DMSO to obtain a 0.5 M stock solution that is stable for a few months at room temperature. Note that PMSF has a very short half-life in aqueous solutions and it should be added to the buffers just before usage.
6. Lysis buffer: 50 mM Tris-HCl (pH 8.0), 10 mM EDTA (pH 8.0), 1% SDS, 1 mM PMSF and protease inhibitor cocktail.
7. Cell scraper (for example Greiner Bio-One cat. 10516762).

2.2 Sonication

1. Sonicator and 15 mL sonication tubes (we use the Diagenode bioruptor Plus and Diagenode 15 mL sonication tubes). We have also used 15 mL tubes from other brands (Falcon, Greiner) and the shearing needs to optimized for each brand.
2. PCR purification columns for purification of samples for the sonication check (we use Qiagen).
3. Agarose gel electrophoresis equipment.
4. Nanodrop or similar to check DNA concentration.

2.3 Chromatin Immunoprecipitation and DNA Cleanup

1. IP dilution buffer: 20 mM Tris-HCl (pH 8.0), 2 mM EDTA (pH 8.0), 1% Triton X-100, 150 mM NaCl, 1 mM PMSF, and protease inhibitor cocktail.
2. Wash buffer 1: 20 mM Tris-HCl (pH 8.0), 2 mM EDTA (pH 8.0), 1% Triton X-100, 150 mM NaCl, 0.1% SDS, 1 mM PMSF.
3. Wash buffer 2: 20 mM Tris-HCl (pH 8.0), 2 mM EDTA (pH 8.0), 1% Triton X-100, 500 mM NaCl, 0.1% SDS, 1 mM PMSF.
4. Wash buffer 3: 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0), 0.25 M LiCl, 0.5% NP-40, 0.5% deoxycholate.
5. TE: 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0).
6. Elution buffer: 25 mM Tris-HCl (pH 7.5), 5 mM EDTA (pH 8.0), 0.5% SDS.
7. Proteinase K (10 mg/mL) and RNase A (10 mg/mL).
8. Phase lock tubes Phase Lock Gel Heavy (Quantabio) or similar product.
9. 3 M Na Acetate solution pH 4.8.
10. GlycoBlue™ Coprecipitant (Thermo Fisher, cat # AM9516) or similar product.
11. Phenol–chloroform–isoamyl alcohol (25:24:1).
12. Chloroform, molecular grade.
13. 100% ethanol and 70% ethanol, molecular grade.

3 Methods

3.1 Cross-Linking and Cell Harvesting

1. Start with one ~80% confluent 145 cm^2 dish (~5×10^7 cells per dish). Add 1/10th volume of your medium of the cross-linking mix directly to the medium (final formaldehyde concentration 1%). Mix by gentle rotation and incubate at room temperature (RT) for 10 min, best slowly shaking (*see* **Note 1**).

2. Quench the cross-linking reaction by addition of 2.5 M glycine to a final concentration of 125 mM and incubate with shaking for 5 min at RT.
3. Wash the cells twice with 10 mL cold PBS.
4. Add 750 µL lysis buffer to the plate, spread it over the plate and incubate at 4 °C for 3–5 min. Scrape dishes thoroughly with a cell scraper and transfer the lysate to a 15 mL tube. Incubate the lysate 20 min on ice (*see* **Note 2**).

3.2 Sonication

1. If SDS has precipitated, the lysate needs to be warmed to RT briefly to dissolve the crystals (*see* **Note 3**).
2. Sonicate lysate to shear DNA to a fragment size of 200–300 bp. For the Diagenode bioruptor use these settings: high output, 30 s on/30 s off, 4 °C water bath temprerature. The required sonication times range between 10 and 30 min (*see* **Note 4**).
3. Centrifuge lysate at 2800 × *g* for 10 min, RT (to avoid SDS precipitation), to pellet cell debris. A properly sheared lysate should give only a rather small pellet. Transfer supernatant to a new tube (*see* **Note 5**) and determine the DNA concentration.
4. To check shearing efficiency decrosslink a sample of the sonicated chromatin and analyze the fragment sizes by DNA agarose gel electrophoresis. We recommend to perform this check for each experiment, even when the sonication conditions were established previously. Dilute 20 µL sheared chromatin to 100 µL with MilliQ water and add 10 µL RNase A and 10 µL proteinase K. Incubate for 1 h at 37 °C, then overnight at 65 °C. Purify the DNA using a standard PCR purification kit. Load 10 µL of the purified DNA on a 1.5% agarose gel together with a suitable DNA size marker.

3.3 Immunoprecipitation

1. Use approximately 1 mg DNA per sample and dilute the sample 1:5 in IP dilution buffer.
2. Remove 50 µL of diluted chromatin to serve as input sample and store it at −20 °C until further use (Subheading 3.4, **step 3**).
3. Optional step to reduce unspecific binding: Preclear the diluted chromatin by incubating with beads (*see* **steps 5** and **6**) for 1 h and remove the beads again.
4. Add the selected antibodies to the lysate. The optimum amount of antibodies needs to be determined individually but ~10 µg antibody per 1 mg of DNA often works well (*see* **Note 6**). Incubate the samples over night at 4 °C with rotation.
5. Prepare the beads (slurry contains 50% beads and 50% buffer) by resuspending the beads and wash 100 µL bead slurry per sample with IP dilution buffer. For this, wash the beads 3 times

with IP dilution buffer (1 mL each time) and resuspend the beads to have again 100 μL beads resuspension per sample (*see* **Note 7** for information on beads). *Optional step*: To reduce unspecific binding to the beads, wash the beads with 0.1 mg/mL BSA dissolved in IP dilution buffer. Wash 3 times with 1 mL solution per 100 μL bead slurry.

6. Add 100 μL bead resuspension per sample and incubate for 2 h at 4 °C with rotation.
7. Remove the beads from the sample. In case of Dynabeads use a suitable magnet, otherwise spin the beads down for 5 min at 180 × *g*. Remove the supernatant carefully without aspirating beads.
8. Perform the following washing steps. Use 1 mL buffer per sample and incubate the beads for 5 min at 4 °C under rotation.
   (a) 2× IP dilution buffer.
   (b) 2× wash buffer 1.
   (c) 2× wash buffer 2.
   (d) 1× wash buffer 3 (from here incubate at RT).
   (e) 2× TE buffer (to reduce the salt and detergent concentration before the elution step).

3.4 Elution and Reversal of Cross-Links

1. Remove the liquid from the washing step as much as possible and add 200 μL elution buffer. Incubate the samples for 20 min at 65 °C, flick the tubes sometimes during incubation.
2. Centrifuge the beads 5 min at 2800 x *g* and transfer the eluate to a new tube without taking any beads along. Repeat **step 1**.
3. Pool the eluates from the two steps. Thaw the input sample and fill it up to 400 μL with elution buffer. From here on the input is treated identical to the samples.
4. Add 25 μL proteinase K and 25 μL RNAse to the eluates and the input sample. Incubate for 1 h at 37 °C, then overnight at 65 °C.

3.5 DNA Cleanup

1. Centrifuge the required amount of phase lock tubes 1 min 15,000 × *g* at RT (important, otherwise the resin is too viscous) to make sure the phase lock resin is at the bottom of the tube. Transfer the samples to phase lock tubes and add 1/10 of volume 3 M Na Acetate solution (pH 4.8) and 2 μL GlycoBlue™.
2. Add 1 volume of phenol–chloroform–isoamyl alcohol (25:24:1) to your sample and shake heavily for about 1 min, do not vortex. Centrifuge at 15,000 × *g* for 5 min at RT. The phenol-containing phase is now trapped in by the resin.

3. Add 1 volume chloroform to the samples and shake again heavily, do not vortex. Centrifuge at 15,000 × *g* for 5 min at RT. Only the aqueous phase containing the DNA remains above the resin and can be transferred to a new 1.5 mL tube.
4. Add 2.5 volumes 100% ethanol and store the sample over night at −20 °C to precipitate the DNA. Alternatively, place the tube at −80 °C for at least 1 h.
5. Centrifuge the samples at 15,000 × *g* and 4 °C for 45 min. A blue DNA/coprecipitant pellet should be nicely visible. Carefully remove the supernatant without disturbing the pellet.
6. Wash the pellet with 750 µL 70% ethanol to remove coprecipitated salt and centrifuge the sample at 15,000 × *g* and 4 °C for 10 min.
7. Remove all liquid and dry the pellet until all ethanol has evaporated. If available, a SpeedVac Vacuum Concentrator could be used but overdrying of the pellet should be avoided.
8. Dissolve the pellet in 30–50 µL TE buffer.

3.6 Analysis and Preparation for Sequencing

1. Proceed to an analysis of the samples by qPCR (recommended as quality control to test ChIP enrichment before preparing the samples for sequencing).
2. Prepare the sequencing library according to the instructions of the library preparation kit used.

4 Notes

1. Formaldehyde is used to cross-link the proteins to the DNA. The effciciency of cross-linking depends on temperature and incubation time. Excessive cross-linking reduces antigen accessibility and sonication efficiency. Note that some epitope-tags, for example the Flag-tag, can be masked by formaldehyde cross-linking. In this case the use of a 3 × Flag-tag was helpful.
2. When using cells that are easily lifting off from the culture plate, e.g. HEK293 cells, harvest the cells before cross-linking by spraying them off the plate with the culture medium and transfer them into a 50 mL tube. The same procedure can also be used for cells growing in suspension. Adapt the volumes for cross-linking and quenching according to **step 1**. After quenching the cross-linking reaction, pellet the cells by centrifugation (5 min, 180 × *g*). Wash 3 times with cold PBS, the same volume as the culture medium. Then resuspend the pellet in ChIP lysis buffer (750 µL per 1 × 10^7 cells). Proceed to Subheading 3.2, **step 1**.

3. If crystals reappear during sonication, set the thermostat that controls the temperature of the sonication bath to 7 °C.
4. Optimal conditions for sonication are highly dependent on the number and type of cells, volume of sample, duration of sonication and sonicator power setting used. Sonication needs to be optimized for each cell line. We recommend to perform a time course experiment to determine the optimal conditions before starting your actual experiment. For this, use the same amounts of cells and buffer as for your main experiment and sonicate the sample for 0, 10, 20, and 30 min and take a sample at each time point that will be processed as described under Subheading 3.2, **step 3**.
5. The sheared chromatin can be snap frozen in liquid nitrogen and stored at −80 °C for up to 3 months. Avoid multiple freeze-thaws.
6. Table 1 shows the antibodies that have been used to efficiently ChIP the cohesin complex.
7. The most commonly used beads are Protein A or Protein G Dynabeads™ (ThermoFisher, Protein A cat. 10002D; Protein G cat. 10004D) since they allow a straightforward separation of the beads from the sample by using a magnet. However, we have successfully used Affi-Prep Protein A Support beads for ChIP-sequencing. Although the beads need to be separated from the sample by centrifugation, they have certain advantages. We successfully immobilized antibodies on the beads by cross-linking with DMP (Dimethyl pimelimidate dihydrochloride; Sigma-Aldrich, cat. 80490) [15] for sequential CHIP experiments. Further, the costs are considerably lower than magnetic beads. For the choice of the beads, the origin specie of the antibodies and the antibody isotype are very important. In Table 2, we list the affinity of protein A and G beads to different immunoglobulin isotypes.

Acknowledgments

We thank Valentina Casa for discussions about differences between ChIP-protocols and Thomas van Staveren for input on the manuscript. M.M.G. was supported by a was supported by a Dutch Cancer Society (KWF) grant EMCR 2015-7857 to K.S.W.

Table 2
Antibody compatibility table

Species	Ig Subclasss	Protein A	Protein G
Human	IgG1, IgG2, IgG4	+++	+++
	IgG3	+	+++
	IgD	–	–
	IgD	+	–
	Fab	+	+
	ScF_v	+	–
Mouse	IgG1	+	++
	IgG2a, IgG2b, IgG2c	+++	+++
	IgM	–	–
Rat	IgG1	+	++
	IgG2a	–	+++
	gG2b	–	+
	gG2c	+++	+++
Goat	IgG1	+	+++
	IgG2	+++	+++
Sheep	IgG1	+	+++
	IgG2	+++	+++
Cow/bovine	IgG1	+	+++
	IgG2	+++	+++
Horse	IgG(ab)	+	–
	IgG(c)	+	–
	IgG(T)	–	+++
Rabbit	Total IgG	+++	+++
Dog	Total IgG	+++	+
Cat	Total IgG	+++	+
Pig	Total IgG	+++	+
Guinea pig	Total IgG	+++	+
Chicken	Total IgG	–	–

References

1. Nishiyama T (2018) Cohesion and cohesin-dependent chromatin organization. Curr Opin Cell Biol 58:8–14. https://doi.org/10.1016/j.ceb.2018.11.006. S0955-0674(18)30166-2
2. van Ruiten MS, Rowland BD (2021) On the choreography of genome folding: a grand pas de deux of cohesin and CTCF. Curr Opin Cell Biol 70:84–90. https://doi.org/10.1016/j.ceb.2020.12.001
3. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, Yahata K, Imamoto F, Aburatani H, Nakao M, Imamoto N, Maeshima K, Shirahige K, Peters JM (2008) Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature 451(7180):796–U793. https://doi.org/10.1038/nature06634
4. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering CH, Mirny L, Spitz F (2017) Two independent modes of chromatin organization revealed by cohesin

removal. Nature 551(7678):51–56. https://doi.org/10.1038/nature24281

5. Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, Huang X, Shamim MS, Shin J, Turner D, Ye Z, Omer AD, Robinson JT, Schlick T, Bernstein BE, Casellas R, Lander ES, Aiden EL (2017) Cohesin loss eliminates all loop domains. Cell 171(2):305–320.e324. https://doi.org/10.1016/j.cell.2017.09.026. S0092-8674(17)31120-0 [pii]
6. Pugacheva EM, Kubo N, Loukinov D, Tajmul M, Kang S, Kovalchuk AL, Strunnikov AV, Zentner GE, Ren B, Lobanenkov VV (2020) CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. Proc Natl Acad Sci U S A 117(4): 2020–2031. https://doi.org/10.1073/pnas.1911708117
7. Hansen AS, Pustova I, Cattoglio C, Tjian R, Darzacq X (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. elife 6:e25776. https://doi.org/10.7554/eLife.25776
8. Wutz G, Varnai C, Nagasaka K, Cisneros DA, Stocsits RR, Tang W, Schoenfelder S, Jessberger G, Muhar M, Hossain MJ, Walther N, Koch B, Kueblbeck M, Ellenberg J, Zuber J, Fraser P, Peters JM (2017) Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. EMBO J 36(24):3573–3599. https://doi.org/10.15252/embj.201798004
9. Viny AD, Bowman RL, Liu Y, Lavallee VP, Eisman SE, Xiao W, Durham BH, Navitski A, Park J, Braunstein S, Alija B, Karzai A, Csete IS, Witkin M, Azizi E, Baslan T, Ott CJ, Péer D, Dekker J, Koche R, Levine RL (2019) Cohesin members Stag1 and Stag2 display distinct roles in chromatin accessibility and topological control of HSC self-renewal and differentiation. Cell Stem Cell 25(5):682–696.e688. https://doi.org/10.1016/j.stem.2019.08.003
10. Busslinger GA, Stocsits RR, van der Lelij P, Axelsson E, Tedeschi A, Galjart N, Peters JM (2017) Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. Nature 544(7651):503–507. https://doi.org/10.1038/nature22063
11. Holzmann J, Politi AZ, Nagasaka K, Hantsche-Grininger M, Walther N, Koch B, Fuchs J, Durnberger G, Tang W, Ladurner R, Stocsits RR, Busslinger GA, Novak B, Mechtler K, Davidson IF, Ellenberg J, Peters JM (2019) Absolute quantification of cohesin, CTCF and their regulators in human cells. elife 8:e46269. https://doi.org/10.7554/eLife.46269
12. Cabianca DS, Casa V, Bodega B, Xynos A, Ginelli E, Tanaka Y, Gabellini D (2012) A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. Cell 149(4): 819–831. https://doi.org/10.1016/j.cell.2012.03.035
13. Zuin J, Franke V, van Ijcken WFJ, van der Sloot A, Krantz ID, van der Reijden M, Nakato R, Lenhard B, Wendt KS (2014) A cohesin-independent role for NIPBL at promoters provides insights in CdLS. PLoS Genet 10(2):e1004153. https://doi.org/10.1371/journal.pgen.1004153
14. Parenti I, Diab F, Gil SR, Mulugeta E, Casa V, Berutti R, Brouwer RWW, Dupe V, Eckhold J, Graf E, Puisac B, Ramos F, Schwarzmayr T, Gines MM, van Staveren T, van IJcken WFJ, Strom TM, Pie J, Watrin E, Kaiser FJ, Wendt KS (2020) MAU2 and NIPBL variants impair the heterodimerization of the cohesin loader subunits and cause cornelia De Lange syndrome. Cell Rep 31(7):107647. https://doi.org/10.1016/j.celrep.2020.107647
15. Casa V, Moronta Gines M, Gade Gusmao E, Slotman JA, Zirkel A, Josipovic N, Oole E, van IJcken WFJ, Houtsmuller AB, Papantonis A, Wendt KS (2020) Redundant and specific roles of cohesin STAG subunits in chromatin looping and transcriptional control. Genome Res 30(4):515–527. https://doi.org/10.1101/gr.253211.119

Chapter 8

Epi-Decoder: Decoding the Local Proteome of a Genomic Locus by Massive Parallel Chromatin Immunoprecipitation Combined with DNA-Barcode Sequencing

Maria Elize van Breugel and Fred van Leeuwen

Abstract

The genome in a eukaryotic cell is packaged into chromatin and regulated by chromatin-binding and chromatin-modifying factors. Many of these factors and their complexes have been identified before, but how each genomic locus interacts with its surrounding proteins in the nucleus over time and in changing conditions remains poorly described. Measuring protein–DNA interactions at a specific locus in the genome is challenging and current techniques such as capture of a locus followed by mass spectrometry require high levels of enrichment. Epi-Decoder, a method developed in budding yeast, enables systematic decoding of the proteome of a single genomic locus of interest without the need for locus enrichment. Instead, Epi-Decoder uses massive parallel chromatin immunoprecipitation of tagged proteins combined with barcoding a genomic locus and counting of coimmunoprecipitated barcodes by DNA sequencing (TAG-ChIP-Barcode-Seq). In this scenario, DNA barcode counts serve as a quantitative readout for protein binding of each tagged protein to the barcoded locus. Epi-Decoder can be applied to determine the protein–DNA interactions at a wide range of genomic loci, such as coding genes, noncoding genes, and intergenic regions. Furthermore, Epi-Decoder provides the option to study protein–DNA interactions upon changing cellular and/or genetic conditions. In this protocol, we describe in detail how to construct Epi-Decoder libraries and how to perform an Epi-Decoder analysis.

Key words Epi-Decoder, DNA barcodes, ChIP, Proteome, Chromatin

1 Introduction

Multiple cellular processes such as transcription, replication, and DNA repair involve interactions of the genome with many different proteins and protein complexes in a coordinated manner. How these interactions are coordinated under changing cellular conditions and across different locations in the genome is largely unknown because systematically measuring protein–DNA interactions at a specific locus in the genome is challenging [1–3]. A commonly used technique to determine the local proteome is

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_8,

capture of the DNA region(s) followed by mass spectrometry (capture-MS). However, this general strategy requires high levels of enrichment of the locus of interest combined with quantitative mass spectrometry of small amounts of protein [1–3]. As an alternative approach, independent of locus enrichment and mass spectrometry, we developed a technique in budding yeast called Epi-Decoder [4, 5], which is based on genomics by exploiting the power of DNA barcoding and massive parallel DNA sequencing [6–9]. Epi-Decoder enables decoding the proteome of a single barcoded genomic locus by utilizing parallel chromatin immunoprecipitation of tagged proteins, DNA sequencing of coimmunoprecipitated DNA barcodes, and DNA-barcode counting to infer the occupancy of each tagged protein at the barcoded locus (outlined in Figs. 1 and 2). Compared to capture-MS, this technique enables working with low quantities of starting material because DNA barcodes can be amplified by PCR prior to sequencing.

The first step in setting up an Epi-Decoder analysis is to choose a genomic locus of interest (*see* Fig. 1). This can be a gene of interest, a noncoding region, a promoter or other regulatory element, an origin of replication, or any other region of interest. A random DNA barcode is then inserted at this locus of interest by making use of the CRISPR-Cas9 system. In the protocol described here, only one barcode is inserted, but inserting multiple barcodes is also possible [10–12]. By picking thousands of transformed yeast colonies, a "barcoder" library is constructed. Within this library, each well in an ordered array contains a yeast clone with a different DNA barcode sequence inserted at the common locus of interest. By pooling columns and rows of the barcoder library and sequencing each pool with unique index primers, the coordinates of all individual barcodes in the barcoder library can be uncovered. Placing a selection marker close to the locus of interest prior to inserting the barcodes enables selection for the locus that contains the barcode during subsequent genetic crosses.

The short DNA barcodes are embedded in or proximal to the chromatin structure of the locus of interest. This property enables the use of barcodes to report on the composition of chromatin. For example, when a barcoder library is combined with a library of genetic mutations, it can be used to systematically survey in parallel how each mutant affects a particular chromatin feature of interest [13, 14]. For Epi-Decoder, the barcoder library is combined with a yeast library of tagged proteins, in which each clone represents a different protein tagged with a common epitope tag that can be used for chromatin immunoprecipitation. Here we use the Tandem Affinity Purification (TAP) tag which contains a protein A domain that can be used for pool down by immunoglobulins (IgG) or specific anti-TAP antibodies during chromatin immunoprecipitation. Other epitope-tag libraries can also be used [15–19].

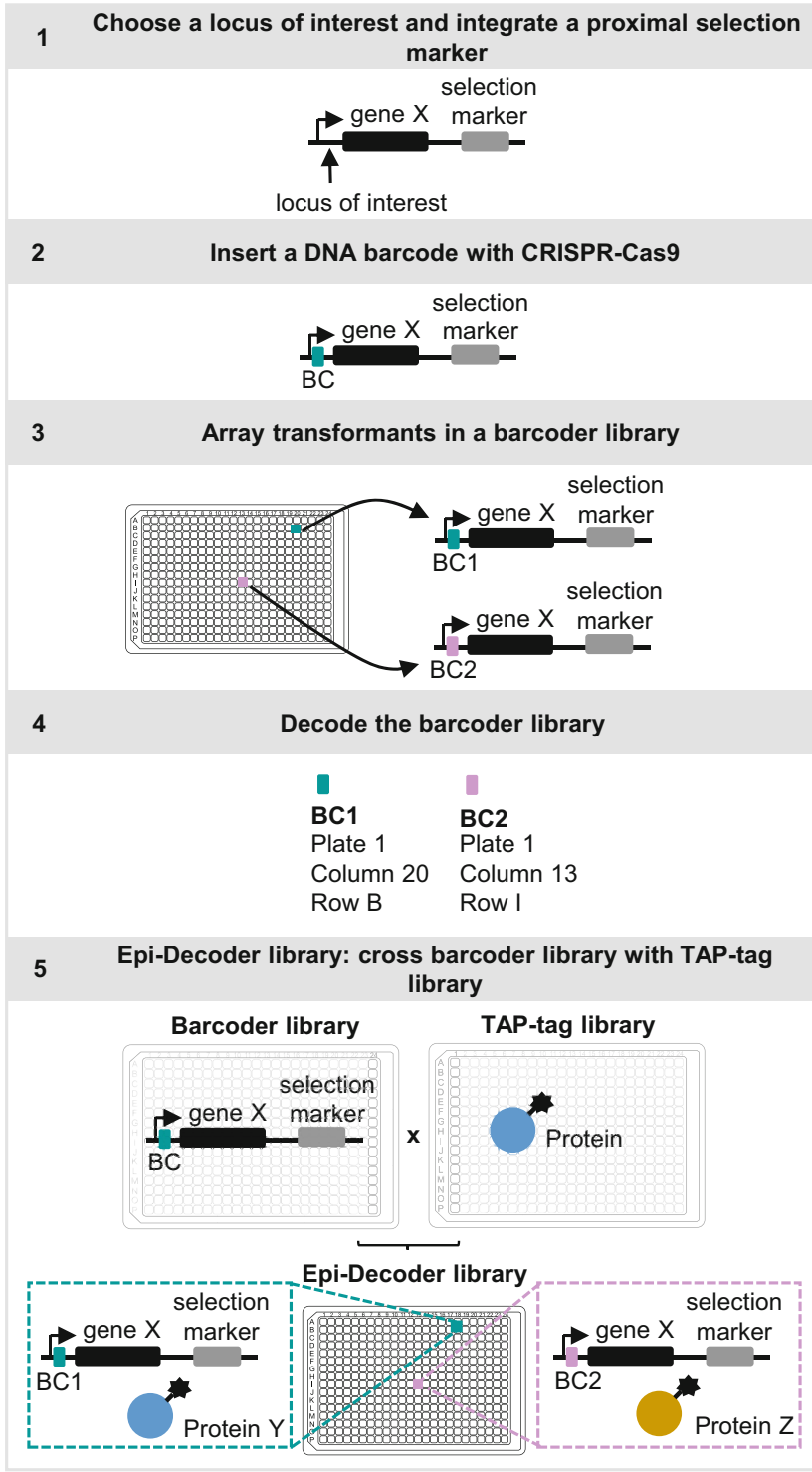


Fig. 1 Systematic overview of the steps required to obtain an Epi-Decoder library. (1) In this example, the locus of interest is the promoter region of gene X. Proximal to the locus of interest, a selection marker is integrated.

To obtain the Epi-Decoder library, the barcoder library is crossed with the TAP-tag library [20] by using synthetic genetic array (SGA) technologies [21, 22]. In the resulting Epi-Decoder library (diploid or haploid), each yeast clone contains a unique combination of one tagged protein and one DNA barcode at the locus of interest. This resulting Epi-Decoder library can be used to systematically decode the local proteome of a single locus of interest (*see* Fig. 2). Briefly, the Epi-Decoder library is pooled into a single pool which is facilitated by the unique barcode–protein combinations. Cells are treated with formaldehyde to cross-link proteins to the DNA and subsequently lysed to collect chromatin. Chromatin is sonicated to obtain 150–600 bp fragments. Chromatin fragments bound to TAP-tagged proteins are then immunoprecipitated by using rabbit IgG antibodies. Subsequently, protein–DNA cross-links are reversed and barcodes from the input and immunoprecipitated DNA are amplified to attach sequencing adapters and indices. After deep sequencing and barcode counting, IP over input is calculated which is a measure for protein occupancy for each protein associated with its barcoded region.

We previously applied Epi-Decoder to a barcoded HO-KanMX locus, harboring barcodes in intergenic regions [23, 18], and to a native locus by inserting a barcode in the coding sequence of the *ADE2* gene, close to the transcription start site. At these loci we identified hundreds of chromatin binders that could be assigned to histones, known DNA- and chromatin-binding proteins, transcription and RNA-processing factors, and factors involved in cellular metabolism and protein folding [4, 5].

Epi-Decoder can be applied in a wide variety of situations, for example in determining the proteome of multiple genomic loci simultaneously, or in different cellular conditions or mutant backgrounds. This protocol is an extensive description of all the steps required to complete an Epi-Decoder analysis. Where applicable, additional notes are given to provide the necessary tips and tricks to successfully complete this protocol.

Fig. 1 (continued) (2) By using the CRISPR-Cas9 system, a DNA barcode is inserted at the locus of interest in thousands of clones. (3) Resulting transformants are picked and arrayed in 384-well plates which results in the barcoder library. Each well contains a strain that harbors a unique DNA barcode. (4) To decode the barcoder library, pools are made and indexed using pool-specific index primers. By sequencing the individual pools, the coordinates of each barcode in the barcoder library can be determined. (5) Using Synthetic Genetic Array (SGA) methodology, the barcoder library and TAP-tag library are crossed to generate a diploid or haploid Epi-Decoder library in which each clone contains one unique barcode and one unique tagged protein

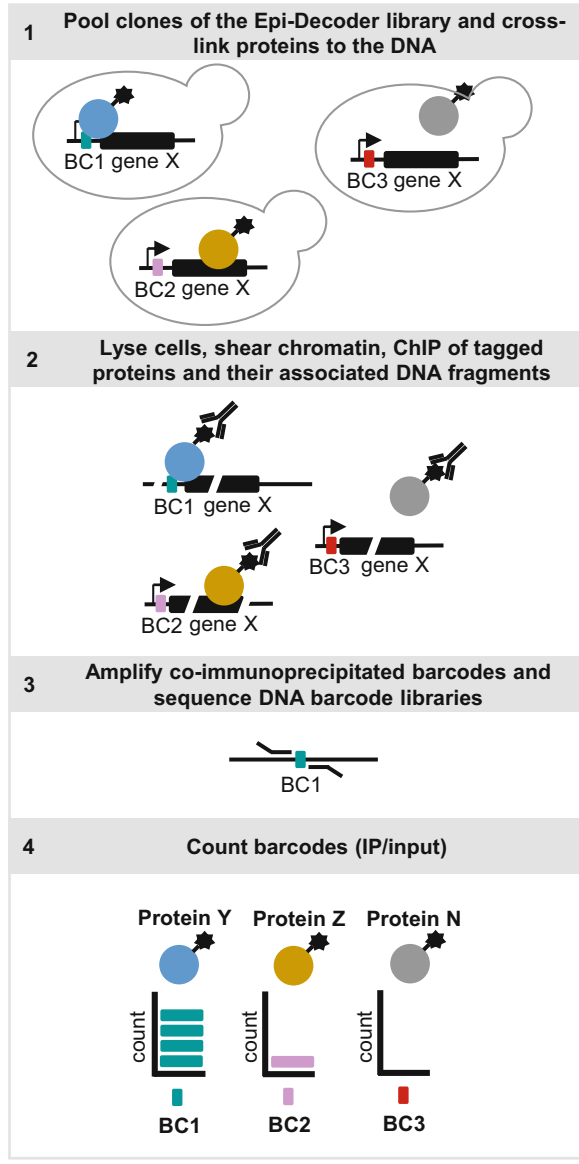


Fig. 2 Overview of the steps required to systematically decode the local proteome of a single locus of interest. (1) After combining the clones of an Epi-Decoder library into a single pool, proteins are cross-linked to the DNA. (2) Following cell lysis, DNA is sheared by sonication. By using a TAP-specific antibody, TAP-tagged proteins and potential DNA barcodes cross-linked to them are pulled down. (3) After reversing crosslinks, the barcodes are amplified using a PCR protocol that also results in the preparation of libraries for deep sequencing. (4) Barcodes are extracted from the sequencing reads and counted. Barcode counts (IP/Input) serve as a readout for protein binding to the barcoded locus

2 Materials

2.1 Yeast Strains, Plasmids and Oligos

2.1.1 Yeast Strains

To create the Epi-Decoder library, a TAP-tag library is crossed with a barcoder library. For this purpose, the commercially available TAP-tag library [20] is used (*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 yfg-TAP*-SkHIS3MX6) (*see* **Note 1**). Other protein tag libraries can also be used if the tags are compatible with chromatin immunoprecipitation protocols. We previously constructed the following Epi-Decoder libraries: NKI2560 (Diploid *MAT*a/α *CAN1/can1Δ::HphMX LYP1/lyp1Δ::STE3pr-LEU2 his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 ura3Δ0/ura3Δ0 met15Δ0/met15Δ0 ADE2/BC-ADE2-NatMX yfg/yfg-TAP*-SkHIS3MX6) and NKI4217 (Haploid *MATα can1Δ::HphMX lyp1Δ::STE3pr-LEU2 his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 HO::BC-KanMX-BC yfg-TAP*-SkHIS3MX6) [5]. These Epi-Decoder libraries are based on previously constructed barcoder libraries: NKI2559 (*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 BC-ADE2-NatMX*) [5] and the barcoder library, obtained from the Nislow and Andrews labs (*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 HO::BC-KanMX-BC*) [18, 23]. Libraries are stored in 384-well glycerol freezer stock plates. Barcoder libraries are available upon request. To make a custom barcoder library, we recommend using strain NKI4212 (*MATα can1Δ::HphMX lyp1Δ::STE3pr-LEU2 his3Δ1 leu2Δ0 ura3Δ0 met15Δ0*) [5]. This strain was derived from strain Y8205 [24] by replacing the *STE2*pr-*Sp_his5* cassette at the *CAN1* locus, originally designed to select for *MAT*a haploid cells by growth on plates lacking histidine, by HphMX. Removing this histidine selection marker enables selection during SGA for the TAP-SkHIS3MX6 allele using histidine prototrophy (*see* **Note 2**). Because of this adjustment, one can select for diploids, or *MAT*α haploid cells using growth on media lacking leucine, selecting for *LEU2* expression driven by *MAT*α-specific *STE3* promoter.

2.1.2 Plasmids

To select for the locus of interest that contains the barcode during genetic crosses, a selection marker is inserted close to the locus of interest. In this protocol, the NatMX selection marker is used which can be amplified from pFvL99 [25]. pFvL99 is available upon request. The marker can be integrated by homologous recombination using a standard yeast transformation protocol [26, 27]. To introduce a DNA barcode at the locus of interest, the pML104 vector (Addgene Plasmid #67638) expressing Cas9 and containing a guide RNA (gRNA) expression cassette and *URA3* marker can be used. This plasmid can be manipulated by inserting within the gRNA expression cassette a short DNA sequence encoding a gRNA targeting the locus of interest, following the instructions in [28] and as explained below.

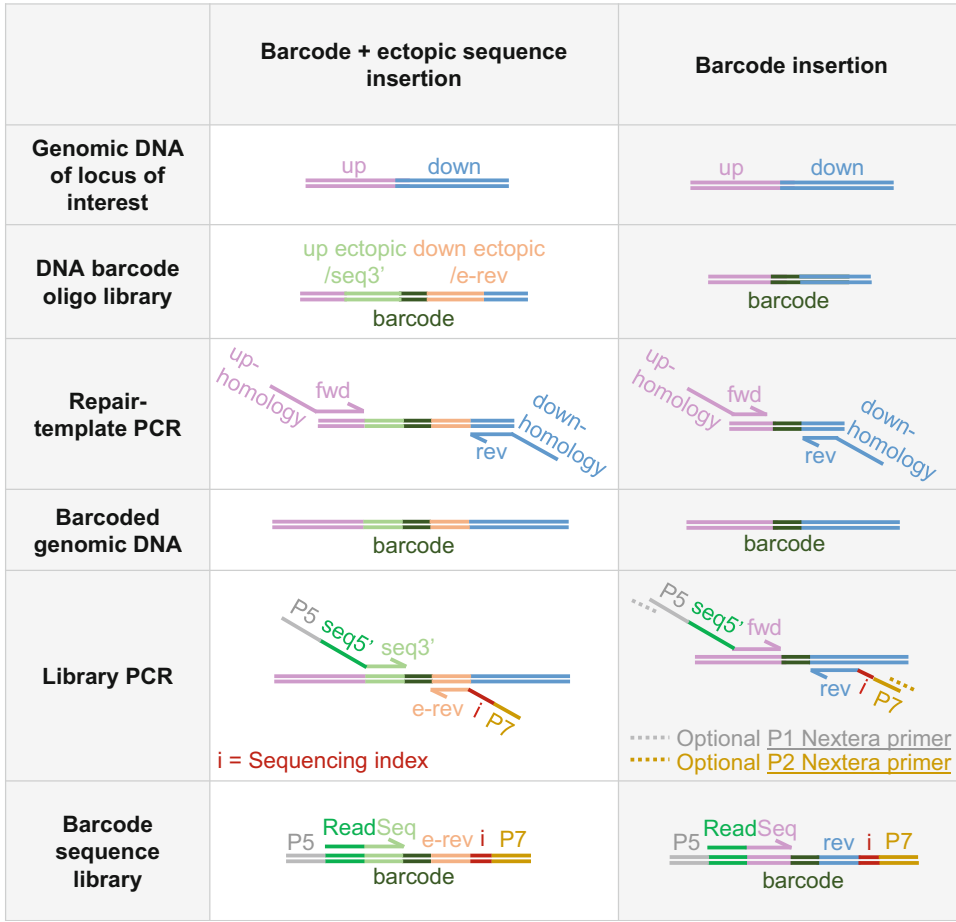


Fig. 3 Overview of the different PCR steps and oligos required to construct an Epi-Decoder library and perform an Epi-Decoder analysis. Two different approaches can be taken, either inserting a DNA barcode flanked by ectopic sequences at the locus of interest or only a DNA barcode embedded in the native context. Considerations for design of the barcoded locus and sequencing libraries are described in **Note 3**

2.1.3 Oligos

A DNA barcode oligo library and PCR primers are required to generate repair templates for yeast barcoder construction (*see* Subheading 3.1.3). To prepare Epi-Decoder and barcoder libraries for sequencing on the Illumina platform, primers are used to amplify the barcodes and at the same time append adapters, a target sequence for a sequencing primer, and an index sequence for multiplexing (*see* Fig. 3 and **Note 3**).

2.2 Yeast Drugs and Media

After autoclaving, media for plates need to cool down to approximately 55 °C before adding additional supplements. After adding all the ingredients, mix thoroughly and pour plates in a sterile environment. Optional: add 10 mg tetracycline to 1 L media for the liquid and agar plates to avoid growth of bacteria. Agar plates that are used for arrayed libraries have a volume of 50 mL (*see* **Note 4**).

1. CloNat (Nourseothricin): Dissolve 100 mg/mL in water, filter-sterilize, and store in 1 mL aliquots at −80 °C.
2. Canavanine: Dissolve 20 mg/mL in water, filter-sterilize, and store in 3 mL aliquots at −20 °C.
3. Thialysine: Dissolve 100 mg/mL in water, filter-sterilize, and store in 500 μL aliquots at −20 °C.
4. 5-Fluoro Orotic acid (5-FOA). Use as powder.
5. 10× Tyrosine (0.5 g/L): Dissolve 0.5 g tyrosine in 1 L water and filter-sterilize.
6. 10× YNB: Dissolve 17 g Yeast Nitrogen Base (YNB w/o amino acids and w/o ammonium sulfate) in 1 L water. Stir with low heat into solution and filter-sterilize.
7. 10× Ammonium Sulfate: Dissolve 50 g ammonium sulfate in 1 L water and filter-sterilize.
8. 10× MSG: Dissolve 10 g monosodium glutamic acid in 1 L water and filter-sterilize.
9. YC-7aa dropout solution (100×) [29]: Dissolve 10 g arginine, 5 g aspartic acid, 5 g isoleucine, 5 g phenylalanine, 5 g proline, 5 g serine, 10 g threonine, and 5 g valine in 1 L water and filter-sterilize.
10. YC-8aa dropout solution (100×; YC-7aa without arginine): Dissolve 5 g aspartic acid, 5 g isoleucine, 5 g phenylalanine, 5 g proline, 5 g serine, 10 g threonine, and 5 g valine in 1 L water and filter-sterilize.
11. YEPD medium: Dissolve 10 g yeast extract and 20 g Bacto peptone in 900 mL water. Autoclave and add 100 mL 20% glucose.
12. YEPD plates: Dissolve 10 g yeast extract, 20 g Bacto peptone, and 22 g agar in 900 mL water and autoclave. Add 100 mL 20% glucose and pour plates.
13. YEPD + CloNat plates: Use the YEPD medium from YEPD plates, but now add 1 mL of CloNat (100 mg/mL) stock solution after cooling down to 55 °C.
14. Barcode insertion selection plates (YC-URA + CloNat): Dissolve 22 g agar in 529 mL water and autoclave. Add 100 mL 20% glucose, 100 mL 10× YNB, 100 mL 10× MSG, 10 mL 100× YC-7aa dropout solution, 100 mL 10× tyrosine, 10 mL tryptophan (10 mg/mL), 10 mL adenine (1 mg/mL), 10 mL lysine (10 mg/mL), 10 mL methionine (5 mg/mL), 10 mL leucine (10 mg/mL), 10 mL histidine (5 mg/mL), and 1 mL of CloNat (100 mg/mL) stock solution.
15. Loss of CRISPR plasmid plates (5-FOA + CloNat): Dissolve 22 g agar in 509 mL water and autoclave. Add 100 mL 20% glucose, 100 mL 10× YNB, 100 mL 10× MSG, 10 mL 100×

YC-7aa dropout solution, 100 mL 10× tyrosine, 20 mL uracil (1 mg/mL), 10 mL tryptophan (10 mg/mL), 10 mL adenine (1 mg/mL), 10 mL lysine (10 mg/mL), 10 mL methionine (5 mg/mL), 10 mL leucine (10 mg/mL), 10 mL histidine (5 mg/mL), 1 mL of CloNat (100 mg/mL) stock solution, and 1.0 g of 5-fluoroorotic acid (5-FOA).

16. Diploid selection plates (YC-HIS + CloNat): Dissolve 22 g agar in 439 mL water and autoclave. Add 100 mL 20% glucose, 100 mL 10× YNB, 100 mL 10× MSG, 10 mL 100× YC-7aa dropout solution, 100 mL 10× tyrosine, 100 mL uracil (1 mg/mL), 10 mL tryptophan (10 mg/mL), 10 mL adenine (1 mg/mL), 10 mL lysine (10 mg/mL), 10 mL methionine (5 mg/mL), 10 mL leucine (10 mg/mL), and 1 mL of CloNat (100 mg/mL) stock solution.
17. Sporulation plates (SPO): Dissolve 22 g agar in 986.5 mL water and autoclave. Add 12.5 mL 40% potassium acetate and 1 mL 20% raffinose. It is important to not autoclave the potassium acetate.
18. Haploid selection #1 plates (*MAT*α mating type haploids and TAP-tag): Dissolve 22 g agar in 456.5 mL water and autoclave. Add 100 mL 20% glucose, 100 mL 10× YNB, 100 mL 10× ammonium sulfate (50 g/L), 100 mL 10× tyrosine (0.5 g/L), 10 mL 100× YC-8aa dropout solution (when using canavanine, drop out arginine in YC-7aa), 10 mL adenine (1 mg/mL), 10 mL tryptophan (10 mg/mL), 100 mL uracil (1 mg/mL), 10 mL methionine (5 mg/mL), 3 mL canavanine (20 mg/mL), and 0.5 mL thialysine (100 mg/mL).
19. Haploid selection #2 plates (select for *MAT*α haploids, TAP-tag, and the barcoded locus): Dissolve 22 g agar in 455.5 mL water and autoclave. Add 100 mL 20% glucose, 100 mL 10× YNB, 100 mL 10× MSG, 100 mL 10× tyrosine (0.5 g/L), 10 mL 100× YC-8aa dropout solution, 10 mL adenine (1 mg/mL), 10 mL tryptophan (10 mg/mL), 100 mL uracil (1 mg/mL), 10 mL methionine (5 mg/mL), 3 mL canavanine (20 mg/mL), 0.5 mL thialysine (100 mg/mL), and 1 mL CloNat (100 mg/mL).
20. 30% glycerol solution (autoclaved) for yeast stocks.

2.3 Buffers

Prepare all buffers in autoclaved Milli-Q water and store buffers that contain protease inhibitors at 4 °C.

1. Buffer A: 2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris–HCl pH 8.0, 1 mM EDTA-NaOH pH 8.0.
2. TE: 10 mM Tris–HCl pH 8.0, 1 mM EDTA.
3. Fix solution: 11% formaldehyde methanol-free, 50 mM Hepes–KOH pH 7.5, 100 mM NaCl, 1 mM EDTA.

4. TBS: 25 mM Tris pH 7.9, 150 mM NaCl, 2.5 mM KCl.
5. Breaking buffer: 100 mM Tris–HCl pH 7.9, 20% glycerol, protease inhibitor cocktail EDTA-free.
6. FA buffer: 50 mM Hepes–KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, protease inhibitor cocktail EDTA-free.
7. PBS pH 7.4: 155 mM NaCl, 3 mM Na_2HPO_4, 1.1 mM KH_2PO_4.
8. FA-HS buffer: 50 mM HEPES-KOH pH 7.5, 500 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate, protease inhibitor cocktail EDTA-free.
9. RIPA: 10 mM Tris-HCl pH 8.0, 250 mM LiCl, 0.5% NP-40, 0.5% Na-deoxycholate, 1 mM EDTA.
10. Elution buffer: 50 mM Tris-HCl pH 8.0, 10 mM EDTA, 1% SDS.

2.4 Reagents

1. Restriction reagents: restriction enzymes *Swa*I and *Bcl*I and corresponding 10× restriction enzyme buffer (e.g., NEBuffer 3.1), 10× SuRE/Cut Buffer M.
2. Ligation reagents: T4 ligase and 10× T4 ligase buffer.
3. PCR reagents: 5× Phusion HF buffer, Phusion high-fidelity DNA polymerase, 10 mM dNTPs, 1 kb plus DNA ladder.
4. Transformation reagents: 50% (w/v) PEG6000, 100 mM LiAc, 1 M LiAc, 5 mg/mL salmon sperm DNA (ss-DNA).
5. DNA extraction reagents: phenol–chloroform–isoamyl alcohol (25:24:1), 4 M ammonium acetate, 100% ethanol.
6. Chromatin immunoprecipitation reagents: 4 M Tris-HCl pH 8.0, 100 mM PMSF, 10 mg/mL RNase A, 10 mg/mL proteinase K, 1 mg/mL Rabbit IgG, 3 M ammonium sulfate, 0.02% sodium azide, protease inhibitor cocktail tablets (complete, EDTA-free, Roche).
7. Kits: DNA gel extraction kit (e.g., QIAquick), PCR purification kit (e.g., QIAquick), plasmid miniprep kit (e.g., PureLink Quick Plasmid Miniprep kit).

2.5 Equipment

1. Hamilton Microlab STAR liquid handling system.
2. Matrix® WellMate®.
3. ROTOR HDA (Singer Instruments).
4. Singer Plusplates™ (Singer Instruments).
5. Singer Repads™ (Singer Instruments).
6. Microplate 384-well polystyrene plates to store glycerol stocks of libraries in and to transfer single colonies into.
7. 384-well microplate aluminum sealing tape.

8. Thermocycler.
9. Tabletop centrifuge.
10. Bead beater.
11. 0.5 mm Zirconia/silica beads (Biospec).
12. 2 mL screwcap tubes for bead beating.
13. Bioruptor® Pico sonication device (Diagenode).
14. 1.5 mL Bioruptor® Pico microtubes with caps (Diagenode).
15. Dynabeads™ M-270 Epoxy (Invitrogen).
16. Dynabeads™ Magnet Particle Concentrator (MPC™).
17. Gel Doc system.

3 Methods

3.1 Design and Construction of the Barcoder Library

3.1.1 Preparation of the Yeast Strain for the Barcoder Library

To select for the barcoded locus during genetic crossing with the TAP-tag library, a selection marker should be placed in the genome proximal to the barcoded locus of interest. We recommend to use strain NKI4212 and to use a dominant drug selection marker such as NatMX or KanMX. This step can be omitted when using diploid cells. However, complementary markers (for example NatMX or HphMX in the barcoder strain and SkHIS3MX6 in the TAP-tagged strain) are required for selection of diploids. The selection marker can be inserted using standard procedures for gene targeting by homologous recombination [26, 27].

3.1.2 Generating the Cas9/gRNA Vector

The next step is to insert a DNA barcode at the locus of interest by using CRISPR-Cas9. For this the pML104 plasmid is used which expresses Cas9 and is manipulated to contain a gRNA targeting the locus of interest.

1. Isolate the pML104 vector from *E. coli* cells with a PureLink™ Quick Plasmid Miniprep Kit (Invitrogen). Make sure to use *E. coli* cells that are Dam methyl transferase deficient to enable cutting with the methylation-sensitive *Bcl*I restriction enzyme.
2. Digest the pML104 vector in a two-step digestion:
   (a) Add 1 μL *Swa*I (10 units), 5 μL 10× NEBuffer 3.1, 1 μg DNA in a total volume of 50 μL and incubate for 15 min at 25 °C.
   (b) Heat-inactivate the digestion for 20 min at 65 °C and then add the second enzyme: 1 μL *Bcl*I (10 units). Incubate 15 min at 50 °C.
3. Purify the digested plasmids by loading them on a 1% agarose gel and isolate them from the gel with a gel purification kit.

4. Hybridize the gRNA oligonucleotides (*see* **Note 5** for design tips) that target the locus of interest. Mix 1 μL oligo1 (100 μM), 1 μL oligo2 (100 μM), 1 μL 10× SuRE/Cut Buffer M and 7 μL water, and hybridize in a thermocycler: 100 °C for 2 min followed by a decrease of 0.1 °C every second to a final temperature of 4 °C.
5. To ligate the gRNA oligonucleotides in the vector backbone, incubate the mixture containing 1 μL 1:50 hybridized oligonucleotides, 100 ng cut vector, 2 μL 10× T4 ligase buffer and 1 μL T4 ligase in a total volume of 20 μL for 1 h at room temperature.
6. Transform 4 μL of the ligation mix in chemically competent *E. coli* cells to amplify the gRNA-containing pML104 vector. Verify insertion of the gRNA with Sanger sequencing.

3.1.3 Preparation of the Barcode-Containing Repair Template

A barcoded repair template, containing sequences flanking the Cas9 induced cut site (*see* Fig. 3), is used to generate a barcoder library. The design should be such that upon integration of the repair template, the Cas9 cut site and/or the PAM motif are absent to avoid recutting by Cas9. Barcoded repair templates can be prepared in various ways. A simple approach is to use a DNA oligonucleotide library as a template for PCR (*see* **Note 6**). This library consists of DNA oligos that have an *n*-bp random barcode sequence, an optional additional unique flanking sequence, and homology arms with the Cas9 cut site (*see* Fig. 3). We recommend using barcodes of 10–20 bp to ensure high complexity. The homology arms can be further extended by including additional homology arms in the primers to amplify the barcoded oligo library and convert it into a double-stranded DNA-repair template. Note that at least one of the sequences flanking the barcode will be used for sequencing of the barcodes in Epi-Decoder analysis. It is therefore important to consider the requirements for the sequencing primer when designing the barcoded locus. More details about the primers and the design can be found in **Note 3**.

1. PCR to amplify the barcode from the DNA oligonucleotide library (5 μL DNA oligonucleotide library (0.01 μM), 10 μL 5× Phusion HF buffer, 1 μL dNTPs (10 mM), 2.5 μL primer 1 (10 μM), 2.5 μL primer 2 (10 μM), 0.5 μL Phusion™ high-fidelity DNA polymerase in a total volume of 50 μL) in a thermocycler: 15 s at 98 °C, 14 cycles (10 s at 98 °C, 25 s at the annealing temperature which depends on the oligo and 20 s at 72 °C) followed by 72 °C for 5 min.
2. Purify the PCR product using a PCR purification kit and check on a 1% agarose gel.

3.1.4 Construction of the Barcoder Library

Transform the NatMX-marked strain with the Cas9/gRNA containing vector and repair template and array the resulting transformants in 384-well plates containing liquid YC media (with MSG instead of ammonium sulfate) lacking uracil and containing CloNat (YC-URA + CloNat). Yeast cells that undergo a Cas9-mediated cut will die unless the double-strand break is repaired by homology-directed repair. Therefore, the vast majority of transformants carrying the Cas9/gRNA vector will have undergone a successful barcode integration. The arrayed transformants together make the barcoder library that can be used to mate with the TAP-tag library to generate the diploid or haploid Epi-Decoder library. Table 1 provides an overview for the required SGA steps in generating the barcoder and Epi-Decoder libraries.

Table 1
Overview of the SGA steps required to generate a barcoder and Epi-Decoder library. For each step, required media, whether the media is solid or liquid, the plate-format and time scale is noted. All steps are carried out at 30 °C except for sporulation, which is done at 23 °C

	Action	Media	Plate	Format	Days	Note
Generating the barcoder library	Array transformants and select for plasmid	YC-URA + CloNat	Liquid	384	1–5	
	Growth without plasmid selection	YEPD + CloNat	Solid	384	1	
	Growth without plasmid selection	YEPD + CloNat	Solid	384	1	
	Select for plasmid loss	CloNat + 5-FOA	Solid	384	3	
	Growth	YEPD + CloNat	Solid	384	1	
	Barcoder library glycerol stocks	YEPD + CloNat	Liquid	384	2	
Generating the Epi-Decoder library	Growth of barcoder and TAP-tag libraries	YEPD + selection	Solid	1536	1	
	Mating	YEPD	Solid	1536	1	
	Diploid selection	YC-HIS + CloNat	Solid	1536	1	
	Diploid selection	YC-HIS + CloNat	Solid	1536	1	
	Presporulation	YEPD	Solid	1536	1	
	Sporulation	SPO	Solid	1536	5	Pin 4×
	Haploid selection 1	Haploid selection #1	Solid	1536	2	Pin 4×
	Haploid selection 1	Haploid selection #1	Solid	1536	1	
	Haploid selection 2	Haploid selection #2	Solid	1536	1	
	Haploid selection 2	Haploid selection #2	Solid	1536	1	
	Epi-Decoder library glycerol stocks	YEPD + CloNat	Liquid	384	2	

1. Grow the NatMX-marked strain overnight in 5 mL YEPD.
2. Next morning, dilute the culture to $OD_{660} = 0.1$ in 15 mL and grow for 5 h to log-phase.
3. Harvest the culture in a 50 mL Eppendorf tube at 840 × *g* for 5 min.
4. Pour off medium and resuspend cells in 10 mL sterile water, spin again.
5. Pour off water and resuspend cells in 1 mL 100 mM LiAc and transfer to a 1.5 mL Eppendorf tube.
6. Pellet cells at top speed for 15 s and remove LiAc.
7. Resuspend cells in 500 μL 100 mM LiAc.
8. Pipette 50 μL sample into a new Eppendorf tube, pellet cells at top speed for 15 s and remove LiAc.
9. Add the transformation mix: 240 μL PEG6000 (50% w/v, *see* **Note 7**), 36 μL 1.0 M LiAc, 35 μL ss-DNA (5 mg/mL, boiled 5 min and chilled on ice), 500 ng Cas9/gRNA containing vector, 3–6 μg repair template (*see* **Note 8**) in a total volume of 360 μL.
10. Vortex the mixture for 1 min.
11. Put the mixture 30 min at 30 °C followed by a 30-min heat shock at 42 °C.
12. Microfuge at 6800 × *g* for 15 s and remove the transformation mix.
13. Resuspend cells in 250 μL sterile water and plate on YC-URA + CloNat plates with glass beads. Aim for approximately 50 colonies per plate to obtain a high number of transformants while avoiding, as much as possible, overlapping colonies since these will result in clones with a mix of barcodes.
14. Grow colonies for 2–4 days at 30 °C.
15. Pick a minimum of 5000 single colonies (*see* **Note 9**) and move them to 384-well "Barcode insertion selection" plates (YC-URA + CloNat media + tetracycline). CloNat and tetracycline are added to prevent growth of contaminating fungi and bacteria.
16. Plate the arrayed strains with the ROTOR on agar plates containing YEPD + CloNat to remove the selection for the *URA3*-containing Cas9/gRNA vector. Replicate the colonies one more time to YEPD + CloNat.
17. Subsequently plate on "Loss of CRISPR plasmid plates" (5-FOA + CloNat) to select for Cas9/gRNA plasmid loss, and subsequently plate on YEPD + CloNat. Confirm that the *URA3* vector was lost by plating on YC-URA media.

18. Transfer the colonies from the solid YEPD + CloNat plates to 384-well liquid plates containing 30 μL YEPD + CloNat and grow yeast to saturation for 2 days.
19. Add 30 μL 30% glycerol and store the libraries at −80 °C for safe storage until Subheading 3.1.5.

3.1.5 Decoding the Barcode Sequences and Locations in the Barcoder Library

The resulting barcoder library requires decoding to identify the coordinates of the individual barcodes. These coordinates are needed to assign individual barcodes to TAP-tagged proteins after crossing the barcoder library with the arrayed TAP-tag library.

1. Use a Hamilton robotics liquid handling workstation (*see* **Note 10**) to make pools of strains that are located in the same row or column (rows and columns of multiple plates can be combined when plates are oriented in a 4 × 4 grid, *see* **Note 11** and Fig. 5). By pooling the strains in this way, each barcode is represented in 1 row pool and 1 column pool. To make pools, take 15 μL culture from each well and collect the pools in reservoirs. Transfer the liquid from the reservoirs to an Eppendorf tube.
2. Extract DNA from each pool using phenol–chloroform–isoamyl alcohol:
   (a) Resuspend pool pellet in 500 μL of water. Transfer cells to a 2 mL tube with screw cap and collect by a 5 s spin. Pour off supernatant.
   (b) Add 200 μL Buffer A.
   (c) Add 200 μL phenol–chloroform–isoamyl alcohol (25:24:1).
   (d) Lyse the cells by bead beating 3 min with 200 μL Zirconia/silica beads.
   (e) Add 200 μL TE and spin the cells for 5 min at maximum speed.
   (f) Transfer the aqueous layer to a new tube.
   (g) Add 1 mL 100% ethanol (room temperature), invert the tube and spin 2 min at 16,200 × *g*.
   (h) Discard supernatant and resuspend pellet in 400 μL TE.
   (i) Precipitate the DNA by adding 10 μL 4 M ammonium acetate, mix and then add 1 mL 100% ethanol (room temperature). Invert the tube to mix and spin 2 min at 16,200 × *g*.
   (j) Air dry the pellet and resuspend in 50 μL TE.
3. Amplify the barcodes from all pools using PCR and make sequencing libraries similar to what is described in Subheading 3.3.5. Make sure to incorporate a pool-specific unique index

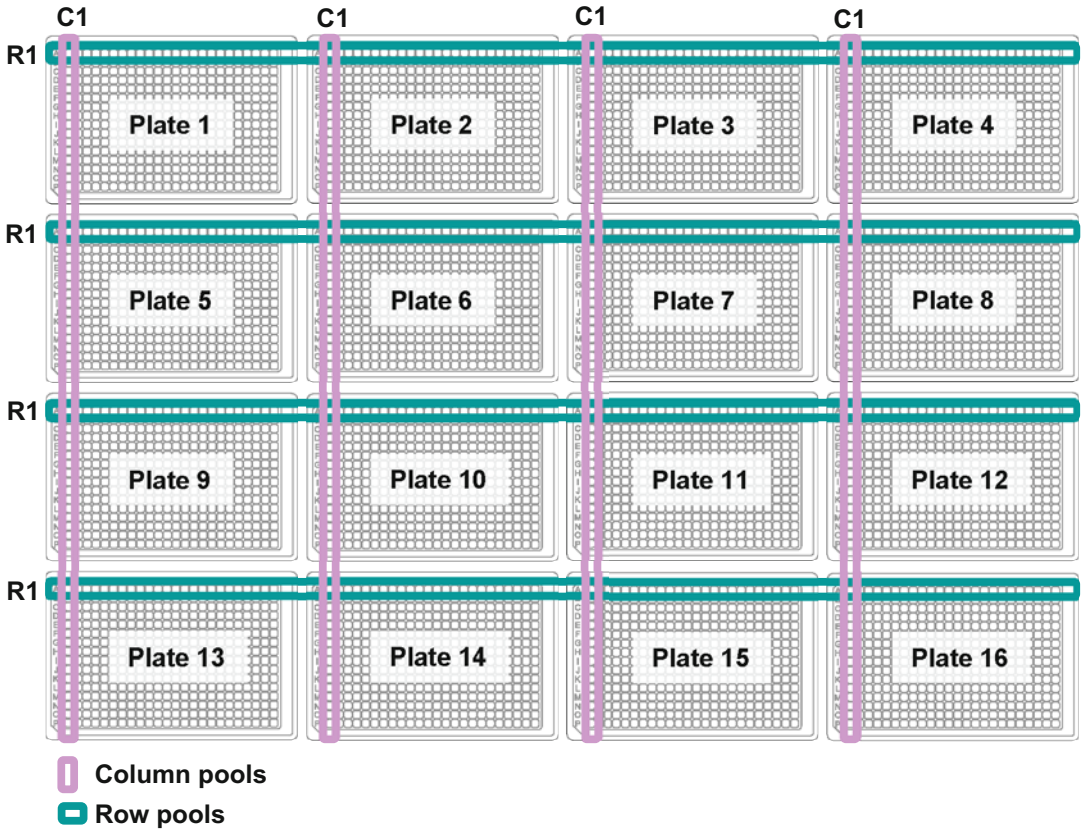


Fig. 5 Schematic overview of a cost-effective pooling strategy. Assume a total of 16 384-well plates. In this example, combine the first columns of all plates (C1) in a pool (pink). Repeat this for C2–C24. Combine the first rows of all plates (R1) in a pool (green) and repeat this for R2-R16. Pool the 384 wells from plate 1 and repeat this for all plates. This will result in a total of 24 + 16 + 16 = 56 pools

sequence in each PCR product. This way, each pool can be identified by its own index and all pools (rows and columns) can be mixed for sequencing (*see* **Note 11**). If the indexed barcoder library PCR products correspond to samples of different complexity (i.e., a different number of clones), the mixing should be adjusted such that the average count for each barcode in the final sequencing sample is the same for every indexed library.

4. Perform high-throughput sequencing and sort reads by their pool-specific indices similar as described in Subheading 3.3.5. By searching an identified barcode in a single-row pool and single-column pool, a coordinate on the original 384-well plate can be appointed to that specific barcode. Reject barcodes that have multiple coordinates and locations to which multiple barcodes are assigned.

5. Rearray the barcoder library with the ROTOR to get rid of barcodes that occur more than once, to eliminate strains that lack barcodes, and to make the barcoder library and TAP-tag library layout compatible.
6. Make new glycerol stocks of the barcoder library in 384-format for long-term storage.

3.2 Generating the Epi-Decoder Library

The Epi-Decoder library is obtained by genetic crossing of the *MAT*α barcoder library and *MAT*a TAP-tag library. This is done using SGA methodology and the ROTOR machine (*see* Table 1).

1. Thaw the 384-well format glycerol stocks of the TAP-tag and barcoder libraries and mix the glycerol stock 5 times with the ROTOR by using the function "wet mixing." It is important that the pins from the ROTOR scrape the bottom of the 384-well plate for equal mixing of the glycerol stocks.
2. Plate fresh copies of both the *MAT*a TAP-tag library and *MAT*α barcoder library on YEPD + selection plates and incubate overnight.
3. Combine the 384-well format plates of the barcoder library in a 1536-well format. Do the same for the TAP-tag library.
4. Mate each of the *MAT*a TAP-tag library plates with a plate of the barcoder *MAT*α strains on YEPD plates and incubate overnight. We recommend to mate the barcoder and TAP-tag plates in three different configurations to obtain three versions of the Epi-Decoder library. Using this approach, each TAP-tagged strain is linked to three different barcodes, which will help to avoid or identify possible barcode-specific effects.
5. Select for diploids by pinning on YC-HIS + CloNat plates and incubate overnight. Repeat this step. Epi-Decoder analysis can be performed with the heterozygous diploid library at this step [5] when ectopic sequences flanking the barcode are used (*see* **Note 3**). To generate haploid Epi-Decoder libraries, proceed to the next step.
6. If diploid selection was performed on synthetic media, plate on presporulation rich media plates (YEPD) and incubate overnight.
7. To generate a haploid library, spot the resulting diploids on sporulation plates and incubate for 5 days at 23 °C in a high humidity box to obtain haploids (*see* **Note 12**).
8. Select for haploids containing the *MAT*α mating type locus and TAP-tag (Haploid selection #1), and subsequently select for *MAT*α haploids containing the barcoded locus and the TAP-tag (Haploid selection #2). Repeat each selection step twice (*see* **Note 13**).

9. Verify the barcodes of a few strains by colony PCR and Sanger sequencing, and the TAP-tags by colony PCR and expression of the TAP-tagged proteins by western blot (*see* **Note 14**). Grow the library in YEPD + CloNat and make glycerol stocks of the Epi-Decoder library in 384-format for long term storage. Storage in the arrayed format will enable retrieving individual clones from the library. At this point, all the clones of a library can also be combined into one liquid pool (*see* Subheading 3.3.1). Make sure to keep the different configurations of the Epi-Decoder libraries separate since the different versions share the same set of barcodes. Aliquots of each pooled library can be stored as glycerol stocks. For stocks, add 1 volume 30% glycerol and store the libraries at −80 °C.

3.3 Epi-Decoder: TAG-ChIP-Barcode-Seq

The resulting Epi-Decoder library is used to systematically decode the local proteome of a single locus of interest. To do so, the Epi-Decoder library is pooled into a single pool which is facilitated by the unique barcode–protein combinations (*see* Figs. 1 and 2). Chromatin immunoprecipitation is performed and after sequencing of the barcodes, IP over input is calculated which is a measure for protein occupancy.

3.3.1 Cross-Linking and Pellet Collection

1. Thaw the 384-well format glycerol stock of the Epi-Decoder library and mix the glycerol stock 5 times with the ROTOR by using the function "wet mixing."
2. Pin the 384-well format Epi-Decoder library on solid YEPD plates before arraying them to a 1536-well format. This prevents cross contamination and allows each colony to grow equally. Grow plates overnight.
3. Rearray the 384-well format Epi-Decoder library to a 1536-well format on solid YEPD plates and grow overnight.
4. Pour 10 mL YEPD media onto each individual plate and use a cell scraper to scrape the cells from the agar. Collect yeast by transferring the liquid to a 50 mL Falcon tube. Repeat this step a few times until the maximum volume of the Falcon tube is reached.
5. Pool collected yeast from one complete Epi-Decoder library in a total volume of 300 mL (depending on the number of plates) YEPD to an OD_{660} of 0.1.
6. Grow liquid cultures until log phase (OD_{660} of ~0.5) while shaking at a speed of 200 rpm in an orbital shaker (New Brunswick Scientific™ Innova™ 42 Incubator Shakers).
7. Cross-link 15 min with 1/10th of volume of freshly prepared Fix solution (*see* **Note 15**).
8. Quench the cross-linking reaction for 1 min with 750 mM Tris-HCl (*see* **Note 16**).

9. Spin the cultures 10 min 3350 × *g* at 4 °C, discard supernatant and wash the pellet in 30 mL cold TBS.
10. Spin down 5 min 3350 × *g* at 4 °C and discard supernatant.
11. Resuspend pellet in 1 mL cold TBS with 0.2 mM PMSF and transfer cell suspension to a bead-beating tube.
12. Spin 15 s maximum speed and remove supernatant.
13. Pellets can be frozen at −80 °C or directly used to proceed to chromatin preparation.

3.3.2 Chromatin Preparation

Keep all reagents and samples on ice during chromatin preparation. Make buffers beforehand and store them on ice.

1. Add 200 µL cold breaking buffer and 500 µL silica/zirconia beads to the cell pellet.
2. Lyse cells (*see* **Note 17**) by bead-beating in the bead beater using cold blocks for 3 min (change blocks after 1.5 min). Check under a microscope if cells have lysed, if not, bead-beat a bit longer.
3. Add 500 µL cold FA buffer to the lysate and invert a couple of times. Let the beads sink to the bottom of the tube.
4. Transfer suspension with broken cells to a new 1.5 mL tube. Repeat washing the lysate with another 500 mL cold FA buffer and combine the samples.
5. Spin 1 min 16,200 × *g* at 4 °C to pellet chromatin and debris. Discard supernatant.
6. Wash pellet again with 1 mL cold FA buffer, spin 1 min 16,200 × *g* at 4 °C and discard supernatant.
7. Add 450 µL cold FA buffer to the pellet and divide the sample over sonication tubes (maximum volume is 300 µL) for sonication.
8. Sonicate using the Diagenode Pico "Bioruptor" for 10 min at 30-s intervals at 4 °C (*see* **Note 18**).
9. Spin 5 min 16,200 × *g* at 4 °C to pellet debris and transfer supernatant (soluble chromatin) to a new 1.5 mL tube. Fill up to a total volume of 1 mL with cold FA buffer.
10. To check DNA fragments after sonication, add 2 µL RNase A (10 mg/mL) and incubate 45 min at 37 °C. Then, reverse cross-links by adding 2 µL proteinase K (10 mg/mL) and incubate overnight at 65 °C. Load the DNA on a 1% agarose gel, the size should be between 100 and 600 bp (*see* Fig. 4).
11. Store chromatin at −80 °C or proceed to IP.

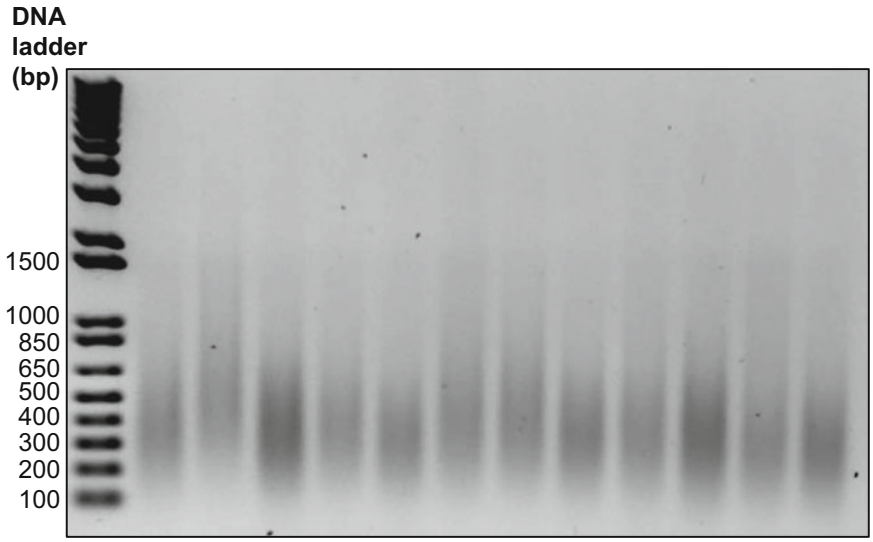


Fig. 4 Example of chromatin size after sonication. Chromatin fragment sizes after 10 min sonication with the Bioruptor Pico from Diagenode (30 s on, 30 s off). Chromatin is sheared to a size between 100 and 600 bp. The displayed DNA ladder is the 1 kb Plus DNA Ladder

3.3.3 Coupling Epoxy-Activated Dynabeads (IgG Dynabeads)

1. Prepare 1 mg/mL of Rabbit IgG in PBS (store at −20 °C for future use).
2. Resuspend 10 mg of epoxy-activated Dynabeads in 600 μL phosphate buffer (0.1 M Na phosphate pH 7.4).
3. Wash twice with 600 μL phosphate buffer.
4. Resuspend beads in 200 μL phosphate buffer.
5. Add 200 μL 1 mg/mL Rabbit IgG and 200 μL 3 M ammonium sulfate.
6. Incubate 24 h on slow-tilt rotation at 37 °C.
7. Place tube on magnet for 4 min and remove supernatant.
8. Wash 4 times 10 min in 1 mL of PBS.
9. Resuspend beads in 1.2 mL of PBS + 0.02% sodium azide.
10. Store at 4 °C for about 2 weeks.

3.3.4 IP, Wash and DNA Clean-Up

1. Add 80 μL IgG Dynabeads to 800 μL chromatin. Take along 80 μL chromatin as input. Rotate both IP and input at 4 °C overnight.
2. Concentrate IP samples on a Dynabeads Magnetic Particle Concentrator. After 30 s, invert twice and wait 30 s. Aspirate off the liquid.
3. Add 950 μL FA buffer and rotate for 5 min at room temperature.

4. Perform the following 5-min wash steps: twice with 950 μL FA buffer, twice with 950 μL FA-HS buffer, twice with 950 μL RIPA and once with 950 μL TE.
5. Resuspend beads in 100 μL elution buffer and incubate 10 min at 65 °C while shaking.
6. Collect the DNA containing elution buffer using the magnet.
7. Add 70 μL TE to IP and input samples and 20 μL elution buffer to the input samples. Treat IP and input samples with 0.5 μL RNase A (10 mg/mL) and incubate 45 min at 37 °C.
8. Add 10 μL Proteinase K (10 mg/mL) and incubate at 65 °C overnight to reverse cross-links.
9. Purify DNA using the QIAquick PCR purification kit (Qiagen) and elute in 30 μL elution buffer.
10. Store DNA at −20 °C.

3.3.5 Library Preparation and Barcode Counting

1. Mix the following reaction: 15 μL DNA (undiluted for IP, 1:50 for input), 8 μL High-Fidelity fusion buffer (ThermoFisher), 0.4 μL dNTPs (10 mM), 0.4 μL Phusion High-Fidelity DNA Polymerase (ThermoFisher), 0.4 μL primer 1 (locus specific sequence primer with the p5 adapter, 10 μM), 0.4 μL primer 2 (pool specific index primer with the p7 adapter, 10 μM) in a total volume of 40 μL. Amplify the barcodes by cycling through the following program: 30 s at 98 °C, 18 cycles (15 s at 98 °C, 20 s at 55 °C, 20 s at 72 °C) and 5 min at 72 °C (*see* **Note 19**). The exact cycling conditions might vary between loci. Note that every IP and input sample should be amplified with a different index primer.
2. Check a small aliquot (5 μL) on a 1.5% agarose gel and generate a picture using the Gel Doc.
   (a) Bands visible: quantify the signal intensity in each lane. If the bands are faint, put a larger volume on a new agarose gel.
   (b) Bands not visible: put samples back in the PCR machine for 5 more cycles and check again on an agarose gel (*see* **Note 20**).
3. Quantify the signal intensity of each sample by using Image Lab, or a similar software. Perform local background subtraction to create background adjusted intensity values. Alternatively, DNA can also be quantified on a Bioanalyzer after cleaning the PCR product.
4. Mix the indexed PCR products in equimolar fashion and consider sample complexity if necessary (*see* **step 3** in Subheading 3.1.5) and run on a 1.5% agarose gel. Excise the DNA fragment (100–150 bp) from the gel and extract with a QIAquick gel extraction kit (Qiagen).

5. Quantify and sequence the purified DNA (single read, >50 bp) on a HiSeq2500/MiSeq or similar platform using one or a mix of custom sequencing primers (*see* Fig. 3).

Extract the barcodes from the sequencing reads using the PerI scripts eXtracting Counting and Linking to Barcode References (XCALIBR). Code and detailed description of XCALIBR are available at https://github.com/NKI-GCF/xcalibr. The output contains a table with counts for each barcode–index combination.

Further processing of the data could include removing barcodes with low counts (e.g., <10), correcting plate-specific differences in counts by normalizing each plate by its median, log2 transform the counts table, matching ORF names to barcode–index combinations or removing factors with low input counts.

4 Notes

1. In this protocol, the commercially available TAP-tag library [20] is used. This library is *MAT*a and crossed with a *MAT*α barcoder library. If preferred, it is possible to use a *MAT*a barcoder library instead of *MAT*α. In this case, the TAP-tag library needs to be converted from *MAT*a to *MAT*α [5] using NKI4212. Note that any other tagged library can be used, as long as the tag can be used for immunoprecipitation.
2. The TAP-tag library has been described to contain the HIS3MX6 selection marker flanking each TAP-tagged gene [20]. HIS3MX6 has been reported to contain the *A. gossypii TEF* promoter and terminator driving expression of the *S. pombe his5* gene [30]. By sequencing several TAP-tagged gene cassettes from the TAP-tag library we determined that the heterologous *HIS3* gene used for the TAP-tag library is derived from *Saccharomyces kluyveri* (*Lachancea kluyveri*) instead of *S. pombe*. Therefore, we refer to the selection marker as SkHIS3MX6. The sequence of the SkHIS3 gene is as follows (5′–3′): ATGGCAGAACCAGCCCAAAAAAAGCAAAAA
CAAACTGTTCAGGAGCGCAAGGCGTTTATCTCCCG
TATCACTAATGAAACTAAAATTCAAATCGC
TATTTCGCTGAATGGTGGTTATATTCAAATAAAA
GATTCGATTCTTCCTGCAAAGAAGGATGACGATG
TAGCTTCCCAAGCTACTCAGTCACAGGTCATCGA
TATTCACACAGGTGTTGGCTTTTTGGATCATATGATC
CATGCGTTGGCAAAACACTCTGGTTGGTCTCT
TATTGTTGAATGTATTGGTGACCTGCACATTGACGAT
CACCATACTACCGAAGATTGCGGTATCGCATTAGGG
CAAGCGTTCAAAGAAG
CAATGGGTGCTGTCCGTGGTGTAAAAAGATTCGG
TACTGGGTTCGCACCATTGGATGAGGCGCTAT

CACGTGCCGTAGTCGATTTATCTAATAGAC
CATTTGCTGTAATCGACCTTGGATTGAAGAGAGAGAT
GATTGGTGATTTATCCACTGAAATGATTCCA
CACTTTTTGGAAAGTTTCGCGGAGGCGGCCAGAAT
TACTTTGCATGTTGATTGTCTGAGAGGTTTCAACGAT
CACCACAGAAGTGAGAGTGCGTT
CAAGGCTTTGGCTGTTGCCATAAGAGAAGCTATTTC
TAGCAATGGCACCAATGACGTTCCCTCAAC
CAAAGGTGTTTTGATGTGA

3. There are several important considerations to take into account for the design of the barcoded locus and the design of the sequence library preparation.
   (a) For the design of a barcoded locus, two different approaches can be taken, either inserting a DNA barcode flanked by ectopic sequences at the locus of interest or only a DNA barcode embedded in the native context (Fig. 3). Using native sequences minimizes the disruption of the genomic locus, but using ectopic sequences can add several benefits to the design. First, ectopic flanking sequences can facilitate allele-specific amplification of the barcodes in diploid cells, in which the wild-type allele lacks the barcode and the ectopic sequences. Second, if the locus of interest is flanked by native sequences with a low melting temperature (Tm), ectopic sequences can be used to insert sequences with a higher melting temperature, which is required for efficient PCR and library sequencing (see also below). Third, if the barcode is flanked by a native genomic sequence, the sequencing of the libraries will have to be performed with custom sequencing primers. If one opts for ectopic sequence flanks of which one corresponds to the standard Illumina ReadSeq primer (or the 3′ part of it), this will allow for using the standard Illumina sequencing primer during high throughput sequencing.
   (b) When choosing the precise location of the barcode at the locus of interest, note that at least one of the sequences flanking the barcode will correspond to (part of) the sequencing primer in high-throughput barcode sequencing (*see* Fig. 3). The flanking sequence will therefore correspond to the "ReadSeq" primer or at least the 3′ part of it (which is extended to a full sequencing primer in the library PCR). The 5′ part of the sequencing primer is provided in the P5 primer that is used to amplify the barcodes and generate the sequencing libraries. Together, the 5′ + 3′ part form the sequencing primer. This sequencing primer can be the Illumina sequencing primer (5′ AC ACTCTTTCCCTACACGACGCTCTTCCGATCT 3′)

or a custom sequencing primer that should have properties that match those of the Illumina sequencing primer as closely as possible: $T_m = 66$ °C, 33 bp, 52% GC, and have a low risk of forming secondary structures. In our experience, samples with different sequencing primers can be multiplexed in one lane or flow cell. We observed that when the GC content and Tm of the flanking sequence is low, extending the sequencing primer to increase the Tm can give successful sequencing results.

(c) Epi-Decoder sequencing libraries are generally generated by a two-step PCR protocol. The first part involves amplification of the barcodes using annealing of only the 3′ ends of the long primers. These short ends have a low Tm. Once product has been formed, it can be amplified by annealing of the full-length oligo, for which a higher Tm can be used. If one or both of the native sequences flanking the barcode corresponds to a primer with low Tm/GC-content, this standard protocol may not work efficiently. In our experience, adding to the PCR mix two shorter primers (*see* Fig. 3, we call these primers P1 and P2 Nextera) corresponding to the most 5′ ends of the long primers allows for using annealing conditions at lower temperature, which can boost the PCR efficiency.

4. The plates for the ROTOR machine should contain a volume of 50 mL media. When pouring these plates, use a pipette to prevent the formation of bubbles. Make sure to pour and dry the plates in an environment where the airflow is minimal to avoid contamination and unequal drying. Prevent over-drying these plates. If the media does not dry up in a leveled surface, the ROTOR can have troubles pinning colonies.

5. When designing a gRNA targeting a locus of interest, make use of multiple online tools that help to minimize off-target effects and increase cutting efficiency. Always confirm the cutting efficiency. In addition, when designing the gRNA sequences for the pML104 vectors, include the following sequences flanking the gRNA sequence (N_{20}):

   (a) Forward (5′ → 3′): GATCNNNNNNNNNNNNNNNNNNNGTTTTAGAGCTAG

   (b) Reverse (5′ → 3′): CTAGCTCTAAAACNNNNNNNNNNNNNNNNNNNN

6. A DNA oligonucleotide library can be ordered at the company IDT (the cheapest option is to order a 4 nM Ultramer DNA Oligo) but other options and companies are also possible. Make sure to enter a 10–20 bp random barcode in the repair template by adding 10–20 "N" symbols (*see* Fig. 3). Dilute the library to 100 μM using Milli-Q water.

7. Prepare the PEG6000 50% w/v solution fresh since evaporation can cause the percentage to change which negatively impacts the transformation efficiency.
8. Determine the optimal amount of repair template to gain a high transformation efficiency. Do not exceed the total volume of the transformation mix (360 μL).
9. Barcodes are coupled to TAP-tagged proteins. The TAP-tag library contains approximately 5000 TAP-tagged proteins, therefore a minimum of 5000 barcodes is required in theory. In our hands, approximately 80% of transformed strains have a unique integrated barcode at the desired locus. This depends on the CRISPR-Cas9 cutting and repair efficiency. In addition, not all barcodes will be sequenced as efficiently or may show too much sequence resemblance to other barcodes. Thus, in practice, pick an excess of colonies (i.e., 6000) to reach the necessary barcode complexity.
10. Other automated systems also work. Alternatively, pools can be made by hand but this requires substantial efforts.
11. Using unique index sequences for each row and column requires a large number of i7-indexed oligos. In case of 16 384-well plates, by forming a grid, this will result in 64 row pools and 96 column pools. The intersection between row number and column number will determine the location of each barcode in the grid. To decode the barcoder library in a more cost-effective way, it is possible to reduce the number of required indexed oligos by combining all common row numbers and common column numbers in pools and adding pools of each plate (*see* Fig. 5). In case of 16 384-well plates, by forming a grid, this will result in 16 common row number pools and 24 common column number pools and 16 plate number pools. The intersection between common row, column and plate number will determine the location of each barcode in the grid. This latter approach also involves fewer PCR library preparations. The number of indexed pools and index primers can be further reduced by using i5 and i7 dual index strategies. Such strategies will usually involve additional sequencing cycles beyond the single read run for the single index libraries.
12. When replicating on sporulation plates, pin 4 times on the same plate using the ROTOR. Do not cover the plates in plastic wrap (which has been reported to negatively affect spore formation) but put the plates in a humidified incubator with wet towels around them. A plastic box with wet towels inside a regular incubator can be used as an alternative.
13. Pin 4 times from the SPO plates to the first haploid selection plates to transfer as many cells as possible. During and after

haploid selection, if colonies appear too small, the plates should be left to grow for an additional day. This may help eliminating unnecessary loss of colonies.

14. Keep in mind that the maximum PCR product size from a "quick and dirty" colony PCR is approximately 1000 bp. Therefore, design primers accordingly.
15. The optimal cross-linking time differs per protein. Cross-linking too short will cause only strong chromatin binders to cross-link. On the other hand, over cross-linking can cause proteins that normally do not bind the chromatin to cross-link. A good starting point is to cross-link 15–20 min. Make sure to use freshly prepared Fix solution since formaldehyde loses its stability quickly after opening the stock solution. We also recommend to use methanol-free formaldehyde to avoid membrane permeabilization during protein–DNA cross-linking.
16. Recent studies indicate that glycine is a poor quencher of formaldehyde [31]. As an alternative, Tris–HCl can be used. To prevent reversing cross-links by Tris, quench for 1 min.
17. The maximum number of cells for the bead-beating process is 1.5×10^9.
18. The maximum volume of the sonication tubes is 300 μL. If needed, divide samples over multiple sonication tubes and pool them afterward.
19. The goal is to obtain a band on a gel that can be quantified using as few cycles as possible to prevent PCR bias and jackpot effects. To do so, use as much DNA as possible and keep the number of PCR cycles to a minimum. The exact PCR conditions for a library preparation depend on the sequence context of the integrated barcodes and will therefore have to be optimized for each barcoded locus.
20. If bands are not visible when using more than 25 cycles, repeat the PCR, but now use only 5 μL DNA. It is possible that chemicals that came from the purification columns are inhibiting the PCR reaction.

Acknowledgments

We would like to thank Tessy Korthout, Deepani Poramba-Liyanage, and Ila van Kruijsbergen for developing and optimizing the Epi-Decoder approach, and Tibor van Welsem and Desiré García Pichardo for reading the manuscript and helpful discussions. This work was supported by ZonMW TOP grant 91218022 and Dutch Research Council grant NWO-VICI-016.130.627 to F.v.L.

References

1. Gauchier M, van Mierlo G, Vermeulen M, Déjardin J (2020) Purification and enrichment of specific chromatin loci. Nat Methods 17: 380–389. https://doi.org/10.1038/s41592-020-0765-4
2. Wierer M, Mann M (2016) Proteomics to study DNA-bound and chromatin-associated gene regulatory complexes. In: Human molecular genetics, vol 25. Oxford University Press, Oxford. https://doi.org/10.1093/hmg/ddw208
3. Kan SL, Saksouk N, Déjardin J (2017) Proteome characterization of a chromatin locus using the proteomics of isolated chromatin segments approach. Methods Mol Biol 1550: 19–33. https://doi.org/10.1007/978-1-4939-6747-6_3
4. Poramba-Liyanage DW, Korthout T, Cucinotta CE, van Kruijsbergen I, van Welsem T, El Atmioui D, Ovaa H, Tsukiyama T, van Leeuwen F (2020) Inhibition of transcription leads to rewiring of locus-specific chromatin proteomes. Genome Res 30:635. https://doi.org/10.1101/gr.256255.119
5. Korthout T, Poramba-Liyanage DW, van Kruijsbergen I, Verzijlbergen KF, van Gemert FPA, van Welsem T, van Leeuwen F (2018) Decoding the chromatin proteome of a single genomic locus by DNA sequencing. PLoS Biol 16:e2005542. https://doi.org/10.1371/journal.pbio.2005542
6. Shendure J, Lieberman Aiden E (2012) The expanding scope of DNA sequencing. Nat Biotechnol 30:1084–1094. https://doi.org/10.1038/nbt.2421
7. Kebschull JM, Zador AM (2018) Cellular barcoding: lineage tracing, screening and beyond. Nat Methods 15:871. https://doi.org/10.1038/s41592-018-0185-x
8. Liszczak G, Muir TW (2019) Nucleic acid-barcoding technologies: converting DNA sequencing into a broad-spectrum molecular counter. Angew Chem Int Ed 58:4144. https://doi.org/10.1002/anie.201808956
9. Trauernicht M, Martinez-Ara M, van Steensel B (2020) Deciphering gene regulation using massively parallel reporter assays. Trends Biochem Sci 45:90–91. https://doi.org/10.1016/J.TIBS.2019.10.006
10. Mans R, Wijsman M, Daran-Lapujade P, Daran J-M (2018) A protocol for introduction of multiple genetic modifications in Saccharomyces cerevisiae using CRISPR/Cas9. FEMS Yeast Res 18:foy063. https://doi.org/10.1093/femsyr/foy063
11. Malcı K, Walls LE, Rios-Solis L (2020) Multiplex genome engineering methods for yeast cell factory development. Front Bioeng Biotechnol 8:1264. https://doi.org/10.3389/fbioe.2020.589468
12. Jaffe M, Sherlock G, Levy SF (2017) iSeq: a new double-barcode method for detecting dynamic genetic interactions in yeast. G3 7: 143–153. https://doi.org/10.1534/g3.116.034207
13. Poramba-Liyanage DW, Korthout T, van Leeuwen F (2019) Epi-ID: systematic and direct screening for chromatin regulators in yeast by barcode-ChIP-Seq. Methods Mol Biol 2049: 87–103. https://doi.org/10.1007/978-1-4939-9736-7_5
14. Vlaming H, Molenaar TM, van Welsem T, Poramba-Liyanage DW, Smith DE, Velds A, Hoekman L, Korthout T, Hendriks S, Altelaar AFM, van Leeuwen F (2016) Direct screening for chromatin status on DNA barcodes in yeast delineates the regulome of H3K79 methylation by Dot1. elife 5:e18919. https://doi.org/10.7554/eLife.18919
15. Yofe I, Weill U, Meurer M, Chuartzman S, Zalckvar E, Goldman O, Ben-Dor S, Schütze C, Wiedemann N, Knop M, Khmelinskii A, Schuldiner M (2016) One library to make them all: streamlining the creation of yeast libraries via a SWAp-Tag strategy. Nat Methods 13:371–378. https://doi.org/10.1038/nmeth.3795
16. Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK (2003) Global analysis of protein localization in budding yeast. Nature 425:686–691. https://doi.org/10.1038/nature02026
17. Weill U, Yofe I, Sass E, Stynen B, Davidi D, Natarajan J, Ben-Menachem R, Avihou Z, Goldman O, Harpaz N, Chuartzman S, Kniazev K, Knoblach B, Laborenz J, Boos F, Kowarzyk J, Ben-Dor S, Zalckvar E, Herrmann JM, Rachubinski RA, Pines O, Rapaport D, Michnick SW, Levy ED, Schuldiner M (2018) Genome-wide SWAp-Tag yeast libraries for proteome exploration. Nat Methods 15: 617–622. https://doi.org/10.1038/s41592-018-0044-9
18. Douglas AC, Smith AM, Sharifpoor S, Yan Z, Durbic T, Heisler LE, Lee AY, Ryan O, Gottert H, Surendra A, van Dyk D, Giaever G, Boone C, Nislow C, Andrews BJ (2012) Functional analysis with a barcoder yeast gene overexpression system. G3 2(10):1279–1289. https://doi.org/10.1534/g3.112.003400

19. Suter B, Fontaine J-F, Yildirimman R, Raskó T, Schaefer MH, Rasche A, Porras P, Vázquez-Álvarez BM, Russ J, Rau K, Foulle R, Zenkner M, Saar K, Herwig R, Andrade-Navarro MA, Wanker EE (2013) Development and application of a DNA microarray-based yeast two-hybrid system. Nucleic Acids Res 41:1496–1507. https://doi.org/10.1093/nar/gks1329
20. Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. Nature 425:737–741. https://doi.org/10.1038/nature02046
21. Kuzmin E, Rahman M, VanderSluis B, Costanzo M, Myers CL, Andrews BJ, Boone C (2021) tau-SGA: synthetic genetic array analysis for systematically screening and quantifying trigenic interactions in yeast. Nat Protoc 16(2):1219–1250. https://doi.org/10.1038/s41596-020-00456-3
22. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H, Andrews B, Tyers M, Boone C (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science (New York, NY) 294:2364–2368. https://doi.org/10.1126/science.1065810
23. Yan Z, Costanzo M, Heisler LE, Paw J, Kaper F, Andrews BJ, Boone C, Giaever G, Nislow C (2008) Yeast Barcoders: a chemogenomic application of a universal donor-strain collection carrying bar-code identifiers. Nat Methods 5:719–725. https://doi.org/10.1038/nmeth.1231
24. Tong AH, Boone C (2006) Synthetic genetic array analysis in Saccharomyces cerevisiae. Methods Mol Biol 313:171–192. https://doi.org/10.1385/1-59259-958-3:171
25. Verzijlbergen KF, Menendez-Benito V, van Welsem T, van Deventer SJ, Lindstrom DL, Ovaa H, Neefjes J, Gottschling DE, van Leeuwen F (2010) Recombination-induced tag exchange to track old and new proteins. Proc Natl Acad Sci U S A 107(1):64–68. https://doi.org/10.1073/pnas.0911164107
26. Gietz RD, Woods RA (2006) Yeast transformation by the LiAc/SS carrier DNA/PEG method. Methods Mol Biol 313:107–120. https://doi.org/10.1385/1-59259-958-3:107
27. Cold Spring Harbor Laboratory (2015) Methods in yeast genetics and genomics: a CSHL course manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. ISBN 978-1-621821-34-2
28. Laughery MF, Hunter T, Brown A, Hoopes J, Ostbye T, Shumaker T, Wyrick JJ (2015) New vectors for simple and streamlined CRISPR-Cas9 genome editing in Saccharomyces cerevisiae. Yeast 32(12):711–720. https://doi.org/10.1002/yea.3098
29. van Leeuwen F, Gottschling DE (2002) Assays for gene silencing in yeast. Methods Enzymol 350:165–186
30. Wach A, Brachat A, Alberti-Segui C, Rebischung C, Philippsen P (1997) Heterologous HIS3 marker and GFP reporter modules for PCR-targeting in Saccharomyces cerevisiae. Yeast 13(11):1065–1075. https://doi.org/10.1002/(SICI)1097-0061(19970915)13:11<1065::AID-YEA159>3.0.CO;2-K
31. de Jonge WJ, Brok M, Kemmeren P, Holstege FCP (2019) An extensively optimized chromatin immunoprecipitation protocol for quantitatively comparable and robust results. bioRxiv:835926. https://doi.org/10.1101/835926

Chapter 9

A Protocol for Studying Transcription Factor Dynamics Using Fast Single-Particle Tracking and Spot-On Model-Based Analysis

Asmita Jha and Anders S. Hansen

Abstract

Single-particle tracking (SPT) makes it possible to directly observe single protein diffusion dynamics in living cells over time. Thus, SPT has emerged as a powerful method to quantify the dynamics of nuclear proteins such as transcription factors (TFs). Here, we provide a protocol for conducting and analyzing SPT experiments with a focus on fast tracking ("fastSPT") of TFs in mammalian cells. First, we explore how to engineer and prepare cells for SPT experiments. Next, we examine how to optimize SPT experiments by imaging at low densities to minimize tracking errors and by using stroboscopic excitation to minimize motion-blur. Next, we discuss how to convert raw SPT data into single-particle trajectories. Finally, we illustrate how to analyze these trajectories using the kinetic modeling package Spot-On. We discuss how to use Spot-On to fit histograms of displacements and extract useful information such as the fraction of TFs that are bound and freely diffusing, and their associated diffusion coefficients.

Key words Single-particle tracking, Transcription factors, Live-cell imaging, Fluorescence microscopy, spaSPT, Spot-On, Diffusion, Single-particle trajectories, Single-molecule, Diffusion coefficient

1 Introduction

DNA-binding proteins such as transcription factors (TFs) play key roles in essentially all nuclear processes including gene regulation, DNA repair, and replication. TFs diffuse throughout the nucleus as they search for and bind their cognate DNA binding sites and recruit cofactors, chromatin remodelers, and general transcriptional machinery before dynamically dissociating from chromatin to begin a new cycle [1] (Fig. 1). Much of our current understanding of TFs has come from structural, biochemical, and genomics approaches. For example, structural methods such as cryo-EM have revealed how DNA-binding domains interact with DNA at atomic resolution, biochemical reconstitution approaches have revealed hierarchical and sequential binding of the general transcription

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_9,

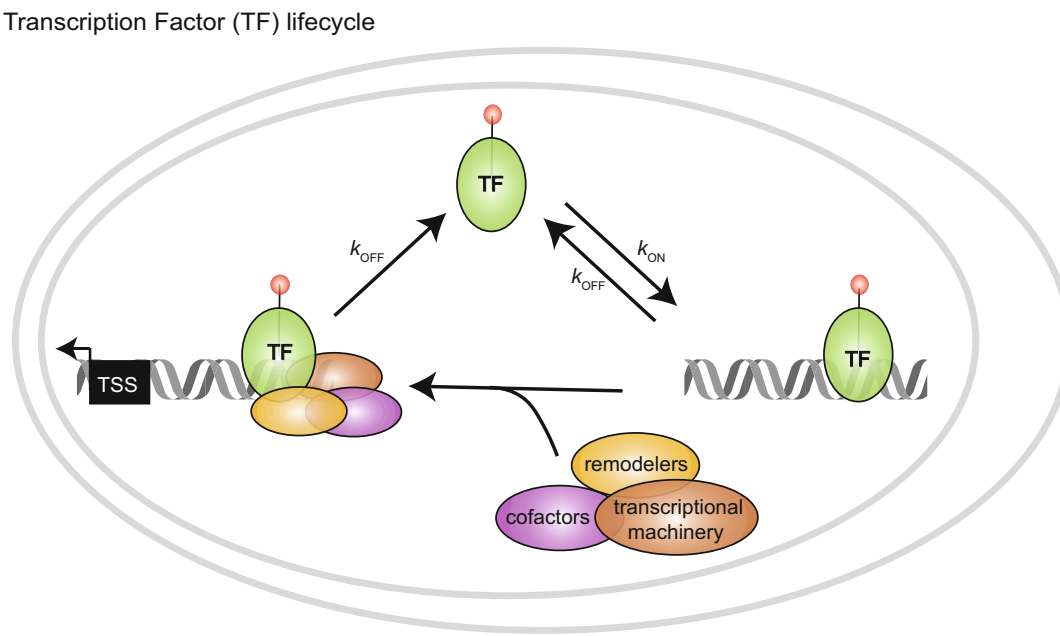


Fig. 1 Outline of the dynamic life cycle of TFs. TFs undergo a dynamic life cycle inside the nucleus and can exist in multiple states. They diffuse, search for and bind to cognate DNA-binding sites, recruit cofactors and the general transcriptional machinery, and dissociate in search for the next DNA-binding site

factors, and genomic studies such as ChIP-Seq have shown where in the genome TFs bind [2]. However, many aspects of the dynamic TF life cycle inside living cells such as diffusion, target search mechanisms, DNA residence times, and clustering cannot be captured with these static, single snapshot approaches. Since understanding TF dynamics is essential for understanding TF regulation and function, live-cell imaging has thus emerged as a powerful tool to overcome these limitations and to track the real-time kinetics of a TF's dynamic life cycle.

Early work using live-cell imaging methods such as fluorescence recovery after photobleaching (FRAP) and fluorescence correlation spectroscopy (FCS) revealed DNA-binding of nuclear proteins to be highly dynamic [3–5]. In FRAP, a region of interest is photobleached and the rate of fluorescence recovery to the region of interest is subsequently observed. By monitoring how quickly bleached proteins exit the photobleached region and are replaced by unbleached proteins, dynamic protein parameters like diffusion coefficients and residence times can be estimated [6]. For example, a stably DNA-bound protein would be replaced at a slower rate and thus exhibit a slow FRAP recovery. FCS, on the other hand, measures the change in fluorescence in a small volume of interest. By analyzing the temporal correlation in fluorescence fluctuations and fitting kinetic models, one may infer diffusion coefficients, TF concentration, DNA binding and other parameters [7]. However,

because both FRAP and FCS probes bulk TF diffusion, target search, DNA binding, and DNA unbinding for many of TF molecules simultaneously, analysis of FRAP and FCS data requires complex reaction–diffusion modeling. Previous work and benchmarking approaches have demonstrated that conceptually distinct FRAP and FCS models sometimes fit experimental data equally well, which can make it challenging to quantitatively interpret FRAP and FCS data [5, 6, 8, 9].

Single-particle tracking (SPT) overcomes these limitations by enabling direct observation of individual fluorescently labelled proteins in single cells in real time [10]. In SPT, TFs are localized in each frame and then connected across frames to form trajectories. Through analysis of these SPT trajectories, we can then separate proteins into subpopulations based on their distinct diffusive behaviors, thus illuminating each aspect of the TF life cycle (Fig. 1) [1]. For example, since chromatin is a slow-moving scaffold, DNA binding of TFs can be observed as a change in the diffusion coefficient from a freely diffusing state ($D \sim 1$–$10\ \mu m^2/s$ for most TFs) to a slow-moving bound state ($D \sim 0.001$–$0.05\ \mu m^2/s$). Furthermore, by following the DNA-bound TFs over time, the residence time can be estimated [8, 11–13]. Once the bound fraction and residence time have been determined, the TF search time, how long a TF searches on average for a cognate site, can be calculated [14]. Moreover, anomalous diffusion and TF clustering can be inferred [15]. As such, SPT makes it possible to directly observe and quantify each aspect of the TF life cycle in living cells.

Recent applications of SPT have revealed how anomalous diffusion and transient trapping by protein clusters accelerate the TF target search mechanism [16] and suggested that longer TF residence times result in higher transcriptional output [17, 18]. Other SPT applications have focused on specific protein(s) such as the preinitiation complex assembly [19], TALEN and Cas9 nucleases [20], and the Polycomb proteins [21, 22]. Other SPT studies have quantified TF binding in in mitosis [23, 24] and how low-complexity domains affect TF dynamics [25]. Finally, SPT approaches have now matured to the point where single TF tracking inside living *Drosophila* and mouse embryos is possible [26].

At a high level, SPT methods applied to TFs and related proteins fall into at least three classes: "fastSPT," "slowSPT," and "all-in-one SPT." "fastSPT" approaches such as single particle tracking photoactivated localization microscopy (sptPALM) [27] and stroboscopic photoactivation SPT (spaSPT) [28] utilize imaging at high frame rates (~50–250 Hz) to track both bound and fast-diffusing TFs. Analysis of "fastSPT" data can reveal diffusion mechanisms, bound fractions, the number of diffusive states and more, but photobleaching rates are generally too high to infer residence times. Second, "slowSPT" uses long-exposure times to blur out fast-diffusing proteins and selectively focuses on slow-

diffusing, presumably chromatin-bound TFs [11, 29, 30]. Thus, slowSPT makes it possible to measure the residence time of the DNA-bound subpopulation, but cannot report on fast-diffusing subpopulations. "All-in-one SPT" approaches combine short exposures with variable dark times to attempt to simultaneously quantify the entire TF life-cycle including diffusion, number of states, and residence time [8, 12, 30, 31].

Here, we focus on "fastSPT," specifically spaSPT experiments. We will discuss how to optimize experimental and acquisition parameters, and how to analyze the resulting SPT data using Spot-On, a kinetic modeling framework that makes it possible to extract diffusion coefficients, the number of diffusive states, and the bound fraction from single-particle trajectories acquired from SPT experiments [28]. SPT experiments have four key steps: (1) cell preparation, (2) imaging, (3) trajectory generation, and (4) trajectory analysis (Fig. 2).

The first step of an SPT experiment is cell preparation. To be able to track single proteins, we must achieve sparse and bright fluorescent labeling. Typically, a TF is tagged as a genetically encoded fusion protein. Here, endogenous tagging using genome-editing is preferable, since it can avoid artifacts often associated with transient overexpression [14, 32]. Traditional fluorescent proteins such as GFP are not well-suited for SPT since SPT requires sparsity. Instead, photoswitchable proteins such as mEos and Dendra or self-labeling tags such as SNAP-Tag or HaloTag are preferred [27, 31]. HaloTag combined with bright organic dyes is the most popular approach since it combines superior photostability and brightness with high specificity and control over labeling density. Controlling labeling density is essential; if too few in-focus proteins are labeled, we obtain no trajectories, but if too many are

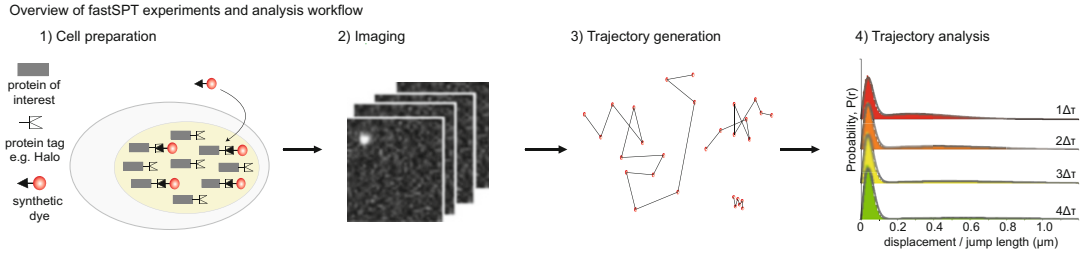


Fig. 2 Overview of the key steps involved in conducting a "fastSPT" experiment and analyzing the data using Spot-On. A fastSPT experiments has four main steps. (1) Cell preparation: cells expressing a tagged protein of interest are labeled with a synthetic dye; (2) Imaging: fluorescence microscopy is then used to observe the movement of single labeled proteins (this figure was adapted from Video 2 from ref. 28 with permission). (3) Trajectory generation: particles are localized in each frame of the movies and tracked across frames to obtain SPT trajectories; (4) Trajectory analysis: SPT trajectories are analyzed using Spot-On to extract information about the diffusion coefficients and the bound and free subpopulations (shown: simulated SPT data with 50% bound and 50% free with $D_{\mathrm{FREE}} = 4\ \mu m^2/s$ at 100 Hz)

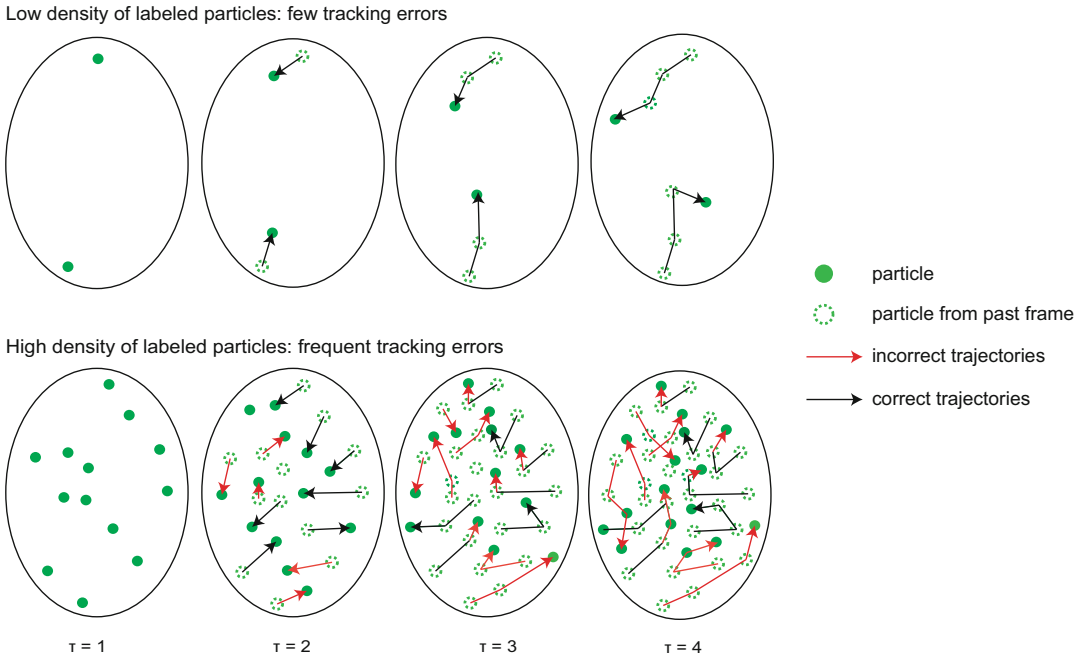


Fig. 3 High particle densities result in frequent tracking errors (misconnections). Top panel: at low particle densities, particle trajectories can be clearly distinguished resulting in few misconnections. Bottom panel: at high particle densities, particle trajectories frequently overlap resulting in tracking errors (misconnections shown in red) when localizations are connected across frames during the tracking step

labeled, their paths will cross which leads to tracking errors (Fig. 3). Utilizing the HaloTag together with cell-permeable dyes such as Janelia Fluor (JF) dyes make it possible to control labeling density in two ways (Fig. 4) [33–35]. First, if "regular" JF dyes are used such as JF_{549} or JF_{646} [34], one can obtain a desired labeling density by titrating labeling time (typically 15–30 min) and dye concentration (typically ~1 pM to 5 nM depending on TF expression level). Second, one can control density using photoactivatable JF dyes, such as PA-JF_{549} and PA-JF_{646} [35] which only become fluorescent upon photoactivation using 405 nm illumination. With these dyes, one typically uses a higher labeling density (typically ~5 nM to 100 nM depending on TF expression level) to label many TFs and photoactivates a small fraction. The use of PA-dyes is recommended since it makes it possible to track TFs at very low densities such that tracking errors are minimized (Fig. 3) and facilitates simultaneous acquisition of thousands of trajectories by continuously photoactivating new subsets of TFs to compensate for photobleaching [27, 28]. With "regular" JF dyes one generally faces a hard trade-off between low density (few trajectories, few tracking errors) and high density (many trajectories, many tracking errors). However, PA-JF dyes are less cell-permeable, less chemically stable, and more prone to labeling artifacts especially

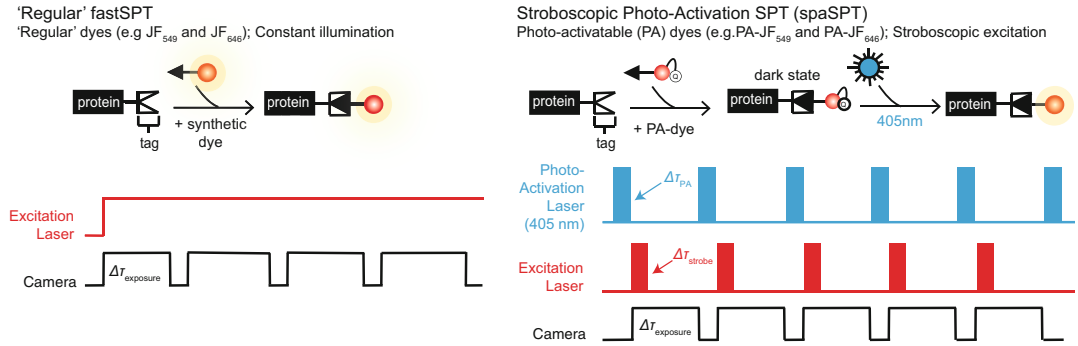


Fig. 4 Overview and comparison of fastSPT with "regular" dye and spaSPT. Left: overview of "regular" fastSPT. Here, the protein of interest is labeled with a regular dye that is continuously fluorescent (e.g., JF_{549} or JF_{646}) and excited with constant illumination from the excitation laser. Right: overview of stroboscopic photoactivation SPT (spaSPT). Here, the protein of interest is labeled with a photoactivatable (PA) dye that exists in a dark state, but which can be stochastically photoactivated into a fluorescent state using 405 nm illumination. This allows careful control of the density of fluorescent particles, and photoactivation of new proteins as existing ones photobleach which make it possible to obtain large numbers of trajectories, yet at low density. Stroboscopic pulsing of the excitation laser is used to minimize motion-blurring of fast-diffusing proteins and pulsing of the photoactivation laser during the camera read time is used to minimize background fluorescence

for low-to-moderately expressed proteins (unpublished observations). Thus, careful labeling control experiments should be performed if using PA-JF dyes.

Once cells expressing a tagged TF have been mounted on the microscope, we can proceed to the second step, imaging. In general, successful SPT acquisition requires a microscope with a high numerical aperture (NA) objective, a sensitive camera, and sufficiently powerful excitation lasers [9]. Most SPT studies use Highly Inclined and Laminated Optical Sheet (HILO) illumination since it conveniently reduces out-of-focus background fluorescence, thereby increasing the signal-to-noise ratio [36]. However, other modalities are also suitable for SPT, and a full discussion of suitable microscope modalities is beyond our scope. Here, we will focus specifically on how to optimize stroboscopic photoactivation SPT (spaSPT) imaging acquisition, though several considerations apply to SPT in general.

First, since chromatin-bound TFs are largely immobile, they produce a diffraction limited emission spot as expected from a point source, which can be precisely localized [37]. In contrast, detecting and localizing fast-diffusing TFs is challenging because as a frame is acquired, fast-diffusing TFs move and spread their emission photons across many pixels resulting in an imaging artifact known as *motion blur* (Fig. 5; [28, 38, 39]). For example, for a typical pixel size of 100 nm and TF $D = 3\ \mu m^2/s$, 53% of TFs would move at least 3 pixels during a $\Delta\tau = 10$ ms acquisition time (100 Hz) assuming

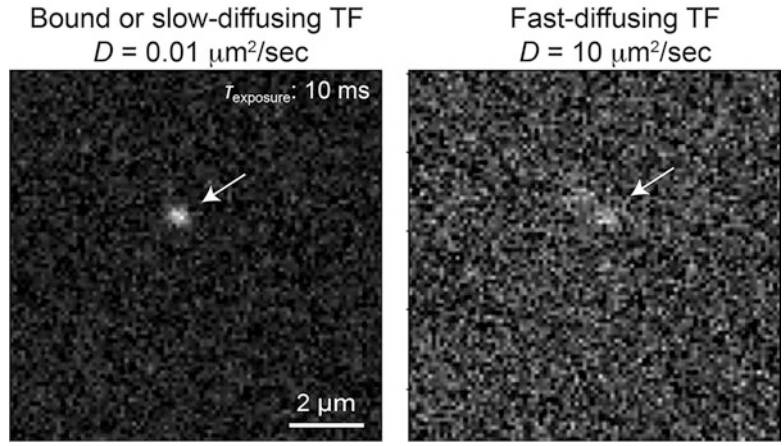


Fig. 5 Illustration of motion-blurring of fast-diffusing particles. To illustrate the concept of motion-blurring, we simulated 2D Brownian motion with a timestep of 1 μs for a bound or slow-diffusing TF (Left: $D = 0.01$ μm^2/s) and for a fast-diffusing TF (Right: 10 μm^2/s) with a 10 ms exposure time with a pixel size of 110 nm. We used an Airy disc, following the Fraunhofer diffraction pattern for a circular aperture, as the point spread function and added realistic Poissonian photon shot noise, read noise, and dark current noise. Whereas bound and slow-diffusing particles are easily detected, detection and precise localization of motion-blurred fast-diffusing particles is extremely challenging which leads to bias

Brownian motion ($P(r > r_{\mathrm{MAX}}) = 1 - \exp\left(-r_{\mathrm{MAX}}^2/4D\Delta t\right)$). Since most localization algorithms assume diffraction limited emissions from an immobile point source [40], such motion blur can lead to both undercounting of the fast-diffusing subpopulation and imprecise localization [28, 41]. Stroboscopic excitation, whereby the excitation laser is pulsed, makes it possible to reduce motion blurring (Fig. 4). For example, using either a 2 ms or 1 ms excitation pulse, would reduce the fraction of TFs that move at least 3 pixels to 2.35% or 0.06%, respectively (100 nm pixels, $D = 3$ μm^2/s). Thus, stroboscopic excitation makes it possible to minimize motion blurring, though it requires sufficiently powerful excitation lasers to generate enough signal during the short exposure.

Second, photoactivation (405 nm) and excitation laser (e.g., 561 or 633 nm) powers should be optimized in spaSPT [28]. To minimize photobleaching, the excitation laser power should be set to the lowest power that gives sufficient signal-to-noise to reliably and precisely localize particles. To minimize tracking errors, but still obtain sufficient trajectories, a mean number of ~1–2 in-focus fluorescent particles per nucleus per frame is typically optimal. To achieve this, the 405 nm photoactivation laser power can be tuned: too high power will lead to too many activated fluorescent particles resulting in tracking errors; too low power, and there will be too few particles to track. If continuous photoactivation at low power is

used it will contribute background fluorescence. Pulsing the 405 nm photoactivation laser during the brief camera read time between frames conveniently avoids this (Fig. 4).

Third, we must optimize the frame rate. If the frame rate is too fast, TF displacements between frames will be difficult to distinguish from the localization uncertainty. If the frame rate is too slow, fast diffusing particles will defocalize (move out of the axial detection range of ±~350 nm) before we can track them. The average displacement, assuming 2D Brownian motion, between frames is given by $\sqrt{4D\tau}$. For a typical TF with $D \sim 3\ \mu m^2/s$, this translates to ~350 nm displacement for a frame rate of 100 Hz and ~250 nm displacements for a frame rate of 200 Hz which is substantially greater than typical 1D localization uncertainties of ~20–40 nm. Thus, for most TFs, frame rates of 100–200 Hz are optimal.

Once the movies have been acquired using optimized acquisition parameters we can proceed to the third step, trajectory generation [42]. Here we provide a brief discussion of trajectory generation; for an in-depth discussion please refer to [40, 42]. Trajectory generation consists of two steps: (1) localizing particles in each frame and (2) connecting the localized particles from frame to frame to form trajectories. First, sufficient signal-to-noise and low motion-blur is required for particle detection and precise particle localization [37, 42]. Localization involves first filtering and thresholding images to identify particles, followed by precise sub-pixel localization of the XY-coordinates. Most algorithms use point spread function (PSF) fitting to achieve this localization, though weighted centroid estimation is more robust to high motion-blurring [41]. Second, once the particles have been localized in each frame, they are connected across frames in the tracking step to generate trajectories (XY coordinates for each timepoint). Tracking algorithms vary from relatively simple like the nearest-neighbor and the Hungarian algorithms [43] to more complex such as the Multiple-Target Tracing [44] and u-track [45]. Some of these algorithms are conveniently available through ImageJ plugins such as TrackMate and the MOSAICsuite [43, 46]. Notably, if the SPT data is of high quality and the particle density is low (~<1–2 particles per frame), the choice of tracking algorithm plays a relatively minor role. For a tracking algorithm comparison, please *see* [40].

After single-particle trajectories have been generated, we can proceed to the fourth step, trajectory analysis. Here we focus on fastSPT analysis. One approach which we refer to as MSD_i uses mean square displacement (MSD) analysis to estimate the diffusion coefficient of each trajectory, plots a histogram of diffusion coefficients ($Log(D)$), and then extracts subpopulations by fitting probability distributions to this histogram. Other methods attempt to estimate both the subpopulations and the transitions between them using Hidden Markov Modeling and/or Bayesian approaches [47–

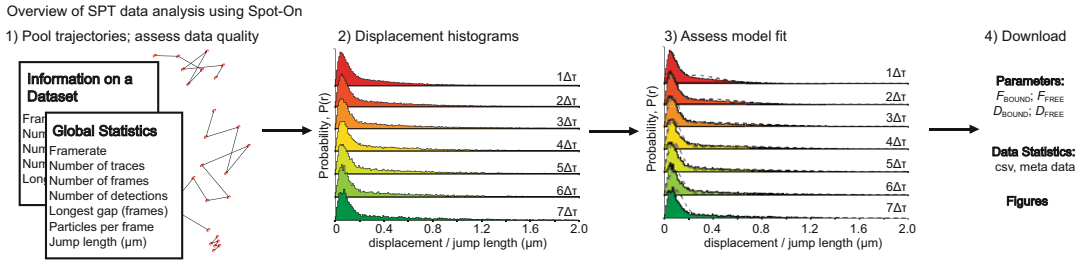


Fig. 6 Steps involved in analyzing single-particle trajectories using Spot-On. Schematic of the Spot-On web-interface workflow: (1) upload single-cell datasets of pooled trajectories and assess global SPT data statistics; (2) generate histograms of displacements (jump lengths); (3) fit either a two-state or three-state model to the data and assess the fit; (4) download the fitted parameters

50]. However, these methods do not account for defocalization [51], which leads to an overestimation of the bound subpopulation, and in benchmarking studies MSD_i approaches perform quite poorly [28]. These limitations can be overcome by pooling trajectories, fitting displacement histograms as a function of time, and then modeling defocalization as a function of the inferred diffusion coefficient of each subpopulation (Fig. 6). This approach was elegantly introduced by Mazza et al. in 2012 [8]. We subsequently simplified, expanded, and benchmarked this approach as Spot-On [14, 28]. Spot-On is available open-source in MATLAB and Python, as well as a convenient "no coding required" drag-and-drop web-interface, https://SpotOn.Berkeley.edu/.

The Spot-On web-interface is divided into three main sections (1) uploading single-particle trajectories, (2) generating histograms of displacements for multiple time points, and (3) fitting the displacement histograms to a kinetic model in order to estimate subpopulation sizes and their associated diffusion coefficients (Fig. 6). First, single-particle trajectories are uploaded to Spot-On and summary statistics are displayed (number of traces, their length, number of frames, etc.). Once the trajectories have been uploaded and assessed they can be used to generate a displacement histogram for multiple timepoints. After the displacement histogram has been generated, Spot-On proceeds to fit the histogram to a kinetic model using Brownian motion under steady state conditions without state transitions (i.e., it is assumed that transitions between the bound and free states are negligible in each individual trajectory). Spot-On offers fitting to two kinetic models: a two-state or a three-state model (Fig. 7). The two-state model considers a bound and free subpopulation and uses least-squares fitting to estimate three parameters: the bound fraction (F_{BOUND}), the bound diffusion coefficient (D_{BOUND}), and the free diffusion coefficient (D_{FREE}); the free subpopulation is given by $1 - F_{\mathrm{BOUND}}$. The three-state model considers one bound and two free subpopulations and uses least-squares fitting to estimate five parameters:

the bound fraction (F_{BOUND}), the bound diffusion coefficient (D_{BOUND}), the slower free fraction (F_{SLOW}), the slow free diffusion coefficient (D_{SLOW}), and the faster free diffusion coefficient (D_{FAST}); the faster free subpopulation is given by $1 - F_{\mathrm{BOUND}} - F_{\mathrm{SLOW}}$. A key advantage of Spot-On is that it accounts for defocalization due to 2D imaging of 3D motion [51], since axially diffusing particles will gradually exit the focal plane (±~350 nm). The rate of defocalization depends on the time interval between frames and the diffusion coefficient, leading to under-counting of the free subpopulations. Spot-On not only corrects for this bias, but the observed rate of defocalization, Z_{corr}, is used as additional information to estimate the free diffusion coefficients with higher confidence [8, 14, 28] (Fig. 7). Spot-On can also optionally fit the 1D localization error, σ (standard deviation of localization uncertainty). Finally, the user can download figures as well as the data and inferred parameters from Spot-On directly (Fig. 6).

We end by briefly discussing 2- vs. 3-state model selection and useful control SPT experiments. First, is a 2-state or 3-state model

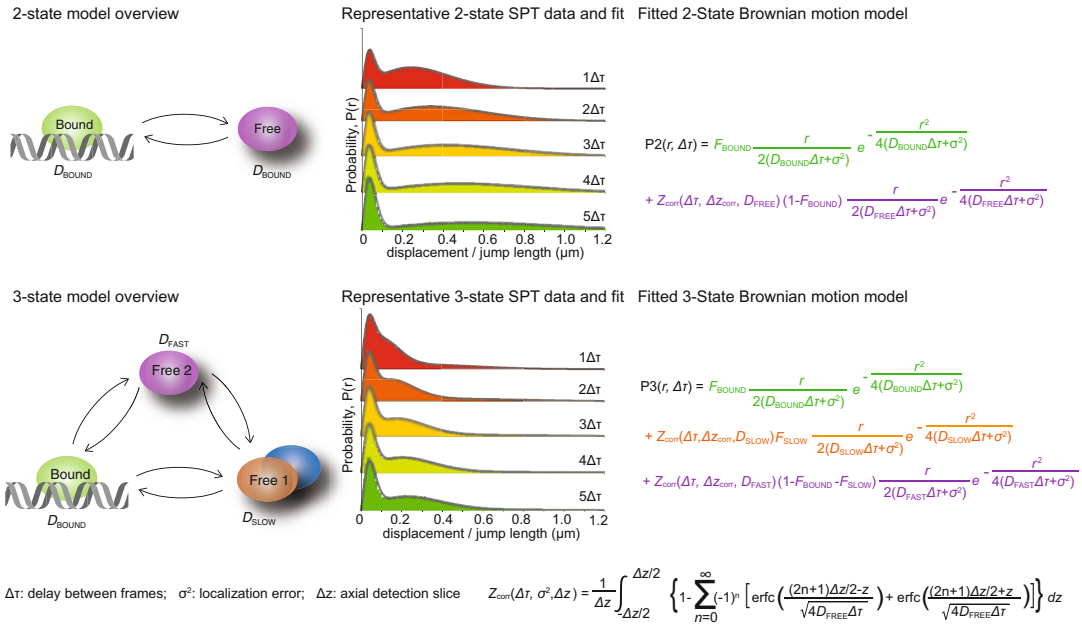


Fig. 7 Overview of two-state and three-state models implemented in Spot-On. Top: The two-state model implemented in Spot-On models a chromatin-bound and free subpopulation while assuming Brownian motion. Representative data, fits, and the underlying model are shown. Middle: The three-state model implemented in Spot-On models a chromatin-bound and two free subpopulations corresponding to a slower and a faster free state while assuming Brownian motion. Bottom: Definitions and defocalization correction implemented in Spot-On. The datasets used to illustrate the models and fits were simulated using simSPT [28] with the following parameters for the 3-state model: $D_{\mathrm{BOUND}} = 0.01$ μm^2/s; $F_{\mathrm{BOUND}} = 0.25$; $D_{\mathrm{SLOW}} = 0.25$ μm^2/s; $F_{\mathrm{SLOW}} = 0.50$; $D_{\mathrm{FAST}} = 6.0$ μm^2/s; $F_{\mathrm{FAST}} = 0.25$; $\sigma = 25$ nm. To illustrate the 2-state model, the following parameters were used: $D_{\mathrm{BOUND}} = 0.01$ μm^2/s; $F_{\mathrm{BOUND}} = 0.2$; $D_{\mathrm{FREE}} = 3.0$ μm^2/s; $F_{\mathrm{FREE}} = 0.80$; $\sigma = 25$ nm

better? Given the higher number of free parameters, a 3-state model will always fit the data better. In particular, since diffusion inside the nucleus is generally non-Brownian and anomalous unlike the underlying Spot-On model, a slight mismatch between the data and a model fit is expected. Therefore, a slight mismatch between the data and 2-state model is not necessarily evidence for two freely diffusive states. We therefore generally favor the 2-state model unless the fit is quite poor or unless there are biological and mechanistic reasons to support the two free diffusive states in the three-state model. For example, components of the general transcriptional machinery such as Cyclin T1 and TBP can freely diffuse either as monomers or part of a larger multiprotein complex, thus motivating and justifying two distinct diffusive states in the three-state model [19, 52].

Finally, inclusion of controls is essential for validating SPT approaches. At a minimum, we suggest a "free" and "bound" control. An ideal "free" control is HaloTag fused to a nuclear localization signal (Halo-NLS). Halo-NLS should exhibit a minimal bound fraction (<15%) and exhibit a fast diffusion coefficient ($D \sim 8$–$12\ \mu m^2/s$); a substantially higher bound fraction or slower diffusion coefficient is a sign of too high motion blurring (note that the positively charged NLS affords some DNA binding to Halo-NLS [53]). Similarly, an ideal "bound" control is a stably bound protein such as a histone. Histone H2B (H2B-Halo) is a popular choice and should show a high bound fraction (>70%; some unbound H2B is expected if overexpressed from a non-cell cycle regulated promoter). Inclusion of Halo-NLS and H2B-Halo controls thus makes it possible to validate the "dynamic range" of TF behaviors that can be quantified. Furthermore, if a TF has a well-defined DNA-Binding Domain (DBD), we also suggest a ΔDBD-TF-Halo control.

In the following protocol, we discuss step-by-step how to conduct and analyze SPT experiments using mouse embryonic stem cells (mESCs) expressing an endogenous genetically encoded TF-Halo fusion protein as an example. This protocol can be modified depending on the cell line, protein of interest, fluorescent label, or microscope in use.

2 Materials

Below we described the required reagents and resources for the four main steps of a fastSPT experiment (1) reagents for cell preparation, (2) equipment for microscopy, (3) code for trajectory generation, and (4) analysis using Spot-On.

2.1 Reagents Needed for Cell Preparation

Cell preparation reagents are highly cell-type specific. Here we use reagents specific to mESCs that express a Halo-tagged TF as an example. All of the following reagents must be prepared in a biosafety cabinet, practicing strict sterile technique.

1. Growth Media: In order to prepare your growth medium, combine the following reagents: Knockout DMEM 1×, 15% Fetal Bovine Serum, 2 mM GlutaMAX Supplement, 1 mM MEM nonessential amino acids solution, 1000 U/mL LIF, 0.1 mM 2(β)-mercaptoethanol, 100 U/mL penicillin–streptomycin. Store at 4 °C.
2. Matrigel: dilute according to manufacturer's instructions prior to cell plating. Store aliquots at −20 °C. After being diluted in a serum-free medium, store at 4 °C (*see* **Note 1**).
3. Imaging dish: 35 mm dish, No. 1.5 Coverslip, 14 mm Glass Diameter, uncoated (*see* **Note 2**).
4. Trypsin-EDTA (0.05%), phenol red. Store at −20 °C.
5. Sterile 1× Phosphate Buffered Saline pH 7.4.
6. Biosafety Cabinet with Laminar Flow.
7. Tissue Culture (TC) incubator set to 37 °C and 5.5% CO_2.
8. Phenol-red free imaging Media: DMEM without phenol red, 15% fetal bovine serum, 2 mM GlutaMAX Supplement, 1 mM MEM nonessential amino acids solution, 1000 U/mL LIF, 0.1 mM 2(β)-ME, 100 U/mL penicillin–streptomycin. Store at 4 °C (*see* **Note 3**).
9. Dimethyl sulfoxide, sterile filtered.
10. Synthetic Dyes: Halo or SNAP dyes (e.g., PA-JF_{646} or PA-JF_{549}). We recommend storing dyes at 1000× the desired concentration in DMSO at −20 °C in single-use aliquots to minimize freeze–thaw cycles [34, 35] (*see* **Note 4**).

2.2 Microscope Set-Up

Many microscope modalities are suitable for SPT, including wide-field microscopes. Here we use as our example a custom-built Nikon TI Microscope, implementing highly inclined illumination [36] that we previously used [14]. Key components include the following.

1. Live-cell incubation chamber heated to 37 °C that maintains a humidified atmosphere at 5.5% CO_2.
2. A high-NA objective. For HILO, we used a 100×/NA 1.49 Oil-immersion TIRF objective (Nikon apochromat CFI Apo TIRF 100× Oil).
3. Powerful excitation lasers matched to the desired fluorophores. We used 561 nm (1 W, Genesis, Coherent) for (PA)-JF_{549}; 633 nm (1 W, Genesis, Coherent) for (PA)-JF_{646}; 405 nm (140 mW, OBIS, Coherent) for photoactivation.

4. A fast and sensitive camera. Most EM-CCD and back-illuminated high quantum efficiency sCMOS cameras are suitable. We used an iXon Ultra 897 EM-CCD camera (Andor) (*see* **Note 5**).
5. Emission filters that match the fluorophores. We used: JF_{549}/PA-JF_{549}: Semrock 593/40 nm band-pass filter; JF_{646}/PA-JF_{646}: Semrock 676/37 nm bandpass filter.
6. Control of laser intensity. Rapid control (<100 μs) of laser intensity at multiple wavelengths is essential for stroboscopic excitation. We achieved this using an AOTF (AA Opto-Electronic, France, AOTFnC-VIS-TN) and DAQ card (National Instruments, NI-DAQ PCI-6723).
7. Microscope control software. We used Nikon Elements.

2.3 Localization and Tracking

Once raw SPT movies have been acquired, particles must be localized in each frame (localization) and then tracked between frames to form trajectories (tracking). Popular and user-friendly algorithms and implementations to achieve this include MTT [44], u-track [45], TrackMate [43], and the MOSAICsuite [46]. We used the MTT algorithm implemented in MATLAB (*see* **Note 6**). For a performance comparison of tracking algorithms, please *see* [40].

2.4 Analysis Using Spot-On

To analyze trajectory data using Spot-On, use either the web-interface, the MATLAB, or the Python version (*see* **Note 7**).

3 Methods

3.1 Cell Preparation

The following steps should be carried out in a biosafety cabinet and everything must be kept sterile. The steps apply to mESCs that express an endogenous genetically encoded TF-Halo fusion protein This protocol can be adjusted for the cell line, dye, or fluorophore in use.

1. Grow cells for seeding on tissue culture dishes until they are at 70–80% confluency.
2. Coat the glass bottom 35 mm imaging dish with Matrigel—Add 1 mL diluted Matrigel per imaging dish, spread and incubate at 37 °C and 5.5% CO_2 for 30–60 min (*see* **Note 8**).
3. Aspirate all of the media from the culture dish and wash cells with PBS. Gently swirl the PBS to ensure all residual media has been removed.
4. Aspirate PBS and add just enough 0.05% Trypsin-EDTA to cover the bottom of the culture dish and place in TC incubator for ~3 min.
5. Remove cells from the incubator and check if all the cells have thoroughly dissociated using a light microscope.

6. After cells have dissociated from culture dish, quench with normal culture medium, resuspend cells, pipet up and down with a P1000 pipette until all cell clumps have been broken up into single cells (*see* **Note 9**).
7. Transfer the desired number of cells to a 15 mL Falcon tube and centrifuge at 300 × *g* for 3 min. Enough cells should be used so that plated cells are ~70% confluent after overnight growth on the MatTek dish.
8. While cells are spinning down, remove Matrigel from **step 1** and add cell medium to the 35 mm imaging dish.
9. Remove tube from centrifuge and aspirate supernatant, leaving cell pellet.
10. Resuspend cell pellet in cell medium.
11. Add cells to the imaging dish at the appropriate density for the cell line in use. After adding cells to the imaging dish, gently swirl the dish to evenly distribute cells.
12. Place in TC incubator and grow overnight.

Day of imaging After seeding imaging dishes the day before and verifying using a tissue culture microscope that they look healthy and are at ~70% confluency, we can proceed to dye labeling and imaging.

1. Prior to preparing cells for imaging, turn on the microscope and environmental chamber leaving enough time for the chamber to equilibrate to 37 °C and 5.5% CO_2 before imaging.
2. Prepare three 15 mL Falcons tubes: one with PBS; one with regular medium; and one with phenol red free Imaging Medium. Place these in the 37 °C water bath.
3. Remove the falcon with regular medium from the 37 °C water bath and make a dilution of the synthetic dye (e.g., Halo or SNAP compatible JF dye) to the desired concentration. Pipet up and down to mix (*see* **Note 10**).
4. Remove medium from the imaging dish and add medium with the desired concentration of synthetic dye and place in TC incubator for 15 min.
5. Wash 1: Remove Halo-dye medium and add prewarmed PBS, remove PBS, and add prewarmed medium and place in incubator for 5 min.
6. Wash 2: remove medium and add prewarmed PBS, remove PBS, and add prewarmed imaging medium without phenol red (more/longer washes may be necessary for PA-JF dyes) (*see* **Note 11**).
7. Cells are now ready to be imaged and can be stored in the TC incubator until the microscope is ready.

3.2 Imaging

The specific imaging protocol will be highly dependent on the microscope used, the desired SPT experiment, and a number of other factors. We briefly comment on some of the main steps below for fastSPT experiments.

1. Add immersion oil to the objective, then load the imaging dish with labeled cells on the prewarmed microscope.
2. Move the objective up until cells are in focus using either bright-field or fluorescence to focus on the cells.
3. If using HILO illumination, move stage to center the cell to be studied in the field-of-view. Modulate the TIRF angle until optimal HILO illumination is achieved (maximal signal-to-background ratio and even illumination of the whole nucleus).
4. If optimizing laser acquisition settings, then record a short movie (~500 frames) at the desired frame rate (typically ~100–200 Hz) changing only one parameter at a time. If using photoactivation, adjust 405 nm intensity and/or pulse duration until the desired density of particles is achieved (typically ~1–2 in-focus particles per nucleus per frame). If optimizing the main excitation laser (e.g., 561 nm for JF_{549}), record multiple short movies for different excitation powers and stroboscopic pulse durations, analyze the movies by generating trajectories, and overlay trajectories on raw movies. Choose an excitation setting that gives sufficient signal-to-noise that the localization algorithm misses almost no particles visible by eye in the raw images. Spending significant time iteratively optimizing acquisition settings is usually well worth the effort.
5. Once acquisition settings have been optimized, record fastSPT movies one cell at a time. After centering the field-of-view around a cell and optimizing the HILO angle (the optimal angle may need to be adjusted for each cell), crop a just big enough ROI around the nucleus of interest. Photobleach particles if necessary if the initial density is too high. Then record a fastSPT movie. Our default spaSPT acquisition parameters for most mammalian TFs are: 30,000 frames at 134 Hz, using 1 ms stroboscopic excitation (561 or 633 nm, 1 W, 100% AOTF power), and pulsing the photoactivation laser (405 nm, 140 mW, typically 1–4% AOTF) during the ~0.45 ms camera read-out time between frames.
6. Move at least two full field-of-views away and begin the next movie. We typically collect 6–8 movies per cell line per condition per day for at least three biological replicates performed on different days (at least 18–24 cells in total). Recording multiple cells is necessary to average over cell-to-cell and biological variation (e.g., cell cycle phase if cells are unsynchronized) and to obtain robust results.

7. Once finished with one cell line or condition, clean objective and mount a new imaging dish with a different cell line or condition.
8. Leave it at least 15 min to thermally equilibrate.
9. Then begin the next round of movies.
10. After imaging is complete, transfer all the raw SPT data, clean the objective, and turn off the microscope.

3.3 Trajectory Generation

Please *see* Subheading 2.3 for recommended localization and tracking algorithms. Below, we briefly outline the recommended steps after a day of SPT data acquisition.

1. Make sure to visually inspect SPT movies and visually assess the quality and reliability of the localization and tracking for a few movies by overlaying trajectories on the raw SPT movies.
2. Optimize localization and tracking algorithm parameters if necessary, but make sure to use consistent parameters for all conditions and replicates.
3. Once localization and tracking settings have been finalized, batch process all of the acquired SPT movies if possible.

3.4 Trajectory Analysis with Spot-On

Once trajectories have been generated, we can proceed to analysis. Here we specifically focus on how to analyze fastSPT data with Spot-On's web-interface. Please refer to the Spot-On paper [28] and the documentation available at https://SpotOn.berkeley.edu/SPTGUI/docs/latest for a more complete discussion.

1. Go to https://SpotOn.berkeley.edu/ and click "*Start spotting!*"
2. In "1. Select format" pick the format used for your SPT trajectories (*see* **Note 12**) and drag and drop your data into "3. Select datasets".
3. Make sure through "Uploaded datasets" that the files were successfully uploaded and assess "Global statistics" on the bottom right, which will display metadata for your uploaded SPT data (*see* **Note 13**).
4. Proceed to the "Kinetic Modeling" tab.
5. Under "Dataset selection" include all the datasets you would like to analyze. Click "all" if all the data are from the same condition.
6. Scroll down to "Jump length histograms" and inspect the histograms of displacements. Under "Display dataset" click through each cell to inspect that the data looks reasonable. Click "Show pooled jump length distribution" if you would like to combine the data from each single cell. Some noise is expected, but if the histograms are too sparse, the fitting is less likely to be accurate.

7. Scroll back up to "Parameters" and "Jump length distribution" and choose the desired values for "Bin width," "Number of timepoints," "Jumps to consider," "Use entire trajectories," and "Max jump" (*see* **Note 14** for a brief discussion of how to choose these parameters).
8. Next, proceed to "Model fitting." Choose between the two-state and three-state models, upper and lower bounds on the diffusion coefficients, whether to infer "Localization error" from the data (choose "fit from the data" or to predefine it (default is 35 nm or 0.035 μm)). Choose whether to use the *Z*-correction and if so, specify its value (default is 700 nm or 0.7 μm, which is reasonable for most high NA objectives). Finally, choose whether to use PDF or CDF fitting, whether to fit each single cell or only the merged displacement histogram of all of the cells, and the number of fitting iterations (*see* **Note 15** for a brief discussion of how to choose these).
9. Click "Fit kinetic model." This may take a few minutes.
10. If single-cell fitting was performed, scroll down to "display dataset" under "Jump length histograms" and scroll through each single cell and assess the quality of the fit and the cell-to-cell variation. This way any potentially problematic datasets can be identified (*see* **Note 16**). Once each single cell has been assessed, click "show pooled jump length distribution" to see the pooled data and fit.
11. Spot-On will display the fitted parameters for each single cell (if single cell fitting was chosen) and the global fit parameters: D_{BOUND}, D_{FREE} (D_{SLOW}, D_{FAST}, if 3-state model), F_{BOUND}, F_{FREE}, (F_{SLOW}, F_{FAST}, if 3-state model), σ (if localization error was fitted), and fitting parameter (I_2, AIC, BIC; *see* **Note 17**).
12. Iterate through the various options until a desired fit has been obtained.
13. Then scroll to the bottom of the page and click "Mark for download" and enter a name and description.
14. Next scroll back to the top of the page and click the "Download" tab. Here you can download individual figures (SVG, PDF, PNG, EPS) or you can click "Download all (zip)" to obtain a copy of the fitted parameters, raw data, as well as the figures.

4 Notes

1. When preparing Matrigel make sure everything is done on ice. Thaw individual aliquots on ice for 30 min prior to diluting in serum-free medium. Coating of glass with 0.1% gelatin is also appropriate, though in our experience adherence can be poorer.

2. A cover glass (e.g., Marienfeld-High-Precision 1.5H cover glasses, 0117650) mounted in an Attofluor Cell chamber (ThermoFisher, A7816) can also be used instead of MatTek imaging dishes. For single molecule imaging wash the 25 mm circular cover glasses in isopropanol, then plasma clean and store the cover glasses in isopropanol at 4 °C until use. They can be stored for >6 months at 4 °C.
3. It is essential to use medium without phenol red for fluorescence imaging to avoid excessive background fluorescence.
4. Janelia Fluor dyes can be inquired about at dyes.janelia.org or purchased from Promega.
5. One can minimize localization uncertainty by choosing the objective magnification and camera pixel size such that the pixel size approximately matches the PSF standard deviation [37].
6. Our Matlab version of the MTT algorithm can be accessed here https://gitlab.com/tjian-darzacq-lab/SPT_LocAndTrack.
7. The web-interface can be found at https://spoton.berkeley.edu/SPTGUI/; the Matlab version at https://gitlab.com/tjian-darzacq-lab/spot-on-matlab; and the Python version at https://gitlab.com/tjian-darzacq-lab/Spot-On-cli.
8. If extra Matrigel dishes are coated, they can be sealed with Parafilm and stored in 4 °C for 2–4 days. It is recommended to prepare imaging dishes with Matrigel fresh.
9. Pipet up and down ~10–15 times until cells are dissociated into a single cell suspension. Check under a light microscope to ensure that they are in a single-cell suspension. If mESCs are passaged in clumps, they may differentiate.
10. Optimization of the dye concentration is typically required. For optimizing SPT experiments, we recommend a dye titration experiment using logarithmically spaced concentrations. Labeling will depend on protein concentration, cell type, incubation time, and must thus be optimized for each cell line. For regular Halo-JF dyes, we typically use between ~1 pM and ~5 nM labeling. For photoactivatable Halo-JF dyes, we typically use ~5 nM to ~100 nM. For SPT, complete labeling is neither necessary nor desired. But if complete labeling is desired, 500 nM JF-Halo dye is typically sufficient as shown in [54].
11. When using "regular" JF-HaloTag dyes, two short 5-min washes are generally sufficient. However, for PA-JF dyes, more washes and/or longer than 5-min washes may be required. The optimal washing protocol can be both dye and cell-type specific. As a control, we recommend labeling and washing a wild-type cell that does not express HaloTag and making sure that negligible dye remains in this negative control.

12. Click on "learn more" to see the details of the format. If your trajectory format is not identical to any of the supported format, it will be necessary to first write a script to convert it to one of the Spot-On supported formats. Sample files for each support format are available.
13. More data is always better, but we recommend having at least 6 single cells per condition and at least a few thousand trajectories with at least 3 detections (*see* Fig. 3—figure supplement 12 in [28] for a quantification of how the robustness of the Spot-On fit depends on the number of trajectories). It is also worth paying close attention to "Particles per frame"—if this number is too high, the SPT data is likely to contain frequent tracking misconnections.
14. For a full discussion of how to choose these parameters, please *see* Appendix 2 in [28] and the documentation available at https://spoton.berkeley.edu/SPTGUI/docs/latest. Here, we provide brief guidance:

    Bin width: Bin width used to make displacement histograms and used for PDF-fitting. Default is 10 nm and is generally reasonable unless you have very sparse data. 1 nm is the default setting for CDF-fitting, since CDF-fitting is more robust and less prone to binning artifacts.

    Number of timepoints: How many timepoints to consider in the displacement histogram. If you allow N time points, this corresponds to considering displacements with a maximal time-delay of up to $(N-1)\Delta t$. Generally, displacement histograms become sparser at large time-delays and we generally do not recommend considering time-delays much above 50–60 ms.

    Max jump: the maximal displacements that will be considered in the analysis. This should be larger than the largest displacements in the data. Generally, 3–5 μm is reasonable.

    Jumps to Consider and "Use entire trajectories": If use entire trajectories is set to Yes, all displacement data will be used. If it is set to No, only up to the indicated value of Jumps to consider is used. For example, if Jumps to consider is set to 4 and 8 timepoints, for each trajectory, 4 displacements (if possible) will be used to compute the displacement histogram such that a trajectory of nine frames will contribute four displacements to $1\Delta t$, four displacements to $2\Delta t$, . . ., and two displacements to $7\Delta t$. This is a semiempirical way of correcting for additional biases toward bound molecules, and if there is no bias toward bound molecules in the raw data, "Use entire trajectories" should be set to Yes. This is a subtle choice and please *see* Appendix 2 referenced above for a more complete discussion.

15. As noted above, please *see* Appendix 2 in [28] and the documentation available at https://spoton.berkeley.edu/SPTGUI/docs/latest for a full discussion. It is described briefly below.

    Kinetic model: this choice is discussed in the main text. We recommend starting with the two-state model, and only considering the three-state model if the two-state fit is quite poor and/or there are biochemical and mechanistic reasons to suspect two distinct freely diffusive states.

    Upper and lower bounds on fitted diffusion coefficients: Defaults are [0.0005–0.08 $\mu m^2/s$] for D_{BOUND} and [0.15–0.25 $\mu m^2/s$] for D_{FREE}. Please *see* Appendix 2 in [28] for a full discussion, but briefly, it is important to pay attention to these and make sure Spot-On does not infer a D at the min or max. Also, $D_{BOUND} = 0.08\ \mu m^2/s$ is almost certainly too high for DNA binding and could indicate that the specified localization error is too small and/or problems with microscope stability. It is very useful to perform SPT on a histone control to assess what D_{BOUND} to expect from the bound population.

    Localization Error: this is the 1D standard deviation of the localization uncertainty. If this can be estimated independently and specified, it will improve the robustness of the fit. If it is fitted from the data, please note that it is mainly fitted from the bound subpopulation and that it is not well-fitted if the bound subpopulation is negligible. If the localization error is incorrectly specified, typically the fit to the bound subpopulation will be poor.

    Z correction and dZ: since SPT generally involves 2D imaging of 3D motion, we must correct for defocalization. On most SPT microscopes, the axial detection range is ~700 nm—if particles move out of this range, they generally cannot be detected. Using ~700 nm is generally safe, but please *see* [28] for advice on how to experimentally measure it. In some organisms such as some yeasts and bacteria, the cell is so small, that the observation slice is comparable to the axial detection range, in which case the Z correction should be set to "No," since there is no defocalization.

    Model fit: You can either fit the PDF or CDF of the displacement histogram. Generally, CDF-fitting is more robust since it is less susceptible to binning noise, especially for moderately sparse datasets. However, the two approaches give equivalent results for sufficiently large SPT datasets, and comparing PDFs and fits is generally more intuitive.

    Perform single cell fit: We generally recommend fitting each single cell and assessing each single cell fit. This can be a great way of identifying potentially problematic single cell movies and for assessing cell-to-cell variation. The only downside is that it will take significantly longer for Spot-On to run.

Iterations: Spot-On uses least-squares fitting, which is subject to trapping in local minima during optimization. For each fit iteration Spot-On will generate a random initial guess for each fitted parameter and proceed with optimization for a hard-coded number of steps or until convergence. To avoid trapping in local minima, multiple iterations of this are repeated. For the 2-state model, three iterations are typically more than enough to ensure that the global minima is identified. For 3-state model fitting, or if the fit looks poor, it may be worth increasing the number of fit iterations. The only downside to increasing the number of iterations is a slower fit.

16. Problematic dataset refers to potential outliers in the overall experimental dataset. For example, if an unhealthy cell or a mitotic cell was accidentally chosen, or if the particle density was too high, or is the acquisition settings were chosen poorly (improper TIRF angle, etc.). Looking at each single cell as well as the overall population can be a great way to assess cell-to-cell variation and to assess the robustness of conclusions.
17. BIC and AIC are information criteria that can be used to compare the "goodness of fit" for different models, while penalizing models with more parameters. However, since Spot-On models protein diffusion as Brownian, which it never truly is in cells, we note that using BIC or AIC to compare the goodness of fit of the 2-state and 3-state models can be misleading.

Acknowledgments

We thank Domenic Narducci, Miles Huseyin, Jin Yang, Hugo Brandão, Viraat Goel, Sarah Nemsick, Shdema Filler-Hayut, Michele Gabriele, Jyothi Mahadevan, Meagan Esbin, Maxime Woringer, and Thomas Graham for insightful comments on the manuscript. We would like to acknowledge Davide Mazza, whose 2012 paper introduced the kinetic modeling framework that was ultimately implemented in Spot-On in a modified form, Maxime Woringer who codeveloped Spot-On and led the development of the web-interface and the Python version and who has been maintaining the web-interface, the Tjian-Darzacq lab for discussions during the development of Spot-On and for hosting the web-interface, and Luke Lavis for the development and sharing of Janelia Fluor dyes. We thank Domenic Narducci for the code to simulate the concept of motion-blurring in Fig. 5. This work was supported by the National Institutes of Health under grant numbers R00GM130896, DP2GM140938, and UM1HG011536.

References

1. Lionnet T, Wu C (2021) Single-molecule tracking of transcription protein dynamics in living cells: seeing is believing, but what are we seeing? Curr Opin Genet Dev 67:94–102
2. Cramer P (2019) Organization and regulation of gene transcription. Nature 573:45–54
3. Phair RD, Misteli T (2000) High mobility of proteins in the mammalian cell nucleus. Nature 404:604–609
4. McNally JG, Müller WG, Walker D, Wolford R, Hager GL (2000) The Glucocorticoid Receptor: Rapid Exchange with Regulatory Sites in Living Cells. Science 287: 1262–1265. https://doi.org/10.1126/science.287.5456.1262
5. Mueller F, Stasevich TJ, Mazza D, McNally JG (2013) Quantifying transcription factor kinetics: at work or at play? Crit Rev Biochem Mol Biol 48:492–514
6. Mueller F, Mazza D, Stasevich TJ, McNally JG (2010) FRAP and kinetic modeling in the analysis of nuclear protein dynamics: what do we really know? Curr Opin Cell Biol 22:403–411
7. Politi AZ, Cai Y, Walther N, Hossain MJ, Koch B, Wachsmuth M, Ellenberg J (2018) Quantitative mapping of fluorescently tagged cellular proteins using FCS-calibrated four-dimensional imaging. Nat Protoc 13:1445–1464
8. Mazza D, Abernathy A, Golob N, Morisaki T, McNally JG (2012) A benchmark for chromatin binding measurements in live cells. Nucleic Acids Res 40:e119–e119
9. Shen H, Tauzin LJ, Baiyasi R, Wang W, Moringo N, Shuang B, Landes CF (2017) Single particle tracking: from theory to biophysical applications. Chem Rev 117:7331–7376
10. Goulian M, Simon SM (2000) Tracking single proteins within cells. Biophys J 79:2188–2198
11. Chen J, Zhang Z, Li L, Chen B-C, Revyakin A, Hajj B, Legant W, Dahan M, Lionnet T, Betzig E, Tjian R, Liu Z (2014) Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell 156:1274–1285
12. Garcia DA, Fettweis G, Presman DM, Paakinaho V, Jarzynski C, Upadhyaya A, Hager GL (2021) Power-law behavior of transcription factor dynamics at the single-molecule level implies a continuum affinity model. Nucleic Acids Res 49:6605. https://doi.org/10.1093/nar/gkab072
13. Reisser M, Hettich J, Kuhn T, Popp AP, Große-Berkenbusch A, Gebhardt JCM (2020) Inferring quantity and qualities of superimposed reaction rates from single molecule survival time distributions. Sci Rep 10: 1758
14. Hansen AS, Pustova I, Cattoglio C, Tjian R, Darzacq X (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. elife 6:e25776
15. Metzler R, Jeon J-H, Cherstvy AG, Barkai E (2014) Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. Phys Chem Chem Phys 16:24128–24164
16. Hansen AS, Amitai A, Cattoglio C, Tjian R, Darzacq X (2019) Guided nuclear exploration increases CTCF target search efficiency. Nat Chem Biol 16:257. https://doi.org/10.1038/s41589-019-0422-3
17. Loffreda A, Jacchetti E, Antunes S, Rainone P, Daniele T, Morisaki T, Bianchi ME, Tacchetti C, Mazza D (2017) Live-cell p53 single-molecule binding is modulated by C-terminal acetylation and correlates with transcriptional activity. Nat Commun 8:313
18. Popp AP, Hettich J, Gebhardt JCM (2021) Altering transcription factor binding reveals comprehensive transcriptional kinetics of a basic gene. Nucleic Acids Research, 49(11), pp.6249–6266
19. Nguyen VQ, Ranjan A, Liu S, Tang X, Ling YH, Wisniewski J, Mizuguchi G, Li KY, Jou V, Zheng Q, Lavis LD, Lionnet T, Wu C (2020) Spatio-Temporal Coordination of Transcription Preinitation Complex Assembly in live Cells. bioRxiv. https://doi.org/10.1101/2020.12.30.424853
20. Jain S, Shukla S, Yang C, Zhang M, Fatma Z, Lingamaneni M, Abesteh S, Lane ST, Xiong X, Wang Y, Schroeder CM, Selvin PR, Zhao H (2021) TALEN outperforms Cas9 in editing heterochromatin target sites. Nat Commun 12:606
21. Huseyin MK, Klose RJ (2021) Live-cell single particle tracking of PRC1 reveals a highly dynamic system with low target site occupancy. Nat Commun 12:887
22. Tatavosian R, Duc HN, Huynh TN, Fang D, Schmitt B, Shi X, Deng Y, Phiel C, Yao T, Zhang Z, Wang H, Ren X (2018) Live-cell single-molecule dynamics of PcG proteins imposed by the DIPG H3.3K27M mutation. Nat Commun 9:2080
23. Teves SS, An L, Hansen AS, Xie L, Darzacq X, Tjian R (2016) A dynamic mode of mitotic

bookmarking by transcription factors. elife 5: e22280

24. Deluz C, Friman ET, Strebinger D, Benke A, Raccaud M, Callegari A, Leleu M, Manley S, Suter DM (2016) A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. Genes Dev 30:2538. http://genesdev.cshlp.org/content/early/2016/12/05/gad.289256.116.abstract
25. Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, Heckert A, Banala S, Lavis L, Darzacq X, Tjian R (2018) Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science 361:eaar2555
26. M. Mir, A. Reimer, M. Stadler, A. Tangara, A. S. Hansen, D. Hockemeyer, M. B. Eisen, H. Garcia, X. Darzacq, Y. L. Lyubchenko, Springer New York, NY, 2018; Single Molecule Imaging in Live Embryos Using Lattice Light-Sheet Microscopy. https://doi.org/10.1007/978-1-4939-8591-3_32, pp. 541–559
27. Manley S, Gillette JM, Patterson GH, Shroff H, Hess HF, Betzig E, Lippincott-Schwartz J (2008) High-density mapping of single-molecule trajectories with photoactivated localization microscopy. Nat Methods 5: 155–157
28. Hansen AS, Woringer M, Grimm JB, Lavis LD, Tjian R, Darzacq X (2018) Robust model-based analysis of single-particle tracking experiments with Spot-On. elife 7:e33125
29. Watanabe N, Mitchison TJ (2002) Single-Molecule Speckle Analysis of Actin Filament Turnover in Lamellipodia. Science 295:1083–1086. https://doi.org/10.1126/science.1067470
30. Gebhardt JCM, Suter DM, Roy R, Zhao ZW, Chapman AR, Basu S, Maniatis T, Xie XS (2013) Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. Nat Methods 10:421
31. Presman DM, Ball DA, Paakinaho V, Grimm JB, Lavis LD, Karpova TS, Hager GL (2017) Quantifying transcription factor binding dynamics at the single-molecule level in live cells. Methods 123:76–88
32. Shao S, Xue B, Sun Y (2018) Intranucleus single-molecule imaging in living cells. Biophys J 115:181–189
33. Los GV, Encell LP, McDougall MG, Hartzell DD, Karassina N, Zimprich C, Wood MG, Learish R, Ohana RF, Urh M, Simpson D, Mendez J, Zimmerman K, Otto P, Vidugiris G, Zhu J, Darzins A, Klaubert DH, Bulleit RF, Wood KV (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem Biol 3:373–382
34. 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
35. Grimm JB, English BP, Choi H, Muthusamy AK, Mehl BP, Dong P, Brown TA, Lippincott-Schwartz J, Liu Z, Lionnet T, Lavis LD (2016) Bright photoactivatable fluorophores for single-molecule imaging. Nat Methods 13:985
36. Tokunaga M, Imamoto N, Sakata-Sogawa K (2008) Highly inclined thin illumination enables clear single-molecule imaging in cells. Nat Methods 5:159–161
37. Thompson RE, Larson DR, Webb WW (2002) Precise nanometer localization analysis for individual fluorescent probes. Biophys J 82:2775–2783
38. Elf J, Li G-W, Xie XS (2007) Probing Transcription Factor Dynamics at the Single-Molecule Level in a LivingCell. Science 316: 1191–1194. https://doi.org/10.1126/science.1141967
39. Izeddin I, Récamier V, Bosanac L, Cissé II, Boudarene L, Dugast-Darzacq C, Proux F, Bénichou O, Voituriez R, Bensaude O, Dahan M, Darzacq X (2014) Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. elife 3:e02230
40. Chenouard N, Smal I, de Chaumont F, Maška M, Sbalzarini IF, Gong Y, Cardinale J, Carthel C, Coraluppi S, Winter M, Cohen AR, Godinez WJ, Rohr K, Kalaidzidis Y, Liang L, Duncan J, Shen H, Xu Y, Magnusson KEG, Jaldén J, Blau HM, Paul-Gilloteaux P, Roudot P, Kervrann C, Waharte F, Tinevez J-Y, Shorte SL, Willemse J, Celler K, van Wezel GP, Dan H-W, Tsai Y-S, de Solórzano CO, Olivo-Marin J-C, Meijering E (2014) Objective comparison of particle tracking methods. Nat Methods 11:281–289
41. Deschout H, Neyts K, Braeckmans K (2012) The influence of movement on the localization precision of sub-resolution particles in fluorescence microscopy. J Biophotonics 5:97–109
42. Lee A, Tsekouras K, Calderon C, Bustamante C, Pressé S (2017) Unraveling the thousand word picture: an introduction to super-resolution data analysis. Chem Rev 117: 7276–7330
43. Tinevez J-Y, Perry N, Schindelin J, Hoopes GM, Reynolds GD, Laplantine E, Bednarek SY, Shorte SL, Eliceiri KW (2017) TrackMate:

an open and extensible platform for single-particle tracking. Methods 115:80–90

44. Sergé A, Bertaux N, Rigneault H, Marguet D (2008) Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. Nat Methods 5:687–694
45. Jaqaman K, Loerke D, Mettlen M, Kuwata H, Grinstein S, Schmid SL, Danuser G (2008) Robust single-particle tracking in live-cell time-lapse sequences. Nat Methods 5:695–702
46. Shivanandan A, Radenovic A, Sbalzarini IF (2013) MosaicIA: an ImageJ/Fiji plugin for spatial pattern and interaction analysis. BMC Bioinformatics 14:349
47. Persson F, Lindén M, Unoson C, Elf J (2013) Extracting intracellular diffusive states and transition rates from single-molecule tracking data. Nat Methods 10:265
48. Monnier N, Barry Z, Park HY, Su K-C, Katz Z, English BP, Dey A, Pan K, Cheeseman IM, Singer RH, Bathe M (2015) Inferring transient particle transport dynamics in live cells. Nat Methods 12:838–840
49. Vink JNA, Brouns SJJ, Hohlbein J (2020) Extracting transition rates in particle tracking using analytical diffusion distribution analysis. Biophys J 119:1970–1983
50. Karslake JD, Donarski ED, Shelby SA, Demey LM, DiRita VJ, Veatch SL, Biteen JS (2020) SMAUG: analyzing single-molecule tracks with nonparametric Bayesian statistics. Methods 193:16. https://doi.org/10.1016/j.ymeth.2020.03.008
51. Kues T, Kubitscheck U (2002) Single molecule motion perpendicular to the focal plane of a microscope: application to splicing factor dynamics within the cell nucleus. Single Mol 3:218–224
52. Lu H, Yu D, Hansen AS, Ganguly S, Liu R, Heckert A, Darzacq X, Zhou Q (2018) Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. Nature 558:318–323
53. Mangel WF, McGrath WJ, Xiong K, Graziano V, Blainey PC (2016) Molecular sled is an eleven-amino acid vehicle facilitating biochemical interactions via sliding components along DNA. Nat Commun 7:10202
54. Cattoglio C, Pustova I, Walther N, Ho JJ, Hantsche-Grininger M, Inouye CJ, Hossain MJ, Dailey GM, Ellenberg J, Darzacq X, Tjian R, Hansen AS (2019) Determining cellular CTCF and cohesin abundances to constrain 3D genome models. elife 8:e40164

Chapter 10

Characterization of Mammalian Regulatory Complexes at Single-Locus Resolution Using TINC

Anja S. Knaupp, Ralf B. Schittenhelm, and Jose M. Polo

Abstract

In mammalian cells, multiprotein complexes form at specific genomic regulatory elements (REs) to control gene expression, which in turn is ultimately responsible for cellular identity. Consequently, insight into the molecular composition of these regulatory complexes is of major importance for our understanding of any physiological or pathological cellular state or transition. However, it remains extremely difficult to identify the protein complex(es) assembled at a specific RE in the mammalian genome using conventional approaches. We therefore developed a novel single locus isolation technique based on Transcription Activator-Like Effector (TALE) proteins termed TALE-mediated isolation of nuclear chromatin (TINC). When coupled with high-resolution mass spectrometry, TINC enables the identification and characterization of protein complexes formed at any RE of interest. Using the *Nanog* promoter in mouse embryonic stem cells as proof of concept, this chapter describes in detail the novel TINC methodology as well as subsequent mass spectrometric considerations.

Key words Single-locus pulldown, Epigenetics, Proteomics, Regulatory complex, Transcriptional regulation, Affinity purification, Transcription activator-like effector proteins, Chromatin immunoprecipitation, Mass spectrometry

1 Introduction

Cell type-specific gene expression programs underlie cellular identity and function, and changes to these programs are associated with various diseases including cancer and diabetes. Consequently, insight into how the expression of specific genes is controlled is of major importance for our understanding of numerous physiological and pathological processes including development and carcinogenesis. In mammalian cells, transcription factors bind to specific genomic regulatory elements (REs) and recruit other proteins including cofactors, basal transcription factors and epigenetic modifiers to control gene expression. While recent advances in epigenomic profiling techniques including the development of ATAC-sequencing [1] have significantly increased our

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_10,

understanding of which REs are utilized in which cell type, it remains largely unknown which proteins interact with these REs due to a lack of appropriate techniques.

In vivo protein–DNA interactions are conventionally interrogated by chromatin immunoprecipitation (ChIP), an antibody-based method that allows isolation of DNA-binding proteins such as transcription factors followed by analysis of the associated DNA molecules usually by PCR or next generation sequencing [2]. However, the shortcomings of this invaluable epigenetic technique are: (1) it requires a priori knowledge (or at least an idea) of proteins that interact with the region of interest, (2) it relies on appropriate antibodies and (3) it allows interrogation of only one factor at a time. Consequently, using ChIP to analyze the multi-molecular composition of a regulatory complex formed at a specific RE of interest would currently not be feasible. To overcome these limitations, we have adapted Transcription Activator-Like Effector (TALE) proteins, which are usually utilized for (epi)genomic modification purposes [3], to isolate a specific genomic region of interest for proteomics analyses and have termed this novel epigenetic technique TINC: TALE-mediated isolation of nuclear chromatin [4] (Fig. 1). To minimize the number of factors enriched through off-target binding, we opted to use TALE proteins instead of the CRISPR-Cas9 system as TALEs are associated with significantly lower off-targeting events than the latter genome engineering tool [5].

In brief, TINC utilizes 3 × HA-tagged TALEs, which are designed to bind to a genomic region chosen for further analysis (Fig. 1). Cells of interest are transfected with a mammalian expression vector containing the customized TALEs and TALE-expressing cells are fixed with formaldehyde to reversibly crosslink protein–DNA and protein–protein interactions. The nuclei are then extracted followed by isolation of the chromatin to minimize background contribution from cytoplasmic contaminants as well as from proteins specifically enriched by the TALEs but without target locus-related functions (e.g., proteins that interact with the nascent TALE polypeptide chains during translation). Similar to a conventional ChIP procedure, the chromatin is then sheared by sonication and the TALE-chromatin complexes are enriched by immunoprecipitation targeting the 3 × HA-tag of the TALEs. After several washes followed by elution, nucleic acids and proteins can then be further processed for analysis. For example, next generation sequencing of the isolated DNA can be used to confirm target enrichment as well as identify potential off-target sites of the TALEs. On the other hand, mass spectrometry analyses of the proteins enriched by TINC can be utilized to determine the composition of the protein complexes formed at the genomic target regions. In order to eliminate proteins enriched unspecifically, empty vector transfected cells should be subjected to TINC as a

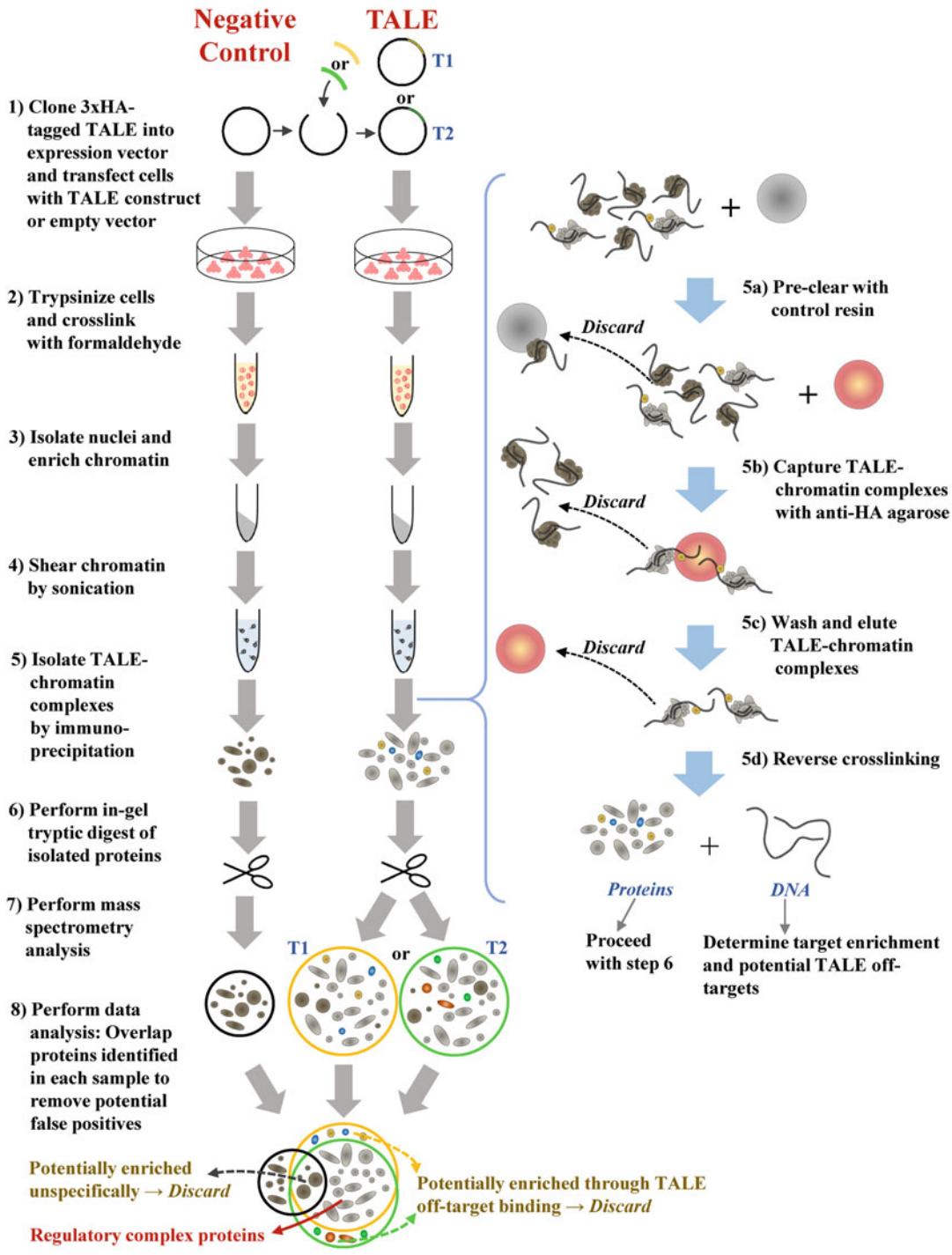


Fig. 1 Schematic of the TINC technique. Mammalian cells are transfected with a construct driving the expression of an epitope-tagged TALE protein designed to bind to a genomic region of interest. TALE expressing cells are then fixed with formaldehyde and the nuclei and the chromatin are isolated before sonication of the chromatin. The target region is then enriched by TALE epitope tag immunoaffinity purification, and nucleic acids and proteins are further purified and analyzed using appropriate techniques including

negative control. To minimize the number of proteins identified due to TALE off-target binding, we recommend performing TINC with two TALEs designed to bind the same genomic region and to consider only proteins as genuine binders that are enriched by both TALEs but not the negative control (Fig. 1). In our validation experiments, we targeted the *Nanog* promoter in mouse embryonic stem cells (ESCs), which allowed us to demonstrate the power of TINC and to provide insight into the multi-molecular composition of the transcriptional complex formed at this key pluripotency RE [4].

Here, we describe the TINC methodology as well as subsequent mass spectrometric approaches that have been utilized recently to analyze the regulatory complex formed at the *Nanog* promoter in ESCs [4] (note that this workflow is suitable to interrogate the protein composition of any genomic region of interest).

2 Materials

Unless otherwise stated, use Milli-Q water (MQ) and analytical grade reagents.

2.1 Generation of TALE-Expressing Cell Lines

1. Customized 3 × HA-tagged TALEs in pEF-DEST51 (*see* **Note 1**).
2. SgrDI restriction enzyme.
3. C57BL/6 ESCs.
4. Lipofectamine 3000.
5. Opti-MEM.
6. Blasticidin.
7. ESC media: KnockOut DMEM supplemented with 15% fetal bovine serum, 1% GlutaMAX Supplement, 1% MEM Non-Essential Amino Acids Solution, 1% Penicillin-Streptomycin, 0.1% 2-Mercaptoethanol, 1000 U/mL Leukemia Inhibitory Factor (LIF), 1 μM PD0325901, and 3 μM CHIR99021.
8. Gelatin-coated tissue culture dishes.
9. Cell culture incubator.

Fig. 1 (continued) next generation sequencing and mass spectrometry, respectively. This allows for validation of target isolation as well as analysis of which proteins are associated with this genomic region. In order to eliminate proteins enriched unspecifically, empty vector transfected cells should be subjected to TINC. Furthermore, we recommend using two different TALEs (e.g., T1 and T2) targeting the same locus in order to eliminate any proteins potentially enriched through TALE off-target binding

10. Anti HA-tag antibody suitable for western blotting.
11. Western blot equipment.
12. Quantitative PCR (qPCR) machine.
13. QPCR reagents and customized primers.

2.2 Crosslinking of the Cells

1. 0.25% trypsin–EDTA.
2. Cell counter.
3. PBS.
4. 37% formaldehyde solution.
5. 1.25 M glycine made up in PBS fresh on the day.
6. Orbital shaker.
7. Centrifuge suitable for processing larger sample volumes at 500 × *g* (*see* **Note 2**).

2.3 Isolation of Nuclei and Chromatin

1. Hydrophobic pipette tips (*see* **Note 3**).
2. Cell lysis buffer: 25 mM Tris-HCl pH 7.4, 0.1% Triton X-100, 85 mM KCl, supplemented with protease inhibitor cocktail prior to usage (1 tablet of Roche cOmplete Protease Inhibitor Cocktail per 50 mL of buffer).
3. 70 μm strainers.
4. RNase A.
5. Heat block or water bath at 37 °C.
6. High-speed centrifuge suitable for processing larger sample volumes at 20,000 × *g* (*see* **Note 4**).
7. SDS buffer: 50 mM Tris-HCl pH 7.4, 10 mM EDTA, 4% SDS.
8. Urea buffer: 10 mM Tris-HCl pH 7.4, 1 mM EDTA, 8 M urea. Prepare fresh on the day.
9. Laboratory balance.

2.4 Shearing of Isolated Chromatin

1. ChIP lysis buffer: 50 mM Tris-HCl pH 8.0, 1% SDS, 10 mM EDTA, supplemented with 1% Protease Inhibitor Cocktail DMSO solution prior to usage.
2. Sonicator such as a Bioruptor NextGen device (Diagenode).
3. 5 M NaCl.
4. Heat block at 100 °C.
5. Benchtop centrifuge to centrifuge samples at 15,000 × *g*.
6. DNA loading dye.
7. 1.5% agarose gel.
8. Electrophoresis buffer and equipment.

2.5 TALE-Chromatin Immunoprecipitation

1. Hydrophobic pipette tips (*see* **Note 3**).
2. ChIP lysis buffer: 50 mM Tris-HCl pH 8.0, 1% SDS, 10 mM EDTA, supplemented with 1% Protease Inhibitor Cocktail DMSO solution prior to usage.
3. Dilution buffer: 16.7 mM Tris-HCl pH 8.0, 165 mM NaCl, 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, supplemented with 1% Protease Inhibitor Cocktail DMSO solution prior to usage.
4. Low salt buffer: 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.5% Na deoxycholate, 0.1% SDS, 1% Nonidet P40 Substitute, 1 mM EDTA, supplemented with 1% Protease Inhibitor Cocktail DMSO solution prior to usage.
5. High salt buffer: 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 0.5% Na deoxycholate, 0.1% SDS, 1% Nonidet P40 Substitute, 1 mM EDTA.
6. TE buffer: 10 mM Tris-HCl pH 8.0, 0.25 mM EDTA.
7. 3 M NaSCN.
8. Tube rotator.
9. Benchtop centrifuge to centrifuge samples at 15,000 × *g*.
10. Centrifuge suitable for processing larger volumes (e.g., 50–100 mL) at 1000 × *g*.
11. Pierce Control Agarose Resin.
12. Pierce Anti-HA Agarose Resin.
13. 50 mL centrifuge tubes (e.g., Falcon).
14. DNA LoBind tubes.
15. 3 kilodalton (kDa) cutoff concentrators such as Amicon Ultra Centrifugal Filter Units.
16. PBS.
17. 5 M NaCl.
18. Heat block at 100 °C.
19. DNA loading dye.
20. 1.5% agarose gel.
21. Electrophoresis buffer and equipment.
22. 4× SDS loading dye: 125 mM Tris-HCl pH 8.3, 40% glycerol, 4% SDS, 400 mM DTT, 0.04% bromophenol blue.

2.6 In-Gel Tryptic Digest

The presence of detergents (such as Triton-X-100 or Tween-20) and human keratin in liquid chromatography–mass spectrometry (LC-MS) samples can significantly reduce the quality of the mass spectrometric results. To minimize contamination, we advise to (1) always wear appropriate personal protective equipment, (2) use only HPLC grade solvents, (3) minimize exposure of the

samples to plastic surfaces, (4) prepare all buffers fresh on the day and (5) store all buffers and reagents in clean, detergent-free glassware.

1. Precast 10-well 4–12% Bis-Tris protein gel.
2. MOPS SDS running buffer.
3. Electrophoresis equipment.
4. InstantBlue Coomassie Protein Stain.
5. Scalpels.
6. Protein LoBind tubes.
7. 50 and 100 mM ammonium bicarbonate.
8. Acetonitrile.
9. 100 mM ammonium bicarbonate, 5 mM DTT.
10. 100 mM ammonium bicarbonate, 200 mM 2-chloroacetamide.
11. Vacuum concentrator such as a CentriVap Concentrator (Labconco).
12. Sequencing grade trypsin.
13. Heat block at 100 °C.
14. Heat block at 65 °C.
15. Heat block at 37 °C.
16. Shaker.
17. Ice.
18. Sonication bath.
19. Spin-vortex.
20. Autosampler vials for LC-MS.

3 Methods

3.1 Generation of TALE-Expressing Cell Lines

Ideally, multiple TALEs targeting the genomic region of interest are designed and tested as follows to identify TALEs with superior specificity and enrichment efficiencies to be used in TINC.

1. Linearize the pEF-DEST51-TALE constructs with SgrDI and transfect C57BL/6 ESCs (or cells of interest) with the linearized constructs using Lipofectamine 3000. As negative control, transfect ESCs with the linearized empty pEF-DEST51 vector.
2. Propagate transfected ESCs in gelatin-coated dishes at 37 °C, 5% CO_2 in ESC media. 48 h post transfection, add blasticidin to the media at a final concentration of 3 μg/mL to eliminate any untransfected cells (*see* **Note 5**). Maintain selective pressure while culturing transfected ESCs.

3. After approximately 2 weeks, manually pick individual colonies, expand and analyze clonal lines for TALE expression by western blot using an anti HA-tag antibody (*see* **Note 6**).
4. Perform ChIP-qPCR to determine the target enrichment efficiency of each TALE (*see* **Note 7**).
5. Confirm that TALE expression and target binding does not change expression of the TALE target gene using RT-qPCR.
6. Proceed to Subheading 3.2 with the two most efficient TALEs as well as the negative control (*see* **Note 8**).

3.2 Crosslinking of the Cells

We usually work with 1×10^9 cells per TINC experiment, however, in our experience, this number can be reduced to 0.5×10^9 when working with a TALE that shows enrichment of approximately 0.4% input or above as determined in **step 4** of Subheading 3.1.

1. Trypsinize cells to cellularize, count and resuspend in PBS at approximately 1×10^6 cells/mL.
2. Add formaldehyde to a final concentration of 1% and incubate shaking at room temperature for 10 min.
3. Add glycine to a final concentration of 0.125 M and incubate shaking at room temperature for 10 min.
4. Collect cells by centrifugation at 500 × *g* for 5 min and discard supernatant (*see* **Note 2**).
5. Resuspend cells in ice-cold PBS at 1×10^6 cells/mL.
6. Collect cells by centrifugation at 500 × *g* for 5 min and discard supernatant.
7. Repeat PBS wash (**steps 5** and **6**).
8. Proceed to Subheading 3.3 (*see* **Note 9**).

3.3 Isolation of Nuclei and Chromatin

In order to minimize cytoplasmic and free nuclear protein contaminant contribution, TINC employs a nuclei isolation and chromatin enrichment step based on Kustatscher et al. [6].

1. Resuspend each pellet composed of 1×10^9 fixed cells in 50 mL of ice-cold cell lysis buffer.
2. Strain through a 70 μm strainer to remove any cell clumps.
3. Centrifuge at 2300 × *g* for 5 min at 4 °C.
4. Remove the supernatant (e.g., cytoplasmic fraction) by pipetting and discard.
5. Resuspend each nuclei pellet in 15 mL of cell lysis buffer, add 300 μL of 10 μg/μL RNase A, mix by pipetting and incubate samples at 37 °C for 30 min (*see* **Note 10**).
6. Centrifuge at 2300 × *g* for 10 min at 4 °C.
7. Remove the supernatant by pipetting and discard.

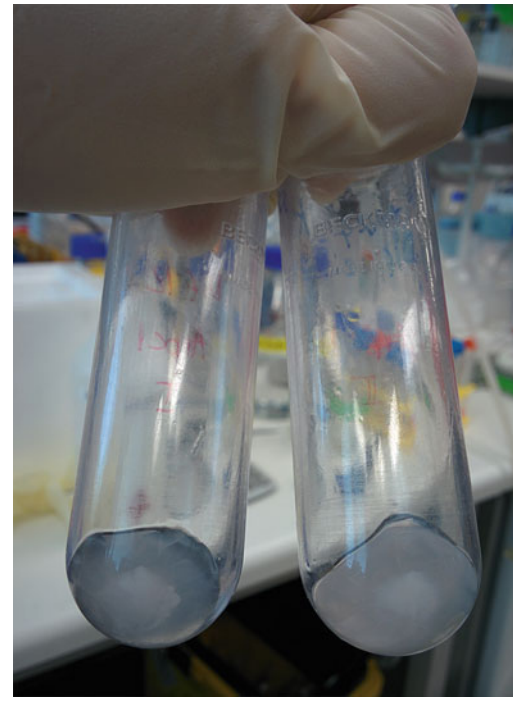


Fig. 2 Examples of the transparent chromatin pellets visible upon high-speed centrifugation performed in Subheading 3.3. Each vial shows a chromatin pellet obtained from approximately 0.5×10^9 ESCs

8. Resuspend each pellet in 10 mL of SDS buffer using hydrophobic pipette tips and incubate at room temperature for 10 min to lyse the nuclei (*see* **Note 3**).
9. Add 30 mL of Urea buffer, mix well by inverting and pass through a 70 μm strainer to remove any final clumps.
10. Centrifuge at 20,000 × *g* for 30 min at 25 °C.
11. The chromatin will be visible as transparent pellets (Fig. 2). Remove the supernatant by pipetting and discard.
12. To remove contaminants, resuspend each chromatin pellet in 10 mL of SDS buffer using hydrophobic pipette tips, add 30 mL of Urea buffer, mix well by inverting and centrifuge at 20,000 × *g* for 30 min at 25 °C.
13. Remove the supernatant by pipetting and discard.
14. To remove Urea, resuspend each pellet in 10 mL of SDS buffer using hydrophobic pipette tips, add an additional 30 mL of SDS buffer, mix well by inverting and centrifuge at 20,000 × *g* for 30 min at 25 °C.
15. Remove the supernatant by pipetting and discard.
16. Weigh each chromatin pellet and proceed to Subheading 3.4.

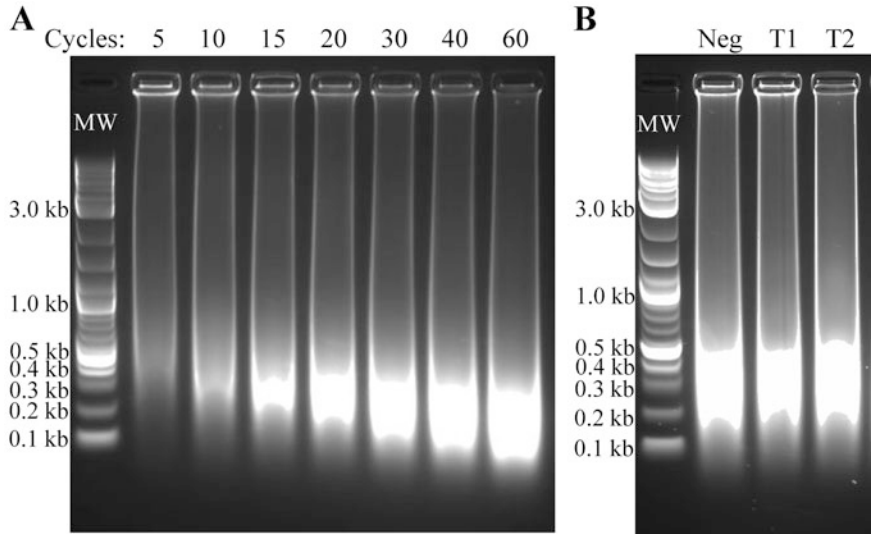


Fig. 3 1.5% agarose gels showing DNA fragment sizes after chromatin shearing. (**a**) Chromatin aliquots were sheared on a Bioruptor NextGen device (Diagenode) for 5, 10, 15, 20, 30, 40, or 60 cycles of 20 s on–30 s off in a test run to determine the optimal cycle number to obtain an average fragment size of 200–500 bp. (**b**) Based on the sonication test run shown in (**a**), the chromatin samples of the negative control as well as TALE 1 (T1) and TALE 2 (T2) were sonicated for 20 cycles of 20 s on–30 s off. These samples show the optimal fragment size for TALE-chromatin immunoprecipitation (Subheading 3.5)

3.4 Shearing of Isolated Chromatin

1. Resuspend chromatin in ice-cold ChIP lysis buffer at approximately 300 mg of chromatin per mL of ChIP lysis buffer.
2. Sonicate the sample to shear the chromatin to an average DNA fragment size of 200–500 bp (*see* **Note 11**). To determine the average chromatin fragment size, decrosslink 20 µL of the sonicated sample: add 2 µL of 5 M NaCl and incubate at 100 °C for 15 min. Spin decrosslinked sample at 15,000 × *g* for 5 min, transfer supernatant into a new tube, add DNA loading dye and separate the sample on a 1.5% agarose gel (Fig. 3).
3. Once an average DNA fragment size of 200–500 bp has been achieved, proceed to Subheading 3.5 (*see* **Note 12**).

3.5 TALE-Chromatin Immunoprecipitation

1. For each sample to be processed, decrosslink 20 µL as described in **step 2** of Subheading 3.4, purify the DNA and determine its concentration to ensure that the same amount of chromatin is used per immunoprecipitation (*see* **Note 13**).
2. Adjust sample volumes to 20 mL with ice-cold ChIP lysis buffer and add 80 mL of ice-cold dilution buffer for a 1:5 dilution (*see* **Note 14**).
3. Remove 50 µL as input sample and put aside on ice.

4. Add 1 mL of settled Pierce Control Agarose Resin to each sample and incubate rotating for 1 h at 4 °C to preclear samples (*see* **Note 15**).
5. Centrifuge at 1000 × *g* for 3 min at 4 °C to collect resin, transfer each supernatant (e.g., sample) into a new tube and discard the resin.
6. Add 1 mL of settled Pierce Anti-HA Agarose Resin to each sample and incubate rotating overnight at 4 °C (*see* **Note 15**).
7. Centrifuge at 1000 × *g* for 3 min at 4 °C to collect resin (e.g., sample) and discard the supernatant by pipetting.
8. Wash samples twice with ice-cold dilution buffer: add 100 mL of dilution buffer to each sample, incubate rotating at 4 °C for 10 min, centrifuge at 1000 × *g* for 3 min at 4 °C to collect resin and discard supernatant by pipetting (*see* **Note 14**).
9. Wash the samples twice with 100 mL of ice-cold low salt buffer as described in **step 8**.
10. Wash the samples twice with 100 mL of ice-cold high salt buffer as described in **step 8**.
11. Resuspend each resin in 3 mL of ice-cold TE buffer and transfer sample into a clean DNA LoBind tube.
12. Incubate rotating at 4 °C for 5 min, centrifuge at 1000 × *g* for 3 min at 4 °C to collect resin and discard supernatant.
13. To elute the TINC-chromatin complexes, resuspend each sample (e.g., resin) in 2 mL of 3 M NaSCN and incubate rotating for 5 min at room temperature.
14. Centrifuge at 1000 × *g* for 3 min at 4 °C to collect resin and transfer supernatants (e.g., samples) into a new DNA LoBind tube.
15. Repeat elution **steps 13** and **14** three times and pool the four elution fractions.
16. Concentrate the samples using 3 kDa cutoff concentrators to approximately 1 mL.
17. Dilute the samples 1:20 with PBS to dilute the NaSCN concentration and continue concentrating to a total volume of 50 µL.
18. Remove 1 µL, add 49 µL of PBS and process along with the input sample (**step 3**) as described in **step 1** to reverse the crosslinking and to purify the DNA and proceed to determine target enrichment efficiency by qPCR.
19. To the remaining sample, add 17 µL of 4× SDS loading dye, incubate at 100 °C for 30 min and proceed to Subheading 3.6 for protein analysis (*see* **Note 16**).

3.6 In-Gel Tryptic Digest (See Note 17)

1. Load the samples from **step 19** of Subheading 3.5 (*see* **Note 18**) onto a 4–12% Bis-Tris protein gel and run a short gel at 200 V in MOPS SDS running buffer to stack the proteins (*see* **Note 19**).
2. Rinse the gel with MQ and incubate in 50 mL of InstantBlue Coomassie Protein Stain shaking at room temperature for 1 h to visualize proteins.
3. Wash the gel with MQ shaking at room temperature until the background is destained (e.g., change MQ several times).
4. Excise the protein bands from the gel using a scalpel (Fig. 4a).
5. Cut each band into six cubes of approximately the same size and transfer the gel cubes into a Protein LoBind tube (Fig. 4b).
6. To equilibrate the gel pieces, add 500 µL of 100 mM ammonium bicarbonate and incubate for 10 min at room temperature with gentle agitation.
7. Remove the supernatant by pipetting and discard.

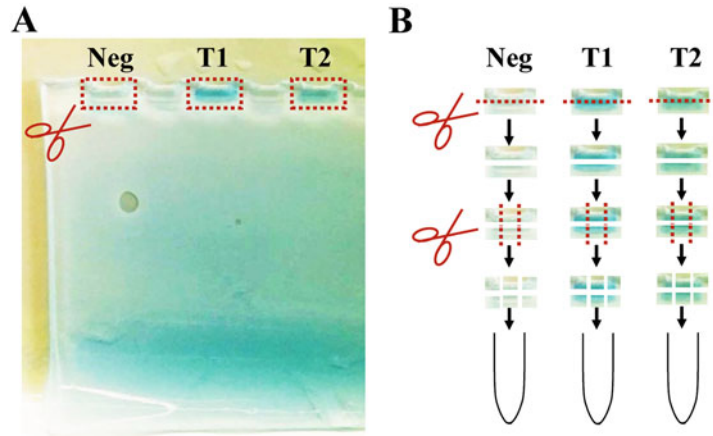


Fig. 4 Example of a short gel. (**a**) Protein samples isolated using TINC from empty vector transfected cells (e.g., Neg) and two different TALEs (e.g., T1 and T2) were loaded onto a NuPAGE 4–12% Bis-Tris Protein Gel. To minimize cross-contamination, samples were loaded into every second well leaving a well empty between samples. The gel was then run at 200 V for approximately 5 min in NuPAGE MOPS SDS Running Buffer to stack the proteins. The gel was then trimmed to cut off the wells before staining with InstantBlue Coomassie Protein Stain and destaining with MQ as described in Subheading 3.6. The protein bands were then excised as indicated by the dotted red lines. For all three samples, gel pieces of similar size were cut out and further processed, even though no obvious band was visible in the negative control. (**b**) The three gel pieces from A were then further cut as indicated by the dotted red lines to obtain six cubes of approximately the same size for each sample. These gel cubes were then transferred into Protein LoBind tubes to be equilibrated with 100 mM ammonium bicarbonate (**step 6** of Subheading 3.6)

8. To destain, add 500 μL of 50 mM ammonium bicarbonate, 50% acetonitrile to the gel pieces and incubate for 20 min at room temperature with gentle agitation.
9. Remove the supernatant by pipetting and discard.
10. Repeat **steps 8** and **9** an additional three times (total of 4), add 500 μL of 50 mM ammonium bicarbonate, 50% acetonitrile to the gel pieces and incubate at room temperature with gentle agitation.
11. Remove the supernatant by pipetting and discard.
12. Add 150 μL of 100 mM ammonium bicarbonate, 5 mM DTT to the gel pieces and incubate for 30 min at 65 °C (*see* **Note 20**).
13. Put the samples on ice for approximately 1 min to cool to room temperature.
14. Add 15 μL of 100 mM ammonium bicarbonate, 200 mM 2-chloroacetamide to each sample and incubate for 30 min at room temperature in the dark (*see* **Note 21**).
15. Remove the supernatant by pipetting and discard.
16. Wash the gel pieces with 500 μL of 50% acetonitrile, 50 mM ammonium bicarbonate shaking at room temperature for 10 min.
17. Remove the supernatant by pipetting and discard.
18. Dehydrate the gel pieces with 500 μL of 100% acetonitrile shaking at room temperature for 10 min.
19. Remove the supernatant by pipetting and discard.
20. Repeat **steps 18** and **19**.
21. Dry the samples for approximately 15 min in a vacuum concentrator until dry.
22. Add 50 μL of pre-chilled 100 mM ammonium bicarbonate, 5% acetonitrile containing 12.5 ng/μL sequencing grade trypsin to each sample (*see* **Note 22**).
23. Incubate on ice for 30 min.
24. Repeat **step 22** and incubate on ice for 2.5 h.
25. Place the samples at 37 °C and digest overnight.
26. Add 50 μL of 50% acetonitrile, 5% formic acid to each sample to quench the digest and vortex for 10 min at room temperature (*see* **Note 23**).
21. Transfer digested peptides (e.g., supernatant) into a clean Protein LoBind tube.
22. Add 100 μL of 50% acetonitrile, 5% formic acid to the gel pieces and vortex for 10 min at room temperature (*see* **Note 23**).

23. Transfer digested peptides into a clean Protein LoBind tube.
24. Add 200 µL of 100% acetonitrile to the gel pieces and vortex for 10 min at room temperature (*see* **Note 23**).
25. Transfer digested peptides into a clean Protein LoBind tube.
26. Pool corresponding fractions (e.g., extracted peptides) (*see* **Note 24**).
27. Dry samples to completion in a vacuum concentrator.
28. Resuspend each sample in 10 µL of 2% acetonitrile, 0.1% formic acid and incubate in a sonication bath for 15 min.
29. Spin-vortex the samples for 30 cycles (10 s spin, 10 s vortex).
30. Transfer the tryptic peptides into autosampler vials and proceed to Subheading 3.7.

3.7 Mass Spectrometry

The samples should be analyzed by liquid chromatography electrospray ionization high-resolution tandem mass spectrometry on any instrument capable of acquiring ms2 spectra or higher (*see* **Note 25**). The following protocol has been optimized for an Orbitrap Fusion Tribrid mass spectrometer (Thermo Scientific) coupled to a nanoLC Ultimate 3000 LC system (Thermo Scientific) equipped with a Dionex UltiMate 3000 RS autosampler (Thermo Scientific), an Acclaim PepMap 100 trap column (100 µm × 2 cm, nanoViper, C18, 5 µm, 100 Å; Thermo Scientific), and an Acclaim PepMap RSLC analytical column (75 µm × 50 cm, nanoViper, C18, 2 µm, 100 Å; Thermo Scientific).

1. Load samples via the Acclaim PepMap 100 trap column onto the Acclaim PepMap RSLC analytical column.
2. Elute tryptic peptides from the column at a flow rate of 250 nL/min using the following gradient (excluding pre- and post-equilibration steps): 2.5–7.5% buffer B in 1 min, 7.5–37.5% buffer B in 120 min, 37.5–42.5% buffer B in 3 min, and 37.5–99% buffer B in 5 min.
3. Acquire mass spectrometric data using the following parameters (Thermo Xcalibur 4.4.16.14; *see* **Note 26**):
   (a) Orbitrap full ms1 scan: resolution: 120,000; Scan range: 375–1575 *m/z*; AGC target: custom; Normalised ACG Target (%): 250; Maximum Injection Time Mode: Custom; Maximum Injection Time: 54 ms.
   (b) Ms2 selection: fixed cycle time: 2 s; Intensity Threshold: 5e4; Include change states: 2–7.
   (c) Orbitrap ms2 scans: Isolation window: 1.4 *m/z*; HCD Collision Energy: 32%; Resolution: 30,000; AGC target: custom; Normalised ACG Target (%): 400; Maximum Injection Time Mode: Custom; Maximum Injection Time: 54 ms.

3.8 Mass Spectrometric Data Analysis

Depending on the goal of the experiment and the acquisition method used, a plethora of software packages are available for both qualitative (e.g., Byonic [ProteinMetrics], Mascot [MatrixScience]) and quantitative (e.g., MaxQuant [7], Proteome Discoverer [Thermo Scientific], Spectronaut [Biognosys], Peaks [Bioinformatics Solutions]) analyses as well as for further downstream bioinformatic pipelines (e.g., LFQ-Analyst [8], Perseus [9]; *see* **Note 27**). Proper mass spectrometric quantification approaches based on ms^n signal intensities/areas are preferred over older quantification techniques such as spectral counting.

1. Search acquired .raw files against the murine UniProtKB/SwissProt database (v2017_07) appended with the TALE protein sequence using Byonic (Protein Metrics) considering a false discovery rate (FDR) of 1% using the target-decoy approach (*see* **Note 28**).
2. Specify carbamidomethylation of cysteine residues as a fixed modification. Select oxidation of methionine and acetylation of protein N-termini as variable modifications.
3. Use trypsin as the enzymatic protease and allow up to two missed cleavages.
4. To minimize false positive protein identifications, we suggest only considering proteins that were identified by at least two unique peptides.
5. We recommend performing at least three replicates for each TINC run including the control(s). If the different replicates have not been acquired in the same mass spectrometric batch and thus a proper quantification on ms1 or ms2 signal intensities cannot be performed, we recommend a present/absent approach. In Knaupp et al. [4], proteins enriched in at least two of the three negative controls were considered as background, and proteins detected in only one of the two different TALE samples as false positives.

4 Notes

1. We identified TALE binding sites using the online tool TALEN Targeter [10, 11] and assembled the TALEs using the interactive capped assembly sequential ligation approach [12] into a TALE backbone with an n-terminal 3 × HA-tag. As expression vector, we selected pEF-DEST51 as it contains the human EF-1α promoter to drive constitutive transgene expression, which is an effective promoter in mouse ESCs [13]. We have also successfully utilized a PiggyBac Transposase expression vector with a CAG promoter for constitutive TALE expression in mouse ESCs. However, an expression vector with a promoter suitable for the cell system of interest should be chosen.

2. We usually spin the samples in 500 mL aliquots in 500 mL centrifuge tubes, however, the samples can also be aliquoted to 50 mL volumes and centrifuged in 50 mL centrifuge tubes.
3. We recommend using uncoated tips such as epT.I.P.S LoRetention tips (Eppendorf) to prevent sample loss as exposure of the chromatin will make the sample very viscous and sticky.
4. We spin 40 mL sample volumes in an Avanti J-30I high speed centrifuge (Beckman Coulter Life Sciences).
5. We recommend performing a blasticidin kill curve to determine the optimal dose for the cells utilized.
6. Alternatively, clonal lines can be obtained by fluorescence-activated cell sorting (FACS) single ESCs onto a layer of irradiated mouse embryonic fibroblasts (MEFs). Using MEFs as feeder cells aids in the recovery of ESCs after FACS. We recommend working with clonal lines as we observed major heterogeneity in TALE expression by blasticidin-resistant pEF-DEST51 transfected ESCs. Such heterogeneity was not noted when TALE-expressing ESCs were created using a PiggyBac Transposase expression vector.
7. ChIP-sequencing can be performed to identify any TALE off-targets.
8. We recommend performing TINC with two different TALEs targeting the same locus in order to eliminate any proteins potentially enriched through TALE off-target binding.
9. Alternatively, fixed cell pellets can be stored at −80 °C for up to 6 months prior to proceeding with Subheading 3.3.
10. A RNA digestion step is included to minimize enrichment of ribosomes, which are a common contaminant of chromatin purifications, as well as proteins linked to nascent RNA. Do not add RNase A if chromatin-linked RNAs are to be analyzed upon regulatory complex isolation using TINC.
11. For us, this is usually achieved after 10–20 cycles of sonication (each cycle = 20 s on/30 s off) on a Bioruptor NextGen device (Diagenode) using 1.5 mL Bioruptor Microtubes (Diagenode) or 15 mL polystyrene conical centrifuge tubes.
12. Alternatively, the sample can be snap-frozen in liquid nitrogen and stored at −80 °C prior to proceeding with the procedure.
13. While the DNA concentrations are usually very similar across samples, we check the DNA concentration of each sample prior to TALE-chromatin immunoprecipitation and adjust sample volumes if required to proceed with the same amount of chromatin across all samples. To purify the DNA, we use a QIAquick PCR Purification Kit and to determine its concentration we perform a Qubit DNA quantification assay according to the suppliers' instructions.

14. We usually split each 100 mL sample into two 50 mL aliquots and use 50 mL centrifuge tubes for sample handling.
15. We recommend washing the resin in dilution buffer to remove any storage buffer prior to sample addition. To do so, add resin to 10 mL of dilution buffer, centrifuge sample at 1000 × *g* for 3 min and discard supernatant.
16. Alternatively, the sample can be snap-frozen in liquid nitrogen and stored at −80 °C prior to proceeding with the procedure.
17. Although in-gel digestions have been used in the original TINC publication [4], other techniques can also be utilized to prepare and clean up the samples for mass spectrometric acquisition including various in-solution or on-column digestion protocols. If a label-based quantification strategy will be pursued, additional steps have to be performed depending on the chosen labelling approach (*see* also **Note 26**). Alternatively, the samples can be subjected to western blot analysis to determine the presence of specific proteins.
18. If the samples were frozen at −80 °C, incubate for approximately 5 min at 100 °C prior to loading onto the gel.
19. The aim is to stop the gel when the entire sample has entered the gel (e.g., all of the blue loading dye has moved from the wells into the gel). For us, this usually takes 3–5 min at 200 V. Please note, we usually only load samples into every second well of the gel to minimize the potential of cross-contamination when loading and/or cutting the gel.
20. Ensure the gel pieces are fully submerged in the buffer. If required, increase the volume of 100 mM ammonium bicarbonate, 5 mM DTT.
21. Adjust volume accordingly depending on the volume utilized in **step 12** of Subheading 3.6.
22. We recommend validating that the pH is >8.0 by pipetting 1 μL on a pH strip. If this is not the case, add additional ammonium bicarbonate buffer until this is achieved. Furthermore, ensure the gel pieces are fully submerged in the buffer. If required, increase the buffer volume.
23. Adjust volume accordingly depending on the volume utilized in **step 22** of Subheading 3.6.
24. The sample can be frozen at −80 °C prior to proceeding with **step 28** of Subheading 3.6.
25. Providing detailed information on suitable LC-MS/MS systems and their operation is far beyond the scope of this article. Researchers with no or only limited expertise in mass spectrometry and proteomic analyses should seek advice and assistance from experts in proteomic laboratories or core facilities.

26. These parameters are optimized for data-dependent acquisition (DDA) mass spectrometry and subsequent label-free quantification based on ms1 signal intensities. Depending on the number of samples, their complexity and personal preference, the samples can also be acquired using data-independent acquisition (DIA) mass spectrometry in addition to various label-based strategies such as tandem mass tag (TMT) labelling (*see* also **Note 17**).
27. Providing detailed information on available software packages as well as subsequent bioinformatic pipelines to identify significantly regulated proteins is beyond the scope of this article. Proteomic and/or bioinformatic core facilities should be approached for assistance if required.
28. Irrespective of the software package(s) used, the search database(s) should be appended with the TALE protein sequences.

Acknowledgments

This work was supported by a National Health and Medical Research Council (NHMRC) grant (APP1069830) to J.M.P. J. M.P. was supported by the Australian Research Council (ARC) Stem Cells Australia Special Initiative, an NHMRC CDF (APP1036587), an ARC Future Fellowship (FT180100674) and a Silvia and Charles Viertel Senior Medical Research Fellowship. A.S.K. was supported by an NHMRC ECF (APP1092280). A.S.K, R.B.S. and J.M.P were supported by an ARC Discovery Project (DP210104029). We thank Flowcore, the MHTP Medical Genomics Facility, Micromon, the Monash Bioinformatics and Histology Platforms. This study used BPA-enabled/NCRIS-enabled infrastructure located at the Monash Proteomics and Metabolomics Facility led by R.B.S. The Australian Regenerative Medicine Institute is supported by grants from the State Government of Victoria and the Australian Government.

References

1. Buenrostro JD, Giresi PG, Zaba LC et al (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
2. Gade P, Kalvakolanu DV (2012) Chromatin immunoprecipitation assay as a tool for analyzing transcription factor activity. Methods Mol Biol 809:85–104
3. Li H, Yang Y, Hong W et al (2020) Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. Signal Transduct Target Ther 5:1
4. Knaupp AS, Mohenska M, Larcombe MR, Ford E, Nguyen T, Lim SM, Chen J, Firas J, Liu X, Nefzger CM, Schroder J, Rossello FJ, Haigh JJ, Lister R, Schittenhelm RB, Polo JM (2020) TINC - a method to dissect transcriptional complexes at single locus resolution - reveals novel Nanog regulators in mouse embryonic stem cells. Stem Cell Rep 15:1246
5. Wang X, Wang Y, Wu X et al (2015) Unbiased detection of off-target cleavage by CRISPR-

Cas9 and TALENs using integrase-defective lentiviral vectors. Nat Biotechnol 33:175–178

6. Kustatscher G, Wills KLH, Furlan C et al (2014) Chromatin enrichment for proteomics. Nat Protoc 9:2090–2099
7. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372
8. Shah AD, Goode RJA, Huang C et al (2020) LFQ-analyst: an easy-to-use interactive web platform to analyze and visualize label-free proteomics data preprocessed with MaxQuant. J Proteome Res 19:204–211
9. Tyanova S, Temu T, Sinitcyn P et al (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 13:731. https://doi.org/10.1038/nmeth.3901
10. Doyle EL, Booher NJ, Standage DS et al (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. Nucleic Acids Res 40: W117–W122
11. Cermak T, Doyle EL, Christian M et al (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Res 39:e82
12. Briggs AW, Rios X, Chari R et al (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. Nucleic Acids Res 40:e117
13. Zeng X, Chen J, Sanchez JF et al (2003) Stable expression of hrGFP by mouse embryonic stem cells: promoter activity in the undifferentiated state and during dopaminergic neural differentiation. Stem Cells 21:647–653

Chapter 11

Profiling Protein–DNA Interactions Cell-Type-Specifically with Targeted DamID

Owen J. Marshall and Caroline Delandre

Abstract

Targeted DamID (TaDa) is a means of profiling the binding of any DNA-associated protein cell-type specifically, including transcription factors, RNA polymerase, and chromatin-modifying proteins. The technique is highly sensitive, highly reproducible, requires no mechanical disruption, cell isolation or antibody purification, and can be performed by anyone with basic molecular biology knowledge. Here, we describe the TaDa method and downstream bioinformatics data processing.

Key words DamID, Targeted DamID, Gene transcription, Transcription factors, Protein-DNA interactions

1 Introduction

Targeted DamID (TaDa) [1, 2] is a recently developed method for profiling DNA–protein interactions cell-type specifically. The technique is based on DamID, a means of marking protein–DNA interactions by fusing a bacterial DNA adenine methylase (Dam) to any protein of interest [3, 4]. Adenine methylation—common in prokaryotes—is extremely sparse in eukaryotes [5], and Dam-fusion proteins leave enriched methylation at GATC sequences in close proximity to where they bind. Following genomic DNA isolation, methylated GATC sites are cut with methylation-sensitive restriction enzymes, enriched via ligation-mediated PCR and sequenced using next-generation sequencing. Resulting binding profiles can be used to identify transcription factor binding sites (e.g., [6]), gene transcription (using the proxy of RNA polymerase occupancy, e.g., [1]), chromatin state profiling (e.g., [7]) and chromatin accessibility profiling (e.g., [8]) in a semiquantitative manner. Notably, DamID can be used to profile the association of proteins that do not bind DNA directly (such as

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_11,

Lamin B [9]) as well as transcription factors that lack DNA-binding domains (O. Marshall, unpublished data) and noncoding RNAs [10].

As high levels of adenine methylation at GATC sites are toxic in eukaryotes [1], TaDa uses a bicistronic transcript and the phenomenon of spontaneous ribosomal reinitiation [11] to massively reduce the translation rates of Dam-fusion proteins [1]. The result is a DamID system that can be driven cell type–specifically by inducible bipartite expression systems such as GAL4/UAS [12]. Methylation marks are laid down gradually over the period of induction (which can be a period from as short as 16–48 h or longer) prior to tissue harvesting, and represent a form of profiling free of the potential artefacts associated with chemical fixation or cellular disruption. Importantly, the extremely low translation rates of Dam-fusion proteins mean that the concentration of these proteins are at effectively undetectable levels, and do not induce cell-fate changes or cause over-expression phenotypes. Despite this, the technique is highly sensitive, generating reproducible binding profiles from as few as 10,000 total driven cells.

Dam is a promiscuous adenine methylase, and will methylate all open chromatin regions in eukaryotes [13]. Dam-fusion proteins display similar behavior, and their methylation signature thus represents a combination of open chromatin methylation together with enrichment of methylation around specific binding sites of the protein of interest [13]. In order to account for this, all TaDa experiments with Dam-fusion proteins are carried out alongside a Dam-only control, and specific signal isolation is carried out in software [13]. An additional advantage of the methylation of open chromatin is that Dam-only TaDa experiments provide a simple and easy means to profile chromatin accessibility cell-type specifically (a feature that has been termed Chromatin Accessibility TaDa, or CATaDa) [8].

In *Drosophila melanogaster,* TaDa profiling is generally performed via a genetic cross between a GAL4 driver line and both an inducible Dam-fusion protein and a Dam-only control. A separate line is required to profile each DNA-associated protein, and preexisting lines are available for RNA polymerase profiling and chromatin protein profiling. New TaDa lines can be created by cloning and transgenesis using the pUAST-attB-LT3-NDam [1] or pTaDaG2 [14] base vectors. Alternatively, the FlyORF-TaDa system allows for the conversion of the extensive FlyORF library of lines to TaDa lines via a genetic cross [15].

Key to the success of TaDa is its ease of use. Cell-type-specificity is generated via genetic drivers rather than through mechanical disruption and sorting, and no antibodies are required. The technique is extraordinarily resilient and reliable, requiring only the most basic experience with genomic DNA isolation, restriction enzyme digestion and the PCR to generate highly reproducible

binding profiles. The procedure rarely if ever fails in our research group, and can be performed by undergraduate students with little previous laboratory experience.

Although this chapter has been written with experiments performed in *D. melanogaster* in mind, the technique has been adapted to both mammalian systems [16, 17] and recently to *Caenorhabditis elegans* [18], and, except for the initial tissue isolation steps, the following protocol can also be used on material generated from these systems.

2 Materials

2.1 Tissue Preparation

1. For fly head isolation, three woven wire test sieves of 150, 425, and 710 μm aperture size, respectively.
2. Phosphate Buffered Saline (PBS): place PBS tablet into water according to manufacturer's recommendations and autoclave.
3. 0.5 M Ethylenediaminetetraacetic acid (EDTA).
4. 1.5 mL homogenizing pestle: wipe and store in 100% ethanol after every use.

2.2 DNA Extraction

1. Quick-DNA Miniprep Kit (Zymo Research) or similar kit.
2. Agarose.
3. Tris Acetate EDTA (TAE) buffer.
4. SYBR™ Safe DNA Gel Stain (Thermo Fisher Scientific) or ethidium bromide.

2.3 DpnI Digestion

1. DpnI (20,000 units/mL).
2. CutSmart® Buffer (New England Biolabs), or buffer compatible with DpnI.
3. NucleoSpin Gel and PCR Cleanup Kit (Macherey-Nagel) or similar kit.

2.4 Adaptor Ligation

1. 50 μM dsAdR: combine 50 μL AdRt (100 μM) and 50 μL AdRb (100 μM) (*see* Table 1), incubate at 95 °C for 2 min in a removable heat block, and remove the heating block and allow to cool to room temperature (at least 45 min).
2. Adaptor Ligation Buffer: 2 μL of 10× T4 Ligase Reaction Buffer, 0.8 μL of dsAdR (*see* Subheading 2.4, **item 1**), 1.2 μL H_2O. Make a 100× batch and store at −20 °C in 40 μL aliquots.
3. T4 DNA Ligase (400,000 units/mL).

Table 1
Targeted DamID adaptor and primer sequences (* = phosphorothioate linkage)

Name	Sequence
AdRt	CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGA
AdRb	TCCTCGGCCG
DamID-PCR	GGTCGCGGCCGAGGATC
NGS-PCR1	AATGATACGGCGACCACCGA*G
NGS-PCR2	CAAGCAGAAGACGGCATACGA*G

2.5 DpnII Digestion

1. DpnII Digestion Buffer: 4 μL DpnII Buffer, 15 μL H_2O. Make a 100× batch and store at −20 °C in 190 μL aliquots.
2. DpnII (10,000 units/mL).

2.6 PCR Amplification

1. 50 μM DamID PCR primer (*see* Table 1).
2. PCR Buffer: 32 μL 5× MyTaq HS Buffer (Meridian Bioscience), 2.5 μL DamID-PCR primer, 83.5 μL H_2O. Make a 50× batch and store at −20 °C in 1180 μL aliquots.
3. MyTaq™ HS DNA Polymerase (Meridian Bioscience), or similar polymerase and buffer [2].
4. NucleoSpin Gel and PCR Cleanup Kit (Macherey-Nagel) or similar kit.
5. Spectrophotometer (e.g., NanoDrop™).

2.7 Sonication and Removal of DamID Adaptors

1. Sonicator.
2. 1.5 mL tubes compatible with sonicator (we use the Bioruptor® Plus sonicator and TPX microtubes from Diagenode).
3. DNA analyzer for sizing and quality control of DNA samples (e.g., TapeStation System; Agilent).
4. Genomic DNA ScreenTape and Reagents (Agilent) or equivalent for other system.
5. AlwI (10,000 units/mL).
6. CutSmart® Buffer (New England Biolabs) or buffer compatible with AlwI.

2.8 Sequencing Library Preparation

2.8.1 DNA Cleanup

1. Sera-Mag Speedbeads carboxyl magnetic beads.
2. Dilute Sera-Mag beads in polyethylene glycol (PEG) solution: in a 15 mL tube, make a solution of 3 g PEG 8000, 3 mL of 5 M NaCl, 150 μL of 1 M Tris−HCl pH 8.0, 30 μL of 0.5 M EDTA. Add H_2O to 14 mL and mix by inversion until PEG is dissolved. Mix Sera-Mag Speedbeads container well to ensure beads are in suspension and add 300 μL to the PEG solution. Mix by inversion and fill with H_2O up to 15 mL. Store at 4 °C.

3. Magnetic stand. To make a homemade magnetic stand (*see* **Note 1**): 75 × 10 × 3 mm N42 Neodymium magnet (e.g., FIRST4MAGNETS®; cat. no. F75103-2), 20 µL filter tip box insert, and 75 × 10 × 2.5 mm plastic spacer (supplied with a pair of the strip magnets). Magnetic stands can also be purchased (e.g., NEBNext® Magnetic Separation Rack, New England Biolabs).
4. Resuspension Buffer: 10 mM Tris−HCl pH 8.0, 0.1 mM EDTA. Make a 15 mL solution and store at −20 °C in 1 mL aliquots.

2.8.2 Concentration Adjustment

1. Qubit™ Fluorometer (Thermo Fisher Scientific) or similar instrument.
2. Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific).

2.8.3 End Repair

1. End Repair Enzyme mix: 1.14 µL T4 DNA Polymerase (3000 units/mL), 0.23 µL DNA Polymerase I, Large (Klenow) Fragment (5000 units/mL), 1.14 µL T4 Polynucleotide Kinase (10,000 units/mL). Make a 50× batch and store at −20 °C.
2. End Repair Buffer: 3 µL 10× T4 Ligase Reaction Buffer, 1.2 µL 10 mM dNTPs, 3.3 µL H_2O. Make a 50× batch and store at −20 °C in 35 µL aliquots.

2.8.4 Adenylation of 3′ Ends and Adaptor Ligation

1. Klenow Fragment 3′ to 5′ exo- (5000 units/mL).
2. Quick Ligation™ Kit (New England Biolabs) or similar kit.
3. Sequencing adaptors (*see* Table 2): resuspend adaptors at 100 µM in Tris-EDTA buffer solution with 50 mM NaCl. Mix 25 µL adaptor index and 25 µL universal primer. Incubate at 95 °C for 2 min in a removable heat block, and remove the heating block and allow to cool to room temperature (it will take at least 45 min).

2.8.5 DNA Cleanup

Same reagents as Subheading 2.8.1.

2.8.6 DNA Fragment Enrichment

1. NEBNext Ultra II Q5 Master Mix (New England Biolabs), or other High Fidelity DNA Polymerase, optimized for amplification of NGS libraries.
2. PCR Primer Cocktail: mix 25 µL PCR1 primer, 25 µL PCR2 primer (*see* Table 1), 50 µL H_2O. Store at −20 °C.

2.8.7 DNA Cleanup

Same reagents as Subheading 2.8.1.

2.8.8 Library Quality Control

Same reagents as Subheading 2.7, **items 3** and **4**, and Subheading 2.8.2.

Table 2
NGS adaptor sequences ([Phos] = 5′ phosphorylation, * = phosphorothioate linkage)

Name	Barcode	Sequence
Universal	N/A	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCT TCCGATC*T
Index 1	ATCAGG	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTTCTGCTT*G
Index 2	CGATGT	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTTCTGCTT*G
Index 3	TTAGGC	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTTCTGCTT*G
Index 4	TGACCA	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCAATCTCGTATGCCGTCTTCTGCTT*G
Index 5	ACAGTG	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCCGTCTTCTGCTT*G
Index 6	GCCAAT	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATATCTCGTATGCCGTCTTCTGCTT*G
Index 7	CAGATC	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCATCTCGTATGCCGTCTTCTGCTT*G
Index 8	ACTTGA	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACACTTGAATCTCGTATGCCGTCTTCTGCTT*G
Index 9	GATCAG	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACGATCAGATCTCGTATGCCGTCTTCTGCTT*G
Index 10	TAGCTT	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACTAGCTTATCTCGTATGCCGTCTTCTGCTT*G
Index 11	GGCTAC	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGCTACATCTCGTATGCCGTCTTCTGCTT*G
Index 12	CTTGTA	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACCTTGTAATCTCGTATGCCGTCTTCTGCTT*G
Index 13	AGTCAA	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGTCAACAATCTCGTATGCCGTCTTCTGCTT*G
Index 14	AGTTCC	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACAGTTCCGTATCTCGTATGCCGTCTTCTGCTT*G
Index 15	ATGTCA	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACATGTCAGAATCTCGTATGCCGTCTTCTGCTT*G
Index 16	CCGTCC	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACCCGTCCCGATCTCGTATGCCGTCTTCTGCTT*G
Index 18	GTCCGC	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTCCGCACATCTCGTATGCCGTCTTCTGCTT*G
Index 19	GTGAAA	[Phos]GATCGGAAGAGCACACGTCTGAACTCCAGTCACGTGAAACGATCTCGTATGCCGTCTTCTGCTT*G

3 Methods

As TaDa uses PCR amplification from small amounts of starting material, avoiding cross-contamination is crucial. From Subheading 3.1 onward, wear gloves, use filter tips, and use small aliquots of autoclaved MilliQ (or nuclease-free) H_2O to avoid contamination.

3.1 Tissue Preparation

1. Set up an appropriate cross to induce the TaDa system with a tissue-specific GAL4 driver, together with tub-GAL80ts (in flies); or with a Cre-specific driver (in mammals). Consider the induction time, the tissue and the cell-type being profiled. (*See* **Note 2**).
2. (If using flies) grow the progeny at 18 °C until induction is required.
3. (If using flies) induce by shifting progeny to 29 °C for 16–48 h (*see* **Note 3**).
4. After inducing the TaDa system, collect enough material for at least 10,000 driven cells. For isolated tissues, dissect in PBS; for embryos, wash in PBS; and in both cases transfer to a 1.5 μL tube, gently pellet and remove supernatant. For adult flies, transfer into 15 μL conical tubes cooled in dry ice. Samples may be stored at −20 or −80 °C at this point (*see* **Note 4**). If isolating adult fly heads, place the three sieves on top of each other with the largest aperture size sieve on top and the smallest at the bottom. Place the tube of frozen flies on dry ice, vortex, and add the flies to the top sieve. Shake or tap the sieves vigorously and check the middle sieve; it should contain the fly heads (*see* **Note 5**). Fly heads can be stored in 1.5 μL tubes at −20 °C until ready to use.
5. Take samples in 1.5 μL tubes from the freezer, and add 75 μL H_2O and 20 μL 0.5 M EDTA.
6. If working with embryos or heads, homogenize samples with a pestle by inserting it into the tube, twisting and pushing down with the pestle ~20 times, and progress immediately to the next step.
7. If working with dissected tissues, progress immediately to the next step.

3.2 DNA Extraction

All the reagents for this section are included in the Quick-DNA Miniprep Plus Kit (Zymo Research). An equivalent DNA preparation kit can be used. Be very gentle with samples prior to DpnI digestion (Subheading 3.3, **step 2**), as any genomic DNA shearing will result in broken ends that can potentially ligate DamID adaptors. Do not vortex but mix samples gently by inversion and/or very gentle flicking.

1. Prior to Subheading 3.1, **step 2**, make a Prot K Master Mix by combining 10 μL Proteinase K and 95 μL Solid Tissue Buffer (blue buffer) per sample, and vortexing to mix well.
2. Add the Prot K Master Mix to the sample and flick gently to mix.
3. Digest at 56 °C for 1–3 h in a heat block (*see* **Note 6**).
4. Cool samples to room temperature, add 400 μL Genomic Binding Buffer, and mix by gentle inversion.
5. Add samples to spin columns (*see* **Note 7**) on a vacuum manifold (*see* **Note 8**) and draw through.
6. Add 400 μL DNA Pre-Wash Buffer and draw through.
7. Add 700 μL gDNA Wash Buffer and draw through.
8. Add 200 μL gDNA Wash Buffer and draw through.
9. Transfer spin column to a collection tube and spin at maximum speed for 2 min.
10. Transfer spin column to a new 1.5 mL tube, add 50 μL DNA Elution Buffer, and leave for at least 30 min.
11. Centrifuge at >6000 × *g* for 1 min to elute the DNA.
12. Optional: run 1 μL on a 0.8% agarose get to check quality (*see* **Note 9**).

3.3 DpnI Digestion

All the reagents for the cleanup (**step 3**) are included in the NucleoSpin Gel and PCR Cleanup Kit (Macherey-Nagel). An equivalent PCR clean up kit can be used.

1. Transfer 43.5 μL of the sample to a new 1.5 mL tube.
2. Add 5 μL CutSmart Buffer and 1.5 μL DpnI, mix very gently, and digest at 37 °C overnight (*see* **Note 10**).
3. Add 100 μL Buffer NTI to the sample, transfer to a spin column, and draw through with a vacuum manifold.
4. Add 700 μL Buffer NT3 and draw through.
5. Repeat **step 4**.
6. Transfer spin column to a collection tube and centrifuge at 11,000 × *g* for 1 min.
7. Transfer spin column onto a new 1.5 mL tube, add 32 μL H_2O, incubate for 1 min at room temperature, and centrifuge at 11,000 × *g* for 1 min to elute DNA.

3.4 Adaptor Ligation

1. Transfer 15 μL of each sample into a PCR tube (*see* **Note 11**).
2. Store remaining sample at −20 °C or lower for future use if required (*see* **Note 12**).

3. Add 4 μL Adaptor Ligation Buffer and 1 μL T4 DNA Ligase. Mix well.
4. Incubate for 2 h at 16 °C followed by 10 min at 65 °C to heat-inactivate (*see* **Note 13**).

3.5 DpnII Digestion

1. Add 19 μL DpnII Digestion Buffer and 1 μL DpnII. Mix well.
2. Incubate for 2 h at 37 °C followed by 20 min at 65 °C (*see* **Note 13**).

3.6 PCR Amplification

1. Add 118 μL PCR Buffer and 2 μL MyTaq™ HS DNA Polymerase. Mix well.
2. Split into 4 × 40 μL reactions in a PCR strip (*see* **Note 14**).
3. Run the following PCR program (*see* **Note 15**):

1×	68 °C	10 min
1×	94 °C	30 s
1×	65 °C	5 min
1×	68 °C	15 min
3×	94 °C 65 °C 68 °C	30 s 1 min 10 min
17×	94 °C 65 °C 68 °C	30 s 1 min 2 min
1×	68 °C	5 min

4. Clean up using the NucleoSpin Gel and PCR Cleanup Kit (*see* Subheading 3.3). Use 320 μL of Buffer NTI and wash column with Buffer NT3 three times (instead of two) if using MyTaq™ HS DNA Polymerase. Elute by adding 32 μL H_2O to the spin column and leaving for at least 5 min before centrifuging.
5. Measure DNA concentration on a Qubit or similar high sensitivity instrument (*see* **Note 16**).
6. Run 1 μL on a 1% agarose gel (or on a Genomic DNA ScreenTape using the TapeStation System) to check for amplification of a broad 400 bp–2 kb range of fragments (Fig. 1a).

3.7 Sonication and Removal of DamID Adaptors

1. Dilute 2 μg of sample in 90 μL H_2O in 1.5 mL Bioruptor® Plus TPX microtubes (*see* **Note 17**).
2. Add 10 μL CutSmart Buffer and mix well. Cool on ice.
3. Sonicate in Bioruptor® Plus: add 500 mL ice to the chamber and top up to the designated mark with water, cool tube holder on ice, sonicate for 5 cycles of 27 s on–27 s off on high power (*see* **Note 18**).

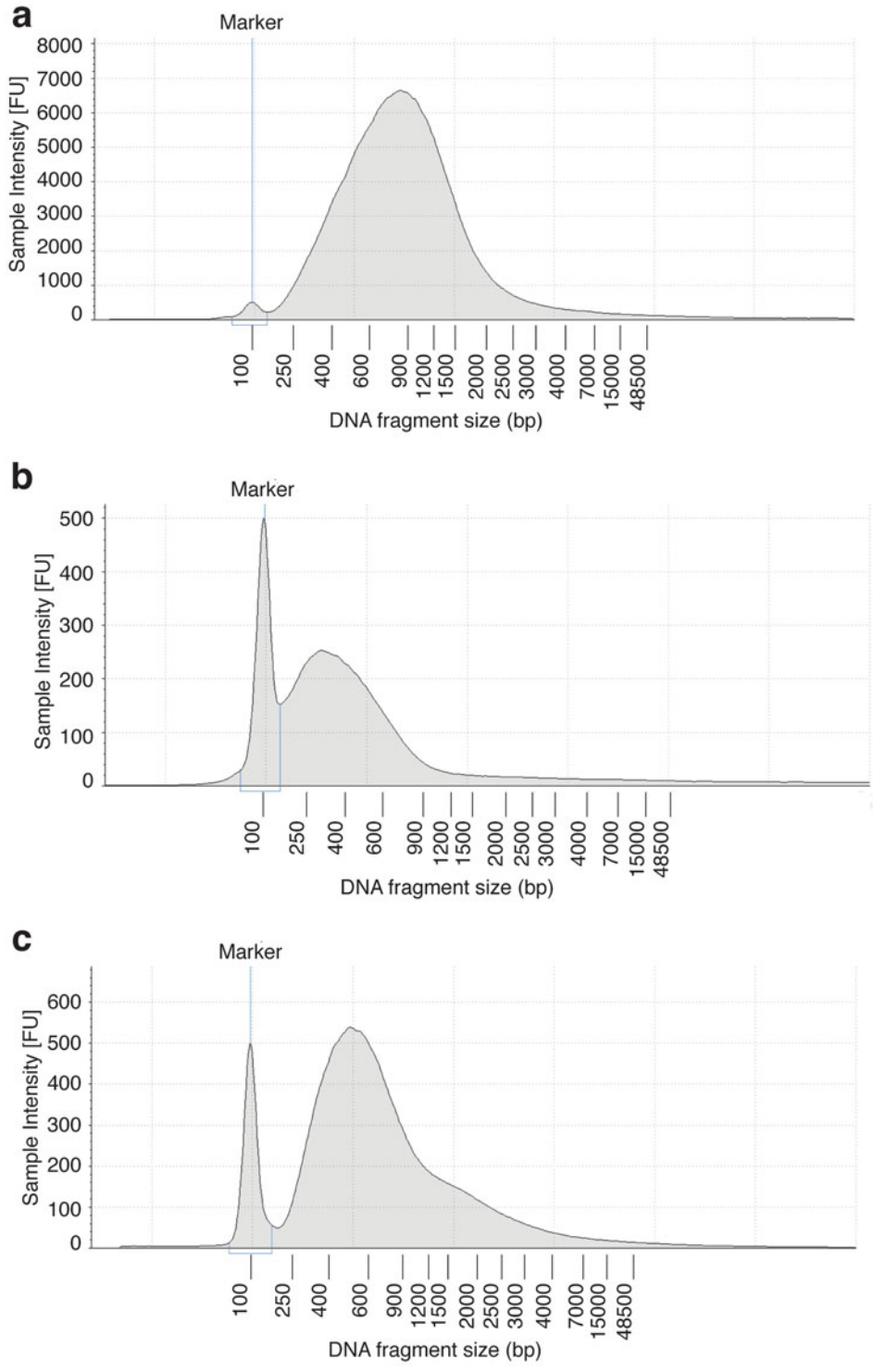


Fig. 1 Profiles of DNA fragments created by the TaDa technique. Example TapeStation plots showing the DNA size and quality for the same sample (**a**) before sonication (Subheading 3.6, **step 5**), (**b**) after sonication (Subheading 3.7, **step 4**), and (**c**) at the final library quality check (Subheading 3.8.9, **step 1**)

4. Check fragment size is ~300 bp (Fig. 1b) with a Genomic DNA ScreenTape using the TapeStation System, or equivalent (*see* **Note 19**).
5. Add 1 μL AlwI and digest at 37 °C overnight to remove DamID adaptors (*see* **Note 20**).

3.8 Sequencing Library Preparation

3.8.1 DNA Cleanup

1. Transfer 70 μL (or 90 μL, if concentration prior to sonication was low) of sample into a PCR tube and add 105 μL (or 135 to 90 μL) Sera-Mag beads. Mix well (*see* **Note 21**).
2. Incubate at room temperature for 10 min.
3. Place on magnetic stand for 10 min (or until the supernatant is clear).
4. Remove supernatant and wash twice with 190 μL 80% ethanol, 30 s each time.
5. Wait 5 min for the beads to air-dry.
6. Resuspend in 25 μL Resuspension Buffer and remove from magnetic stand. Mix well and incubate at room temperature for 2 min.
7. Place on magnetic stand for 5 min (or until the supernatant is clear).
8. Transfer 22.5 μL of the supernatant into a new PCR tube.

3.8.2 Concentration Adjustment

1. Use 1 μL of sample to measure DNA concentration using the Qubit™ dsDNA HS Assay Kit (or kit for equivalent instrument) following manufacturer's instructions.
2. Dilute samples to (no more than) 500 μg of DNA in 20 μL Resuspension Buffer in a PCR tube.

3.8.3 End Repair

1. Add 7.5 μL End Repair Buffer and 2.5 μL End Repair Enzyme mix. Mix well.
2. Incubate for 30 min at 30 °C followed by 20 min at 37 °C.

3.8.4 Adenylation of 3′ Ends

1. Add 0.75 μL Klenow Fragment 3′ to 5′ exo- and mix well.
2. Incubate for 30 min at 37 °C and proceed immediately to the adaptor ligation.

3.8.5 Adaptor Ligation

1. Add 2.5 μL Quick Ligase and 2.5 μL of the relevant adaptor (*see* **Note 22**).
2. Incubate for 10 min at 30 °C.
3. Add 5 μL 0.5 M EDTA.

3.8.6 DNA Cleanup

1. Add 40 μL Sera-Mag beads and mix well.
2. Incubate at room temperature for 10 min.

3. Place on magnetic stand for 5 min (or until the supernatant is clear).
4. Remove supernatant and wash twice with 190 µL 80% ethanol, 30 s each time.
5. Wait 5 min for the beads to air-dry.
6. Resuspend in 52.5 µL Resuspension Buffer and remove from magnetic stand. Mix well and incubate at room temperature for 2 min.
7. Place on magnetic stand for 5 min (or until the supernatant is clear).
8. Transfer 50 µL of the supernatant into a new PCR tube.
9. Add 50 µL Sera-Mag beads and mix well to start the second round of cleanup (*see* **Note 23**).
10. Incubate at room temperature for 10 min.
11. Place on magnetic stand for 5 min (or until the supernatant is clear).
12. Remove supernatant and wash twice with 190 µL 80% ethanol, 30 s each time.
13. Wait 5 min for the beads to air-dry.
14. Resuspend in 22.5 µL Resuspension Buffer and remove from magnetic stand. Mix well and incubate at room temperature for 2 min.
15. Place on magnetic stand for 5 min (or until the supernatant is clear).
16. Transfer 20 µL of the supernatant into a new PCR tube.

3.8.7 DNA Fragment Enrichment

1. Add 5 µL PCR Primer Cocktail and 25 µL NEBNext Ultra II Q5 Master Mix. Mix well.
2. Run the following PCR program:

1×	**98 °C**	**30 s**
6×	98 °C 60 °C 72 °C	10 s 30 s 30 s
1×	72 °C	5 min

3.8.8 DNA Cleanup

1. Add 50 µL Sera-Mag beads and mix well.
2. Incubate at room temperature for 10 min.
3. Place on magnetic stand for 5 min (or until the supernatant is clear).
4. Remove supernatant and wash twice with 190 µL 80% ethanol, 30 s each time.

5. Wait 5 min for the beads to air-dry.
6. Resuspend in 32.5 μL Resuspension Buffer and remove from magnetic stand. Mix well and incubate at room temperature for 2 min.
7. Place on magnetic stand for 5 min (or until the supernatant is clear).
8. Transfer 30 μL of the supernatant into a new tube.

3.8.9 Library Quality Control

1. Check DNA quality (Fig. 1c) with a Genomic DNA ScreenTape using the TapeStation System (*see* **Note 24**).
2. Measure library concentration using the Qubit™ dsDNA HS Assay Kit following manufacturer's instructions.
3. Pool samples to give a final DNA concentration of 20 nM.
4. Run on a next-generation sequencer to obtain at least 20 million reads per index (for *Drosophila melanogaster* samples) or >50 million reads per index with mammalian samples. Both single-end (SE) and paired-end (PE) sequencing data can be processed; however, we recommend PE sequencing if available cost-effectively.

3.9 Processing of Sequencing Data

1. Download sequencing data in FASTQ format (compressed fastq.gz files are fine) to a single directory. If possible, naming the files such that the sample name is at the start of the filename will simplify downstream processing. Dam-only control samples should ideally start with "Dam" (*see* **Note 25**).
2. Download and install the damidseq_pipeline software [13], freely available online from https://owenjm.github.io/damidseq_pipeline. Detailed installation and usage instructions, along with a test dataset, are available from the website.
3. Run damidseq_pipeline in the directory (use the --paired flag if using paired-end sequencing) (*see* **Note 26**). The final outputs are normalized Log_2(Dam-fusion/Dam-alone) binding profiles in BEDGRAPH format.
4. The resulting binding profiles can be viewed using browser software such as IGV (Integrative Genomics Viewer) [19], or publication-quality binding profile figures can be generated using pyGenomeTracks [20].
5. If profiling transcription factor binding, significantly enriched peaks called on false discovery rate (FDR) may be identified using find_peaks, freely available from https://github.com/owenjm/find_peaks. Peaks are outputted in GFF format, and can be visualized using the software described in Subheading 3.9, **step 4**, or used for subsequent processing and analysis.

6. If profiling gene expression through RNA polymerase occupancy, genes with significantly enriched polymerase occupancy (a proxy for gene expression) can be called from the RNA Polymerase II ratio files using the polii.gene.call Rscript, freely available from https://github.com/owenjm/polii.gene.call.
7. If comparing gene expression from different conditions, we recommend the use of NOISeq [21] on antilog-transformed occupancy data generated by polii.gene.call. An example of this workflow can be found in [22].

4 Notes

1. A homemade magnetic stand may be built by taping together the magnet with the plastic spacer under a 20 μL filter tip box insert (acting as a PCR strip rack) so the magnet will be near the tip of the PCR tubes once inserted in the stand. Use tape on the side of the PCR strip to ensure its edge remains firmly in contact with the magnet. Strength N42 works well, but there should be no harm in using more powerful magnets (resulting in a faster purification process).
2. We strongly recommend testing driver specificity under experimental conditions before proceeding with TaDa. If using the original TaDa constructs, we recommend testing with a UAS-inducible marker (e.g., membrane-bound GFP or nuclear RFP) inserted in the same targeted insertion site as the TaDa construct. If using the TaDaG2 vectors [14] induced cells are marked with membrane-bound GFP during TaDa induction, and a proportion of the experimental collection can be sacrificed for microscopy during dissection.
3. The choice of induction depends on the characteristics of the cell type being profiled. We typically use 16 h for embryonic neuroblasts, 24 h for larval cells and 48 h for adult neurons. In general, longer induction times increase the signal–noise ratio, but may be inappropriate for profiling specific stages in developmentally important cell types. Profiling for <12 h is not recommended.
4. Some degree of tissue dissection is recommended in most instances, although it is possible to use whole animals with highly specific GAL4 drivers ([13] and J. Newland, C. Delandre and O. Marshall, unpublished data). We do recommend caution and careful scrutiny of both the driver expression and the resulting data if not dissecting tissues, however. Note, in particular, that many GAL4 lines drive expression in the salivary glands and these will contribute significantly to the resulting binding profiles if not removed.

5. Check the top sieve under the microscope to assess how many fly heads actually separated from the rest of the body. If it is lower than 90%, try vortexing longer (we usually vortex three times for 10 s each and cool them on dry ice in between). We get the highest rates of fly head removal when vortexing while moving the conical tube so the sides are well shaken.
6. Check under the microscope that the tissue has been completely digested; only tissues such as cuticle and mouth hooks should be visible.
7. When working with fly heads, try to reduce the number of heads being transferred onto the column to prevent blockage.
8. If using a vacuum manifold, use disposable connectors (e.g., VasConnectors; QIAGEN) to avoid cross-contamination. DamID is an extremely sensitive technique that involves the PCR amplification of material, and we have observed significant signal contamination (especially from bacterial plasmids) when not using disposable connectors. If not using a vacuum manifold, centrifugation steps can be used as per the manufacturer's instructions without issue; however, we strongly recommend the use of a manifold for time-saving and ease-of-use considerations.
9. There should be a single band on the top of the gel and not a smear. Although it is sensible to perform this check the first time TaDa is carried out (or if using difficult tissues such as the gut) in general we rarely observe significant DNA degradation at this step.
10. The digestion can be reduced to 2 h if required (with some potential loss of sensitivity). DpnI can also be optionally heat-inactivated at 80 °C for 20 min, although it is effectively removed in the subsequent DNA cleanup step.
11. We recommend the use of 8-well PCR strips for medium throughput applications, and use fresh strip caps after every opening to avoid contamination. There are generally very little or undetectable amounts of DNA at this point, as uncut genomic DNA (which should be the majority of DNA in most use cases) will not pass through the spin column. The purified, cut DNA should only come from induced cells. If working with large amounts of tissue and/or if concerned about DNA shearing, the sample concentration can be measured and a maximum of 750 ng should be used going forward. However, such a high yield is unusual in practice. Note that undetectable yields at this stage can produce excellent data.
12. If the yield at Subheading 3.6, **step 5** is low (e.g., <1 μg) it is possible to repeat the processing on this stored aliquot and merge the two repeats at Subheading 3.7, **step 1**. We rarely find that this is necessary. Otherwise, we recommend retaining this aliquot for verification/backup purposes.

13. Unless indicated, we typically use a PCR machine for the enzyme incubations.
14. Splitting the reaction into 40 μL aliquots provides optimum reaction efficiency—we find that final yields are lower if the reaction is not split.
15. Both the initial long extension for 10 min and the subsequent long cycles are required. A 3 min extension time, as present in the final 17 cycles, should amplify all fragments below 5 kb; these represent 99.97% of all GATC fragments in the *Drosophila melanogaster* genome.
16. We find that trace carryover amounts of the MyTaq PCR buffer can significantly affect quantitation via a spectrophotometer, and we recommend using a DNA-chelating-dye-based quantitation method (such as a Qubit), especially if yields are expected to be low. Note that the yield of DNA at this point will greatly depend on both the starting number of cells and the extent of DNA binding exhibited by the protein in question. We have seen extremely low amplification values (e.g., <20 ng total yield) at this stage generate excellent quality final binding profiles.

    In terms of total expected yields, broad chromatin binding proteins (e.g., Polycomb, Brm, Lamin) can generate 2× to 10× the amount of DNA amplification observed in the Dam-only control. DNA polymerase components (e.g., RpII215 or RpII18) typically generate 0.5× to 1×, respectively, the amount of DNA amplification observed in the Dam-only control. The yield from transcription factors greatly depends on the specificity of binding.
17. If the yield from Subheading 3.6, **step 5** is <2 μg, use the total amount of sample at this point.
18. All sonicators are different and conditions should be optimized for the specific device. Note that sonicators typically provide custom-made sample tubes, and these must be used for reproducible fragmentation (the variation when using standard 1.5 mL tubes is unusably large in our experience). If possible, we recommend the use of modern sonicators with inbuilt water cooling—if using these, no ice is required, but prechilling of the unit is necessary.
19. Slight variations in fragment size are acceptable. Very large fragments (>600 bp) may impede clustering efficiency and sequencing yields, and we aim to avoid this (although we have never observed a sequencing library to fail in practice). It is acceptable to perform additional cycles of sonication if the fragment size is deemed too large at this step.
20. AlwI is a Type IIS restriction enzyme that cuts four nucleotides downstream from a GGATC site. Using this enzyme avoids the initial low-diversity of most library fragments commencing

with GATC, and is important for optimal cluster detection efficiency on many sequencing platforms. However, the use of ordered flow cells may alleviate this problem—and in these cases the enzyme Sau3A1 may be used instead to retain the full length of the GATC fragment. Other than reduced read numbers when retaining the initial GATCs, we do not observe any significant difference in TaDa profiles when cutting with either enzyme.

21. Mix by pipetting ~20 times, flick the PCR strip, and pulse down. Rapid mixing of beads and sample is important for even DNA precipitation.
22. If multiplexing four or fewer libraries, selecting adaptors with barcodes that are too similar may result in a reduced number of reads passing the filter. In this case, the preferred indexes to use (in order) are 4, 7, 6, and 8 (i.e., if multiplexing two samples, use indices 4 and 7). When using more than four libraries, we simply use library indices in numerical order and have never observed any issues with reduced read numbers on multiple NGS platforms.
23. This second cleanup step is required to ensure complete removal of sequencing adaptor dimers; should these be present in any library they will out-compete genuine library sequences in the multiplex when hybridizing to the flow cell.
24. The peak should have an average size of 600 bp, with no secondary peak present between 100 and 200 bp (that would indicate unremoved adaptors). A secondary peak twice the size of the original peak could mean the PCR reaction (Subheading 3.8.7, **step 2**) has been exhausted; this generally happens when performing eight or ten (or more) cycles. This peak results from concatemers of the amplified product. In such cases, reduce the number of PCR cycles. It is unclear whether these secondary peaks affect sequencing read numbers, and although we try to avoid them we have not observed any issues when they are present. They will not affect the downstream data quality.
25. Dam-only control files can also be manually specified on the command-line with the --dam = [filename] flag, but will be automatically detected if a single file starting with "Dam" is present.
26. If multiple replicates have been multiplexed, we recommend using the --just_align flag to generate BAM files from all samples, and then run the pipeline again on each set of BAM files from each replicate. Ideally, alignment rates should be above 90% (95% is typical).

References

1. Southall TD, Gold KS, Egger B, Davidson CM, Caygill EE, Marshall OJ, Brand AH (2013) Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: assaying RNA Pol II occupancy in neural stem cells. Dev Cell 26:101–112. https://doi.org/10.1016/j.devcel.2013.05.020
2. Marshall OJ, Southall TD, Cheetham SW, Brand AH (2016) Cell-type-specific profiling of protein–DNA interactions without cell isolation using targeted DamID with next-generation sequencing. Nat Protoc 11:1586–1598. https://doi.org/10.1038/nprot.2016.084
3. van Steensel B, Delrow J, Henikoff S (2001) Chromatin profiling using targeted DNA adenine methyltransferase. Nat Genet 27:304–308. https://doi.org/10.1038/85871
4. van Steensel B, Henikoff S (2000) Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat Biotechnol 18:424–428. https://doi.org/10.1038/74487
5. Luo GZ, He C (2017) DNA N 6-methyladenine in metazoans: functional epigenetic mark or bystander? Nat Struct Mol Biol 24:503–506. https://doi.org/10.1038/nsmb.3412
6. Doupé DP, Marshall OJ, Dayton H, Brand AH, Perrimon N (2018) Drosophila intestinal stem and progenitor cells are major sources and regulators of homeostatic niche signals. Proc Natl Acad Sci 115:12218–12223. https://doi.org/10.1073/pnas.1719169115
7. Marshall OJ, Brand AH (2017) Chromatin state changes during neural development revealed by in vivo cell-type specific profiling. Nat Commun 8:2271. https://doi.org/10.1038/s41467-017-02385-4
8. Aughey GN, Estacio Gomez A, Thomson J, Yin H, Southall TD (2018) CATaDa reveals global remodelling of chromatin accessibility during stem cell differentiation in vivo. elife 7:1–22. https://doi.org/10.7554/elife.32341
9. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 453:948–951. https://doi.org/10.1038/nature06947
10. Cheetham SW, Brand AH (2018) RNA-DamID reveals cell-type-specific binding of roX RNAs at chromatin-entry sites. Nat Struct Mol Biol 25:109–114. https://doi.org/10.1038/s41594-017-0006-4
11. Luukkonen BG, Tan W, Schwartz S (1995) Efficiency of reinitiation of translation on human immunodeficiency virus type 1 mRNAs is determined by the length of the upstream open reading frame and by intercistronic distance. J Virol 69:4086–4094
12. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development 118:401–415
13. Marshall OJ, Brand AH (2015) damidseq_pipeline: an automated pipeline for processing DamID sequencing datasets. Bioinformatics 31:3371–3373. https://doi.org/10.1093/bioinformatics/btv386
14. Delandre C, McMullen JPD, Marshall OJ (2020) Membrane-bound GFP-labelled vectors for Targeted DamID allow simultaneous profiling of expression domains and DNA binding. bioRxiv 1–4. https://doi.org/10.1101/2020.04.17.045948
15. Aughey GN, Delandre C, Southall TD, Marshall OJ (2020) FlyORF-TaDa allows rapid generation of new lines for *in vivo* cell-type specific profiling of protein-DNA interactions in *Drosophila melanogaster*. bioRxiv 2020.08.06.239251. https://doi.org/10.1101/2020.08.06.239251
16. Tosti L, Ashmore J, Tan BSN, Carbone B, Mistri TK, Wilson V, Tomlinson SR, Kaji K (2018) Mapping transcription factor occupancy using minimal numbers of cells in vitro and in vivo. Genome Res 28:592–605. https://doi.org/10.1101/gr.227124.117
17. Cheetham SW, Gruhn WH, van den Ameele J, Krautz R, Southall TD, Kobayashi T, Surani MA, Brand AH (2018) Targeted DamID reveals differential binding of mammalian pluripotency factors. Development 145:dev170209. https://doi.org/10.1242/dev.170209
18. Katsanos D, Barkoulas M (2020) Tissue-specific transcription factor target identification in the Caenorhabditis elegans epidermis using targeted DamID. bioRxiv
19. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. Nat Biotechnol 29:24–26. https://doi.org/10.1038/nbt.1754

20. Ramírez F, Bhardwaj V, Arrigoni L, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T (2018) High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat Commun 9(1):189. https://doi.org/10.1038/s41467-017-02525-w
21. Tarazona S, Furió-Tarí P, Turrà D, Di Pietro A, Nueda MJ, Ferrer A, Conesa A (2015) Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res 43(21):e140. https://doi.org/10.1093/nar/gkv711
22. Hatch HAM, Belalcazar HM, Marshall OJ, Secombe JA (2021)KDM5-Prospero transcriptional axis functions during early neurodevelopment to regulate mushroom body formation. Elife. 2021 Mar 17;10:e63886. https://doi.org/10.7554/eLife.63886. PMID: 33729157; PMCID: PMC7997662

Chapter 12

Genome-Wide Mapping and Microscopy Visualization of Protein–DNA Interactions by pA-DamID

Tom van Schaik, Stefano G. Manzo, and Bas van Steensel

Abstract

Several methods have been developed to map protein–DNA interactions genome-wide in the last decades. Protein A-DamID (pA-DamID) is a recent addition to this list with distinct advantages. pA-DamID relies on antibody-based targeting of the bacterial Dam enzyme, resulting in adenine methylation of DNA in contact with the protein of interest. This ^{m6}A can then be visualized by microscopy, or mapped genome-wide. The main advantages of pA-DamID are an easy and direct visualization of DNA that is in contact with the protein of interest, unbiased mapping of protein–DNA interactions, and the possibility to select specific subpopulations of cells by flow cytometry before further sample processing. pA-DamID is particularly suited to study proteins that form large chromatin domains or that are part of distinct nuclear structures such as the nuclear lamina. This chapter describes the pA-DamID procedure from cell harvesting to the preparation of microscopy slides and high-throughput sequencing libraries.

Key words Protein A-DamID (pA-DamID), Protein–DNA interactions, Genome-wide mapping, Fluorescence microscopy, ^{m6}A-tracer, High-throughput sequencing, Chromatin, Nuclear lamina

1 Introduction

Mapping of protein binding sites in the genome is an essential tool in the fields of chromatin biology and gene regulation. Chromatin immunoprecipitation followed by high-throughput sequencing has been the most widely used method for many years, but various alternatives have been developed, each one with its own strengths and weaknesses (reviewed in [1–3]). Protein A-DamID (pA-DamID) is a newly developed alternative, which combines the versatility of antibody-based detection with the principles of the DamID technology [4, 5].

pA-DamID is an implementation of a protein A-based profiling method, following the principles of ChIC and CUT&RUN [6, 7]. In these methods, permeabilized cells are incubated with an antibody against a protein of interest, which in turn is tagged with a fusion of protein A and micrococcal nuclease. Subsequent

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_12,

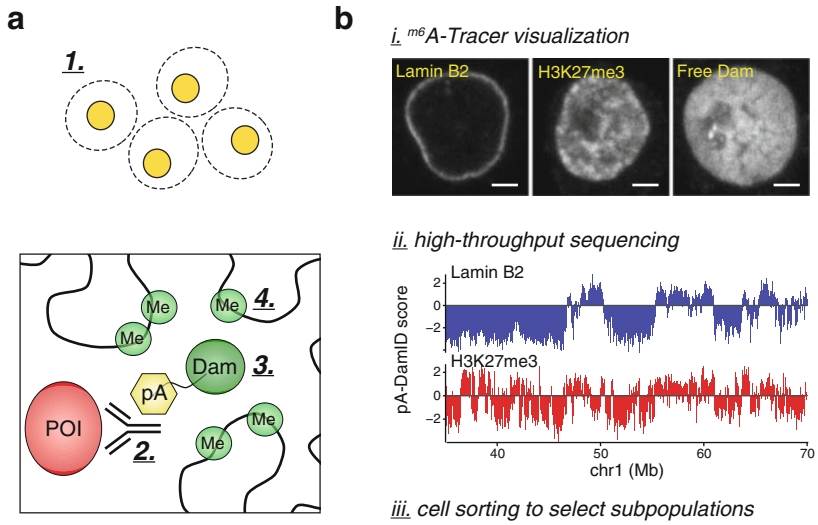


Fig. 1 Overview of the pA-DamID method. (Adapted with permission from [8]). (**a**) An overview of the steps in a pA-DamID experiment. Cells are permeabilized with digitonin (*step 1*) and incubated with a primary antibody against a protein of interest (POI) (*step 2*). This antibody is in turn bound by the pA-Dam fusion protein (*step 3*). Addition of SAM initiates adenine methylation deposition at nearby DNA (*step 4*). (**b**) An overview of the downstream possibilities after a pA-DamID experiment. The ^{m6}A modifications can be visualized in situ with the ^{m6}A-Tracer (*option i*). Example patterns are shown for Lamin B2 (Abcam ab8983, mouse, 1:100 dilution; combined with the bridging antibody Abcam ab6709 (*see* **Note 14**)) and H3K27me3 (CST C36B11, rabbit, 1:100 dilution) in human HAP-1 cells. Additionally, the ^{m6}A distribution of the free Dam control is shown (*see* **Note 9**). Scale bar corresponds to 2 μm. DNA can also be extracted and processed for high-throughput sequencing to infer protein interactions from the ^{m6}A pattern (*option ii*). Example data tracks are shown for the Lamin B2 and H3K27me3 experiments illustrated above. The pA-DamID score is defined as a $\log_2$-ratio of the target over the free Dam control. Finally, samples can be FACS sorted to select specific subpopulations before downstream sample processing (*option iii*) (*see* **Note 17**)

activation of the latter enzyme results in fragmentation of nearby DNA. These fragments are then identified by high-throughput sequencing. pA-DamID instead utilizes a fusion of protein A and DNA adenine methyltransferase (Dam). After addition of its methyl donor *S*-adenosyl-methionine (SAM) this results in ^{m6}A modifications on GATC sequences in proximity of the protein of interest (see Fig. 1a) [8].

The labeled cells can then be processed in three ways (see Fig. 1b). First, the methylated DNA can be visualized in situ with the ^{m6}A-Tracer, which consists of a ^{m6}A binding domain fused to a fluorescent protein [9]. This provides a quick visual check of the subnuclear location and intensity of the ^{m6}A signal. This serves as a quality control step, and in some applications can provide new biological insights. Second, isolated genomic DNA can be used to generate genome-wide binding profiles using the DamID library preparation [10]. Third, specific subpopulations of cells can be selected by fluorescence-activated cell sorting (FACS) prior to the

genome-wide mapping. Due to the sensitivity of the DamID library preparation protocol, only a few thousand sorted cells were sufficient for genome-wide binding profiles of nuclear lamina interactions [8].

pA-DamID should be performed with several controls, as described in more detail in the protocol below. The most important control is a separate sample incubated with free Dam enzyme in the presence of SAM, which is used to control for chromatin accessibility and unspecific binding of Dam. This control is used to normalize the data obtained with pA-Dam, and thus corrects for biases due to variation in chromatin accessibility across the genome, ensuring that pA-DamID specifically captures *bona fide* sites that interact with the protein of interest. A similar control is implemented in conventional DamID [10, 11], but not in alternative protein A-based methods [6, 7, 12]. This makes pA-DamID especially suited for antibodies that target inaccessible DNA such as heterochromatin marks.

Because Dam can only label GATC sequences, which occur in most genomes on average every ~200–400 bp, pA-DamID is particularly suitable for the mapping of proteins that form domains of at least several kb; it may be less suitable for proteins with narrow binding sites such as transcription factors. So far, we have used the method successfully with antibodies that target different types of heterochromatin [8]; other applications need to be tested.

Combined, the control for chromatin accessibility with free Dam, the visual readout of ^{m6}A-marked DNA and the low cell number requirement for library preparation make pA-DamID a powerful method to study protein–DNA interactions of nuclear compartments and chromatin proteins that generally bind in large domains.

2 Materials

Use nuclease free H_2O for all DNA steps.

2.1 pA-DamID Localization and Activation

1. Dam activity mix: 1× MethylTransferase buffer supplemented with 80 μM SAM (both shipped with Dam enzyme if purchased from New England Biolabs).
2. MboI digestion mix: 1× NEB buffer 3 supplemented with 10 mM $MgCl_2$ and 5 units of MboI restriction enzyme (suggested supplier, New England Biolabs).
3. 1 M HEPES–KOH, pH 7.5: dissolve 23.8 g HEPES in 90 mL demineralized H_2O. Mix and adjust pH with KOH pellets to 7.5. Make up to 100 mL with demineralized H_2O. Sterilize by filtration and store up to several months at 4 °C.

4. 5% w/v digitonin: dissolve 12.5 mg digitonin (Millipore) in 250 µL demineralized H_2O (*see* **Notes 1** and **3**). Keep on ice.
5. Dig-Wash buffer: 20 mM HEPES–KOH, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.02% w/v digitonin (*see* **Note 2**), 1× EDTA-free Protease Inhibitor Cocktail in demineralized H_2O. Prepare 50 mL for a routine pA-DamID experiment and keep on ice (*see* **Note 3**).
6. pA-Dam protein: purified pA-Dam protein (*see* **Note 4**).

2.2 Enrichment of ^{m6}A Labeled DNA and Preparation of Illumina Sequencing Library

For all master mixes, prepare 1.1× the required volume.

1. 50 µM DpnI adapter: mix equal volumes of the two adapter oligonucleotides at 50 µM (5′- CTAATACGACTCACTA TAGGGCAGCGTGGTCGCGGCCGAGGA and 5′-TCCTC GGCCGCG) in a microcentrifuge tube. Incubate for 5 min at 95 °C in a heat block. Turn off the heat block while keeping the tube inside to slowly cool the sample to below 50 °C. Aliquot and store at −20 °C.
2. 50 µM Y-shaped adapter: mix equal volumes of the two adapter oligonucleotides at 50 µM (5′- ACACTCTTTCCCTACAC GACGCTCTTCCGATCT and 5′-P- GATCGGAAGAGCA CACGTCT (*see* **Note 5**)) and anneal as described for the DpnI adapter.
3. Indexed P7 primers: 10 µM Illumina TruSeq primers (5′-CA AGCAGAAGACGGCATACGAG [8xN] GTGACTGGAGTT CAGACGTGTGCTCTTCCGATCT , where [8xN] is the sample-specific index sequence).
4. PCR clean-up magnetic beads: magnetic beads for DNA purification steps, for example CleanPCR beads, CleanNA.
5. DpnI digestion mix: 1 µL 10× CutSmart buffer, 0.5 µL DpnI (20 units/µL, suggested supplier, New England Biolabs), and 2 µL H_2O per sample. DpnI may be swapped with shrimp alkaline phosphatase (rSAP, 1 unit/µL, suggested supplier, New England Biolabs) in this mix, see Subheading 3.3, **step 3**.
6. DpnI adapter ligation mix: 0.25 µL of 50 µM DpnI adapter, 2 µL 10× ligation buffer, 0.5 T4 DNA ligase (5 units/µL, suggested supplier, Roche), and 7.25 µL H_2O per sample.
7. Methylation-specific PCR mix: 20 µL 2× MyTaq (Bioline), 1 µL 50 µM PCR primer (5′-NNNNGTGGTCGCGGCCGAG GATC), and 15 µL H_2O per sample.
8. End repair mix: 5 µL 10× end repair buffer, 5 µL dNTP, 5 µL ATP, 1 µL end repair enzyme mix (all reagents from Lucigen), and 9 µL H_2O per sample.

9. Klenow 3′A-overhang mix: 0.1 μL 100 mM dATP, 5 μL NEB buffer 2, 0.5 μL 50 units/μL Klenow Fragment exo- (50 units/μL), and 19.4 μL H_2O per sample.
10. Y-shaped adapter ligation mix: 1 μL 10× ligase buffer, 0.5 μL Y-shaped adapter, 0.5 μL T4 DNA ligase (5 units/μL), and 1.5 μL H_2O per sample.
11. Index PCR mix: 10 μL 2× MyTaq (Bioline), 0.5 μL 10 μM P5-primer (5′- AATGATACGGCGACCACCGAGATCTA CACTCTTTCCCTACACGACGCTCTTCCGATCT), and 1 μL H_2O per sample.

2.3 Visualization of ^{m6}A Labeled DNA

1. ^{m6}A-Tracer protein*:* purified ^{m6}A-Tracer protein (*see* **Note 4**).
2. 2% w/v formaldehyde: mix 13.25 mL PBS with 757 μL of 37% formaldehyde stabilized with methanol.
3. 0.5% v/v NP40: dilute 1 mL NP40 with 9 mL of PBS for a 10% NP40 stock that keeps several months at room temperature. Dilute 500 μL of this 10% stock with 9.5 mL PBS for a 0.5% NP40 solution.
4. 1% w/v BSA: dissolve 0.1 g BSA in 10 mL of PBS. Mix well to dissolve completely.
5. 1 mg/mL 4′,6-diamidino-2-phenylindole (DAPI): Dissolve 1 mg DAPI in 1 mL H_2O. Aliquot and store at −20 °C for several months.
6. Mounting medium: mounting medium to preserve fluorescence, for example Vectashield, Vector Laboratories.

2.4 Equipment and Labware

1. Refrigerated centrifuge (4 °C) for 1.5 mL microcentrifuge tubes.
2. Tube rotator for 1.5 mL microcentrifuge tubes at 4 °C, to gently rotate tubes during incubation steps (~20 rotations/minute).
3. Suction system to remove supernatant from 1.5 mL microcentrifuge tubes.
4. Heat block for 1.5 mL microcentrifuge tubes between 37 and 95 °C.
5. NanoDrop spectrophotometer or other labware to measure DNA concentration, for example a Qubit (Invitrogen).
6. PCR machine.
7. Illumina sequencing machine, for example, HiSeq 2500. Different adapters/primers may be required for different machines.

3 Methods

3.1 pA-Dam Activity Testing

1. The Dam activity of pA-Dam should be determined in activity units (*see* **Note 6**), and can be estimated by comparison with a calibration series of known activities. Incubate dilutions of NEB Dam enzyme (0.25, 1, 4 and 16 NEB units) and dilutions of pA-Dam protein (typically between 0.001 and 2 μL) with 500 ng of adenine unmethylated plasmid (*see* **Note 7**) in 20 μL of Dam activity mix for 30 min at 37 °C, followed by heat inactivation for 15 min at 65 °C.
2. To 10 μL of this reaction, add 40 μL MboI digestion mix. Incubate for 1 h at 37 °C and run 20 μL of the digestion on an 1% agarose gel.
3. Estimate the activity of pA-Dam by comparing the extent of MboI protection with NEB Dam units (*see* Fig. 2).

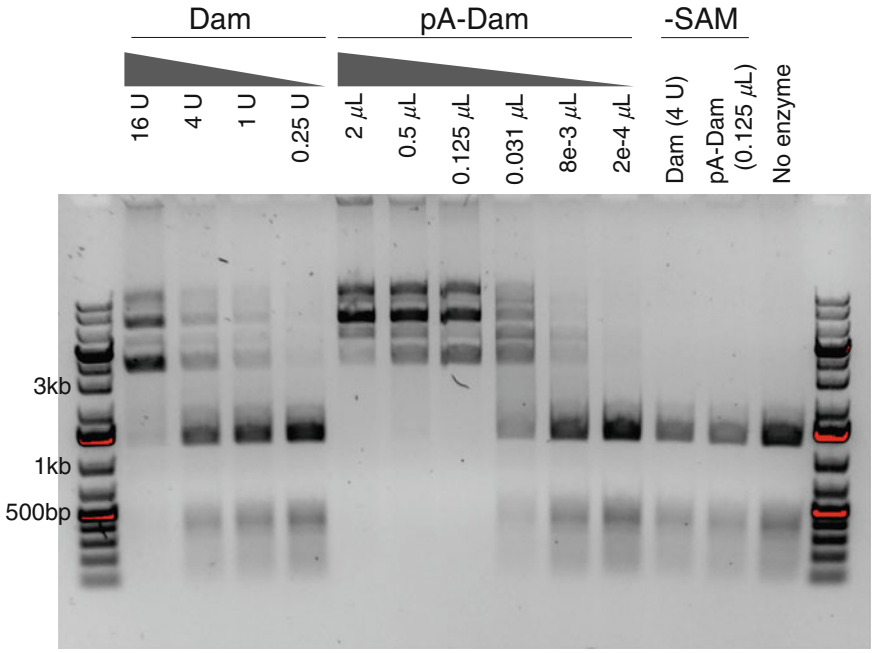


Fig. 2 Gel analysis to calibrate pA-Dam activity. An example gel image used to estimate Dam activity of purified pA-Dam protein. Unmethylated plasmid is incubated with consecutive dilutions of NEB Dam enzyme and pA-Dam protein (1.2 mg/μL). The amount of ^{m6}A methylation can be visualized by the extent of protection from the MboI restriction enzyme. A similar digestion pattern indicates a similar Dam activity, which for example can be seen between 1 unit of NEB Dam (U in the figure) and 8 × 10e−3 μL of pA-Dam. We thus estimated the activity of pA-Dam to be about 120 NEB units/μL. This corresponds to approximately 100 units/mg protein. Controls devoid of SAM should be unprotected and thus fully digested

3.2 pA-Dam Localization and Activation

Carry out all pA-DamID steps on ice unless otherwise specified. Use a P1000 pipette for 200 µL volumes.

1. Harvest cells (*see* **Note 8**).
2. For each epitope that needs to be mapped or visualized, prepare one million of unfixed cells (*see* **Notes 10** and **11**). Take control conditions into account and prepare additional cells accordingly (*see* **Note 9**). Centrifuge cells for 3 min at 500 × *g* in a precooled centrifuge (4 °C) and remove supernatant (*see* **Note 12**). Resuspend the cell pellet in 0.5 mL of ice-cold PBS in 1.5 mL microcentrifuge tubes. Repeat once, followed by resuspension in 0.5 mL of ice-cold Dig-Wash buffer. This washing procedure will be repeated throughout the protocol.
3. Centrifuge cells and resuspend the pellet in Dig-Wash buffer to reach a concentration of 5 M cells/mL. Divide samples between new tubes (200 µL for each antibody or control) and add the antibody of interest. For most commercial antibodies, a dilution of 1:100 is a good starting concentration for a strong signal, but can be optimized for a better dynamic range (*see* **Note 13**). Control experiments are incubated without any antibody. Rotate tubes for 2 h at 4 °C, followed by centrifugation and one wash with Dig-Wash.
4. Optionally, resuspend the cells in 200 µL Dig-Wash with a secondary anti-rabbit antibody (*see* **Note 14**), rotate for 1 h at 4 °C and wash once with Dig-Wash.
5. Next, resuspend the cells in 200 µL with 20–60 NEB units of pA-Dam. Rotate for 1 additional hour at 4 °C and wash two times with Dig-Wash.
6. Resuspend the cells in 100 µL of Dig-Wash supplemented with 80 µM SAM and incubate for 30 min at 37 °C to induce ^{m6}A methylation. For the Dam-only control, add an additional 0.5 µL of Dam enzyme to the sample and mix gently. At the end of incubation, return samples to 4 °C, centrifuge and remove supernatant (*see* **Note 15**).
7. Continue with library preparation in Subheading 3.3 (*see* **Note 16**), microscopy slides to visualize ^{m6}A methylation in Subheading 3.4 or selecting specific subpopulations by FACS before further sample processing (not discussed here, *see* **Note 17**).

3.3 Enrichment of ^{m6}A Methylated DNA and Preparation of Illumina Sequencing Library

1. Isolate genomic DNA (gDNA) and determine the concentration (*see* **Note 18**).
2. Mix between 10 and 500 ng of gDNA with H_2O to a total volume of 6.5 µL (*see* **Note 19**). Prepare at least one control with DpnI omitted and one control with Ligase omitted (Subheadings 3.3, **steps 3** and **5**). Additionally, include a negative control without any DNA.

3. (a) If cell death occurred or double strand DNA breaks were induced (*see* **Note 20**), replace DpnI in the DpnI digestion mix with 0.5 μL of rSAP and add 3.5 μL of this mix, incubate for 1 h at 37 °C and heat inactivate for 10 min at 65 °C. Then add 0.5 μL of DpnI and continue with the incubation in **step 4**.

   (b) If rSAP was not used, add 3.5 μL of the DpnI digestion mix and continue with the incubation in **step 4**.

4. Incubate for 8 h at 37 °C, followed by 20 min of heat inactivation at 80 °C.

5. On ice, add 10 μL of the DpnI adapter ligation mix to every sample and incubate for 16 h at 16 °C, followed by 10 min of heat inactivation at 65 °C.

6. Add 4 μL of the ligation reaction to 36 μL of Methylation-specific PCR mix in PCR tubes. Amplify for 13–23 PCR cycles with the following cycling scheme: 8 min of 72 °C, 13–23 cycles of 94 °C for 20 s, 58 °C for 30 s, 72 °C for 20 s and a final extension at 72 °C for 2 min (*see* **Note 21**).

7. Run 4 μL of each sample on a 1% agarose gel to check for the presence of amplified material. This will appear as a smear between 250 and 1500 bp (*see* Fig. 3 and **Note 22**).

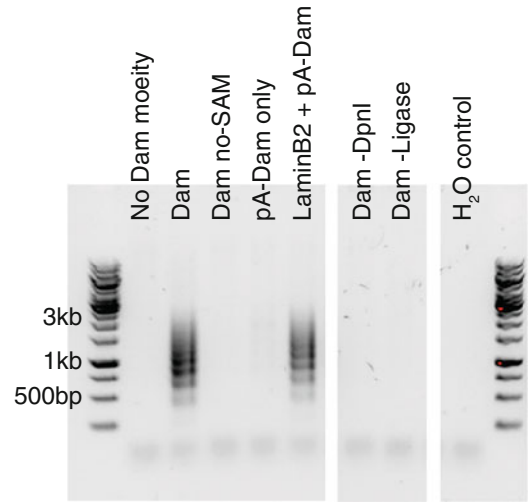


Fig. 3 Gel analysis of amplified ^{m6}A-labeled DNA fragments. (Reproduced with permission from [8]). DNA isolated from a typical pA-DamID experiment was digested with DpnI and ligated to PCR adapters. After 15 PCR cycles, samples were run on an agarose gel to visualize amplification. The negative control, no-SAM control and samples without DpnI or Ligase during the DNA processing should be devoid of signal, and the sample without primary antibody should be significantly weaker than with primary antibody. Samples with clear signal can be further processed for high-throughput sequencing

8. Purify the samples that show a stronger signal compared to the control without primary antibody with PCR clean-up magnetic beads. Add 1.8× volumes of beads, and follow manufacturer's instructions. Elute in 25 μL H_2O (*see* **Note 23**). Alternatively, use a spin column-based PCR purification kit to extract the DNA fragments from the PCR mixture.
9. Add 25 μL of End repair mix and incubate for 45 min at room temperature. Immediately proceed with DNA purification as described in **step 8**.
10. Add 25 μL of Klenow 3′A-overhang mix and incubate for 30 min at 37 °C, followed by 20 min of heat inactivation at 75 °C. Purify DNA as described in **step 8**, but elute in 20 μL of H_2O (*see* **Note 24**). Determine the DNA concentration.
11. Mix 220–250 ng of DNA from **step 10** with H_2O to a total volume of 6.5 μL. Add 3.5 μL of Y-shaped adapter ligation mix and incubate for 16 h at 16 °C, followed by 10 min of heat inactivation at 65 °C. As an additional negative control, include at least one sample with Ligase omitted. Purify DNA as described in **step 8** but elute in 20 μL of H_2O.
12. To 11.5 μL of Index PCR mix, add 8 μL of DNA from **step 11** and 0.5 μL of indexed P7 sequencing primer. Amplify for 9–-12 cycles with the following cycling scheme: 1 min of 94 °C, 9–12 cycles of 94 °C for 30 s, 58 °C for 30 s, 72 °C for 30 s and a final extension at 72 °C for 2 min (*see* **Note 25**).
13. Run 4 μL of each sample on a 1% agarose gel to check for the presence of amplified DNA.
14. Estimate the intensity on gel and pool the samples accordingly to achieve an even representation in the sequencing library. Purify fragments with 1.6× beads as described in **step 8**. Run a small amount of purified DNA on gel to confirm removal of primers.
15. High-throughput sequencing on an Illumina machine (*see* **Note 26**).

3.4 Visualization of ^{m6}A Labeled DNA

All steps are performed at room temperature unless otherwise stated. Keep samples in the dark when fluorescent molecules are present. Prevent drying out of samples by using one hand to remove liquid (i.e., using suction) and adding new liquid with the other hand.

1. Resuspend cells from Subheading 3.2, **step 7** in PBS and incubate on poly-L-lysine–coated coverslips for 30 min on ice, at a cell density of approximately 0.1 million cells/cm^2 (*see* **Notes 27** and **28**).
2. At room temperature, remove PBS and add 2% formaldehyde in PBS, incubate for 10 min and wash with PBS (*see* **Note 29**).

3. Permeabilize cells with 0.5% NP40 in PBS for 20 min (*see* **Note 30**).
4. Block unspecific protein binding sites with 1% BSA in PBS for 30–60 min.
5. Optionally, incubate with primary antibodies in 1% BSA for 1 h (*see* **Note 31**) and wash 3 times with PBS.
6. Incubate with ^{m6}A-Tracer protein and optionally secondary antibodies for 30–60 min (*see* **Note 32**) and wash once with PBS.
7. Stain DNA with 1 μg/mL DAPI for 10 min and wash two times with PBS and one time with demineralized H_2O. Immediately continue with **step 8**.
8. Dry coverslips by touching the rim on a paper tissue and place upside down on a microscopy slides with a drop of mounting medium. Carefully remove excess mounting medium with a tissue.
9. Seal with nail polish and let dry.

4 Notes

1. Use prewarmed (~95 °C) H_2O to dissolve digitonin and mix by pipetting several times up and down. Wear gloves while handling digitonin and be aware of digitonin toxicity concerns.
2. We have successfully used 0.02% w/v digitonin in several human and mouse cell lines, but the optimal concentration can differ in other cell lines. Complete permeabilization is required for unbiased DNA labeling, while too much digitonin can result in nuclear disruption during the pA-DamID protocol. To determine the optimal concentration, incubate and wash cells once with Dig-Wash buffer with multiple digitonin concentrations. Induce ^{m6}A labeling as Dam control (see methods Subheading 3.2) and prepare microscopy slides to visualize ^{m6}A (see methods Subheading 3.4). The optimal digitonin concentration is the lowest concentration with complete and homogenous ^{m6}A labeling. Alternatively, DAPI entry can be used to quickly assess cell permeability, but given the small size of DAPI compared to antibodies this might overestimate the permeabilization.
3. Buffers should remain stable at 4 °C for up to a week, but we prefer to make fresh buffers for every experiment.
4. Plasmids encoding for the pA-Dam and ^{m6}A-Tracer proteins will be shared upon request for protein purification, for example as described in [8]. Additionally, an aliquot of purified protein for initial testing will be shared upon request.

5. This oligonucleotide requires 5′ phosphorylation for the ligation with A-tailed DNA.
6. NEB Dam units are defined as the amount of enzyme required to protect 1 μg of unmethylated Lambda DNA in 1 h at 37 °C in a total reaction volume of 10 μL against cleavage by MboI (see Dam NEB # M0222L). Dam activity units can also be estimated by a direct comparison with NEB Dam enzyme in different conditions as described here. We typically purify pA-Dam protein with an estimated Dam activity around 100 NEB units/mg protein.
7. Unmethylated plasmid DNA can be isolated from Dam-negative bacteria (i.e., NEB catalog # C2925H) and verified by DpnI and DpnII/MboI digestions. Alternatively, other large fragments of unmethylated DNA can be used to assess Dam activity, including mammalian genomic DNA.
8. Adherent cells can be harvested by trypsinization or scraping, while suspension cells can simply be collected and washed with PBS.
9. Three controls are important for pA-DamID experiments. First, a negative control without antibody or Dam added, which should be completely devoid of any ^{m6}A signal. Second, a pA-DamID sample without primary antibody added. This sample can be used to determine potential background binding of pA-Dam. Third, a sample without antibody or pA-Dam, but with Dam enzyme added during the activation step. This latter control should be included for every cell culture condition, as it will be used to control for DNA accessibility and other possible biases.
10. We have tried pA-DamID with formaldehyde fixation (1% formaldehyde for 5–15 min at room temperature) prior to permeabilization with digitonin. However, this resulted in unspecific binding of pA-Dam. pA-DamID might be compatible with other fixation protocols. A consequence of using unfixed cells is that some epitopes are difficult to map with pA-DamID when these are unstable and lost during the permeabilization and washing steps.
11. pA-DamID relies on multiple rounds of centrifugation and supernatant removal to wash the cells. It is possible to work with fewer than one million cells, but this results in near-invisible pellets. We have successfully performed pA-DamID with 0.1 million starting cells. Reduce the cell concentration in **step 3** accordingly to keep working with 200 μL volumes.
12. After 3 minutes of centrifuging, turn the tubes 180° and centrifuge for a few additional seconds at 500 × *g*. This prevents accumulation at the side of the tube and reduces loss of cells during suction.

13. The antibody optimization strategy depends on the protein of interest. Ideally, the enrichment of ^{m6}A-Tracer intensity is used if the protein is localized in distinct nuclear patterns. Alternatively, the optimal antibody dilution can be estimated by selecting the concentration that gives the highest dynamic range after high-throughput sequencing.
14. Protein A has a high binding affinity for rabbit IgG antibodies, but low affinity for antibodies from mouse and some other species. Binding of a secondary rabbit antibody is required in when Protein A has low binding affinity for the primary antibody. We have successfully used 1:100 dilutions of rabbit anti-mouse (Abcam # ab6709) and rabbit anti-goat antibodies (Abcam # ab6697) as bridging antibodies.
15. This is a good moment to take a few cells and look at them with a phase contrast microscope. Suspension cells and trypsinized cells should result in round and intact cells at this stage, while scraped cells remain clustered in aggregates.
16. The pellets can be frozen at −20 °C for several weeks before isolation of genomic DNA.
17. We have performed propidium iodide staining followed by FACS sorting to purify G1, mid-S and G2/M subpopulations [8]. A modified single-cell protocol was used to prepare sequencing libraries from 3000 sorted cells per condition. We assume that other sorting strategies will be feasible as well, for example based on fluorescent proteins, the intensity of ^{m6}A, or labeling of other epitopes.
18. We use commercial kits (Bioline, Qiagen and Invitrogen) to isolate gDNA and determine the concentration with a NanoDrop spectrophotometer.
19. When available, we recommend to use 500 ng of input DNA. A high-quality data set can be obtained from as little as 10 ng, corresponding to gDNA from roughly 1000 diploid human cells. However, the complexity of the sequencing library is generally lower with limited input DNA and results in fewer unique reads. This can be partially rescued by running multiple PCR reactions to utilize all input DNA (i.e., 4× reactions instead of 1×, see methods Subheading 3.3, **step 6**) and combining these afterward. For samples that will be compared directly in downstream analyses, it is important to use the same amount of starting DNA and an identical number of PCR cycles to prevent biases.
20. Apoptotic cells and cells treated with DNA-damaging agents contain fragmented DNA that will be ligated to the adapter and thus amplified independently of DpnI. Such undesired ligation events can be prevented by dephosphorylation of the genomic DNA prior to DpnI digestion.

21. For antibodies against broad histone modifications and the nuclear lamina, we generally achieve strong signals after 15 PCR cycles with 500 ng input DNA. The number of PCR cycles can be increased in case of reduced input DNA or the use of antibodies against less abundant epitopes. Note that the control without primary antibody generally results in a strong smear after 21 PCR cycles with 500 ng of input DNA. This control is thus particularly important for antibodies that bind sparsely on the genome or not efficiently to the epitope. If kept on ice, additional cycles can be added to the PCR reaction after running 4 μL on gel.
22. Here, it is important to verify that the negative control samples (those lacking any Dam or pA-Dam, and those lacking DpnI or Ligase) are devoid of signal, and that the no-antibody control is significantly weaker than the sample with antibody.
23. The amplified PCR fragments were originally ^{m6}A methylated and can be processed for Illumina sequencing as described in the steps below. Alternatively, the fragments can be quantified with quantitative PCR [4] or processed for other purposes.
24. A-tailing on DNA is unstable. Freeze and thaw cycles on A-tailed DNA should be avoided.
25. Optimization is required for this PCR. We typically use 10 PCR cycles, which gives a clear signal on gel. Too many cycles may negatively affect sequencing efficiency.
26. We typically sequence pA-DamID libraries with 65 bp single-end reads, which after adapter removal results in 46 bp for genomic mapping. The sequencing depth ranges between 10 million reads for exploratory analyses to 40 million reads for high-quality profiles.
27. Poly-L-lysine–coated coverslips can be bought or prepared from poly-L-lysine solutions. The electrostatic interactions of poly-L-lysine are required for adhesion of the processed cells to the coverslips.
28. Gently shake the plate under a phase contrast microscope to verify that a large fraction of the cells has attached. Increase binding time if cells are insufficiently attached.
29. After fixation and two PBS washes, cover slips can be stored in PBS at 4 °C for several days, although some epitopes might be too unstable for long-term storage.
30. Even though digitonin is used for cell permeabilization in **step 2** of Subheading 3.2, incubation with NP40 is done here to ensure complete permeabilization.
31. Antibodies are not required to visualize ^{m6}A methylation. For visualization of the targeted protein, it is advised to add more of the antibody used in **step 3** of Subheading 3.2, as pA-Dam

masks the majority of the original antibody. Additional antibody probably results in binding to epitopes that have lost antibody binding during the pA-DamID protocol. Antibodies can be used to visualize other nuclear proteins, but take care not to mix antibody species with the antibody used in **step 3** of Subheading 3.2 to prevent undesired mixing of the protein visualization.

32. We use ^{m6}A-Tracer with a stock concentration of 1.15 mg/mL in a 1:500 dilution [8]. This dilution should be optimized for different batches of ^{m6}A-Tracer, as the fraction of functional protein can differ. The optimal concentration can be determined by staining pA-DamID cells processed with an epitope against a nuclear compartment (i.e., the nuclear lamina) with different concentrations of ^{m6}A-Tracer protein, and selecting the concentration that gives the highest enrichment at the compartment.

Acknowledgments

This work was supported by NIH Common Fund "4D Nucleome" Program grant U54DK107965 (BvS), AIRC-MSCA iCARE2.0 fellowship grant agreement 800924 (SGM) and MSCA Individual fellowship project number 838555 (SGM). The Oncode Institute is supported by KWF Dutch Cancer Society.

References

1. Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet 10(10):669–680. https://doi.org/10.1038/nrg2641
2. Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet 13(12):840–852. https://doi.org/10.1038/nrg3306
3. Klein DC, Hainer SJ (2020) Genomic methods in profiling DNA accessibility and factor localization. Chromosom Res 28(1):69–85. https://doi.org/10.1007/s10577-019-09619-9
4. van Steensel B, Henikoff S (2000) Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. Nat Biotechnol 18(4):424–428. https://doi.org/10.1038/74487
5. Aughey GN, Cheetham SW, Southall TD (2019) DamID as a versatile tool for understanding gene regulation. Development 146(6):dev173666. https://doi.org/10.1242/dev.173666
6. Schmid M, Durussel T, Laemmli UK (2004) ChIC and ChEC; genomic mapping of chromatin proteins. Mol Cell 16(1):147–157. https://doi.org/10.1016/j.molcel.2004.09.007
7. 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
8. van Schaik T, Vos M, Peric-Hupkes D, Hn Celie P, van Steensel B (2020) Cell cycle dynamics of lamina-associated DNA. EMBO Rep 21(11):e50636. https://doi.org/10.15252/embr.202050636
9. Kind J, Pagie L, Ortabozkoyun H, Boyle S, de Vries SS, Janssen H, Amendola M, Nolen LD, Bickmore WA, van Steensel B (2013) Single-cell dynamics of genome-nuclear lamina interactions. Cell 153(1):178–192. https://doi.org/10.1016/j.cell.2013.02.028
10. Vogel MJ, Peric-Hupkes D, van Steensel B (2007) Detection of in vivo protein-DNA interactions using DamID in mammalian cells.

Nat Protoc 2(6):1467–1478. https://doi.org/10.1038/nprot.2007.148

11. Greil F, Moorman C, van Steensel B (2006) DamID: mapping of in vivo protein-genome interactions using tethered DNA adenine methyltransferase. Methods Enzymol 410: 342–359. https://doi.org/10.1016/S0076-6879(06)10016-6

12. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun 10(1):1930. https://doi.org/10.1038/s41467-019-09982-5

Chapter 13

The dCypher Approach to Interrogate Chromatin Reader Activity Against Posttranslational Modification-Defined Histone Peptides and Nucleosomes

Matthew R. Marunde, Irina K. Popova, Ellen N. Weinzapfel, and Michael-C. Keogh

Abstract

Bulk chromatin encompasses complex sets of histone posttranslational modifications (PTMs) that recruit (or repel) the diverse reader domains of Chromatin-Associated Proteins (CAPs) to regulate genome processes (e.g., gene expression, DNA repair, mitotic transmission). The binding preference of reader domains for their PTMs mediates localization and functional output, and are often dysregulated in disease. As such, understanding chromatin interactions may lead to novel therapeutic strategies, However the immense chemical diversity of histone PTMs, combined with low-throughput, variable, and nonquantitative methods, has defied accurate CAP characterization. This chapter provides a detailed protocol for *dCypher*, a novel approach for the rapid, quantitative interrogation of CAPs (as mono- or multivalent Queries) against large panels (10s to 100s) of PTM-defined histone peptide and semisynthetic nucleosomes (the potential Targets). We describe key optimization steps and controls to generate robust binding data. Further, we compare the utility of histone peptide and nucleosome substrates in CAP studies, outlining important considerations in experimental design and data interpretation.

Key words Chromatin binding assay, Histone code, Histone posttranslational modifications, Histone PTMs, Histone PTM binding specificity, Histone peptides, Reader domain, Semisynthetic nucleosomes

1 Introduction

Chromatin is an essential regulatory component of multiple cellular processes, including transcriptional state [1–4] and disease development [5–7]. Its structures are highly dynamic, comprising a complex network of modifications to the DNA (e.g., cytosine methylation) and histone proteins (e.g., lysine methylation/acylation/ubiquitylation, arginine methylation/citrullination, serine phosphorylation; collectively termed posttranslational modifications [PTMs]). These covalent changes are mediated and

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_13,

interpreted by specific chromatin-associated "writers," "readers," and "erasers" to control local genome access and downstream function [8, 9]. This systems-level regulatory information is termed the "histone code," and its elucidation is key to understanding chromatin function [8, 10, 11].

A range of reader domain families and their preference for various PTM classes has been described, including (but not limited to): bromo- and YEATS domains for lysine acylations [12–17]; chromo, TUDOR and PWWP domains for lysine methylations [18–22]; and ubiquitin-dependent recruitment regions (UDRs) for lysine ubiquitylations [23, 24]. Deciphering the binding preference of particular chromatin readers for their histone PTMs (residue and class: e.g., H3 lysine 4 trimethyl [H3K4me3]) is a growing area of research, and will permit the targeting of specific pathways with therapeutic intent [7, 25–27]. However, these efforts have been challenged by the sheer diversity of PTMs [28], which may work alone, in combination, or in opposition, to engage multi-domain chromatin-associated proteins (CAPs) [29, 30]. Thus, methods to interrogate chromatin interactions must be highly efficient and easily modifiable to accommodate different reader domain classes and their modes of engagement, as well as to enable screening against diverse targets.

Historically, histone PTM-binding specificities have been studied using peptide microarrays (e.g., *EpiCypher* EpiTriton™), where libraries of modified histone peptides are spotted onto glass slides [31–33]. The format allows researchers to screen protein QUERIES against hundreds of single or combinatorial PTMs (the potential TARGETS). Histone peptide arrays have been used to characterize many classes of chromatin readers (e.g., chromo-, bromo-, Tudor domains; for a full review see [34]) and modifying enzymes (e.g., lysine methyltransferases G9a [35] and NSD1 [36]). However, the resulting data are largely qualitative, with low sensitivity, narrow signal-to-background windows, and suffer from the high levels of variation inherent to microarrays [37]. This format also requires a large amount of purified Query (μM range), making it difficult to titrate concentrations and explore buffer formulations/cofactor additions. Such optimizations are essential to reduce background, improve assay reliability, and thus begin to generate the quantitative analyses required for cross-Query comparisons.

There is an additional major concern: the historical focus on histone peptides disregards the significance of nucleosome structure in modulating chromatin binding events. In typical portrayals of nucleosome structure, the histone N-terminal tails extend from the nucleosome core and are thus easily accessible. Yet, multiple approaches show the tails often make extensive contacts with nucleosomal DNA [38]. Certain PTMs, such as acetylation and phosphorylation, may act to weaken these contacts, allowing

readers of other PTMs on the same tail to engage their target [38–41]. Further, many chromatin readers and enzymes make multivalent contacts with histone PTMs, nucleosomal DNA (e.g., via AT-hooks or PWWPs [42–45]), and/or the nucleosome core (the H2A/H2B acidic patch being a particular hub [46, 47]). Such multivalent interactions are often involved in histone PTM cross talk and can promote or inhibit chromatin binding. Thus, it is no great surprise that many chromatin-modifying enzymes require a nucleosome substrate for activity (e.g., NSD2 [42], LSD1 [48, 49], and DOT1L [50]), or show dramatically different kinetics to nucleosomes vs. peptides (e.g., SetD8 [51]). As a result of these complexities, most putative chromatin readers domains, and the means by which they act in concert in a given CAP, remain uncharacterized [52, 53], and nucleosome-based data will almost certainly be required for maximal insight.

To address these issues, we developed *dCypher*® as a novel and highly adaptable system for high-throughput CAP profiling. The approach uses chemiluminescent bead-based, no-wash Alpha technology (*see* **Note 1**), and delivers massive gains in sensitivity, flexibility, and throughput relative to histone peptide arrays. Of particular note, *dCypher* is fully compatible with PTM-defined histone peptides and semisynthetic nucleosomes (Fig. 1). In brief, biotinylated peptide or nucleosome substrates (the potential TARGETS) are coupled to streptavidin-coated "Donor" beads, while epitope-tagged proteins (QUERIES; from single domains to

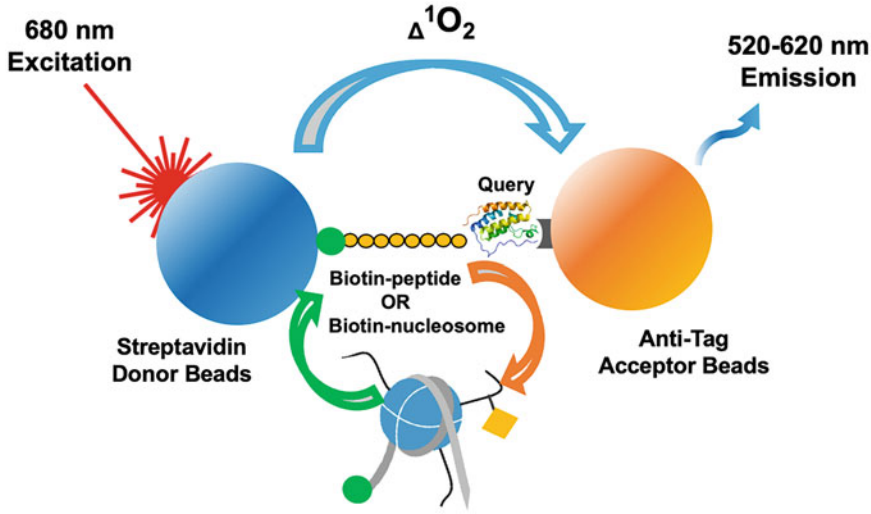


Fig. 1 Depiction of amplified luminescence proximity homogeneous assay (Alpha) technology (*see* **Note 1**). Alpha Donor and Acceptor beads are brought into proximity via [Target: Query] binding. Laser excitation (680 nm) of the Donor releases singlet oxygen that causes emission (520–570 nm) in proximal (within 200 nm) Donor beads; this luminescent signal is directly proportional to the amount of [Donor-Acceptor] complex bridged by the [Target: Query] interaction [23, 54–56]

complexes) are bound to anti-tag "Acceptor" beads (*see* **Note 2**). After mixing potential reactants, the Donor beads are excited at 680 nm, releasing a singlet oxygen that causes emission (520–620 nm) in proximal (within 200 nm) Acceptor beads; this luminescent signal is directly correlated to interaction/binding affinity [23, 54–56]. *dCypher* assays are performed in 384-well plates, enabling high-throughput analysis of potential [Query: Target] interactions.

We have now used *dCypher* to characterize multiple classes of potential chromatin binding domains in mono- or multivalent format against peptide and nucleosome substrates [23, 54–56]. Beyond exploring binding preference, *dCypher* was recently used to characterize a selective/potent inhibitor of nucleosomal H3K36me2/3 binding by the NSD2 N-terminal PWWP domain (representing a potential pathway to a high value therapeutic) [57]. Our extensive studies emphasize the need for rigorous assay optimization when exploring CAP capability, and the central importance of nucleosome context (see below). *dCypher* is uniquely suited for such work. Due to its high sensitivity, we are often able to use 100–1000-fold less Query protein compared to peptide arrays. Indeed, the ability to screen in multiwell plates allows the user to independently titrate the concentration of Query proteins, salt, and potential cofactors/competitors (e.g., free DNA), against potential Targets (nucleosome, peptide, or DNA) in parallel reactions. The resulting data show CAPs can be profoundly impacted by context. As an example, while *dCypher* confirms that the HP1β chromodomain binds all three H3K9 methyl states (me1, me2 and me3) on histone peptides (and makes no discrimination between me3 and me2) [58], it reveals an absolute preference for H3K9me3 nucleosomes (compare Fig. 2a, b). We propose the revised specificity on nucleosome substrates to be the more likely physiological state, and potentially driven by multivalent interactions (enhancing and inhibitory) between histone tails and other nucleosome surfaces. It also has profound implications: methyltransferases or demethylases that convert the H3K9me3 state are now of central importance to mechanistic studies of HP1β function.

Notably, some proteins require a more extensive *dCypher* workflow to reveal their true binding specificity on nucleosomes. In initial assays profiling the DNMT3A PWWP domain, we observed only weak binding to H3K36 methylated nucleosomes (their reported Target [22, 59]). We thus performed an extensive 2D [Query vs. Salt] titration, analyzing the impact of salt (NaCl) concentration on DNMT3A PWWP binding against H3K36me2 (Target) and unmodified (Control) nucleosomes. This showed the domain was highly salt-sensitive, exhibiting PTM selectivity within a narrow range (Fig. 3a). Running the assay at 100 mM NaCl provided the window to probe a large nucleosome panel and identify the exquisite selectivity of DNMT3A PWWP for H3K36me2/3

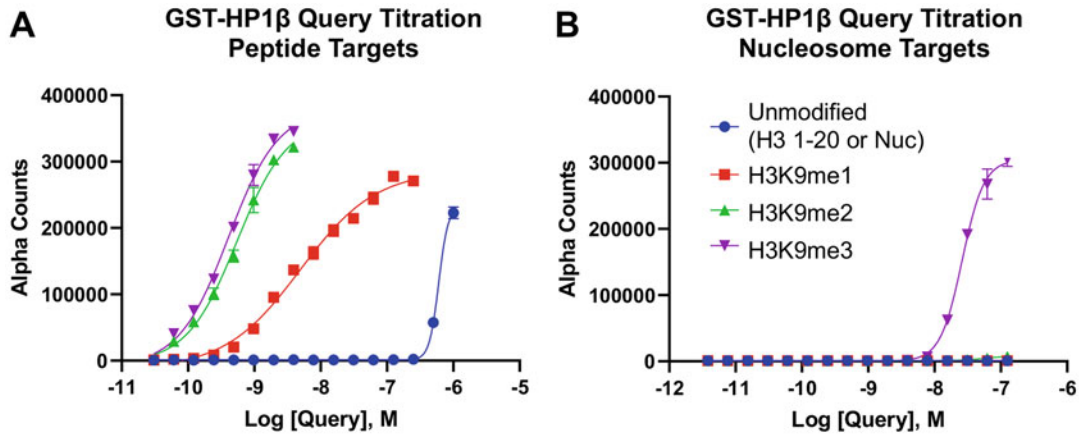


Fig. 2 (**a**) Titration of GST-HP1β chromodomain (UniProt P83916; aa1-185) against peptide Targets (key in **b**) reveals equivalent binding to H3K9me3/me2 and reduced binding to H3K9me1. (**b**) GST-HP1β chromodomain shows a dramatically refined specificity on nucleosomes, binding only H3K9me3 (*see* **Note 8**)

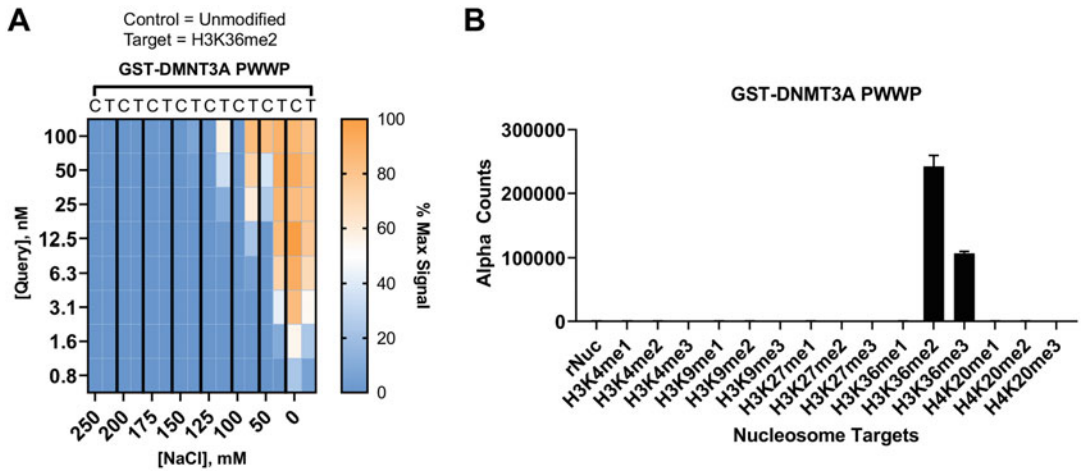


Fig. 3 (**a**) 2D [Query vs. Salt] titration of GST-DNMT3A PWWP (UniProt Q6Y6K1; aa278-432) on nucleosomes identifies NaCl sensitivity (rapid signal drop off >100 mM). Buffer supplemented with 100 mM NaCl was determined as the optimal signal window (note selectivity for H3K36me2 over unmodified nucleosomes). Data is normalized to the maximal Alpha signal on the plate. (**b**) Discovery Screen testing of DNMT3A PWWP in nucleosome assay buffer with 100 mM NaCl identifies a preference for H3K36me2 > me3 and no discernable interaction with all other lysine-methylated nucleosome in the panel (*X*-axis; rNuc, unmodified Nucleosome)

(Fig. 3b). This provided the mechanistic link from specific histone PTMs (H3K36me2/3) to DNMT3A recruitment, and thus how de novo DNA methylation is recruited to intergenic regions in vivo [54]. Furthermore, it explains the similar developmental pathologies associated with loss of function in H3K36 (*NSD1*: Sotos Syndrome) and DNA (*DNMT3A*: Tatton-Brown–Rahman syndrome) methyltransferases.

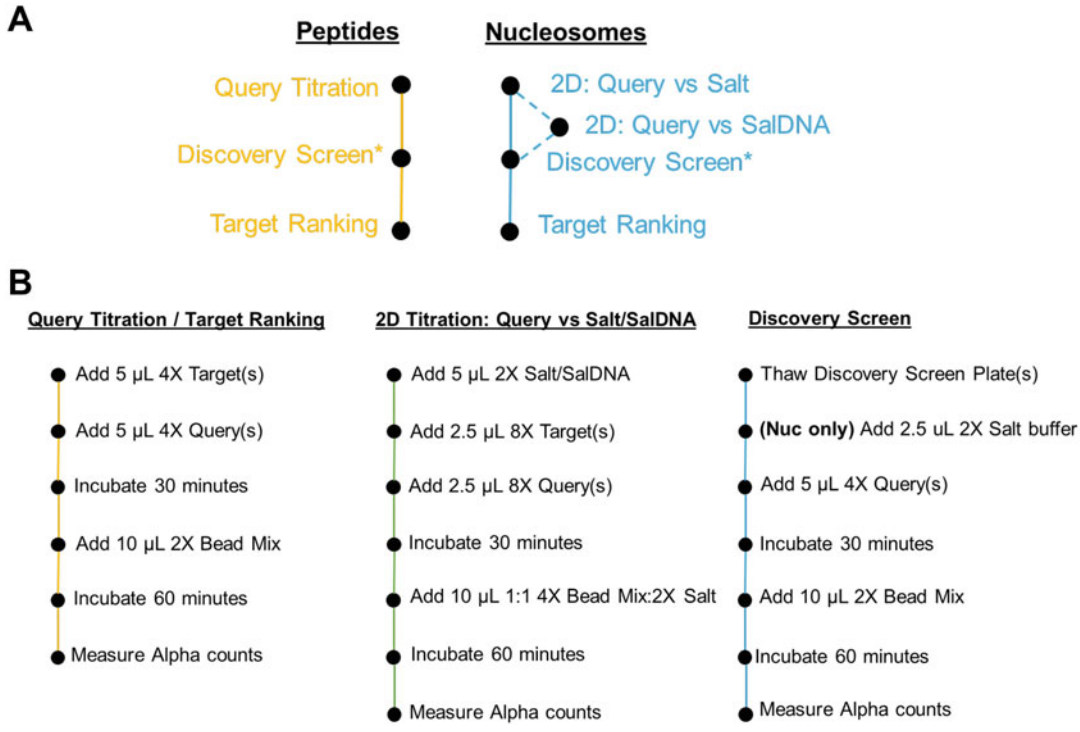


Fig. 4 Stepwise workflow for *dCypher* assay procedures (see Subheadings 3.1–3.5). (**a**) Comprehensive workflow overview to interrogate a chromatin reader. A typical study starts with titrating Query to a predicted histone peptide Target(s) to confirm activity and identify optimal probing concentration (i.e., good signal-over-background, on linear part of binding curve), and then progresses down the peptide branch or moves directly to nucleosomes. For Queries with no known Target a **Discovery Screen** (Subheadings 3.4 and 3.5) to the histone peptide panel at high and low concentrations (chosen from other reader domains of the same family; or *see* **Note 26**) can be performed, and the workflow then restarted with any hits to dial in optimal conditions. (**b**) Experimental guide for various study modules (i.e., order of addition, relative volumes, and incubation times)

This chapter contains a full outline of our *dCypher* pipeline (and its various modules) for testing Queries to PTM-defined peptide and nucleosome Targets (Fig. 4a, b). Of note, peptide-based *dCypher* assays generally do not require an exploration of salt concentration (and thus use a standard buffer). In contrast, salt titrations are always performed when developing nucleosome-based assays, as this often has a profound impact on Query binding (as above for DNMT3A PWWP). We also frequently use exogenous salmon sperm DNA (SalDNA) to interrogate DNA binding by Queries [57], particularly when moving to multidomain (and potentially multivalent) proteins with poorly characterized regions. These experimental modules highlight the complexity inherent to nucleosome studies.

dCypher has proven a powerful approach to interrogate CAP binding, and its application has revealed novel insight to the mechanisms underlying chromatin structure and function

[23, 54–56]. Many results (e.g., Fig. 2) raise important questions about current standards in chromatin methodology, particularly the continued reliance on a reductionist approach of isolated reader domains as Queries and PTM-defined histone peptides as Targets. With its ability to incorporate PTM-defined nucleosomes, *dCypher* will provide the means for physiologically relevant epigenetic discovery.

The *dCypher* workflow has been developed to guide the interrogation of CAP Queries against PTM-defined histone peptides and nucleosome Targets. The approach can be broken into various experimental modules (Fig. 4a), which are optimized to run sequentially in rapid throughput while controlling material consumption (Fig. 4b). Prior to performing the assay, it is necessary to select your protein Query and potential Targets. Epitope-tagged recombinant protein domains (or full-length proteins) can be expressed in-house or obtained from commercial sources. We have found the GST-tag (~220 aa/~26 kDa) to be reliable and produce robust results, but other epitope-fusions (6His and FLAG [DYKDDDDK]) are also compatible (*see* **Note 2**). For biotinylated Targets, PTM-defined histone peptides and semisynthetic nucleosomes are available from *EpiCypher* and were used to develop the *dCypher* platform.

The **first** step of the *dCypher* workflow involves optimization of Query binding to known or predicted PTM-defined targets: this includes exploring protein concentration, buffer conditions (e.g., salt), and potential supplements (e.g., exogenous SalDNA). **Second**, the optimized assay is used to evaluate Query binding against a large panel of potential Targets to determine preference, secondary interactions (e.g., with DNA or the acidic patch [by coupling biotinylated DNA or acid patch mutant nucleosomes to the Acceptor beads]), and the impact of neighboring PTMs. In the **third** and final step, identified Targets are ranked by their relative EC_{50} (EC_{50}^{rel}) values (*see* **Note 3**). This process is usually performed sequentially: first using histone peptides for initial testing and Target confirmation (since this can formally confirm activity of the Query protein based on prior literature); then using nucleosomes as physiological substrates. However, the peptide screen is not required, and may be omitted in favor of focusing on nucleosomes.

2 Materials

A complete list of consumables and equipment needed to perform *dCypher* assays. All peptides and nucleosomes are biotinylated. All solutions are prepared using ultrapure water (deionized to 18 MΩ-cm at 25 °C).

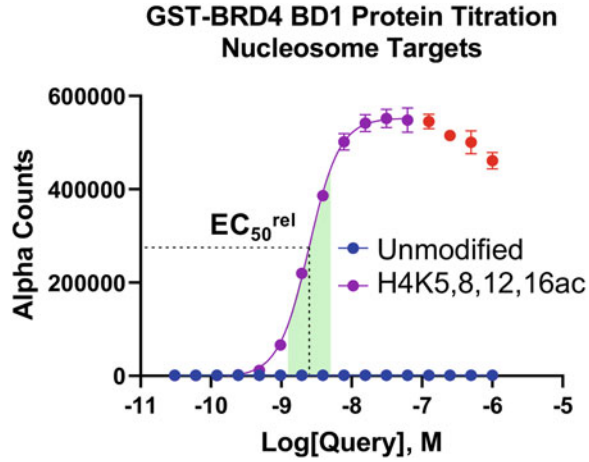


Fig. 6 Titration of GST-BRD4 BD1 (UniProt O60885; aa41-180) against unmodified and H4K5,8,12,16 ac Target nucleosomes. Dashed lines represent the relative EC_{50} (EC_{50}^{rel}, 2.5 nM) for BRD4 BD1 binding to H4K5,8,12,16 ac. Red circles represent assay points removed from analysis due to the hook point being reached (*see* **Note 15**): this indicates bead saturation/declining signal as excess nonbead bound Query is now competing with that on the Acceptor beads for Target binding. Green shaded area represents the optimal probing concentration range (shown is relative EC_{20}–EC_{80}) (*see* **Note 15**)

2.1 General Reagents

1. GST-, 6His-, or FLAG-tagged Query (*see* **Note 2**).
2. GST-HP1β (as in Fig. 2: UniProt P83916; aa1-185).
3. GST-DNMT3A (as in Fig. 3: UniProt Q9Y6K1; aa278-432).
4. GST-BRD4 BD1 (as in Fig. 6: UniProt O60885; aa41-180).
5. Biotinylated Peptides (*see* **Note 4**).
6. Biotinylated Nucleosomes (e.g., *EpiCypher* #16-9001).
7. Poly-L-lysine.
8. Salmon Sperm DNA (SalDNA).
9. Streptavidin Donor Beads (*see* **Note 5**).
10. Glutathione Acceptor Beads.
11. Nickel-Chelate Acceptor Beads.
12. Protein-A Acceptor Beads.
13. Anti-FLAG Antibody.

2.2 Buffers

Assay buffers (compositions as noted) are prepared fresh for each experiment and kept at room temperature (unless otherwise specified).

1. Peptide reconstitution solution: 0.01% BSA in ddH_2O.
2. Peptide assay buffer: 50 mM Tris pH 7.5, 50 mM NaCl, 0.01% Tween 20, 0.01% BSA, 1 mM TCEP, 0.0004% poly-L-lysine (*see* **Note 6**).
3. Nucleosome assay buffer: 20 mM Tris pH 7.5, 0–250 mM NaCl, 0.01% NP-40 alternative, 0.01% BSA, 1 mM DTT.

2.3 Equipment

1. AlphaPlate-384 (Assay Plate; *PerkinElmer* 6005350) or similar product.
2. 384 Deep Well Plate (Dilution plate; *Greiner Bio-One* 781270) or similar product.
3. 1.5 mL Microtubes (Lo-bind).
4. 50 mL conical tubes.
5. Divided Reservoirs.
6. TopSeal A-PLUS (*PerkinElmer* 6050185) or similar product.
7. TopSeal A Black (*PerkinElmer* 6050173) or similar product.
8. MicroAmp Adhesive Film (Storage seals: *Applied Biosystems* 4311971) or similar product.
9. Set of single channel pipettes (0.1–1000 μL).
10. 16-channel pipette (1–10 μL).
11. 16-channel electronic pipette (5–50 μL).
12. Microplate centrifuge.
13. Personal Incubator.
14. EnVision Plate Reader (*PerkinElmer* 2105-0010) (*see* **Note 1**).
15. AlphaScreen Mirror D640as (*PerkinElmer* barcode #444) (*see* **Note 1**).
16. AlphaScreen/AlphaLISA Emission Filter (*PerkinElmer* barcode #244) (*see* **Note 1**).
17. AlphaLISA Emission Filter M615 (*PerkinElmer* barcode #203) (*see* **Note 1**).

3 Methods

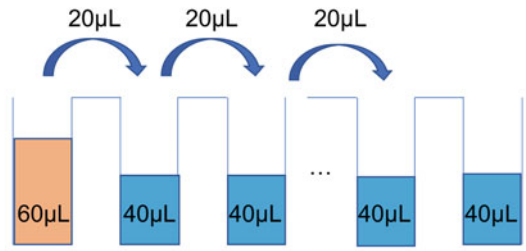
These protocols provide conditions, volumes, and concentrations compatible with GST-tagged protein Queries, although 6HIS or FLAG-tagged proteins may be substituted (*see* **Note 2**). All volume calculations are precise (so desired dead volumes must be added). All procedures are performed at room temperature and incubation steps at 23 °C unless otherwise noted. After each reagent addition, Assay plates should be firmly tapped on the lab bench, a plate seal applied (*see* **Note 7**), and centrifuged briefly (600 × *g* for 10–15 s). It is *highly recommended* that each experiment includes positive control reactions with a similarly tagged Query (*see* **Note 8**).

3.1 Query Titration, Target Ranking

The following protocol describes titration and target ranking experiments using a *GST-tagged* Query and a Target and Control substrate (*see* **Note 9**). For peptide-based assays with a known or suspected Target, we recommend starting with Query titrations (Subheading 3.1: e.g., Fig. 2a). However, when using nucleosomes, a 2D [Query vs. Salt] titration is the suggested first step, as precise salt conditions are often critical for appreciable binding to nucleosome substrates (Subheading 3.2: e.g., Fig. 3a). Query titrations assess binding to an anticipated Target, related PTMs, and a negative control. Results from these assays provide essential information for additional testing: (1) confirm Query functionality and target preference; (2) determine optimal probing concentration (and if salt, competitor DNA, or other additives are needed; *see* Subheadings 3.2 and 3.3); (3) identify any hook point (i.e., maximal Query concentration before bead saturation); and (4) calculate EC_{50} relative (EC_{50}^{rel}) values. Query binding to PTM-defined peptide/nucleosome targets are ranked using their respective EC_{50}^{rel} values, enabling quantitative comparisons across Queries and Targets. This information informs the design of subsequent discovery screens with a larger set of potential Targets (Subheadings 3.4 and 3.5).

1. Plan the desired plate layout and calculate the needed quantities of buffer, Query (or Queries), and Target (or Targets) (*see* **Notes 8–10**).
2. Thaw GST-tagged Queries and peptides/nucleosomes on ice (*see* **Note 11**).
3. Prepare 10 mL of peptide or nucleosome assay buffer immediately before use in a 50 mL conical tube (*see* **Note 6**).
4. Peptide Targets: Prepare 4× (400 nM) peptide dilutions by adding 2.4 μL of each peptide (20 μM) to 117.6 μL peptide assay buffer in 1.5 mL tubes.

   Nucleosome Targets (*see* **Note 12**): Prepare 4× (40 nM) by adding 6.4 μL of each nucleosome (1.5 μM) to 113.6 μL nucleosome assay buffer in 1.5 mL tubes. Prepare these dilutions immediately before use and keep at room temperature.
5. Prepare the highest concentration of Query dilution (4×) by adding 12 μL of Query (20 μM) to 228 μL assay buffer in a 1.5 mL tube.
6. In a Dilution plate (Subheading 2.3), serially dilute the 4× Query (threefold; 20 μL Query into 40 μL assay buffer) to prepare 11 testing concentrations plus a buffer-only control (12th point) (*see* **Note 13**/Fig. 5).
7. Add 5 μL of 4× diluted Target(s)/Control to their assigned well positions in the Assay plate.
8. Add 5 μL of serially diluted Query to the Assay plate.



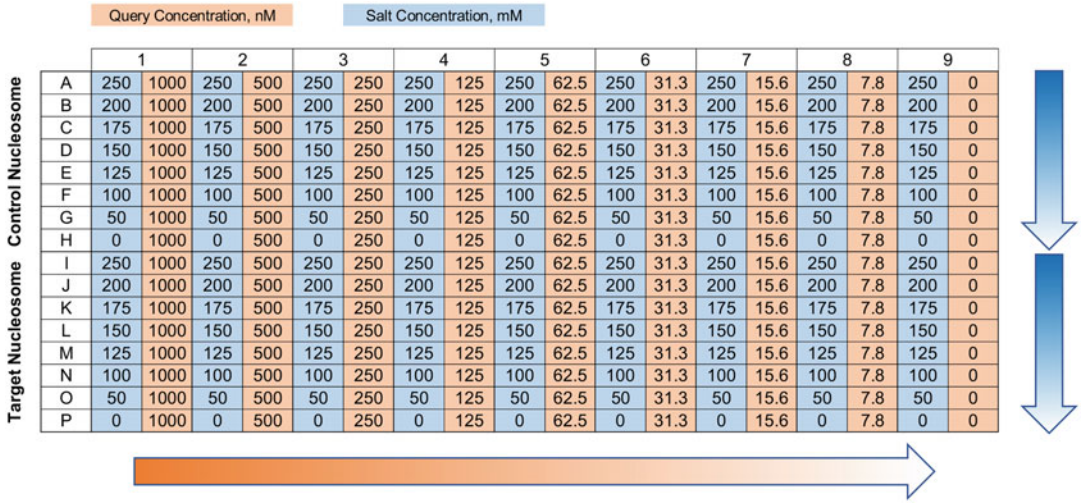
Serial #	1	2	3	11	12
[Query] in 5μL	4 μM	1.3 μM	444 nM	67.7 pM	0
[Query] in 20μL	1 μM	333 nM	111 nM	16.9 pM	0

Fig. 5 Threefold serial dilution of Query in a 384-well deep Dilution plate. Query concentration in 5 μL is when added to the plate; concentration in 20 μL represents that after all reagents have been added

9. Place Assay plate in 23 °C incubator for 30 min.
10. During the incubation, move to a subdued lighting area and prepare a 2× mixture by adding 0.48 μL of 5 mg/mL AlphaLISA Glutathione Acceptor beads (5 μg/mL) and 0.96 μL of 5 mg/mL Alpha Streptavidin Donor beads (10 μg/mL) to 478.6 μL appropriate assay buffer in a 1.5 mL tube (*see* **Notes 2**, **5**, and **14**).
11. Prepare a 4× bead mixture by adding 1.44 μL of 5 mg/mL AlphaLISA Glutathione Acceptor (10 μg/mL) and 2.88 μL of 5 mg/mL Alpha Streptavidin Donor beads (20 μg/mL) to 715.7 μL of appropriate assay buffer in a 1.5 mL tube.
12. Add 10 μL of the bead mixture to each well of Assay plate under subdued lighting.
13. Place Assay plate in 23 °C incubator for 60 min. Biotinylated peptide/nucleosome Targets will bind Donor beads and the GST-tagged Query will couple with Acceptor beads.
14. Remove the plate seal and measure Alpha counts using an *EnVision* plate reader (or similar) using Alpha settings (*see* **Note 1**).
15. Analyze data using *GraphPad Prism* (or similar) to: (1) identify hook point (if present); (2) determine optimal protein probing concentration; and (3) compute EC_{50}^{rel} for each Target (*see* **Note 15**/Fig. 6).

3.2 2D [Query vs. Salt] Titrations

This experimental module is designed to optimize both Query and Salt concentrations for *dCypher* assays. Given the salt sensitivity inherent to Query-nucleosome interactions (e.g., Fig. 3a) this module is a strongly recommended first step for testing Queries



	1		2		3		4		5		6		7		8		9	
A	250	1000	250	500	250	250	250	125	250	62.5	250	31.3	250	15.6	250	7.8	250	0
B	200	1000	200	500	200	250	200	125	200	62.5	200	31.3	200	15.6	200	7.8	200	0
C	175	1000	175	500	175	250	175	125	175	62.5	175	31.3	175	15.6	175	7.8	175	0
D	150	1000	150	500	150	250	150	125	150	62.5	150	31.3	150	15.6	150	7.8	150	0
E	125	1000	125	500	125	250	125	125	125	62.5	125	31.3	125	15.6	125	7.8	125	0
F	100	1000	100	500	100	250	100	125	100	62.5	100	31.3	100	15.6	100	7.8	100	0
G	50	1000	50	500	50	250	50	125	50	62.5	50	31.3	50	15.6	50	7.8	50	0
H	0	1000	0	500	0	250	0	125	0	62.5	0	31.3	0	15.6	0	7.8	0	0
I	250	1000	250	500	250	250	250	125	250	62.5	250	31.3	250	15.6	250	7.8	250	0
J	200	1000	200	500	200	250	200	125	200	62.5	200	31.3	200	15.6	200	7.8	200	0
K	175	1000	175	500	175	250	175	125	175	62.5	175	31.3	175	15.6	175	7.8	175	0
L	150	1000	150	500	150	250	150	125	150	62.5	150	31.3	150	15.6	150	7.8	150	0
M	125	1000	125	500	125	250	125	125	125	62.5	125	31.3	125	15.6	125	7.8	125	0
N	100	1000	100	500	100	250	100	125	100	62.5	100	31.3	100	15.6	100	7.8	100	0
O	50	1000	50	500	50	250	50	125	50	62.5	50	31.3	50	15.6	50	7.8	50	0
P	0	1000	0	500	0	250	0	125	0	62.5	0	31.3	0	15.6	0	7.8	0	0

Fig. 7 Example of a 2D [Query: Salt] titration 384-well plate map (columns 10–24 not shown) for a known Target and Control nucleosome. The Control is expected to represent a nonbinder for Query and is usually an unmodified nucleosome, though this depends on the mode of engagement (e.g., we occasionally use a nucleosome deleted of all histone tails [tailless] or with acid-patch mutations: **Note 9**). Salt concentration is titrated from top to bottom while protein is titrated from left to right (*see* **Note 17**)

to potential nucleosome Target(s). Results from this assay provide essential information to: (1) confirm protein functionality and target preference; (2) identify any hook point; (3) determine optimal probing concentration; (4) determine optimal salt condition; and (5) determine EC_{50}^{rel} values. In our experience this module is not generally required for peptide-based *dCypher* studies (which use a standard buffer).

1. Plan the desired plate layout and calculate the needed quantities of buffer, Query (or Queries), and Target (or Targets) (see **Notes 16** and **17**/Fig. 7).
2. Thaw biotinylated nucleosomes and GST-tagged Queries on ice.
3. Prepare 1 mL of 8 different 2× salt nucleosome assay buffers *with* 500, 400, 350, 300, 250, 200, 100, and 0 mM NaCl in 1.5 mL tubes. Transfer to a Dilution plate for easier pipetting (*see* **Note 17**).
4. Prepare 1× nucleosome assay buffer *without* salt.
5. Prepare 8× (80 nM) nucleosome dilutions by adding 9.6 μL of 1.5 μM nucleosome to 170.4 μL 1× nucleosome assay buffer *without* salt in 1.5 mL tubes.
6. Prepare the highest 8× (8 μM) Query dilution and controls by adding 36 μL of 20 μM Query to 54 μL 1× nucleosome assay buffer *without* salt in a 1.5 mL tube (*see* **Notes 17** and **18**).

7. Transfer Query solution to a Dilution plate, serially dilute 8 times (twofold; 45 μL Query into 45 μL buffer) in 1× nucleosome assay buffer *without* salt. Add a ninth point as a buffer-only control (*see* **Notes 13** and **19**).
8. In an Assay plate, add 5 μL of each 2× salt condition (from Dilution plate) where the concentration is fixed left to right and decreasing top to bottom (*see* **Note 17**/Fig. 7).
9. Add 2.5 μL of 8× nucleosomes to assigned wells (72 wells total/nucleosome).
10. Add 2.5 μL of serially diluted Query (from Dilution plate) where the concentration is decreasing left to right and fixed top to bottom (*see* **Note 17**/Fig. 7). The final Query concentrations will be 1 μM (1×) to 7.8 nM.
11. Place Assay plate in 23 °C incubator for 30 min.
12. During the 30-min incubation, move to a subdued lighting area and prepare a 4× bead mixture by adding 1.44 μL of 5 mg/mL AlphaLISA Glutathione Donor beads (10 μg/mL) and 2.88 μL of 5 mg/mL Alpha Streptavidin Donor beads (20 μg/mL) to 715.7 μL of nucleosome assay buffer *without* salt in a 1.5 mL tube (*see* **Note 14**).
13. Prepare a 1:1 mixture of 4× bead mix plus each 2× salt condition by mixing 45 μL of each 2× salt condition with 45 μL of the 4× bead mix for a total of 8 bead–salt mixes. Move mixes to a Dilution plate for easier transfers.
14. Add 10 μL of the bead–salt mixes in the same method as **step 8**.
15. Place Assay plate in 23 °C incubator for 60 min. Biotinylated nucleosome will bind Donor beads and the GST-tagged Query will couple with Acceptor beads.
16. Remove the plate seal and measure Alpha counts using an *EnVision* plate reader (or similar) using Alpha settings (*see* **Note 1**).
17. Analyze data using *GraphPad Prism* (or similar) to identify: (1) hook point (if present); (2) optimal probing concentration; (3) optimal salt condition; and (4) compute EC_{50}^{rel} for each Target (*see* **Notes 16** and **20**).

3.3 2D [Query vs. Salmon Sperm DNA (SalDNA)] Titration

The 2D [Protein vs. SalDNA] protocol is used in scenarios where Queries are thought to be contacting nucleosomal and/or linker DNA and masking histone PTM interactions [57]. DNA engagement typically manifests as equal binding to all nucleosomes regardless of PTM (assuming the same DNA length is used). If DNA binding is suspected, titrate Query against biotinylated nucleosome (s) and DNA Targets. If DNA binding is observed, the following procedure can be used to challenge the Query-DNA interaction and reveal underlying histone PTM interactions (if they exist).

Results from this assay provide essential information to: (1) confirm protein functionality and PTM target preference; (2) identify hook point (if present); (3) determine optimal probing concentration; (4) determine optimal SalDNA condition; and (5) determine EC_{50}^{rel} values.

1. Plan the desired plate layout and calculate the needed quantities of buffer, Query (or Queries), and Target (or Targets) (similar to **Note 18**).
2. Thaw biotinylated nucleosome(s) and Query on ice.
3. Prepare 1× nucleosome assay buffer using the optimal salt condition (Subheading 3.2: If unclear from initial studies supplementing with 150 mM NaCl is a good starting point).
4. Prepare 2× SalDNA dilution by adding 0.3 μL of 10 mg/mL SalDNA stock to 149.7 μL nucleosome assay buffer in a 1.5 mL tube.
5. Serially dilute 2× SalDNA by adding 50 μL SalDNA into 100 μL buffer (threefold). Prepare a total of 7 serial dilutions, including the 2× stock (i.e., 20, 6.67, 2.22, 0.74, 0.25, 0.08, 0.027, and 0 μg/mL) plus a buffer-only control (similar to **Note 13**).
6. Prepare 8× (80 nM) nucleosome dilutions by adding 9.6 μL of 1.5 μM nucleosome to 170.4 μL nucleosome assay buffer in 1.5 mL tubes.
7. Prepare the highest 8× (8 μM) Query dilution and controls, by adding 36 μL of 20 μM Query to 54 μL nucleosome assay buffer in a 1.5 mL tube (*see* **Note 19**).
8. Transfer Query solution to a dilution plate, serially dilute 8 times (twofold; 45 μL Query into 45 μL buffer) in nucleosome assay buffer and add a ninth point as a buffer-only control (similar to **Note 13**, *see* **Note 19**).
9. Add 5 μL of 2× SalDNA serial dilution where the concentration is fixed left to right and decreasing top to bottom over 9 columns (similar to **Note 17**).
10. Add 2.5 μL of 8× nucleosomes to assigned wells.
11. Add 2.5 μL of 8× serially diluted Query where the concentration is decreasing left to right and fixed top to bottom (similar to **Note 17**).
12. Place Assay plate in 23 °C incubator for 30 min.
13. During the 30-min incubation, move to a subdued lighting area and prepare a 4× bead mixture by adding 1.44 μL of 5 mg/mL AlphaLISA Glutathione Donor beads (10 μg/mL) and 2.88 μL of 5 mg/mL Alpha Streptavidin Donor beads (20 μg/mL) to 715.7 μL of nucleosome assay buffer in a 1.5 mL tube (*see* **Note 14**).

14. Prepare 1:1 mixture by combining 45 μL of each 2× SalDNA concentration with 45 μL of the 4× bead mix for a total of 8 bead–SalDNA mixes. Transfer mixes to a Dilution plate for easier transfers.
15. Add 10 μL of the bead–SalDNA mixes in the same method as **step 9**.
16. Place Assay plate in 23 °C incubator for 60 min. Biotinylated nucleosomes will bind Donor beads and the GST-tagged Query will couple with Acceptor beads.
17. Remove the plate seal and measure Alpha counts using an *EnVision* plate reader (or similar) using Alpha settings (*see* **Note 1**).
18. Analyze data using *GraphPad Prism* (or similar) to identify: (1) hook point (if present); (2) optimal probing concentration; (3) optimal SalDNA condition; and (4) EC_{50}^{rel} for each Target (*see* **Notes 15** and **21**).

3.4 Preparation of Discovery Screen Plate(s)

This protocol details the production of Discovery Screen plates (typically prepared in batch for greatest efficiency). The resulting peptide/nucleosome plates are intended to be stored at −80 °C, thawed once, and used to screen Queries against a broad set of targets (*see* **Notes 22–25**). Peptides and nucleosomes are *not* recommended for simultaneous testing given their different buffer requirements.

1. Generate a Discovery Screen plate map of the intended peptide/nucleosome targets and determine the number of plates to be prepared. Using each target concentration, calculate the volumes required (*see* **Note 22**).
2. Thaw all biotinylated peptides and/or nucleosome stocks on ice.

   Peptides: Prepare a modified peptide assay buffer *without* poly-L-lysine.

   Nucleosomes: Prepare a modified nucleosome assay buffer *without* NaCl. Keep buffers on ice.
3. Prepare 400 nM peptide or 80 nM nucleosome dilutions.

   Peptides: Add 2 μL of 20 μM peptide to 98 μL of modified peptide assay buffer.

   Nucleosomes: Add 2.7 μL of 1.5 μM nucleosome to 47.3 μL modified nucleosome assay buffer. Keep each diluted target on ice (*see* **Note 23**).
4. Transfer Targets to Assay plate(s).

   Peptides: Transfer 5 μL of each in duplicate to 10 Assay plate(s).

   Nucleosomes: Transfer 2.5 μL of each in duplicate to 10 Assay plate(s).

5. Once all material is added, tap microplates firmly on bench, carefully apply a storage seal, and centrifuge each Assay plate to settle any droplets (*see* **Note 24**).
6. Prepared plates can be stored for up to 3 months at −80 °C (*see* **Note 25**).

3.5 Discovery Screen

At this stage, optimal buffer conditions and probing concentrations have been identified for the Queries of interest (Subheadings 3.1–3.3). Profiling Queries using the Discovery Screen plates (Subheading 3.4) will provide a breadth of binding data to many PTM targets (100 in this example) at a single Query concentration. It is recommended to quantitatively rank targets (using EC_{50}^{rel}) from discovery screens by titration testing (Subheading 3.1). In cases with no binding target (i.e., Subheading 3.5 is being entered blind), we suggest testing with both high and low Query concentrations (*see* **Note 26**) and then restarting the workflow with identified Targets (to optimize the system).

1. Plan the desired plate layout and calculate the needed quantities of buffer and Query (or Queries).
2. Thaw Discovery Screen plate(s) on ice and centrifuge (600 × *g* for 1 min) to settle any droplets. Adjust to room temperature for about 10 min.
3. Thaw Queries on ice.
4. Prepare peptide assay buffer or nucleosome assay buffer with optimal salt and DNA (if latter is required).
5. Nucleosomes only: Prepare 500 μL of 2× salt nucleosome assay buffer.
6. Nucleosomes only: Add 2.5 μL of 2× salt nucleosome assay buffer to each well with substrates or buffer control.
7. Prepare 1 mL of Query at 4× the optimal probing concentration in assay buffer (*see* **Note 15**).
8. Add 5 μL of 4× Query dilution to all wells containing substrates or buffer.
9. Place plate in 23 °C incubator for 30 min.
10. During the incubation, move to a subdued lighting area and prepare a 2× bead mixture by adding 4 μL of 5 mg/mL AlphaLISA Glutathione Donor beads (5 μg/mL) and 8 μL of Alpha Streptavidin Donor beads (10 μg/mL) to 1988 μL assay buffer in a tube (*see* **Note 14**).
11. Add 10 μL of the bead mixture to each well under subdued lighting.
12. Place Assay plate in 23 °C incubator for 60 min. Biotinylated nucleosomes will bind Donor beads and the GST-tagged Query will couple with Acceptor beads.

13. Remove the plate seal and measure Alpha counts using an *EnVision* plate reader (or similar) using Alpha settings (*see* **Note 1**).
14. Analyze data using *GraphPad Prism* (or similar) to identify Potential Targets (by signal-over-background: e.g., Fig. 3b).
15. Identified Targets are titration tested (Subheading 3.1) under optimized conditions (Subheadings 3.2 and 3.3) to rank order (*see* **Note 27**).

4 Notes

1. Amplified luminescent proximity homogenous assay (Alpha, *PerkinElmer*) technology is a bead-based, no-wash chemiluminescent approach. The no-wash and signal amplification elements provide dramatically enhanced sensitivity relative to fluorescence-based histone peptide arrays. Prior to performing Alpha-based experiments, it is critical to ensure instrumentation/optics are compatible with the intended Acceptor beads (AlphaScreen or AlphaLISA). We use an EnVision 2104 instrument (Subheading 3.3) equipped with the Alpha 680 nm laser, AlphaScreen mirror (*PerkinElmer* barcode #444), and the AlphaScreen/AlphaLISA Emission filters (barcodes #244 and #203 respectively). The specific emission filter requirements are due to different luminescent chemistries on each Acceptor bead: AlphaScreen uses rubrene (broad ~520–620 nm emission), while AlphaLISA utilizes europium (narrow 615 nm emission). Because of this, AlphaScreen requires the #244 emission filter (570 nm/100 nm bandwidth) for accurate measurement, while AlphaLISA can be measured with the #244 or #203 (615 nm/8.5 nm bandwidth) emission filters.

   Although we perform *dCypher* assays using each Acceptor bead type, AlphaLISA tend to emit brighter, and can provide several-fold improvement in assay sensitivity. However, this comes at a price, with AlphaLISA Acceptor beads costing significantly more than AlphaScreen. Alpha streptavidin Donor beads are compatible with each Acceptor assay format (AlphaScreen and AlphaLISA). Of note, Alpha beads are much smaller in diameter (~250 nm) relative to typical bead-based assays (usually >5 μm). Due to this small size, Alpha beads will remain in suspension for the duration of the assay and do not require resuspension prior to signal measurement.
2. *dCypher* is optimized for use with GST-, 6His-, and FLAG-tagged Queries. Other epitope tags (or primary antibodies) can be used but must first be optimized. It is important to consider

that tags have the potential to modify Query behavior (e.g., GST can induce dimerization). Each tag requires a unique combination of detection reagents.

(a) GST-tags: Use 2.5 μg/mL glutathione Acceptor beads and 5 μg/mL Alpha Streptavidin Donor beads. Glutathione Acceptor beads are only available in AlphaLISA format.

(b) 6His-tags: Use 5 μg/mL Nickel-chelate Acceptor beads and 10 μg/mL Alpha Streptavidin Donor beads. Nickel-chelate Acceptor beads are available as either AlphaScreen or AlphaLISA formats.

(c) FLAG-tags: Use 1:400 anti-FLAG antibody, 5 μg/mL Protein A Acceptor beads and 10 μg/mL Alpha Streptavidin Donor beads. Protein A Acceptor beads are available as either AlphaScreen or AlphaLISA formats.

3. To compare and rank targets we use a four-parameter logistical (4PL) model and compute the relative EC_{50} (EC_{50}^{rel}) values for each target. Although K_d values are typically used for reporting binding affinity, specific conditions must be met to determine a K_d when using Alpha technology: a Query concentration at least 5× below bead binding saturation and 10× excess of fixed target. A competition assay can be performed to determine binding K_d of each interaction but will require case-by-case optimization to ensure sufficient signal-to-background of the Query and Target. It is important not to over-interpret EC_{50}^{rel} values, as they are defined as the concentration of Query required to elicit a response halfway between the maximal and baseline along the concentration–dose response curve. Further, we report EC_{50} values as relative EC_{50} because a stable maximal response (100% ± 5%) control is not included during data generation: as such we cannot ensure saturation.

4. Lyophilized peptides are dissolved in peptide reconstitution buffer (Subheading 2.2). Typically, all peptides are resuspended to the same concentration (we generally use 20 μM) to aid experimental planning.

5. Streptavidin Donor beads are light-sensitive and should only be handled under subdued lighting. After beads have been added to Assay plates, these should be covered with a black or other nontransparent seal to protect from light exposure.

6. Peptide assay buffer may turn slightly cloudy at room temperature but this will not impact assay performance. It is *not recommended* to store peptides long term in buffer containing poly-L-lysine.

7. A new plate seal is applied after each addition to prevent accidental cross-contamination between assay wells. Clear seals are

typically used for incubations prior to bead additions to the plate. After beads have been added, only use black or nontransparent seals (*see* **Note 5**).

8. A positive control Query should be included in each experiment to verify the assay system. The ideal control will use the same tag as the Query under interrogation. GST-, 6His-, and FLAG- tagged HP1β are commercially available and work best when paired with H3 (aa1-20; *EpiCypher* #12-0001) and H3K9me3 (aa1-20; *EpiCypher* #12-0012) peptides or unmodified (Control; *EpiCypher* #16-0006) and H3K9me3 (Target; *EpiCypher* #16-0315) nucleosomes.
9. Queries are usually assayed in duplicate or triplicate against a PTM-defined suspected Target and Control. The latter represents a predicted nonbinder for the Query, and in the case of nucleosome-based assays is usually an unmodified nucleosome, though this depends on the mode of engagement (e.g., could also be deleted of all histone tails [tailless; e.g., *EpiCypher* #16-0027], with acid-patch mutations [e.g., *EpiCypher* #16-0029, #16-0030 and #16-0031], or free DNA [e.g., *EpiCypher* #18-0005]).
10. Query titrations are typically performed at different 12 concentrations in duplicate and in two- or threefold dilution increments to cover a wide range and ensure upper and lower plateaus are captured. The 12th point is always a buffer control to assess assay background signal.
11. Peptides can be thawed at room temperature and then placed on ice. However, nucleosomes should always be thawed on ice, which will occur quickly due to the glycerol in their storage buffers.
12. Never vortex or sonicate nucleosomes to mix. Instead, gently pipet up and down until homogenous and flash centrifuge to settle any droplets on cap or sides of tube.
13. 4× Query serial dilutions are usually prepared in 384 deep well plates (Dilution plates). Sixteen-channel pipettes (if available) greatly increase efficiency and allow all dilutions to be handled simultaneously [Fig. 5].
14. Prior to adding Alpha Donor and Acceptor beads, vortex on high for ~10 s to ensure they are completely mixed and flash centrifuge to settle any droplets on cap or sides of tube.
15. Query Titration data is usually analyzed by generating nonlinear regression XY plots in *GraphPad Prism*. When determining EC_{50}^{rel}, the max and min plateaus must be visible for accurate quantification. Often a hook point is reached when using Alpha technology (Fig. 6): this indicates bead saturation/declining signal as excess nonbead bound Query is now competing with

that on the Acceptor beads for Target binding. Data points beyond the hook point must be removed for proper analysis. Sometimes in order to achieve a proper curve fit after excluding the hooked data points, a maximum signal constraint will need to be used in *GraphPad Prism*, particularly with sharp hook points (signal rapidly decreases as Query concentration increases). Each Query will have a unique hook point but molecular weight (MW) is a general predictor, where higher MW proteins tend to hook at lower molar concentrations. Optimal probing concentrations balance signal-to-background and being within the linear range of the sigmoidal curve: the EC_{50}^{rel} or EC_{80}^{rel} are typically selected. In the example [GST-BRD4 BD1: Fig. 6], the EC_{50}^{rel} can be computed for H4K5,8,12,16 ac as 2.5 nM (dotted line) but is nondeterminable for unmodified nucleosome (Control) as no binding was detected. The optimal probing concentration can be selected from a range, shown in green, which represents the EC_{20}^{rel} – EC_{80}^{rel}.

16. When performing 2D [Query vs. Salt] titrations, the typical 1× (final) concentrations are 250, 200, 175, 150, 125, 100, 50, and 0 mM NaCl (*see* Figs. 3 and 7). Antibodies to FLAG-tagged queries often show high nonspecific nucleosome interaction at low salt, so test 250, 225, 200, 175, 150, 125, 100, and 50 mM NaCl.
17. When planning/preparing each of the eight individual 2× salt buffers and Query serial dilutions, it is recommended to prepare the material in a 384 deep-well Dilution plate for easier transfer by 16-channel pipette to the Assay plate. For adding 2× salt buffer to the Assay plate, pipet the column of salt dilutions from left to right. For transferring 8× Query serial dilutions, pipet the column of serially diluted Query from top to bottom (Fig. 7).
18. In general, the highest Query concentration tested is 1 μM final in 20 μL (8× = 8 μM), though this may not be possible depending on the Query hook point (*see* **Note 15**).
19. Queries are usually diluted in twofold increments for 2D Titration to nucleosomes. In our experience Queries consistently display higher EC_{50}^{rel} concentrations with nucleosomes compared to peptides.
20. Choosing the optimal salt concentration is a combination of balancing signal-to-background of Targets vs. Controls, EC_{50}^{rel} values, and the resulting reagent consumption (considering all experimental modules). The general trend is greater salt stringency will decrease Query binding, which in some cases will help separate Targets from Controls. If this cannot be achieved and DNA binding is the suspected cause, a 2D [Query vs. SalDNA] titration may be necessary.

21. Choosing the optimal SalDNA concentration is similar to choosing optimal salt concentration (*see* **Note 19**). However, for SalDNA the ideal concentration is when maximum signal-to-background between Target and Control is achieved (usually when the DNA interaction is nearly abrogated by SalDNA). In some cases, the DNA interaction is essential for Query: nucleosome engagement and cannot be separated.
22. When preparing Discovery Screen plates, an additional dead volume of at least 10% is factored in to ensure the desired number of plates are prepared. Each substrate is usually prepared in duplicate. Typically, 5 μL of 400 nM peptide and 2.5 μL of 80 nM nucleosome is added per well. Peptides and nucleosomes are not recommended for simultaneous testing given their different buffer requirements.
23. Nucleosomes are prepared in a no-salt buffer for Discovery Screen plates to provide flexibility to adjust to the optimal salt concentration for any given Query. If desired, nucleosomes can be prepared with salt up to 250 mM NaCl (any higher may impact their long-term stability).
24. For proper storage of Discovery Screen plates, carefully apply a storage plate seal, centrifuge to settle any droplets, and store at −80 °C for up to 3 months. The recommended storage plate seals (*see* Subheading 2.3) use a high-bond pressure-sensitive adhesive. To create an extra tight seal, use a pen or marker cap to apply directed pressure around the plate perimeter. If an alternative plate seal is to be used, it is critical that the adhesives are designed for −80 °C.
25. Do not perform more than one freeze-thaw of Discovery Screen plates. They are to be used immediately after thawing.
26. For Queries of ~25–50 kDa with no known or suspected targets, use a high concentration of 1 μM and low concentration of 10 nM for a peptide discovery screen/a high concentration of 1 μM and low concentration of 50 nM for a nucleosome discovery screen. The high and low concentrations may require adjustment based on the protein size to prevent hook point issues (*see* **Note 15**). It is also recommended to start with a nucleosome assay buffer supplemented with 150 mM NaCl.
27. Using *GraphPad Prism*, the data can be organized into columns to analyze and visualize the discovery screen data. Alternatively, Targets can be rank ordered by max signal in an Excel file. A third method to visualize large quantities of ranking data is a heat map (e.g., Fig. 3). It is recommended that a target ranking experiment be performed to quantitatively rank any Targets of interest by their EC_{50}^{rel}.

Acknowledgments

This work was supported by US National Institutes of Health (NIH) grants (R44GM116584, R44GM117683 and R44CA214076) to *EpiCypher*.

References

1. Brownell JE, Zhou J, Ranalli T, Kobayashi R, Edmondson DG, Roth SY, Allis CD (1996) Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. Cell 84(6):843–851. https://doi.org/10.1016/s0092-8674(00)81063-6
2. Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. Cell 128(4):707–719. https://doi.org/10.1016/j.cell.2007.01.015
3. Utley RT, Ikeda K, Grant PA, Cote J, Steger DJ, Eberharter A, John S, Workman JL (1998) Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. Nature 394(6692):498–502. https://doi.org/10.1038/28886
4. Taunton J, Hassig CA, Schreiber SL (1996) A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. Science 272(5260):408–411. https://doi.org/10.1126/science.272.5260.408
5. Mirabella AC, Foster BM, Bartke T (2016) Chromatin deregulation in disease. Chromosoma 125(1):75–93. https://doi.org/10.1007/s00412-015-0530-0
6. Portela A, Esteller M (2010) Epigenetic modifications and human disease. Nat Biotechnol 28(10):1057–1068. https://doi.org/10.1038/nbt.1685
7. Valencia AM, Kadoch C (2019) Chromatin regulatory mechanisms and therapeutic opportunities in cancer. Nat Cell Biol 21(2):152–161. https://doi.org/10.1038/s41556-018-0258-1
8. Jenuwein T, Allis CD (2001) Translating the histone code. Science 293(5532):1074–1080. https://doi.org/10.1126/science.1063127
9. Strahl BD, Allis CD (2000) The language of covalent histone modifications. Nature 403(6765):41–45. https://doi.org/10.1038/47412
10. Lee JS, Smith E, Shilatifard A (2010) The language of histone crosstalk. Cell 142(5):682–685. https://doi.org/10.1016/j.cell.2010.08.011
11. Smith E, Shilatifard A (2010) The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. Mol Cell 40(5):689–701. https://doi.org/10.1016/j.molcel.2010.11.031
12. Jacobson RH, Ladurner AG, King DS, Tjian R (2000) Structure and function of a human TAFII250 double bromodomain module. Science 288(5470):1422–1425. https://doi.org/10.1126/science.288.5470.1422
13. Dey A, Chitsaz F, Abbasi A, Misteli T, Ozato K (2003) The double bromodomain protein Brd4 binds to acetylated chromatin during interphase and mitosis. Proc Natl Acad Sci U S A 100(15):8758–8763. https://doi.org/10.1073/pnas.1433065100
14. Dhalluin C, Carlson JE, Zeng L, He C, Aggarwal AK, Zhou MM (1999) Structure and ligand of a histone acetyltransferase bromodomain. Nature 399(6735):491–496. https://doi.org/10.1038/20974
15. Li Y, Wen H, Xi Y, Tanaka K, Wang H, Peng D, Ren Y, Jin Q, Dent SY, Li W, Li H, Shi X (2014) AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation. Cell 159(3):558–571. https://doi.org/10.1016/j.cell.2014.09.049
16. Andrews FH, Shinsky SA, Shanle EK, Bridgers JB, Gest A, Tsun IK, Krajewski K, Shi X, Strahl BD, Kutateladze TG (2016) The Taf14 YEATS domain is a reader of histone crotonylation. Nat Chem Biol 12(6):396–398. https://doi.org/10.1038/nchembio.2065
17. Li Y, Sabari BR, Panchenko T, Wen H, Zhao D, Guan H, Wan L, Huang H, Tang Z, Zhao Y, Roeder RG, Shi X, Allis CD, Li H (2016) Molecular coupling of histone crotonylation and active transcription by AF9 yeats domain. Mol Cell 62(2):181–193. https://doi.org/10.1016/j.molcel.2016.03.028
18. Botuyan MV, Lee J, Ward IM, Kim JE, Thompson JR, Chen J, Mer G (2006) Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. Cell 127(7):1361–1373. https://doi.org/10.1016/j.cell.2006.10.043
19. Musselman CA, Avvakumov N, Watanabe R, Abraham CG, Lalonde ME, Hong Z, Allen C, Roy S, Nunez JK, Nickoloff J, Kulesza CA,

Yasui A, Cote J, Kutateladze TG (2012) Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. Nat Struct Mol Biol 19(12):1266–1272. https://doi.org/10.1038/nsmb.2435

20. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, Kouzarides T (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature 410(6824):120–124. https://doi.org/10.1038/35065138
21. Pray-Grant MG, Daniel JA, Schieltz D, Yates JR 3rd, Grant PA (2005) Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature 433(7024):434–438. https://doi.org/10.1038/nature03242
22. Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J Biol Chem 285(34): 26114–26120. https://doi.org/10.1074/jbc.M109.089433
23. Weinberg DN, Rosenbaum P, Chen X, Barrows D, Horth C, Marunde MR, Popova IK, Gillespie ZB, Keogh M-C, Lu C, Majewski J, Allis CD (2021) Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands. Nat Genet 53(6): 794–800. https://doi.org/10.1038/s41588-021-00856-5
24. Wilson MD, Benlekbir S, Fradet-Turcotte A, Sherker A, Julien JP, McEwan A, Noordermeer SM, Sicheri F, Rubinstein JL, Durocher D (2016) The structural basis of modified nucleosome recognition by 53BP1. Nature 536(7614):100–103. https://doi.org/10.1038/nature18951
25. Arrowsmith CH, Schapira M (2019) Targeting non-bromodomain chromatin readers. Nat Struct Mol Biol 26(10):863–869. https://doi.org/10.1038/s41594-019-0290-2
26. Dawson MA (2017) The cancer epigenome: concepts, challenges, and therapeutic opportunities. Science 355(6330):1147–1152. https://doi.org/10.1126/science.aam7304
27. Zaware N, Zhou MM (2017) Chemical modulators for epigenome reader domains as emerging epigenetic therapies for cancer and inflammation. Curr Opin Chem Biol 39:116–125. https://doi.org/10.1016/j.cbpa.2017.06.012
28. Huang H, Lin S, Garcia BA, Zhao Y (2015) Quantitative proteomic analysis of histone modifications. Chem Rev 115(6):2376–2418. https://doi.org/10.1021/cr500491u
29. Garske AL, Oliver SS, Wagner EK, Musselman CA, LeRoy G, Garcia BA, Kutateladze TG, Denu JM (2010) Combinatorial profiling of chromatin binding modules reveals multisite discrimination. Nat Chem Biol 6(4):283–290. https://doi.org/10.1038/nchembio.319
30. Tauber M, Fischle W (2015) Conserved linker regions and their regulation determine multiple chromatin-binding modes of UHRF1. Nucleus 6(2):123–132. https://doi.org/10.1080/19491034.2015.1026022
31. Rothbart SB, Krajewski K, Strahl BD, Fuchs SM (2012) Peptide microarrays to interrogate the "histone code". Methods Enzymol 512: 107–135. https://doi.org/10.1016/B978-0-12-391940-3.00006-8
32. Bua DJ, Kuo AJ, Cheung P, Liu CL, Migliori V, Espejo A, Casadio F, Bassi C, Amati B, Bedford MT, Guccione E, Gozani O (2009) Epigenome microarray platform for proteome-wide dissection of chromatin-signaling networks. PLoS One 4(8):e6789. https://doi.org/10.1371/journal.pone.0006789
33. Matthews AG, Kuo AJ, Ramon-Maiques S, Han S, Champagne KS, Ivanov D, Gallardo M, Carney D, Cheung P, Ciccone DN, Walter KL, Utz PJ, Shi Y, Kutateladze TG, Yang W, Gozani O, Oettinger MA (2007) RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. Nature 450(7172):1106–1110. https://doi.org/10.1038/nature06431
34. Mauser R, Jeltsch A (2019) Application of modified histone peptide arrays in chromatin research. Arch Biochem Biophys 661:31–38. https://doi.org/10.1016/j.abb.2018.10.019
35. Rathert P, Dhayalan A, Murakami M, Zhang X, Tamas R, Jurkowska R, Komatsu Y, Shinkai Y, Cheng X, Jeltsch A (2008) Protein lysine methyltransferase G9a acts on non-histone targets. Nat Chem Biol 4(6):344–346. https://doi.org/10.1038/nchembio.88
36. Kudithipudi S, Lungu C, Rathert P, Happel N, Jeltsch A (2014) Substrate specificity analysis and novel substrates of the protein lysine methyltransferase NSD1. Chem Biol 21(2): 226–237. https://doi.org/10.1016/j.chembiol.2013.10.016
37. Shah RN, Grzybowski AT, Cornett EM, Johnstone AL, Dickson BM, Boone BA, Cheek MA, Cowles MW, Maryanski D, Meiners MJ, Tiedemann RL, Vaughan RM, Arora N, Sun ZW, Rothbart SB, Keogh MC, Ruthenburg AJ (2018) Examining the roles of H3K4 methylation states with systematically characterized antibodies. Mol Cell 72(1):162–177 e167.

https://doi.org/10.1016/j.molcel.2018.08.015

38. Ghoneim M, Fuchs HA, Musselman CA (2021) Histone tail conformations: a fuzzy affair with DNA. Trends Biochem Sci 46(7):564–578. https://doi.org/10.1016/j.tibs.2020.12.012
39. Morrison EA, Sanchez JC, Ronan JL, Farrell DP, Varzavand K, Johnson JK, Gu BX, Crabtree GR, Musselman CA (2017) DNA binding drives the association of BRG1/hBRM bromodomains with nucleosomes. Nat Commun 8:16080. https://doi.org/10.1038/ncomms16080
40. Morrison EA, Bowerman S, Sylvers KL, Wereszczynski J, Musselman CA (2018) The conformation of the histone H3 tail inhibits association of the BPTF PHD finger with the nucleosome. elife 7:e31481. https://doi.org/10.7554/eLife.31481
41. Stutzer A, Liokatis S, Kiesel A, Schwarzer D, Sprangers R, Soding J, Selenko P, Fischle W (2016) Modulations of DNA contacts by linker histones and post-translational modifications determine the mobility and modifiability of nucleosomal H3 tails. Mol Cell 61(2):247–259. https://doi.org/10.1016/j.molcel.2015.12.015
42. Li Y, Trojer P, Xu CF, Cheung P, Kuo A, Drury WJ 3rd, Qiao Q, Neubert TA, Xu RM, Gozani O, Reinberg D (2009) The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. J Biol Chem 284(49):34283–34295. https://doi.org/10.1074/jbc.M109.034462
43. Sankaran SM, Wilkinson AW, Elias JE, Gozani O (2016) A PWWP domain of histone-lysine N-methyltransferase NSD2 binds to dimethylated Lys-36 of histone H3 and regulates NSD2 function at chromatin. J Biol Chem 291(16):8465–8474. https://doi.org/10.1074/jbc.M116.720748
44. Vermeulen M, Eberl HC, Matarese F, Marks H, Denissov S, Butter F, Lee KK, Olsen JV, Hyman AA, Stunnenberg HG, Mann M (2010) Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell 142(6):967–980. https://doi.org/10.1016/j.cell.2010.08.020
45. Qin S, Min J (2014) Structure and function of the nucleosome-binding PWWP domain. Trends Biochem Sci 39(11):536–547. https://doi.org/10.1016/j.tibs.2014.09.001
46. McGinty RK, Tan S (2016) Recognition of the nucleosome by chromatin factors and enzymes. Curr Opin Struct Biol 37:54–61. https://doi.org/10.1016/j.sbi.2015.11.014
47. Skrajna A, Goldfarb D, Kedziora KM, Cousins EM, Grant GD, Spangler CJ, Barbour EH, Yan X, Hathaway NA, Brown NG, Cook JG, Major MB, McGinty RK (2020) Comprehensive nucleosome interactome screen establishes fundamental principles of nucleosome binding. Nucleic Acids Res 48(17):9415–9432. https://doi.org/10.1093/nar/gkaa544
48. Shi YJ, Matson C, Lan F, Iwase S, Baba T, Shi Y (2005) Regulation of LSD1 histone demethylase activity by its associated factors. Mol Cell 19(6):857–864. https://doi.org/10.1016/j.molcel.2005.08.027
49. Yang M, Gocke CB, Luo X, Borek D, Tomchick DR, Machius M, Otwinowski Z, Yu H (2006) Structural basis for CoREST-dependent demethylation of nucleosomes by the human LSD1 histone demethylase. Mol Cell 23(3):377–387. https://doi.org/10.1016/j.molcel.2006.07.012
50. Feng Q, Wang H, Ng HH, Erdjument-Bromage H, Tempst P, Struhl K, Zhang Y (2002) Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. Curr Biol 12(12):1052–1058. https://doi.org/10.1016/s0960-9822(02)00901-6
51. Strelow JM, Xiao M, Cavitt RN, Fite NC, Margolis BJ, Park KJ (2016) The use of nucleosome substrates improves binding of SAM analogs to SETD8. J Biomol Screen 21(8):786–794. https://doi.org/10.1177/1087057116656596
52. Rothbart SB, Strahl BD (2014) Interpreting the language of histone and DNA modifications. Biochim Biophys Acta 1839(8):627–643. https://doi.org/10.1016/j.bbagrm.2014.03.001
53. Andrews FH, Strahl BD, Kutateladze TG (2016) Insights into newly discovered marks and readers of epigenetic information. Nat Chem Biol 12(9):662–668. https://doi.org/10.1038/nchembio.2149
54. Weinberg DN, Papillon-Cavanagh S, Chen H, Yue Y, Chen X, Rajagopalan KN, Horth C, McGuire JT, Xu X, Nikbakht H, Lemiesz AE, Marchione DM, Marunde MR, Meiners MJ, Cheek MA, Keogh MC, Bareke E, Djedid A, Harutyunyan AS, Jabado N, Garcia BA, Li H, Allis CD, Majewski J, Lu C (2019) The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. Nature 573(7773):281–286. https://doi.org/10.1038/s41586-019-1534-3
55. Jain K, Fraser CS, Marunde MR, Parker MM, Sagum C, Burg JM, Hall N, Popova IK, Rodriguez KL, Vaidya A, Krajewski K, Keogh MC, Bedford MT, Strahl BD (2020)

Characterization of the plant homeodomain (PHD) reader family for their histone tail interactions. Epigenetics Chromatin 13(1):3. https://doi.org/10.1186/s13072-020-0328-z

56. Lloyd JT, McLaughlin K, Lubula MY, Gay JC, Dest A, Gao C, Phillips M, Tonelli M, Cornilescu G, Marunde MR, Evans CM, Boyson SP, Carlson S, Keogh MC, Markley JL, Frietze S, Glass KC (2020) Structural insights into the recognition of mono- and diacetylated histones by the ATAD2B bromodomain. J Med Chem 63(21):12799–12813. https://doi.org/10.1021/acs.jmedchem.0c01178
57. Dilworth D, Hanley RP, Ferreira de Freitas R, Allali-Hassani A, Zhou M, Mehta N, Marunde MR, Ackloo S, Carvalho Machado RA, Khalili Yazdi A, Owens DDG, Vu V, Nie DY, Alqazzaz M, Marcon E, Li F, Chau I, Bolotokova A, Qin S, Lei M, Liu Y, Szewczyk MM, Dong A, Kazemzadeh S, Abramyan T, Popova IK, Hall NW, Meiners MJ, Cheek MA, Gibson E, Kireev D, Greenblatt JF, Keogh MC, Min J, Brown PJ, Vedadi M, Arrowsmith CH, Barsyte-Lovejoy D, James LI, Schapira M (2021) Pharmacological targeting of a PWWP domain demonstrates cooperative control of NSD2 localization. Nat Chem Biol. https://doi.org/10.1038/s41589-021-00898-0. PMID: 34782742
58. Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L, Zhang Y, Bedford MT (2006) Tudor, MBT and chromo domains gauge the degree of lysine methylation. EMBO Rep 7(4):397–403. https://doi.org/10.1038/sj.embor.7400625
59. Rondelet G, Dal Maso T, Willems L, Wouters J (2016) Structural basis for recognition of histone H3K36me3 nucleosome by human de novo DNA methyltransferases 3A and 3B. J Struct Biol 194(3):357–367. https://doi.org/10.1016/j.jsb.2016.03.013

Part III

Chromatin Accessibility

Chapter 14

High-Resolution ATAC-Seq Analysis of Frozen Clinical Tissues

Paloma Cejas and Henry W. Long

Abstract

The ATAC-seq method enables the genome-wide analysis of accessible chromatin revealing transcriptionally active and poised regulatory elements. The ATAC-seq analysis of clinical specimens at a single-cell resolution reveals the cellular composition of the tissue contributing to the understanding of intra-tissue heterogeneity. Here we describe our method for nuclei isolation from frozen specimens with wide applicability across tissue types, producing nuclei suitable for a number of molecular profiling methods including ATAC-seq in bulk and at a single-cell resolution.

Key words Clinical frozen tissues, Nuclei isolation, Chromatin state, ATAC-seq, ChIP-seq, Single-cell analysis

1 Introduction

Chromatin analysis of clinical tissues is enabling the understanding of the molecular mechanisms underlying physiological and pathological processes [1–8]. A plethora of methods is currently available to perform chromatin analysis with ATAC-seq becoming increasingly popular to profile chromatin accessibility due to its straightforward protocol and its applicability to low cell numbers [9]. However, performing ATAC-seq analysis in clinical specimens is still challenging with no described standard methods widely applicable across tissue types. Instead, there are a high number of protocols for tissue-specific cellular disaggregation from fresh samples that include harsh steps like protease treatment and heating which can produce cell damage. Here we describe a straightforward method to perform nuclei isolation from frozen tissues suitable for bulk and single-cell (sc) ATAC-seq analysis. Our protocol for nuclei isolation from frozen tissues simplifies previous methods that apply challenging and time-consuming density gradient centrifugation approaches [10]. We first perform a histological examination of

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_14,

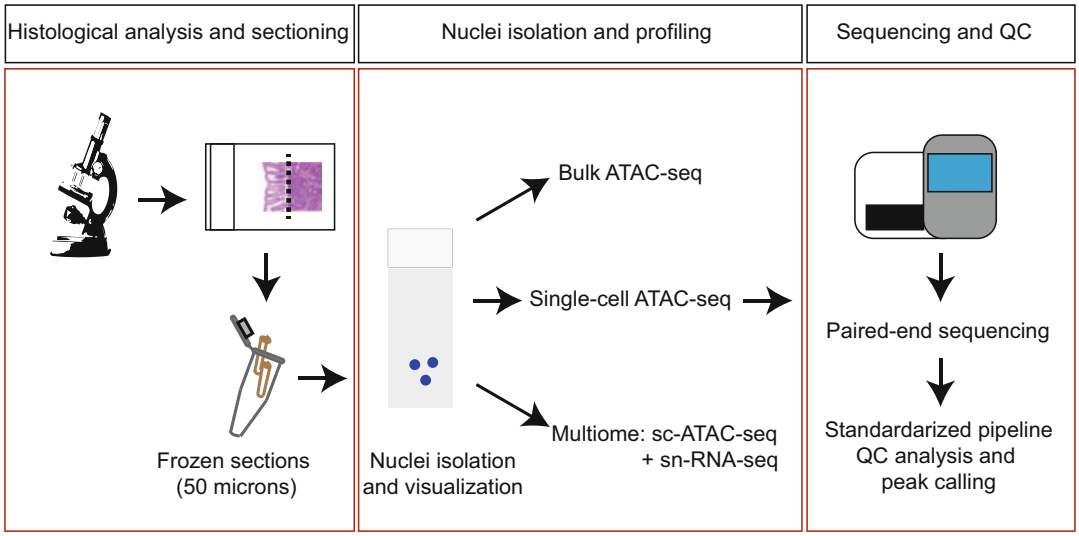


Fig. 1 Schematic of our workflow strategy to perform chromatin accessibility analysis. It consists of an initial histological analysis and tissue sectioning, followed by nuclei isolation yielding material suitable for molecular profiling and analysis

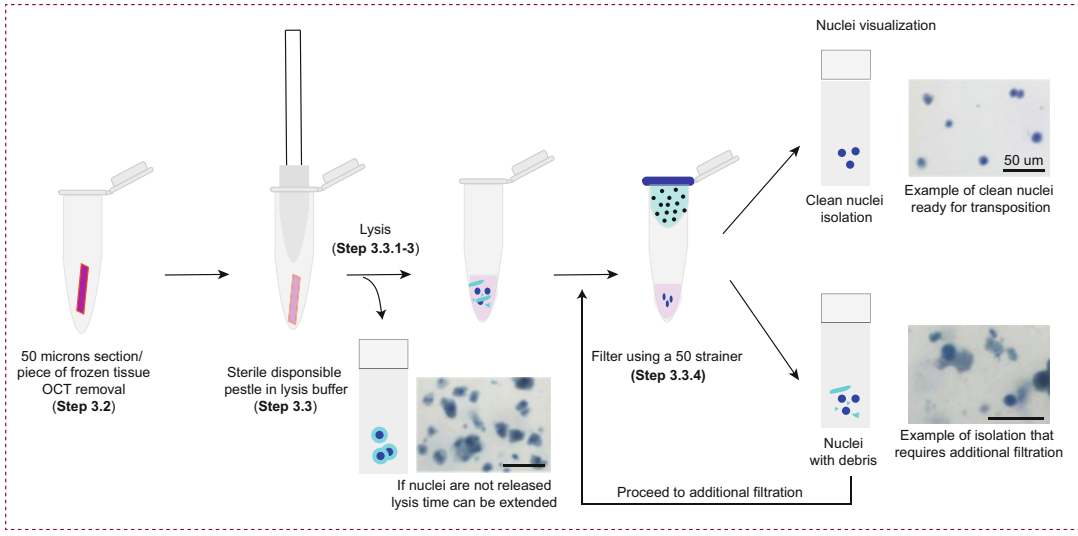


Fig. 2 Schematic of the protocol of nuclei isolation from frozen sections. After homogenization and lysis, the quality of the nuclei should be assessed. If the nuclei are not released, lysis time can be extended. If the resulting nuclei still show debris, the filtration of the nuclei suspension through a 40 μm should be repeated

the tissue, which is particularly important when analyzing clinical samples [3] (Fig. 1), that is followed by the direct homogenization of the frozen tissue, avoiding the disaggregation of fresh tissue (Fig. 2). Our strategy of starting from frozen tissue instead of from fresh tissue, overcomes the requirement of tissue-specific optimizations. In contrast, we use a "one-suits-all" protocol

based on a mechanical method avoiding enzymatic digestion and sample heating that produces a clean nuclear preparation across numerous tissue types. With this protocol, we have produced high-quality results both in bulk and at high-resolution single-cell analysis [11]. Furthermore, the isolated nuclei can be used as the starting material for a wide number of additional methods, including RNA-seq, ChIP-seq, and Hi-CHIP analysis.

2 Materials

When manipulating clinical specimens, wear appropriate protective equipment to avoid injury and cutaneous absorption. Human tissue specimens should be treated as potentially infectious.

2.1 Tissue Freezing

1. Liquid nitrogen.
2. Forceps.
3. Sterile cryovials.
4. Tissue-Tek® O.C.T. Compound.
5. Tissue OCT molds.

2.2 Tissue Sectioning and Histological Analysis

1. Cryostat.
2. Cryostat blades.
3. SuperFrost™ Plus slides for histological examination.
4. Reagents for hematoxylin and eosin staining following standard methods.

2.3 Tissue Homogenization

1. Microcentrifuge, cooling.
2. Pestle and Microtube combo 1.5 mL.
3. Cell mini strainers, 40 μm for 1.5 mL microtube.
4. Lysis buffer: 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$ supplemented with 0.1% NP-40, 0.1% Tween 20 and 0.01% Digitonin. Supplementation should be done right before it is used.
5. Wash buffer: 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween 20.

2.4 Nuclei Counting

1. Countess® II FL Automated Cell Counter.
2. Countess® Cell Counting Chamber Slides.
3. DAPI reagent for nuclear staining.

2.5 ATAC-Seq Transposition (Experiment in Bulk)

1. Heat block at 37 °C.
2. 2× transposition buffer: 20 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 20% Dimethyl Formamide.

3. Transposition mix: 25 μL 2× transposition buffer, 16.5 μL PBS, 0.5 μL 1% Digitonin, 0.5 μL 10% Tween 20, 5 μL H_2O.
4. Illumina Tagment DNA Enzyme and Buffer.
5. Qiagen MinElute® Kit.

2.6 ATAC-Seq Amplification

1. Thermal cycler.
2. Nextera Index Kit (primers 1 and 2) (Illumina Inc).
3. SYBR Green I.
4. NEBNext® High-Fidelity 2× PCR Master Mix.

3 Methods

Frozen specimens and sections should be processed as quickly as possible to avoid tissue autolysis. To avoid cross-contamination, we recommended using a new cryostat blade for each specimen.

3.1 Tissue Freezing

Tissues can be snap-frozen or frozen as OCT embedded blocks. OCT blocks enable the preservation of histological details.

3.1.1 Snap Freezing

1. Collect the tissue in a sterile cryovial (*see* **Note 1**).
2. Immerse the cryovial containing the tissue in liquid nitrogen.
3. Store the vial in a liquid nitrogen storage tank or in a −80 °C freezer for long-term storage.

3.1.2 OCT Embedding and Freezing

1. Before freezing the tissue, arrange cryomolds and label them with a marker.
2. Keep the OCT at room temperature.
3. Put several drops of OCT into the plastic cryomold. Place tissue on top maintaining the correct orientation for cutting. Carefully pour OCT on top of the tissue being careful to avoid bubbles until none of the tissue remains exposed. Hold the cryomold keeping it upright using forceps and introduce it inside liquid nitrogen until the block is frozen. Tissue blocks should be stored at −80 °C.

3.2 Tissue Sectioning

OCT blocks enable the histological analysis that should be performed to evaluate the enrichment in the cell type to study. Macrodissection should be performed to enrich the cells to study, this is particularly important when performing bulk analysis.

1. For the histological analysis, cut a 5–10 μm thick cryostat section and mount it on a Superfrost Plus Slide (*see* **Note 2**) Stain the slides with hematoxylin-eosin (H&E) for histologic examination following standard methods. The tissue can be macrodissected using a sterile blade if necessary.

2. For nuclei isolation, cut 1–2 (50 μm) sections and introduce them into a prechilled 1.5 mL tube.
3. Add 1 mL cold PBS (1×) to remove the OCT.
4. Centrifuge at 1500 × *g* for 1 min at 4 °C and discard the supernatant.
5. Repeat **steps 1** and **2** until OCT is completely eliminated.

3.3 Tissue Homogenization and Cell Lysis

1. After OCT removal, resuspend the tissue section or the snap-frozen tissue in 300 μL cold lysis buffer and dounce the tissue until complete homogenization using disposable pestles in 1.5 mL microtubes (*see* **Note 3**) (Fig. 2).
2. Incubate the tubes for 10 min on ice (*see* **Note 4**).
3. Add 1 mL of wash buffer.
4. Pass the homogenized tissue through a 40 μm cell strainer. The flow-through should contain the nuclei.
5. Centrifuge the nuclei at 800 × *g* for 10 min at 4 °C and remove supernatant.
6. Wash nuclei with 300 μL wash buffer.
7. Repeat **steps 5** and **6** to a total of 2 times.
8. Discard the supernatant and resuspend in 300 μL wash buffer.
9. Count and visualize nuclei with the Countess® II FL Automated Cell Counter counterstained with DAPI reagent. Alternatively, nuclei can be counted using a hemocytometer.
10. If the nuclei preparation still has debris and is not clean enough, repeat **steps 4**–**7** (Fig. 2).
11. At this step, nuclei are ready to proceed to transposition for bulk and sc analysis. Nuclei can be stored at this point resuspended in 90% FBS/10% DMSO at −80 °C or in liquid nitrogen for extended periods. For particular specifications for sc-ATAC-seq *see* **Note 5**.

3.4 Bulk ATAC-Seq

We use the OMNI protocol previously described by Corces et al. [10]. Using this protocol, we have obtained high-quality ATAC-seq results in bulk from 5000–100,000 cells. When starting with lower number of cells the protocol loses robustness and the quality of the resulting signal may drop. At low numbers, the sc strategy can be a better option; sc-ATAC-seq enables the analysis of lower numbers, that could range from as low as several hundred to <10,000 nuclei by commercial microfluidic platforms.

1. Pellet 5000–100,000 nuclei at 800 × *g* for 5 min at 4 °C.
2. Aspirate supernatant without disrupting the pellet.

3. Resuspend the cell pellet in 50 μL of transposition mix by pipetting up and down 6 times and add 2.5 μL of Tn5 transposase (100 nM final) (*see* **Note 6**).
4. Incubate the reaction at 37 °C for 30 min in a thermomixer with 850 rpm mixing (*see* **Notes 7** and **8**).
5. Purify using a Qiagen MinElute Kit and elute the transposed DNA in 12 μL Elution Buffer.

3.5 DNA Amplification

The PCR amplification requires:

1. Amplify the transposed DNA preparing a 50 μL reaction containing: 12 μL of the transposed DNA, 2 μL of 5 μM Customized Nextera PCR Primer 1, 2 μL of 5 μM Customized Nextera PCR Primer 2, 0.3 μL of 100× SYBR Green I, 25 μL of NEBNext High-Fidelity 2× PCR Master Mix, and 8.7 μL of nuclease-free water.
2. Run the following PCR program: 1 cycle of 5 min at 72 °C and 30 s at 98 °C followed by N cycles of 10 s at 98 °C, 30 s at 63 °C, and 1 min at 72 °C.
3. The number of cycles depends on the starting number of transposed nuclei. The previously described protocol (OMNI) recommends the performance of an initial 5 cycles reaction followed by a qPCR side reaction to monitor the amplification reaction to estimate the number of additional cycles to be added [10, 12]. However, to facilitate the calculation of cycles required to amplify: add 15 cycles when starting from <10,000 nuclei; add 13–14 cycles from 10,000–50,000 nuclei; add 12–13 cycles for 50,000–100,000 nuclei.
4. Proceed to library sequencing (*see* **Note 9**), initial analysis of the sequences, and quality control (QC) (*see* **Note 10**).

4 Notes

1. Scissors and forceps should be washed with 70% alcohol in between handling different tissues to avoid cross-contamination.
2. In case it is required, slides can be stored at −80 °C until needed**.**
3. When starting from snap-frozen tissues, a tissue grinder 1 mL dounce can be used to facilitate the homogenization of pieces of tissue. Whenever possible and, to avoid cross-sample contamination, we recommend using disposable pestles that are designed to fit a 1.5 mL microtube. Cell lysis and nuclei isolation may require extended time for certain tissue types. We recommend visualizing the nuclei by staining with methylene blue at 40× microscope magnification (Fig. 2).

4. If lysis is not completed and nuclei are still not completely released, lysis time can be extended.
5. The nuclei isolated with this method are suitable for sc analysis. We have validated its applicability for sc-ATAC-seq, sn-RNA--seq, and Multiome (sc-ATAC-seq and sn-RNA-seq on the same cell) using 10× Genomics technology. However, the nuclei could eventually be used for additional techniques, for example, sci-ATAC [13] and other single-cell methods. For 10× Genomics sc-ATAC-seq analysis, a maximum of 10,000 nuclei can be loaded following the manufacturer's instructions. When performing lineage tracing by analysis of mutations in mitochondrial DNA as described by Lareau et al. [12], we recommend using their protocol that replaces the lysis buffer described in **item 4** of Subheading 2.3 with this buffer: 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% NP-40, 1% BSA) at **step 1** of Subheading 3.3. This buffer retains more of the mitochondrial DNA required for cell traceability.
6. 2.5 μL of the commercial Tn5 enzyme (Illumina Inc) gives an adequate chromatin fragmentation from 5000 up to 10,0000 nuclei. However, if using a lower number of nuclei or a distinct Tn5 enzyme, the amount of enzyme should be optimized.
7. The transposed nuclei at this step can be directly loaded in the 10× Chromium Chip into the Chromium Controller.
8. ATAC-seq can be performed from 1% formaldehyde-fixed cells, in that case, a reverse cross-linking step should be performed right after the transposition. To that aim, the 50 μL volume of transposition mix should be increased to 200 μL with a reverse cross-linking buffer: 50 mM Tris-HCl pH 7.4, 1 mM EDTA, 1% SDS, 0.2 M NaCl, 5 μg/mL proteinase K and incubated at 65 °C in a heat block overnight. After incubation, the DNA should be purified using a Qiagen MinElute® column or any other method of DNA purification.
9. We recommend performing paired-end sequencing for bulk and sc-ATAC-seq analysis. Paired-end sequencing enables assessing the size distribution of the resulting fragments and identifying the precise location of each transposition event.
10. We recommend performing a rigorous quality control with a reproducible analytic pipeline. We apply our designed pipeline ChiLin [14] that includes a number of quality control steps for both the sequencing reads and the peaks identified. For sequencing quality control: read depth, mapping rates, redundancy rates, fraction of reads in peaks (FRIP), fragment size distribution and contamination screening (to assess presence of additional contaminating genomes) should be reported. We recommend performing peak calling by MACS2 [15] using default parameters. Quality control of peaks includes: analysis

of conservation (regulatory elements are conserved across species), assessment of the genome distribution of the peaks, and analysis of the peak overlap with global DNase-sensitive sites (regulatory elements are DNase hypersensitive) [14]. For assessing sequencing bias that can be particularly represented when clinical tissues are analyzed we recommend reading the review by Meyer and Liu [16] . Although further analysis depends on the particular biological question that is assessed, there are useful guidelines by Yan et al. [17].

Acknowledgments

P.C. acknowledges funding from the Ministry of Economy and Competitiveness, Instituto de Salud Carlos III (Institute of Health Carlos III)—PI18-01604. H.W.L. acknowledge support from NIH grant P01 CA163227-06A1 and P01 CA250959-01.

References

1. Cejas P, Drier Y, Dreijerink KMA, Brosens LAA, Deshpande V, Epstein CB et al (2019) Enhancer signatures stratify and predict outcomes of non-functional pancreatic neuroendocrine tumors. Nat Med 25:1260–1265
2. Cejas P, Cavazza A, Yandava CN, Moreno V, Horst D, Moreno-Rubio J et al (2017) Transcriptional regulator CNOT3 defines an aggressive colorectal cancer subtype. Cancer Res 77:766–779
3. Cejas P, Long HW (2020) Principles and methods of integrative chromatin analysis in primary tissues and tumors. Biochim Biophys Acta Rev Cancer 1873:188333
4. Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W et al (2018) The chromatin accessibility landscape of primary human cancers. Science 362(6413):eaav1898. https://doi.org/10.1126/science.aav1898
5. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E et al (2012) The accessible chromatin landscape of the human genome. Nature 489:75–82
6. ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N et al (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583:699–710
7. Singh H, Ha K, Hornick JL, Madha S, Cejas P, Jajoo K et al (2021) Hybrid stomach-intestinal chromatin states underlie human Barrett's metaplasia. Gastroenterology 161(3): 924–939.e11. https://doi.org/10.1053/j.gastro.2021.05.057
8. Font-Tello A, Kesten N, Xie Y, Taing L, Varešlija D, Young LS et al (2020) FiTAc-seq: fixed-tissue ChIP-seq for H3K27ac profiling and super-enhancer analysis of FFPE tissues. Nat Protoc 15:2503–2518
9. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ (2015) ATAC-seq: a method for assaying chromatin accessibility genome-wide. Curr Protoc Mol Biol 109:21.29.1–21.29.9
10. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S et al (2017) An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods 14:959–962
11. Cejas P, Xie Y, Font-Tello A, Lim K, Syamala S, Qiu X et al (2021) Subtype heterogeneity and epigenetic convergence in neuroendocrine prostate cancer. Nat Commun 12(1):5775. https://doi.org/10.1101/2020.09.13.291328
12. Lareau CA, Ludwig LS, Muus C, Gohil SH, Zhao T, Chiang Z et al (2021) Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling. Nat Biotechnol 39: 451–461
13. Cusanovich DA, Hill AJ, Aghamirzaie D, Daza RM, Pliner HA, Berletch JB et al (2018) A single-cell atlas of in vivo mammalian chromatin accessibility. Cell 174:1309–1324.e18

14. Qin Q, Mei S, Wu Q, Sun H, Li L, Taing L et al (2016) ChiLin: a comprehensive ChIP-seq and DNase-seq quality control and analysis pipeline. BMC Bioinformatics 17(1):404. https://doi.org/10.1186/s12859-016-1274-4
15. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE et al (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137
16. Meyer CA, Liu XS (2014) Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat Rev Genet 15:709–721
17. Yan F, Powell DR, Curtis DJ, Wong NC (2020) From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. Genome Biol 21(1):22. https://doi.org/10.1186/s13059-020-1929-3

Chapter 15

Single-Molecule Multikilobase-Scale Profiling of Chromatin Accessibility Using m6A-SMAC-Seq and m6A-CpG-GpC-SMAC-Seq

Georgi K. Marinov, Zohar Shipony, Anshul Kundaje, and William J. Greenleaf

Abstract

A hallmark feature of active cis-regulatory elements (CREs) in eukaryotes is their nucleosomal depletion and, accordingly, higher accessibility to enzymatic treatment. This property has been the basis of a number of sequencing-based assays for genome-wide identification and tracking the activity of CREs across different biological conditions, such as DNAse-seq, ATAC-seq, NOMeseq, and others. However, the fragmentation of DNA inherent to many of these assays and the limited read length of short-read sequencing platforms have so far not allowed the simultaneous measurement of the chromatin accessibility state of CREs located distally from each other. The combination of labeling accessible DNA with DNA modifications and nanopore sequencing has made it possible to develop such assays. Here, we provide a detailed protocol for carrying out the SMAC-seq assay (Single-Molecule long-read Accessible Chromatin mapping sequencing), in its m6A-SMAC-seq and m6A-CpG-GpC-SMAC-seq variants, together with methods for data processing and analysis, and discuss key experimental and analytical considerations for working with SMAC-seq datasets.

Key words Chromatin accessibility, SMAC-seq, Nanopore sequencing, DNA modifications, m6A, EcoGII

1 Introduction

Chromatin accessibility is a key feature of the regulation of gene expression and many other aspects of chromatin biology in eukaryotes. Nearly all eukaryote genomes are packaged by nucleosomes, with each nucleosome being a dimer of two tetramers composed of the four core nucleosomal histones H3, H4, H2A and H2B. Packaging by nucleosomes has a generally inhibitory effect on RNA polymerase activity and to the occupancy of DNA by regulatory

Georgi K. Marinov and Zohar Shipony contributed equally with all other contributors.

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_15,

proteins. Accordingly, active regulatory regions in the genome are characterized by depleted nucleosomal occupancy and increased chromatin accessibility. This has turned out to be a highly useful property enabling the identification of candidate cis-regulatory elements and the tracking of their activity across cell types and conditions.

Mapping accessible chromatin relies on the preferential enzymatic action of various reagents whose access to DNA is occluded by the presence of nucleosomes. Four decades ago it was initially recognized that active cREs are hypersensitive to cleavage by DNase enzymes [1–3]. DNase hypersensitivity remained the primary approach for mapping cREs well into the genomic era, being first coupled to microarrays [4–6], and eventually high-throughput massively parallel sequencing [7–9].

The advent of high-throughput sequencing enabled the development of numerous novel strategies for mapping active CREs. ATAC-seq [10], which relies on the preferential insertion of the Tn5 transposase enzyme into open chromatin, has emerged as the most convenient, versatile, and widely used method for studying the chromatin state of the eukaryotic cell, including down to single cell level [11, 12].

Other methods have also been developed, using restriction enzymes [13], nicking enzymes [14], small molecules [15], viral integration [16], and others.

All of these methods share two common features—they involve fragmentation of DNA and they enrich for accessible DNA during sequencing library generation. Consequently, it is first, not possible to enumerate accessibility states within the cellular population, that is, how often is a given CRE accessible, and second, there is no way to study the relationship between the chromatin states of distant regulatory elements, as the linkage between them is lost during fragmentation.

An alternative strategy to cleavage-based methods is to label accessible DNA with methyltransferase enzymes, then read out methylation states using high-throughput sequencing. This is the basic idea behind the NOMe-seq assay [17] and its later dSMF extension [18]. NOMe-seq uses the GpC methyltransferase M. CviPI to label accessible DNA at GpC positions. Genomic DNA is then subjected to bisulfite readout, providing single-molecule and fractional methylation (and thus accessibility) maps genome-wide. Only M.CviPI can be used in mammalian genomes due to the presence of endogenous CpG methylation, and only the m5C modification can be utilized as this is what can be read out with base pair resolution using short-read sequencing. This presents a limitation, as GpC nucleotides are only found once every ~25 bp in a mammalian genome. In organisms such as Drosophila that do not have endogenous methylation, both a GpC and a CpG methyltransferase (M.SssI) can be used, increasing resolution to ~10 bp

on average, in the form of the dSMF assay. This has allowed the enumeration of protein occupancy states at unprecedented resolution at a single-molecule level [18]. Yet short-read approaches of this kind are still quite limited in their capabilities.

First, these resolution values are averages. In reality genomes contains some quite large stretches with no informative positions (Fig. 1), and not much can be done to address that limitation as long as m5C in GpC/CpG contexts is the only available modification.

Second, it is only possible to analyze fragments no longer than 600 bp due to read-length limitations of short-read sequencers. Even this has been very difficult to achieve, as DNA methylation has traditionally been mapped using bisulfite sequencing, and bisulfite treatment severely degrades DNA to lengths considerably shorter than 600 bp. The introduction of the EM-seq method [19] as an alternative to bisulfite conversion has largely eliminated the degradation issue, but short reads are still short reads, making it impossible to study chromatin states on the scale of many kilobases along the chromatin fiber.

With the advent of long read sequencing technologies, and especially nanopore sequencing, these limitations have been overcome. Nanopore sequencing is capable of reading out arbitrary DNA modifications [20, 21], and of doing so along the length of DNA molecules tens of kilobases long, allowing for the simultaneous capture of the chromatin states of CREs located far apart. This has enabled the development of a qualitatively new class of functional genomic assays [22–24].

The MeSMLR-seq [23] and nanoNOMe [24] assays have adapted the NOMe-seq approach to nanopore sequencing, using a GpC methyltransferase to label accessible DNA, then reading it out using nanopore sequencing. However, while this approach preserves long-range contiguity, it still suffers from the limitations imposed by the density of informative modification positions in the genome (Fig. 1).

In contrast, SMAC-seq [22] uses dense modifications, found once every few nucleotides in the genome. Accessible DNA is enzymatically labeled using a methyltransferase enzyme (or multiple such enzymes), high molecular weight (HMW) DNA is isolated, then subjected to nanopore sequencing, which allows for the direct detection of DNA modifications and thus the assembly of an accessibility map at the single molecule level and on multikilobase scales (Fig. 2). In addition, the dense modifications that SMAC-seq is based on also provide information about nucleosome occupancy/positioning [25] and even transcription factor footprints [26, 27]. Finally, the long reads provided by nanopore sequencing allow chromatin accessibility and nucleosome positioning to be profiled within repetitive regions of the genome that are otherwise not uniquely mappable using short reads.

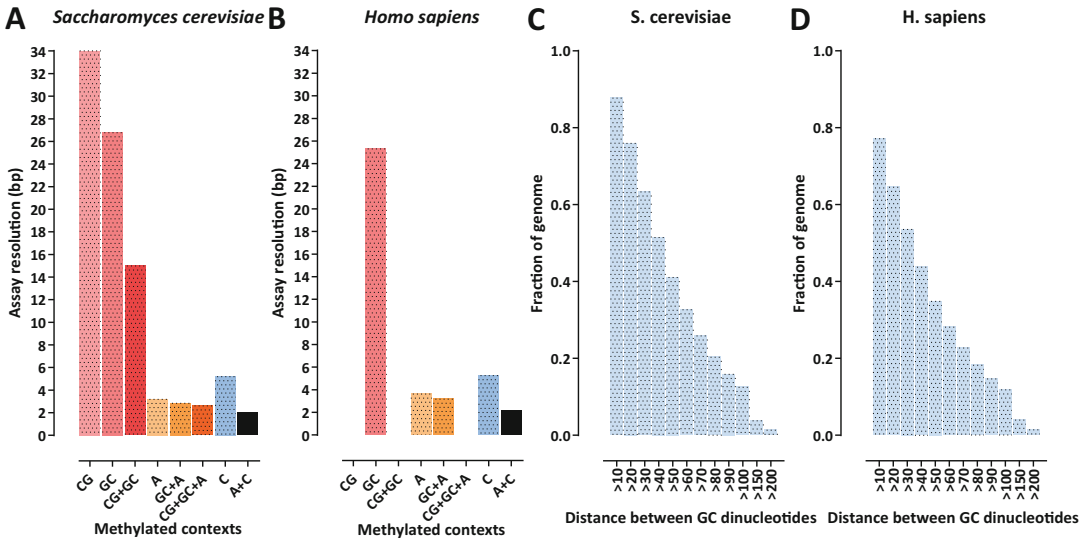


Fig. 1 Overview of the SMAC-seq experimental protocol. Chromatin is treated with m6A and optionally also with CpG and GpC 5mC methyltransferases, which preferentially methylate DNA bases within accessible chromatin. HMW DNA is then isolated and subjected to nanopore sequencing. After read mapping and identifying modified bases, the accessibility state within individual chromatin fibers can be reconstructed

Here, we describe an m6A-SMAC-seq protocol based on the m6A (N6Methyladenosine) methyltransferase EcoGII [28], which labels A bases nonspecifically in all contexts (*see* **Note 1**) as well as a m6A-CpGGpC-SMAC-seq protocol, which uses multiple modifications (m6A and m5C modifications in CpG and GpC contexts) and which can be used in organisms without endogenous DNA methylation. We also describe basic data processing and analysis procedures for working with SMAC-seq datasets.

2 Materials

SMAC-seq uses standard laboratory reagents with the exception of the m6A methyltransferase in the m6A version of the assay (*see* **Note 2**). Other versions of the assay involving different modifications may also require custom reagents.

2.1 SMAC-Seq Buffers and Reagents

1. Nuclei Lysis Buffer: 10 mM Tris–HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1 mM EDTA, 0.5% NP-40.
2. Nuclei Wash Buffer. This is the same as the Lysis Buffer except for the absence of NP-40: 10 mM Tris pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1 mM EDTA.
3. M.CviPI Reaction Buffer: 50 mM Tris-HCl pH 8.5, 50 mM NaCl, 10 mM DTT,

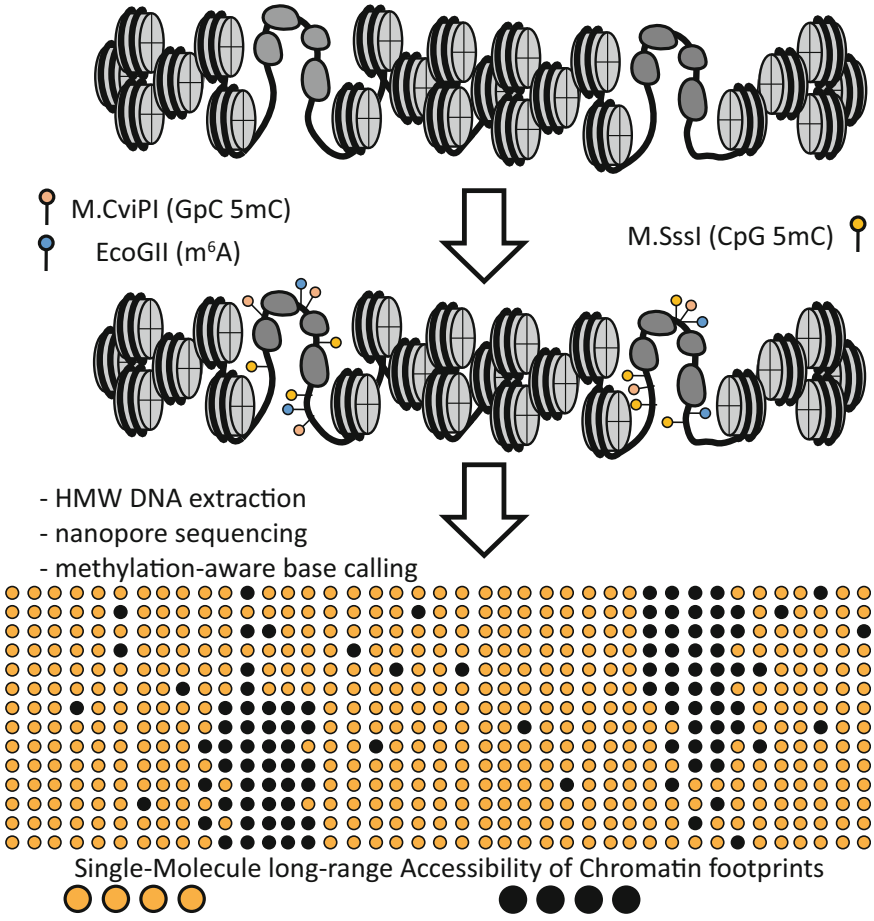


Fig. 2 Impact of the use of dense modifications on the theoretical resolution of methylation-based chromatin accessibility assays. (**a** and **b**) Theoretical average resolution of the SMAC-seq assay for different modification sequence contexts in the *S. cerevisiae* and *H. sapiens* genomes. (**c** and **d**) Limitations of using GpC m5C modifications alone due to the nonuniform distribution of GpC dinucleotides in the genome, which results in many large regions without any informative positions

4. CutSmart Reaction buffer: 1× CutSmart buffer, 0.3 M sucrose.
5. Stop Buffer: 20 mM Tris pH 8.5, 600 mM NaCl, 1% SDS, 10 mM EDTA.
6. Sorbitol Buffer: 1.4 M Sorbitol, 40 mM HEPES-KOH pH 7.5, 0.5 mM $MgCl_2$.
7. 100 T Zymolase
8. M.CviPI methyltransferase.
9. EcoGII methyltransferase (*see* **Note 2**).
10. M.SssI methyltransferase.

11. 32 mM S-adenosylmethionine (SAM)
12. Sucrose solution (prepare a highly concentrated solution, e.g., 2 M).
13. Molecular biology-grade 1 M $MgCl_2$ solution.

2.2 DNA Isolation and Size Selection

1. HMW DNA. We have most often used the MagAttract HMW DNA Kit (Qiagen), but other approaches for isolating HMW can also be applied, such as the NEB Monarch Genomic DNA Purification Kit, the Nanobind CBB Big DNA Kit, and others.
2. Size selection. Several solutions now also exist for HMW size selection that eliminates shorter fragments. We have used the Short Read Eliminator Kit (Circulomics) with fairly consistent levels of success, but equivalent approaches are also applicable.

2.3 Nanopore Sequencing Flow Cells and Reagents

Nanopore and SMAC-seq data can be generated using any of the Oxford Nanopore Technologies (ONT) platforms (Flongle, MinION, GridION, or PromethION). Which one to use is a decision to be made on the basis of the desired output, which in turn is determined by the needed coverage based on genome size, the properties of the genome studied, and so on (*see* **Notes 4** and **5**). ONT offers a variety of library preparation options, the two main ones relevant to SMAC-seq being the following.

1. The Ligation Sequencing Kits are to be used if maximum read length is desired. These require ~1000 ng of input HMW DNA.
2. The Rapid Barcoding Kit uses a transposases to simultaneously fragment DNA and attach adapters to the resulting pieces. Thus it will yield shorter molecules ($\leq$ ~10 kbp) but it allows the pooling of multiple samples in the same run (which is useful if, for example, working with an organism with a small genome and on a PromethION) and works with smaller amount of input HMW DNA (~400 ng).

Which kit is to be used depends on what the optimal choice is with respect to the particular research question and experimental system.

2.4 General Materials and Equipment

1. 1.5-mL microcentrifuge tubes, preferably low protein and DNA binding (*see* **Note 6**).
2. 2 mL, 15 mL and 50 mL tubes.
3. Magnetic stands for 1.5 mL and 2 mL tubes.
4. Thermomixer.
5. Molecular biology-grade 200 proof EtOH.
6. Tabletop centrifuge.
7. Nuclease-free H_2O.

8. 1× PBS.
9. AMPure beads.
10. QuBit fluorometer (ThermoFisher Scientific) or equivalent.
11. QuBit dsDNA HS kit (ThermoFisher Scientific).
12. TapeStation (Agilent) or equivalent.
13. TapeStation Genomic DNA Reagents (Agilent).
14. TapeStation Genomic DNA Screentape (Agilent).

2.5 Computational Resources

The computational analyses described are designed to run on standard Linux systems through the UNIX command line. The maximal memory usage depends on the size of the datasets but is usually less than ~50 GB. However, note that nanopore sequencing datasets can occupy very large amounts of disk space (i.e., many terabytes), thus it is advisable to use a computing system with ample storage (*see* **Note 7**).

2.6 Genomic Sequence and Annotation Files

1. A FASTA file containing the GRCh38 version of the human genome can be downloaded from the UCSC Genome Browser at http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz. Genome files can also be obtained from ENSEMBL (http://ensemblgenomes.org/) and from the NCBI website (http://www.ncbi.nlm.nih.gov/assembly/). However, it has to be noted that in the case of the human genome, reference FASTA files available in public repositories contain alternative haplotype contigs, that is, alternative versions of sequences already present in the assembly. These alternative haplotypes should be removed from reference files before use. The ENCODE Project [29] provides such filtered files from its portal at https://www.encodeproject.org/data-standards/reference-sequences/. The sacCer3 version of the Saccharomyces cerevisiae genome can be obtained from http://hgdownload.cse.ucsc.edu/goldenPath/sacCer3/bigZips/sacCer3.fa.gz
2. Genome annotations in GTF format can be obtained from UCSC, ENSEMBL, NCBI or ENCODE.

2.7 Software Packages

1. UCSC Genome Browser [30, 31] utilities: http://hgdownload.cse.ucsc.edu/admin/exe/.
2. R: https://www.r-project.org/.
3. Python (version 2.7 or higher) https://www.python.org/.
4. TGL Kmeans: https://github.com/tanaylab/tglkmeans.
5. SciPy: https://www.scipy.org/.
6. Matplotlib: https://matplotlib.org/.

7. Minimap2 [32] (version 2.17) https://github.com/lh3/minimap2.
8. Tombo [33] (version 1.5) https://nanoporetech.github.io/tombo/.
9. Albacore https://nanoporetech.com/.
10. Megalodon https://github.com/nanoporetech/megalodon.
11. Guppy https://nanoporetech.com/.
12. Rerio https://github.com/nanoporetech/rerio.
13. tabix: http://www.htslib.org/doc/tabix.html (*see* **Note 8**).
14. Additional scripts: https://github.com/georgimarinov/SMAC-seq-scripts. Contains python scripts for processing and post-processing of SMACseq data used in the examples shown below.

3 Methods

The principle behind the assay and the typical SMAC-seq experimental procedure are outlined in Fig. 2. SMAC-seq consists of the following basic steps:

1. Nuclei isolation.
2. Enzymatic treatment of chromatin.
3. HMW DNA extraction.
4. Nanopore sequencing.
5. Read mapping and calling modified basis.
6. Aggregate and single-molecule accessibility analysis.

We provide several slightly different protocols for working with yeast (*see* **Note 12**) as well as with mammalian and fly cells.

3.1 Nuclei Isolation (Budding Yeast)

Start with 2.5×10^8 yeast cells (the equivalent to 1×10^6 human cells).

1. Spin cells for 1 min at 13000 rpm. Remove supernatant.
2. Wash cells with 100 μL Sorbitol Buffer.
3. Spin cells 1 min at 13000 rpm. Remove supernatant.
4. Resuspend pellet in 200 μL Sorbitol Buffer + 10 mM DTT + 0.5 mg/mL 100T Zymolase.
5. Incubate for 5 min at 30 °C, shaking 300 rpm.
6. Centrifuge for 2 min at 5000 rpm. Remove supernatant.
7. Add 100 μL SB buffer (no DTT) and resuspend gently.

8. Centrifuge for 2 min at 5000 rpm. Remove supernatant.
9. Add 100 μL ice-cold lysis buffer.
10. Incubate on ice for 10 min.
11. Spin down at 5000 rpm for 5 min at 4 °C.
12. Wash with 100 μL cold wash buffer.
13. Spin at 5000 rpm for 5 min at 4 °C.
14. Resuspend in M.CviPI Reaction Buffer (100 μL).

3.2 Enzymatic Treatment of Chromatin for m6A-GpC-CpGSMAC-Seq (Budding Yeast)

1. Add 200 U of M.CviPI and 200 U of EcoGII.
2. Add SAM to a final concentration of 0.6 mM, and sucrose to a final concentration of 300 mM.
3. Incubate at 30 °C for 7.5 min.
4. Add 128 pmol SAM (= 4 μL 32 mM solution) and another 100 U of both enzymes.
5. Incubate at 30 °C for 7.5 min.
6. Add 60 U of M.SssI.
7. Add 128 pmol SAM (= 4 μL 32 mM solution) (*see* **Note 3**).
8. Add $MgCl_2$ to a final concentration of 10 mM.
9. Incubate at 30 °C for 7.5 min.
10. Stop reaction by adding an equal volume of Stop Buffer.

3.3 Nuclei Isolation for Human, Drosophila, and Other Cells Without Cell Walls

Start with 1×10^6 diploid human cells. Scale accordingly according to genome size, variations in cell ploidy, the aimed-for amount of sequencing (*see* **Note 9**).

1. Wash cells with 1× PBS.
2. Centrifuge for 5 min at 500 × *g* at 4 °C. Remove supernatant.
3. Resuspend cells in 200 μL ice-cold Nuclei Lysis Buffer.
4. Incubate on ice for 10 min.
5. Centrifuge for 5 min at 500 × *g* at 4 °C. Remove supernatant.
6. Resuspend nuclei in 200 μL cold Nuclei Wash Buffer.
7. Centrifuge for 5 min at 500 × *g* at 4 °C. Remove supernatant.
8. Resuspend nuclei in 200 μL CutSmart Reaction buffer.

3.4 Enzymatic Treatment of Chromatin for m6A-SMAC-Seq (Human Cells)

1. Add 200 U of EcoGII.
2. Add SAM at 0.6 mM and sucrose at 300 mM.
3. Incubate at 37 °C for 10 min.
4. Stop reaction by adding SDS to a concentration of 0.2%.

3.5 Enzymatic Treatment of Chromatin for m6A-GpC-CpGSMAC-Seq (Drosophila Cells)

1. Add 200 U of M.CviPI and 200 U of EcoGII.
2. Add SAM at a final concentration of 0.6 mM and sucrose at a final concentration of 300 mM.
3. Incubate at 30 °C for 7.5 min.
4. Add 128 pmol SAM (= 4 μL 32 mM solution) and another 100 U of both enzymes.
5. Incubate at 30 °C for 7.5 min.
6. Add 60 U of M.SssI.
7. Add 128 pmol SAM.
8. Add $MgCl_2$ at 10 mM.
9. Incubate at 30 °C for 7.5 min.
10. Stop reaction by adding SDS to a concentration of 0.2%.

3.6 HMW DNA Isolation

Here we describe HMW DNA using the Qiagen MagAttract HMW DNA Kit. Many other kits/protocols can also be used with similar success.

1. Add 20 μL Proteinase K into a 2 mL tube.
2. Add 200 μL of sample.
3. Add 4 μL RNase A solution and 150 μL Buffer AL. Mix by vortexing.
4. Incubate at room temperature for 30 min.
5. Add 15 μL MagAttract Suspension G beads.
6. Add 280 μL Buffer MB and incubate at room temperature for 3 min at 1400 rpm in a Thermomixer.
7. Separate the beads on a magnetic stand, carefully and completely remove the supernatant.
8. Add 700 μL Buffer MW1 and incubate at room temperature for 1 min at 1400 rpm in a Thermomixer.
9. Separate the beads on a magnetic stand, carefully and completely remove the supernatant.
10. Add 700 μL Buffer MW1 and incubate at room temperature for 1 min at 1400 rpm in a Thermomixer.
11. Separate the beads on a magnetic stand, carefully and completely remove the supernatant.
12. Add 700 μL Buffer PE and incubate at room temperature for 1 min at 1400 rpm in a Thermomixer.
13. Separate the beads on a magnetic stand, carefully and completely remove the supernatant.
14. Add 700 μL Buffer PE and incubate at room temperature for 1 min at 1400 rpm in a Thermomixer.

15. Separate the beads on a magnetic stand, carefully and completely remove the supernatant.
16. Add 700 µL nuclear-free H_2O by slowly pipetting on the side of the tube opposite to the beads while on the magnetic stand. Do not disturb the pellet, otherwise DNA loss can ensue.
17. Remove H_2O, and repeat the H_2O wash step.
18. Add an appropriate volume of Buffer AE, that is, 100–200 µL (*see* **Note 10**).
19. Incubate at room temperature for 3 min at 1400 rpm in a Thermomixer.
20. Separate the beads on a magnetic stand, carefully transfer the supernatant to a new DNA lo-bind tube using a wide bore tip.
21. Measure DNA concentration using a Qubit dsDNA HS assay.
22. Evaluate the DNA size distribution profile on the TapeStation using the gDNA screen tape and reagents.
23. Store the DNA at 4 °C (*see* **Note 11**).

3.7 DNA Size Selection

Selection of very HMW DNA using the Circulomics Short Read Eliminator Kit is described here. Use either the SRE or the SRE XL version depending on the properties of the genome studied and the input DNA size distribution. The SRE XL version will remove fragments $\leq$40 kbp while the SRE one will eliminate fragments $\leq$25 kbp.

1. Start with a total volume of 60 µL at DNA concentration between 50 and 150 ng/µL in a 1.5 mL DNA lo-bind tube.
2. Add 60 µL of Buffer SRE or Buffer SRE XL. Mix by tapping.
3. Centrifuge at 10,000 × *g* for 30 min at room temperature.
4. Carefully remove the supernatant without disturbing the DNA pellet (note that the pellet is not visible; always place the tube with the hinge facing outward to ensure reliable positioning of the pellet at the bottom of the tube).
5. Add 200 µL of 70% EtOH (make fresh immediately before use). Do not tap or mix. Centrifuge at 10,000 × *g* for 2 min at room temperature.
6. Carefully remove the supernatant without disturbing the DNA pellet.
7. Repeat the 70% EtOH wash and centrifugation step.
8. Add at least 50 µL Buffer EB and incubate at room temperature for 20 min (*see* **Note 10**).
9. Resuspend well by tapping.
10. Measure DNA concentration using a Qubit dsDNA HS assay.

11. Evaluate the DNA size distribution profile on the TapeStation using the gDNA screen tape and reagents.
12. Store the DNA at 4 °C (*see* **Note 11**).

Example TapeStation results for poor-quality, high-quality and post-size selection HMW DNA are shown in Fig. 3.

3.8 Nanopore Library Construction and Sequencing

Carry out nanopore library construction and sequencing according to the manufacturer's instructions depending on the particular kit and flow cell/sequencer being used.

3.9 Computational Analysis

The basic processing of SMAC-seq data described here consists of the following steps:

1. Initial base calling.
2. Read mapping.
3. Generating modification calls.
4. Compilation of basic data statistics.
5. Generation of aggregate modification scores.
6. Generation of averaged coverage tracks.

Analysis at the single molecule level can be subsequently carried out.

The overall workflow is summarized in Fig. 4.

3.9.1 Read Mapping and Modification Calling

There are two different ways to extract modifications. Historically, SMAC-seq per-read modification calls were extracted using Tombo, which is a non–model-based DNA modification caller for any context. It is no longer updated and requires base calling using the older and less accurate base calling software Albacore. The state-of-art way to call modifications is Megalodon. Megalodon is a command-line tool that combines base calling using Guppy with modified base calling based on pretrained modification calling models in the Rerio package, in which all-context m6A and 5mC models are available. Because of the higher accuracy of these models and the ease of use of Megalodon, this is at present the preferred method for calling modifications.

Calling modifications with Tombo involves the following steps (run these commands for each individual fast5 file in parallel to speed up the process):

1. Base calling using Albacore. Tombo requires that reads are first base-called using Albacore. Running Albacore requires the user to specify the exact type of flow cell and the kit used to build the library, as follows:

```
read_fast5_basecaller.py --flowcell {FLOW_CELL}
--kit {RUN_KIT} -i {FAST5_DIR}
-t {NUMBER_OF_THREADS} -s {OUTPUT_DIR}
-o fastq,fast5 --disable_filtering
```

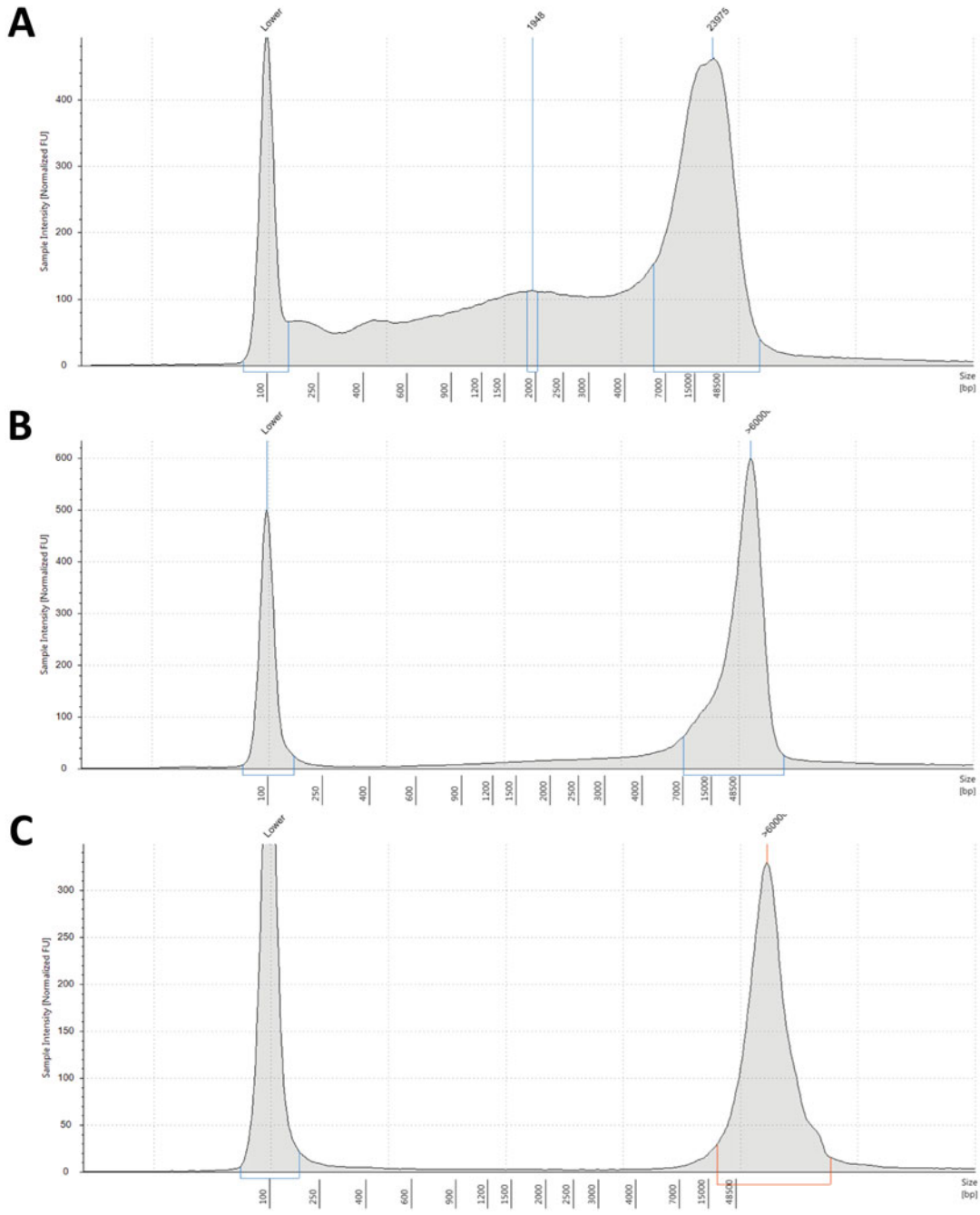


Fig. 3 HMW DNA isolation and size selection for long-read sequencing. It is of critical importance for the success of SMAC-seq experiments (and many other long read-based assays) to use high quality HMW DNA as input to sequencing. Numerous protocols exist for isolating HMW DNA and HMW DNA size selection. Shown are TapeStation gDNA profiles for a DNA sample with poor size distribution (**a**), a DNA sample with good size distribution (**b**), and a DNA sample after size selection using the Circulomics Short Read Eliminator Kit (**c**)

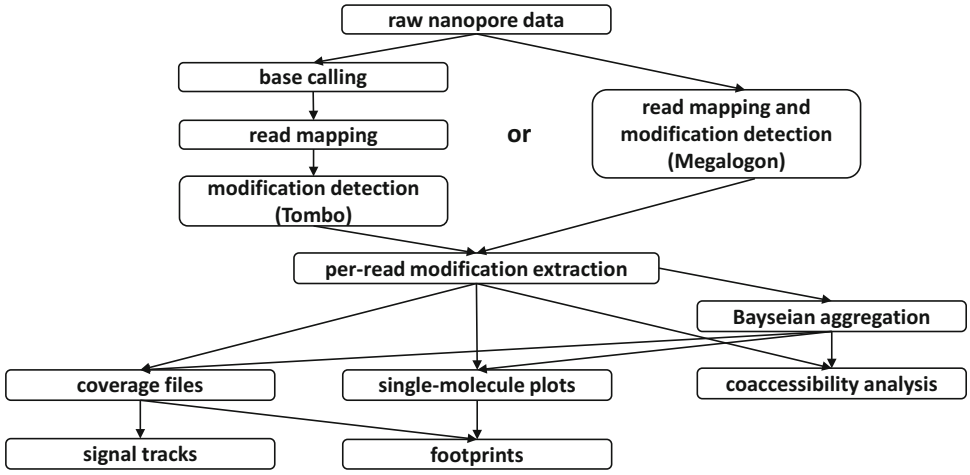


Fig. 4 Summary of the SMAC-seq analysis workflow. For Tombo processing, raw nanopore read traces are first subjected to base calling, mapped to the reference genome, and modified bases are then identified after "resquiggling" of the reads. The newer Megalodon-based processing combines these steps in one. Per-read modification calls are then extracted, and converted into a common file format that allows for downstream tasks to be carried out

2. Read preprocessing. Following base calling at the read level using Albacore, Tombo maps every read to its corresponding fast5 signal track, as follows:

```
tombo preprocess annotate_raw_with_fastqs
--processes {NUMBER_OF_THREADS} --overwrite
--fast5-basedir {FAST5_DIR}
--fastq-filenames {ALBACORE_PRODUCED_FAST5}
```

3. Tombo resquiggling. Next, the reads are mapped and nanopore signal is "resquiggled" against the reference genome as follows (note that Tombo uses minimap2 to carry out the mapping):

```
tombo resquiggle --ignore-read-locks
--processes {NUMBER_OF_THREADS} --overwrite
{FAST5_DIR} {REFERENCE_GENOME}
```

4. Tombo de novo modification calling. To call m6A and 5mC modifications in all contexts we use the de novo mode of Tombo as follows:

```
tombo detect_modifications de_novo
--statistics-file-basename {STATS_FILE_NAME}
--per-read-statistics-basename {MODS_FILE_NAME}
```

```
 --processes {NUMBER_OF_THREADS}
 --multiprocess-region-size 2000000
 --fast5-basedirs {FAST5_DIR}
```

3.9.2 Tombo Extraction

Default Tombo outputs do not include information about modification at the basepair single-molecule level. These need to be extracted using the Tombo Python API using custom-written scripts. Run the TomboSingleReadsExtract-tombo_de_novo-1.5.py script in order to convert Tombo per_read_stats files into text files. The script has multiple options for different sequence contexts, excluding certain sequence contexts, etc.:

```
 python TomboSingleReadsExtract-tombo_de_novo-1.5.py tombo.
per_read_stats genome.fa outfile_prefix
 [-m5C-only] [-m6A-only] [-CG-only] [-CG-CG-only]
 [-GC-only] [-m6A-CG-only] [-m6A-GC-only]
 [-m6A-GC-CG-only] [-doT] [-T-only]
 [-generic bases(comma-separated)]
 [-excludeContext string(...,stringN) radius]
 [-excludeChr chr1[...,chrN]]
 [-chrPrefix string]
```

Example for A positions:

```
python TomboSingleReadsExtract-tombo_de_novo.py
0.tombo.per_read_stats genome.fa 0.tombo.m6A-only
-m6A-only
```

Example for A, CpG and GpC positions:

```
python TomboSingleReadsExtract-tombo_de_novo.py
0.tombo.per_read_stats genome.fa 0.tombo.m6A-GC-CG-only
-m6A-GC-CG-only
```

Run the script for each individual `tombo.per_read_stats` file.

3.9.3 Read Mapping and Modification Calling Using Megalodon

Megalodon is run in one step as follows:

```
megalodon {LOCATION_OF_FAST5_FILES}
--guppy-params "-d {PATH_TO_RERIO_MODELS}"
 --guppy-config res_dna_r941_min_modbases-all-context_v001.
cfg
--outputs basecalls,mod_basecalls,per_read_mods
--reference {REFRENCE_GENOME}
--write-mods-text --output-directory {OUTPUT_DIR}
--guppy-server-path {LOCATION_OF_GUPPY_BIN}
```

For the purposes of downstream single-molecule analysis the `--outputs basecalls,mod_basecalls,per_read_mods` and `--write-mods-text` options need to be specified. These will result in output of per-read modifications in a text format.

3.9.4 Megalodon Per-Read Modification Extraction

To extract the per-read modification, we run the following script:

```
python megalodon-to-single_line.py *.per_read_modified_base_calls.txt
*megalodon.reads.tsv
```

Run the script for each individual Megalodon file.

3.9.5 Merging and Indexing

Merge the converted files into a single file, and sort by coordinates in the same step:

```
cat *.reads.tsv | sort -k1,1 -k2,2n -k3,3n
| bgzip > merged.reads.tsv.bgz
```

Then `tabix`-index the file:

```
tabix -s 1 -b 2 -e 3 merged.reads.tsv.bgz
```

This will create a `tabix`-index —.bgz— file in the following format, with one entry for each read:

1. Column 1: chromosome.
2. Column 2: left-most modified/informative position within the read.
3. Column 3: right-most modified/informative position within the read.
4. Column 4: . character (for legacy reasons).
5. Column 5: nanopore read ID.
6. Column 6: nan (for legacy reasons).
7. Column 7: comma-separated list of modified/informative positions.
8. Column 8: comma-separated list of Tombo probabilities, matching the order of the positions in Column 7.

3.9.6 Calculate Mapping Statistics

Calculate read mapping statistics as follows:

```
python NanoporeTSVMappingStats.py
merged.reads.tsv.bgz
NanoporeTSVMappingStats-merged
```

This will produce a short report with the total number of mapped reads, the total number of mapped bases, the mean mapped read length and the median read length.

3.9.7 Create Coverage File

While the true strength of SMAC-seq lies in the single-molecule analysis, SMAC-seq data can also be highly informative at an aggregate level, which allows for CREs and positioned nucleosomes to be discerned by visualization of average SMAC-seq profiles on a genome browser. For the purpose of such analyses, a coverage file in the style of the output from the popular bisulfite sequencing analysis tool Bismark [34] is created, using the `methylation-reads-tsv-to_coverage.py` script:

```
python methylation_reads_all.tsv threshold outfile
[-stranded +|-] [-minAbsLogLike float]
[-minAbsPValue float]
[-BayesianIntegration window(bp) step alpha beta pseudosam-
plesize] [-N6mAweight pseudosamplesize genome.fa]
[-saveNewSingleMoleculeFile filename]
```

Nanopore DNA modification data is not binary, instead it is recorded as probabilities. It thus has to be binarized at some threshold. We have found, through exploration of the parameter space and comparison to known biological truths, that the most intuitive threshold of 0.5 works optimally [22]. Example:

```
python methylation-reads-tsv-to_coverage.py
merged.reads.tsv.bgz 0.5 merged.cutoff_0.5.coverage
```

Convert the resulting plain text file to a `.bgz` file:

```
cat merged.cutoff_0.5.coverage |
bgzip > merged.cutoff_0.5.coverage.bgz
```

Then `tabix`-index it:

```
tabix -s 1 -b 2 -e 3 merged.cutoff_0.5.coverage.bgz
```

The format of the `coverage` file is as follows:

1. Column 1: chromosome.
2. Column 2: left-most position of the modified/informative sequence context.
3. Column 3: right-most position of the modified/informative sequence context.
4. Column 4: number reads in which the sequence context is methylated.

5. Column 5: number reads in which the sequence context is unmethylated.
6. Column 6: total number of reads.

3.9.8 Bayesian Integration

Even when using m6A, SMAC-seq still does not cover every single nucleotide in the genome, and coverage varies substantially between different locations depending on local sequence content differences. In addition, base calling for ONT data is still far from perfectly accurate (*see* **Note 13**), and detecting modifications is particularly challenging. On the other hand, the biologically meaningful length scale for DNA accessibility is not necessary the individual basepair, but somewhat larger sequence contexts.

For these reasons we often use aggregate accessibility scores over fixed-length windows, which combine information over all available informative positions in the window, thus providing more reliable, even if coarser-grained, views of accessibility patterns. This is done using a simple Bayesian procedure, as follows.

For a given window of width w, specified by coordinates $c,i,i+w$ (where c is the chromosome, and i is the leftmost coordinate of the window), and for all reads $r \in R_{c,i,i+w}$ fully spanning the window, we obtain all Tombo probabilities $p_{r,(c,j)}$ such that $j \in [i, i+w)$ for the assayed sequence contexts on the corresponding genomic strand (*see* **Note 14**). We usually use a Beta prior $B(\alpha,\beta)$, with $\alpha = \beta = 10$, which is updated based on each probability $p_{r,(c,j)}$ for all $j \in [i,i+w)$ (but the prior can be easily changed if necessary, *see* below), in order to obtain a final accessibility score $p_{r,(c,i,i+w)}$ for read r and window $c,i,i+w$.

This Bayesian integration calculation is also carried out using the same `methylation-reads-tsv-to_coverage.py` script. For efficiency of calculation, compute it in parallel on the individual converted tombo files, as follows (for a 10-bp context and (10,10) prior):

```
python methylation-reads-tsv-to_coverage.py
0.tombo 0.5
0.tombo.all0.cutoff_0.5.coverage.BI_w10_a10_b10
-minAbsPValue 0.4 -BayesianIntegration 10 1 10 10 50
-saveNewSingleMoleculeFile
0.tombo.BI_w10_a10_b10.reads.tsv
```

Merge the Bayesian integration files:

```
cat *tombo.BI_w10_a10_b10.reads.tsv
| sort -k1,1 -k2,2n -k3,3n
| bgzip > merged.BI_w10_a10_b10.reads.tsv.bgz
```

Then `tabix`-index the resulting .bgz file:

```
tabix -s 1 -b 2 -e 3 merged.BI_w10_a10_b10.reads.tsv.bgz
```

3.9.9 Filtering Fully Methylated Reads

On occasions, we observe a population of reads that appear as fully methylated across their whole length or over large segments of it. They are most likely derived from dead cells or represent some other undesired artefact. In order to remove such potentially artefactual reads, we obtain a "filtered" read set by removing all reads containing a $\geq$1-kbp stretch that is $\geq$75% methylated (while also filtering out reads shorter than 1 kb).

This operation can be carried out using the `filterFullyMethylatedReads.py` script as follows:

```
 python filterFullyMethylatedReads.py methylation_reads_all.
tsv WindowSize minFraction
 [-keepShort] [-missingBasesFilter genome.fa basecontexts(com-
ma-separated) minFraction
 [-doMBFSet]]
```

3.9.10 Create Genome Browser Tracks

In order to create average-methylation (and thus accessibility) tracks that can be visualized on a genome browser such the UCSC or the WashU ones, use the following script:

```
python coverage_to_wig.py coverage.bgz window step chrField
MfieldID UfieldID chrom.sizes outprefix [-minCov N_reads]
```

where the M and the U fields indicate the column IDs of the numbers of methylated and unmethylated reads, respectively, and the window and step parameters specify the width and the stride used for averaging the signal (i.e., window of 50 and step of 5 means that the average methylation level over 50 bp windows tiling the genome every 5 bp will be outputted).

This script will output two `bedGraph` files—a `coverage.wig` one (which contains the number of reads covering a position) and a `meth.wig` one (which contains the fraction of methylated reads). These can then be converted into `bigWig` files that can in turn be displayed on a genome browser using the `wigToBigWig` program from the UCSC utilities:

```
wigToBigWig meth.wig chrom.sizes meth.wig
```

where the chrom.sizes files contains one line per chromosomes including the chromosome name and its length in bp (tab-separated).

An example of an average SMAC-seq profile is shown in Fig. 5.

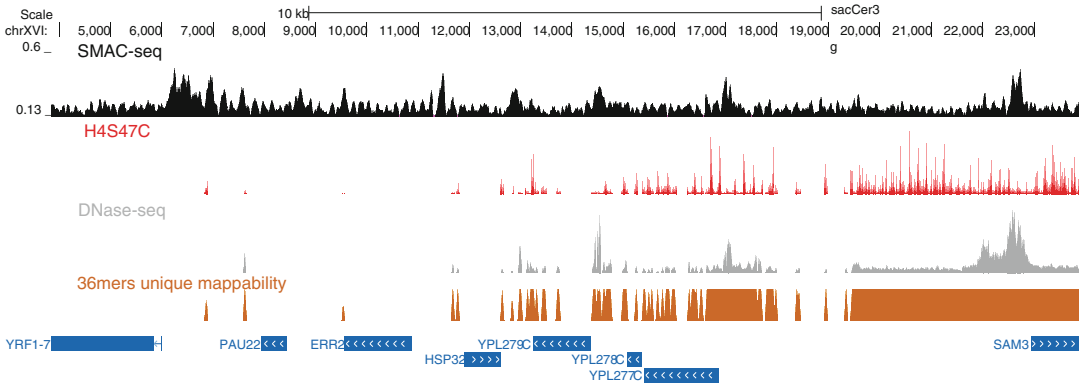


Fig. 5 Examples of average m6A-CpG-GpC-SMAC-seq profiles visualized on the UCSC Genome Browser. Shown is a subtelomeric regions on chrXVI. SMAC-seq signal provides information both about accessible open chromatin measures (peaks in DNAse-seq data) and positioned nucleosomes. The latter are shown here in the form of H4S47C chemical nucleosome mapping [35], which maps the positions of dyads (SMAC-seq signal is enriched on nucleosome linkers, thus the inverse relationship between the two). SMAC-seq, being a long-read assay, also provides information about repetitive regions of the genome (in this case, telomeres, which are not uniquely mappable with short reads as shown by the 36-mer unique mappability track)

3.9.11 Making Metaplots Around a Position

A common analysis task is to generate a metaplot around a given set of genomic features (such as TSSs, positioned nucleosomes, TF binding motifs, and others). The `coverage.bgz` can be used to make such metaplots, as follows, with a variety of parameters (window size, minimal coverage per position, different input file formats, stranded or unstranded, and others):

```
python signalAroundPeaks-nano.py inputfilename chrFieldID
posField strandField radius window coverage.bgz outputfile-
name [-bismark.cov] [-bed] [-minCov N]
[-unstranded] [-narrowPeak]
```

Examples of such plots around yeast transcription starts sites and human occupied CTCF motifs are shown in Fig. 6.

3.9.12 Making Single Molecule Plots

One of the two key strengths of SMAC-seq is the ability to analyze accessibility at the single molecule level. There are many ways to do that, due to the nonbinary nature of raw nanopore data and of the long length of nanopore reads, which allows for/requires analysis at different resolution levels. Single molecule maps can be generated using the continuous modification probability values or they can be binarized.

The `SMAC-footprints-from-methylation-reads-tsv-tabix.py` and `SMAC-footprints-from-methylation-reads-tsv-tabix-kmeans.py` scripts can be used to generate such plots. The first script will apply hierarchical clustering while the second one will use *k*-means (in our experience, we obtain

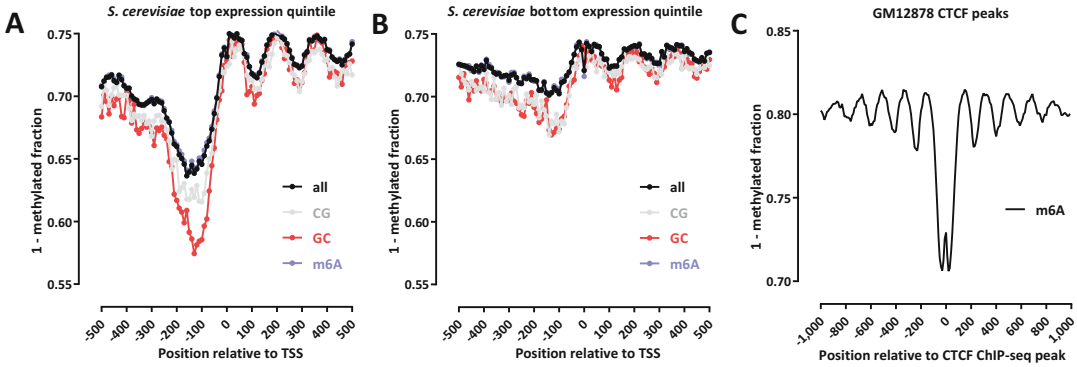


Fig. 6 Examples of average SMAC-seq metaprofiles over predefined genomic features. (**a** and **b**) Average m6-CpG-GpC SMAC-seq profiles for the top 20% and bottom 20% of genes (ranked by expression levels) in *S. cerevisiae*. Profiles are split by modification channel. (**c**) Average m6-SMAC-seq profile around CTCF ChIP-seq peaks in the human GM12878 cell line. CTCF is known to strongly position nucleosomes in the vicinity of its occupancy sites [36]. ChIP-seq peaks were obtained from the ENCODE Project Consortium [29]

decidedly better results using the *k*-means approach). The commands are otherwise the same. There is a wide variety of options regarding the input list of region (which can be in any format), the display (averaging over arbitrary number of basepairs), subsampling of reads, color schemes, binarization or continuous display, and others:

```
python methylation_reads_all.tsv peak_list chrFieldID
leftFieldID rightFieldID strandFieldID tabix_path outfile_-
prefix [-resize factor] [-subset N] [-label fieldID] [-minCov
fraction]
[-minPassingBases fraction] [-minReads N]
[-unstranded] [-minAbsLogLike float]
[-scatterPlot colorscheme minScore maxScore color|none]
[-window bp] [-readStrand +|-]
[-printMatrix] [-deleteMatrix] [-binarize threshold]'
```

The following command will generate binarized single molecule maps retaining only reads that completely span the input set of regions, averaging over 10 bp windows:

```
python SMAC-footprints-from-methylation-reads-tsv-tabix-
kmeans.py
SMAC-seq.reads.tsv.bgz regions.bed 0 1 2 3 tabix
SMAC-seq.regions.binary-0.5-gist_heat.10bp
-window 10 -minCov 1 -binarize 0.5
-scatterPlot gist_heat 0 1.1 w -unstranded
```

An example of such a single-molecule level visualization for yeast m6ACpG-GpC-SMAC-seq data is shown in Fig. 7a.

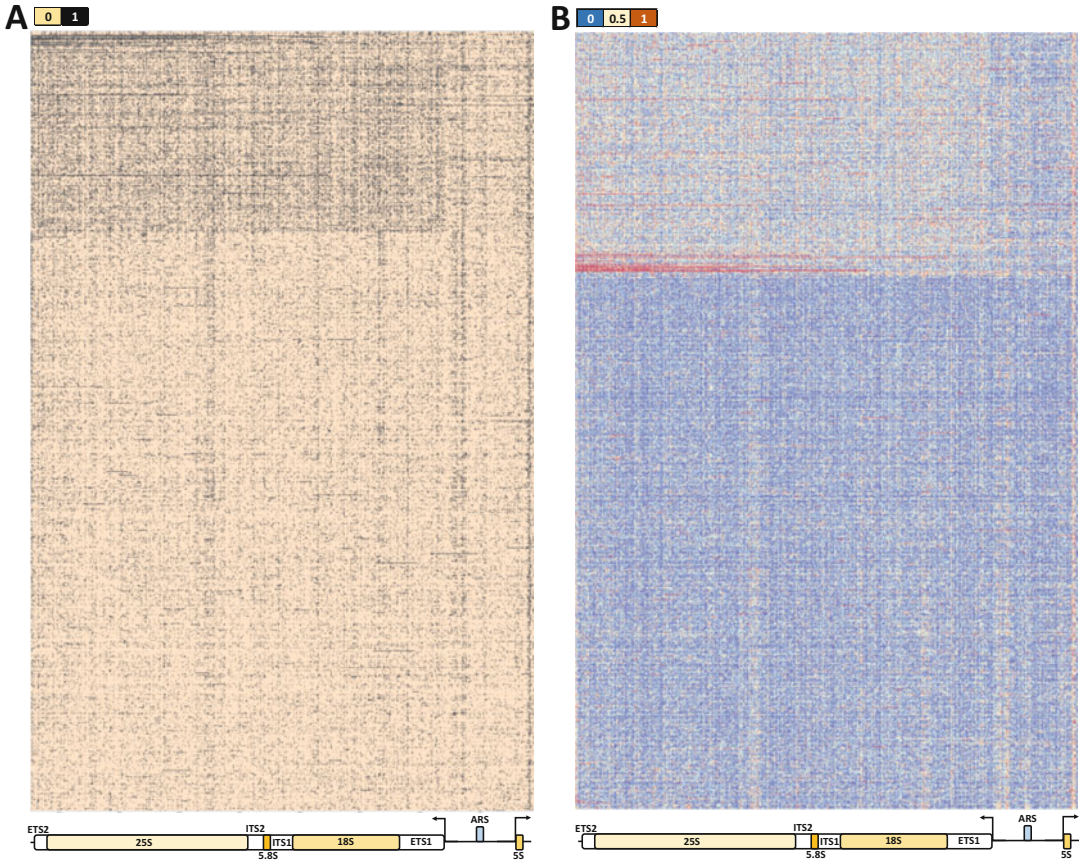


Fig. 7 Examples of single molecule m6A-CpG-GpC-SMAC-seq maps in *S. cerevisiae*. Shown is the yeast rDNA locus, binarized (**a**) and as a continuous display (**b**). Yeast rDNA is organized into multicopy (~150) arrays, consisting of ~9.1 kb units, each containing a copy of the 35S precursor pre-rRNA, transcribed by Pol I, a 5S RNA, transcribed by Pol III, and a replication origin ARS element, located in nontranscribed (NTS) regions of the array. The rDNA chromatin structure adopts two distinct conformations [37, 38]—an inactive nucleosomal state and an extremely highly transcriptionally active, largely devoid of nucleosomes (and thus highly accessible) state. Note that 1000 reads were sampled at random for each plot, and that different samplings are shown in (**a**) and (**b**)

The following command will generate continuous-signal single molecule maps retaining only reads that completely span the input set of regions, averaging over 10 bp windows:

```
 python SMAC-footprints-from-methylation-reads-tsv-tabix-
kmeans.py
 SMAC-seq.reads.tsv.bgz regions.bed 0 1 2 3 tabix
 SMAC-seq.regions.binary-0.5-RdYlBu.10bp
 -window 10 -minCov 1
 -scatterPlot RdYlBu 0 1 w -unstranded
```

An example of such a single-molecule level visualization for yeast m6ACpG-GpC-SMAC-seq data is shown in Fig. 7b.

3.9.13 Calculating NMI Matrices

Finally, another common analysis task when working with SMAC-seq data is to estimate the degree of single-molecule coaccessibility along the chromatin fiber.

To this end, we apply a Normalized Mutual Information as follows. Each chromosome c is split into windows of size w. For each such window $(c,i,i+w)$, the maximum range to the right of it, $(c,j,j+w)$ such that the span $(c,i,j+w)$ is covered by $\geq M$ reads, is identified. All reads spanning $(c,j,j+w)$ are then extracted and subsampled down to M reads (usually $M = 100$). Accessibility scores are then aggregated and binarized for all windows located in the span $(c,j,j+w)$, and for all M reads fully spanning it, resulting in a local coaccessibility matrix LCM of size $M \times (j+w-i)/w$. A Normalized Mutual Information (NMI) score for each pair of columns LCM_k and LCM_l is then calculated as follows:

$$\begin{aligned}\mathrm{MI}(\mathrm{LCM}_k,\mathrm{LCM}_l) = {} & p(0,0)\log_2\left(\frac{p(0,0)}{p_k(0)\,p_l(0)}\right)\\ & +p(1,1)\log_2\left(\frac{p(1,1)}{p_k(1)\,p_l(1)}\right)\\ & +p(0,1)\log_2\left(\frac{p(0,1)}{p_k(0)\,p_l(1)}\right)\\ & +p(1,0)\log_2\left(\frac{p(1,0)}{p_k(1)\,p_l(0)}\right)\end{aligned} \tag{1}$$

While, in principle, mutual information cannot be negative, NMI scores are normalized and rescaled in the interval $(-1,1)$ so that anticorrelated regions are given negative scores (this is done for visualization and interpretation purposes):

$$\mathrm{NMI}(\mathrm{LCM}_k,\mathrm{LCM}_l) = \begin{cases} \dfrac{\mathrm{MI}(\mathrm{LCM}_k,\mathrm{LCM}_l)}{\sqrt{H(\mathrm{LCM}_k)H(\mathrm{LCM}_l)}} & \text{for } p(0,0)+p(1,1) \geq 0.5 \\ -\dfrac{\mathrm{MI}(\mathrm{LCM}_k,\mathrm{LCM}_l)}{\sqrt{H(\mathrm{LCM}_k)H(\mathrm{LCM}_l)}} & \text{for } p(0,0)+p(1,1) < 0.5 \end{cases} \tag{2}$$

where H refers to the entropy of each individual distribution.

To calculate NMI matrices, the `SingleMoleculeCorrelation-NMI-matrix.py` script can be used:

```
python SingleMoleculeCorrelation-NMI-matrix.py
 SMAC-seq.reads.tsv.bgz regions.bed chrFieldID leftField
rightFieldID minCoverage windowsize stepsize tabix_location
outfileprefix
 [-subsample N] [-expectedMaxDist bp] [-label fieldID]
```

Example:

```
python SingleMoleculeCorrelation-NMI-matrix.py
SMAC-seq.reads.tsv.bgz regions.bed 0 1 2 50 1 1200 tabix
NMI.min50cov.1bp.regions.SMAC-seq -expectedMaxDist 1500
```

If running genome-wide, split the genome into overlapping bins for parallelization efficiency, for example, 50 kb in size with a 10 kb stride, and calculate a separate matrix for each, then take the average NMI values for each pair of coordinates for downstream analyses.

An example of the results of NMI analysis for yeast m6A-CpG-GpCSMAC-seq data is shown in Fig. 8.

4 Notes

1. EcoGII deposits m6A modifications without a sequence preference, but it does not do so with perfect efficiency. It is not conclusively established why, that is, whether the presence of neighboring already modified bases prevents further methylation or whether perhaps the enzyme is highly nonprocessive and stays bound to DNA for a prolonged period after completion of the reaction, thus occluding neighboring bases from further enzymatic action. The methylation efficiency reported by NEB is ~50%; however, this is based on a relatively short treatment (5 min). On the other hand, based on the original more detailed study describing EcoGII [28], methylation efficiency seems to be closer to 80% for a prolonged treatment of about an hour. The incubation times during a SMAC-seq experiment would place the expected efficiency somewhere in between these values. Unfortunately, the most straightforward imaginable experiment that would properly establish EcoGII methylation efficiencies in the context of a nanopore-based experiment—nanopore sequencing of naked DNA subjected to EcoGII treatment—is not in fact possible because of the strong bias of the Oxford Nanopore platform against fully methylated templates, which simply do not sequence well and are mostly discarded.
2. EcoGII is commercially available as a solution at relatively low concentration, and if sufficiently many units of it are to be used, the volume needed becomes too large and could interfere with the labeling reaction. For these reasons, we are using a custom-made highly concentrated EcoGII from NEB.
3. SAM is unstable. This is one of the reasons why it is added twice to the reaction, and it is also why it should be handled carefully, avoiding repeated freeze-thaw cycles.

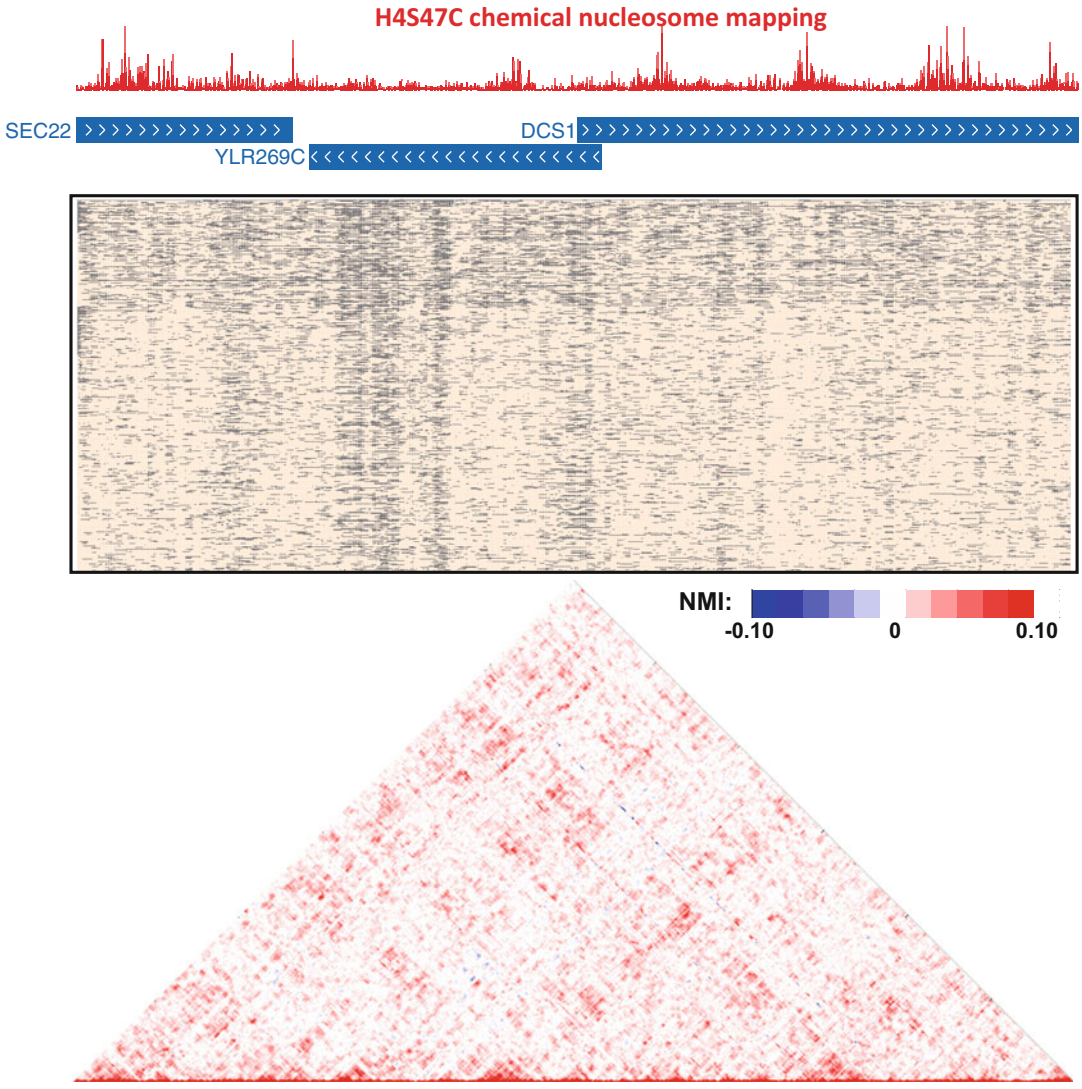


Fig. 8 Example of m6A-CpG-GpC-SMAC-seq coaccessibility maps (measured in terms of NMI) in *S. cerevisiae*. Shown is the promoter region of the DCS1 gene, together with chemical mapping nucleosome positioning data (top), the single-molecule SMAC-seq map (middle), and the coaccessibility map (bottom)

4. ONT offers multiple sequencing platforms, and it is advisable to be familiar with the properties of each and the tradeoffs between them. The MinION is the main ONT sequencing platform, typically generating ~10 Gbp data (although higher throughput runs have also been observed in practice, up to 20 Gbp) and several million reads. In the context of SMAC-seq, a single MinION is often sufficient to generate adequate coverage over a small genome such as that of yeasts.

The Flongle is a miniaturized flow cell that also runs on the MinION instrument, typically generating ~100,000 reads. It is not sufficient for production-scale runs, but as it is priced at ~1/9 of the cost of a MinION flow cell, it is ideal for testing protocol, carrying out QC runs, etc.

The GridION can use either MinION or Flongle flow cells and run five of them in parallel.

The PromethION is the high-throughput ONT sequencer. It uses different flow cells, each of which can generate up to ~100 Gbp of data (and more than ten million reads), and can run 48 such flow cells in parallel at the same time. Each such flow cell is priced at more than twice the cost of a MinION flow cell. To study larger and more complex eukaryotic genomes using SMAC-seq, the throughput of the PromethION becomes necessary, and often multiple such flow cells are needed.

5. It is important to note that "coverage" means very different things in the contexts of genome sequencing and SMAC-seq. Usually, "coverage" refers to how many reads cover a given position in the genome on average. However, the more relevant metric for SMACseq is instead "coverage at length L," that is, how many reads cover two position spread apart at a given instance. One of the main goals of SMAC-seq is to capture the coordinated behavior of distal CREs and this is only possible when sufficiently many single molecules containing both CREs have been sequenced. Across eukaryotes a clear trend is observed—as genome size increases, CREs become spread apart more and more. Thus, while 2 yeast promoters are on average 1.5–2 kb apart, the distance between an enhancer and its cognate promoter in a mammalian genome is often tens of kilobases. Thus, the required sequencing throughput to achieve the same effective "coverage at length *L*" does not scale linearly once the fact that even with careful size selection there are still many more shorter nanopore reads than very long ones is taken into account.
6. Low-binding tubes are preferable in order to minimize DNA loss.
7. The sheer volume of nanopore sequencing data presents a different level of challenge in terms of computational infrastructure compared to short-read sequencing. A single PromethION flow cell can produce 100 Gbp of data within 48 h, and a PromethION instrument can in principle run 48 such flow cells in parallel.

   However, base calls are far from the only information that needs to be stored. For analysis of SMAC-seq datasets (and of DNA modifications in general), the nanopore current signal itself is what is most important, as it is used during the

resquiggling and DNA modification detection steps. Thus, the actual disk space footprint of such a flow cell is between one and two orders of magnitude higher than storing the base calls alone.

In addition, a separate challenge has historically been posed by the number of files. This has changed with more recent versions of the ONT processing software, but historically ONT data has been stored in a large number of individual small files, which could be so large that it reached the limit on the number of files per use that many shared computational clusters have in place, necessitating sequential processing of datasets in batches and cleaning of files in between each.

8. The files containing single-molecule SMAC-seq information can be huge in size, surpassing 1 TB on occasions. Random access is critical for downstream analysis to be practical. The workflows described here achieve this by using `tabix` indexing of coordinate-sorted files.
9. Nanopore sequencing involves no amplification of DNA while having strict constraints on the minimum amount of DNA that is to be used as input to each sequencing run. A typical PromethION run uses at least 1 μg of DNA, but if size selection is to be applied prior to it, this corresponds to several times more input DNA per run. A typical diploid human cell contains ~6 pg of DNA, thus 1×10^6 cells contain ~6 μg of DNA. Multiple PromethION runs are required to obtain good coverage for a mammalian-sized genome, thus tens of micrograms of DNA are needed as input to size selection and then sequencing. Scale up reactions accordingly based on the specifics of the experiment with these considerations in mind.
10. Elution volumes are important for nanopore sequencing. All ONT sequencing kits have a minimum requirement for the amount of input DNA but also a maximum limit to the volume in which it is contained. Concentrating DNA using beads will result in significant losses while doing so by evaporation leads to its degradation. Thus, it is best to have a large amount of DNA in a small volume. However, there is a trade-off between the elution volume and the efficiency of elution—larger elution volumes lead to better overall yields. Thus, the optimal elution volume is to be decided based on the number of cells used for the SMAC-seq reaction and the exact ONT kits that are to be used for sequencing.
11. HMW DNA is stable for a long time at 4 °C, but it is strongly recommended not to freeze it at −20 °C or −80 °C as this will likely result in fragmentation. Also, highly concentrated HWM DNA can sometimes precipitate out of solution after prolonged storage so make sure to inspect tubes before use.

Resuspend by tapping the tubes with your fingers, do not pipette up and down as this is also thought to lead to HMW DNA degradation. In addition, always transfer HMW DNA using wide bore tips to prevent shearing.

12. Yeast (and fungal cells in general) have thick cell walls comprised of polysaccharides, lipids and chitin in various proportions. They present a barrier to the access of most enzymes to the nucleus, thus protocols tailored to such cells involve treatment with zymolyase or chitinase enzymes [39], with the exact details varying depending on the species studied.
13. Nanopore sequencing is a powerful tool for detecting DNA modifications, but discerning modified bases from raw nanopore signal is not yet a fully resolved problem, especially for methylation modifications, which do not provide a huge shift in current signal relative to the unmodified base. Detection of m6A is more challenging that detection of m5C, possibly because a single methyl group changes the overall properties of a purine base to a lesser extent than it does for a pyrimidine. In addition, it should be noted that current implementations of nanopore sequencing do not actually read out a single bases at a time. Instead, they read several bases at a time and the problem of base calling and modification detection is solved not in the small space of bases but in the much larger space of k-mers of size 5 or 6. Base calling errors are therefore at present an unavoidable part of the reality of dealing with nanopore datasets.

    In our experience, the error rate for calling m6A at the level of a single base within a single molecule in the context of SMAC-seq is in the 20–25% range, while that for m5C is somewhere around 15%. However, we expect the performance to improve significantly in the future through a combination of computational and experimental approaches.
14. Unlike CpG and GpC sequence contexts, which are symmetric, and therefore bases that are to be modified are present at the same position on both strand, m6A provides different information on the forward and reverse strand, as it is not a symmetric sequence context. This is a partial limitation of m6A-SMAC-seq, because different profiles can be generated from the two strands in some situations.

Acknowledgments

The authors thank members of the Greenleaf and Kundaje labs for many helpful discussions. This work was supported by NIH grants UM1HG009436 and P50HG007735 (to W.J.G.). W.J.G. is a Chan Zuckerberg investigator. Z.S. is supported by EMBO

Long-Term Fellowship EMBO ALTF 1119-2016 and by Human Frontier Science Program LongTerm Fellowship HFSP LT 000835/2017-L. G.K.M. was supported by the Stanford School of Medicine Dean's Fellowship.

References

1. Wu C (1980) The 50 ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature 286(5776):854–860
2. Keene MA, Corces V, Lowenhaupt K et al (1981) DNase I hypersensitive sites in Drosophila chromatin occur at the 50 ends of regions of transcription. Proc Natl Acad Sci U S A 78:143–146
3. McGhee JD, Wood WI, Dolan M et al (1981) A 200 base pair region at the 50 end of the chicken adult β-globin gene is accessible to nuclease digestion. Cell 27:45–55
4. Dorschner MO, Hawrylycz M, Humbert R et al (2004) High-throughput localization of functional elements by quantitative chromatin profiling. Nat Methods 1:219–225
5. Sabo PJ, Humbert R, Hawrylycz M et al (2004) Genome-wide identification of DNaseI hypersensitive sites using active chromatin sequence libraries. Proc Natl Acad Sci U S A 101:4537–4542
6. Sabo PJ, Kuehn MS, Thurman R et al (2006) Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. Nat Methods 3:511–518
7. Crawford GE, Holt IE, Whittle J et al (2006) Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res 16:123–131
8. Boyle AP, Davis S, Shulha HP et al (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132(2):311–322
9. Thurman RE, Rynes E, Humbert R et al (2012) The accessible chromatin landscape of the human genome. Nature 489(7414):75–82
10. Buenrostro JD, Giresi PG, Zaba LC et al (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
11. Buenrostro JD, Wu B, Litzenburger UM et al (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. Nature 523(7561):486–490
12. Cusanovich DA, Daza R, Adey A et al (2015) Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348(6237):910–914
13. Chereji RV, Eriksson PR, Ocampo J, Clark DJ (2019) DNA accessibility is not the primary determinant of chromatin-mediated gene regulation. bioRxiv:639971
14. Ponnaluri VKC, Zhang G, Estève PO et al (2017) NicE-seq: high resolution open chromatin profiling. Genome Biol 18(1):122
15. Umeyama T, Ito T (2017) DMS-seq for in vivo genome-wide mapping of protein-DNA interactions and nucleosome centers. Cell Rep 21(1):289–300
16. Timms RT, Tchasovnikarova IA, Lehner PJ (2019) Differential viral accessibility (DIVA) identifies alterations in chromatin architecture through large-scale mapping of lentiviral integration sites. Nat Protoc 14(1):153–170
17. Kelly TK, Liu Y, Lay FD et al (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. Genome Res 22(12):2497–2506
18. Krebs AR, Imanci D, Hoerner L, Gaidatzis D et al (2017) Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. Mol Cell 67(3): 411–422.e4
19. Vaisvila R, Ponnaluri VKC, Sun Z et al (2019) EM-seq: detection of DNA methylation at single base resolution from picograms of DNA. bioRxiv:2019.12.20.884692
20. Simpson JT, Workman RE, Zuzarte PC et al (2017) Detecting DNA cytosine methylation using nanopore sequencing. Nat Methods 14: 407–410
21. Rand AC, Jain M, Eizenga JM et al (2017) Mapping DNA methylation with highthroughput nanopore sequencing. Nat Methods 14: 411–413
22. Shipony Z, Marinov GK, Swaffer MP et al (2020) Long-range single-molecule mapping of chromatin accessibility in eukaryotes. Nat Methods 17:319–327
23. Wang Y, Wang A, Liu Z et al (2019) Singlemolecule long-read sequencing reveals the chromatin basis of gene expression. Genome Res 29:1329–1342

24. Aughey GN, Estacio Gomez A, Thomson J et al (2018) CATaDa reveals global remodelling of chromatin accessibility during stem cell differentiation in vivo. eLife 7:e32341
25. Schones DE, Cui K, Cuddapah S et al (2008) Dynamic regulation of nucleosome positioning in the human genome. Cell 132:887–898
26. Hesselberth JR, Chen X, Zhang Z et al (2009) Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat Methods 6(4):283–289
27. Neph S, Vierstra J, Stergachis AB et al (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. Nature 489:83–90
28. Murray IA, Morgan RD, Luyten Y et al (2018) The non-specific adenine DNA methyltransferase M.EcoGII. Nucleic Acids Res 46:840–848
29. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74
30. Kuhn RM, Haussler D, Kent WJ (2013) The UCSC genome browser and associated tools. Brief Bioinform 14:144–161
31. Kent WJ, Zweig AS, Barber G et al (2010) BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics 26: 2204–2207
32. Li H (2016) Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32(14):2103–2110
33. Stoiber MH, Quick J, Egan R, Lee JE, Celniker SE, Neely R, Loman N, Pennacchio L, Brown JB (2017) De novo identification of DNA modifications enabled by genomeguided nanopore signal processing. bioRxiv:094672
34. Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27(11): 1571–1572
35. Brogaard K, Xi L, Wang JP, Widom J (2012) A map of nucleosome positions in yeast at basepair resolution. Nature 486(7404):496–501
36. Fu Y, Sinha M, Peterson CL, Weng Z (2008) The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. PLoS Genet 4(7): e1000138
37. Conconi A, Widmer RM, Koller T, Sogo JM (1989) Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. Cell 57(5):753–761
38. Goetze H, Wittner M, Hamperl S, Hondele M, Merz K, Stoeckl U, Griesenbeck J (2010) Alternative chromatin structures of the 35S rRNA genes in Saccharomyces cerevisiae provide a molecular basis for the selective recruitment of RNA polymerases I and II. Mol Cell Biol 30(8):2028–2045
39. Schep AN, Buenrostro JD, Denny SK et al (2015) Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res 25:1757–1770

Part IV

Genome Structure and Organization

Chapter 16

Circular Chromosome Conformation Capture Sequencing (4C-Seq) in Primary Adherent Cells

Judith Marsman, Robert C. Day, and Gregory Gimenez

Abstract

The three-dimensional structure of the genome is highly organized and is an important aspect of gene regulation. Chromatin interactions can be identified using chromosome conformation capture-based techniques, which rely on proximity ligation. Of these techniques, circular chromosome conformation capture sequencing (4C-seq) is used to identify all chromatin interactions occurring with a single chromosomal location (one versus all). Here we describe a 4C-seq protocol that has been optimized for primary adherent cells, for which the first digestion step is inefficient using standard 4C-seq protocols. It can, however, also be applied to other cell or tissue types. This protocol utilizes a standard DNA library preparation method using a commercial kit, and includes a description of the data processing steps.

Key words 4C-seq, Chromosome conformation capture, Genome organization, Chromatin architecture, Next-generation sequencing

1 Introduction

Chromatin interactions can be detected using chromosome conformation capture (3C)-based techniques, which use proximity ligation of crosslinked chromatin to capture regions that are in close proximity. 3C-based techniques can be applied to identify chromatin interactions on several scales, ranging from the detection of DNA–DNA contacts between several a priori selected chromosomal locations by 3C (one versus one), to capturing all interactions on a genome-wide scale by Hi-C (all versus all) [1]. Here, a protocol for 4C-sequencing (4C-seq) is described, which is used to capture all interactions anchored at a single chromosomal location of interest (one versus all) [2, 3]. 4C-seq can generate high-resolution chromatin interaction profiles viewed from a genomic location(s) of interest, and is therefore useful for studying local chromatin structure at one or several genomic locations. While

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_16,

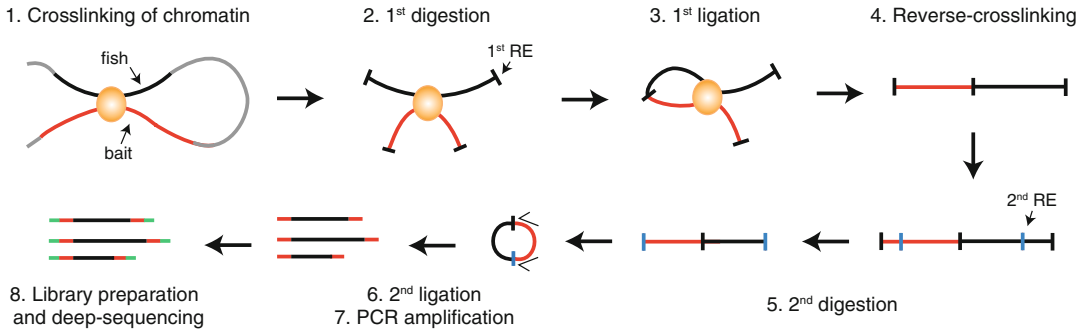


Fig. 1 Experimental steps of 4C-seq. The bait (DNA region of interest to view interactions from) is indicated in red, and the fish (captured fragments) in black. The following steps are performed: (1) chromatin is crosslinked to preserve DNA loops that are held together by proteins (orange circles); (2) the DNA is digested with the primary restriction enzyme (RE), and (3) ligated in dilute conditions; (4) chromatin is reverse–crosslinked; (5) a second digestion is performed with the secondary restriction enzyme (blue lines); (6) a second ligation is performed to circularize fragments, followed by (7) PCR amplification using inverted 4C-seq primers; (8) the resulting 4C-PCR fragments are used as input for library preparation during which sequencing adapters (green) are attached, and are deep-sequenced

technically this is also possible with Hi-C, for larger genomes a very high sequencing depth is required to achieve the same resolution as can be achieved with 4C-seq.

The 4C-seq protocol described here involves the following steps (*see* Fig. 1): (1) crosslinking of chromatin to preserve DNA–DNA interactions, (2) lysis and restriction enzyme digestion, (3) proximity ligation, in which chromatin that is in close proximity has a higher chance of being ligated together compared to chromatin that is further away, (4) reverse–crosslinking and purification of DNA, (5) a second restriction enzyme digestion to reduce fragment sizes, (6) a second ligation and purification of DNA, (7) PCR using bait-specific inverse primers to amplify fragments interacting with your bait of interest (fish), and (8) library preparation and deep sequencing.

This protocol is specifically optimized for primary adherent cells, for which an efficient restriction enzyme digestion is more difficult to achieve compared to other cell or tissue types. However, the protocol could be applied to any cell or tissue type with or without this issue. Previously published 4C-seq protocols worked well for us on suspension cells and some adherent cell types, but were not effective for primary adherent cells (such as human umbilical vein endothelial cells and human aortic smooth muscle cells) [4, 5]. With these cell types, the first digestion step that is crucial for producing a successful 4C-seq library was very inefficient. This might be due to these cell types having a tougher cell and/or nuclear structure, making it more difficult for primary restriction enzymes to access the DNA of crosslinked chromatin. In the described protocol, the lysis and primary restriction enzyme

digestion steps have been optimized, while the rest of the protocol remains similar to previously published 4C-seq protocols. Note that this protocol was not tested on all types of primary adherent cells, and that previously published 4C-seq protocols may work fine for certain primary adherent cell types.

In addition to changes in the lysis and primary digestion steps, the sequencing library preparation method described in this protocol is different from previously published 4C-seq protocols, in that it uses a standard DNA library preparation kit. In previously published 4C-seq protocols, 4C-PCRs are performed with inverse primers containing partial or full 5′ adapter-overhangs [6, 7]. In our hands, this methodology resulted in a different PCR library composition compared to adapters without 5′ overhangs when running PCR fragments on DNA gels (not just a shift in size caused by adapter-overhangs), although we have never compared this using deep-sequencing. To prevent PCR biases occurring because of 5′ adapter overhangs on primers, here, the PCR is performed with standard primers containing no 5′ adapter overhang. To produce sequencing libraries, a standard DNA library preparation kit that can be purchased commercially is used to ligate sequencing adapters. This involves (1) end-repair and A-tailing (depending on the library preparation kit used), (2) ligation of sequencing adapters, and (3) an adapter PCR of 7 cycles. The disadvantage of this method is that adapters will also ligate onto the DNA input used for the 4C-PCRs, leading to sequencing of 4C input fragments. These "background" reads comprise around 10–25% of all reads obtained after sequencing, and will be filtered out during the demultiplexing of reads based on bait-specific primers. Because a proportion of the reads will be discarded, a slightly higher sequencing depth (at least 1.25 million reads per bait) is needed when using this library preparation method, as opposed to 1 million recommended in previously published 4C-seq protocols [7]. We recommend using the described library preparation method if 4C-PCRs with a 5′ adapter overhang results in a different fragment library composition than 4C-PCRs without adapter overhang.

4C-PCR fragments from multiple baits can be pooled together before library preparation, as they can later be demultiplexed based on bait-specific 4C primer sequences contained within the reads. 4C-PCR fragments obtained from multiple samples (replicates and different biological conditions) can also be pooled together by using different barcodes contained within the sequencing adapters for each sample. In this protocol, pooled libraries consisting of multiple baits per library, are first sequenced on a small scale (MiSeq) to check if each bait is represented equally, which is often not the case. Based on the MiSeq run, ratios of bait-specific 4C-PCR inputs are adjusted, and libraries are prepared again. Libraries are then deep-sequenced paired-end on a larger platform (e.g., HiSeq), but can be sequenced single-end too.

We also describe the data processing steps, using available tools that can be run with a basic knowledge of Linux/Unix and R [8]. The steps include (1) demultiplexing of baits using bait-specific primer sequences, (2) selection of paired-end reads containing the forward and reverse 4C-seq primers in the correct orientation (only when performing paired-end sequencing), (3) adapter and quality trimming, (4) trimming of bait sequences, and (5) mapping to the genome (*see* Fig. 2). The mapped reads can then be used to generate read counts per digestion fragment, and for statistical analysis, for which several R packages are available that have been described elsewhere [9–13].

2 Materials

Prepare all solutions using ultrapure water and analytical grade reagents.

2.1 Solutions, Reagents, and Consumables

1. Fetal calf serum (FCS).
2. 1× phosphate buffered saline (PBS).
3. 4% buffered formaldehyde solution in PBS, pH 6.9 (*see* **Note 1**).
4. Lysis buffer: 10 mM Tris–HCl pH 8.0, 10 mM NaCl, 0.2% NP-40, 1 tablet protease inhibitors per 10 mL (*see* **Note 2**).
5. 2.5 M glycine.
6. 10% sodium dodecyl sulphate (SDS) in water.
7. 15% Triton X-100 in water (v/v).
8. Restriction enzymes and corresponding restriction enzyme buffers.
9. 10 mM Tris–HCl pH 7.5.
10. 10 mg/mL Proteinase K (Prot K).
11. Reagents for running a DNA agarose gel: agarose, buffers and a DNA ladder (e.g., 1 kb plus DNA ladder).
12. 100 mM ATP stock: dissolve adenosine 5′-triphosphate disodium salt hydrate in water and store in aliquots at −20 °C (*see* **Note 3**).
13. 10× ligase buffer: 100 mM $MgCl_2$, 500 mM Tris–HCl pH 7.5, 10 mM ATP, 100 mM DTT, adjust pH to 7.5.
14. T4 DNA ligase.
15. 10 mg/mL RNase A.
16. Phenol–chloroform–isoamyl alcohol (25:24:1, v/v) (P/C).
17. 3 M NaOAC pH 5.6.
18. 70% ethanol.

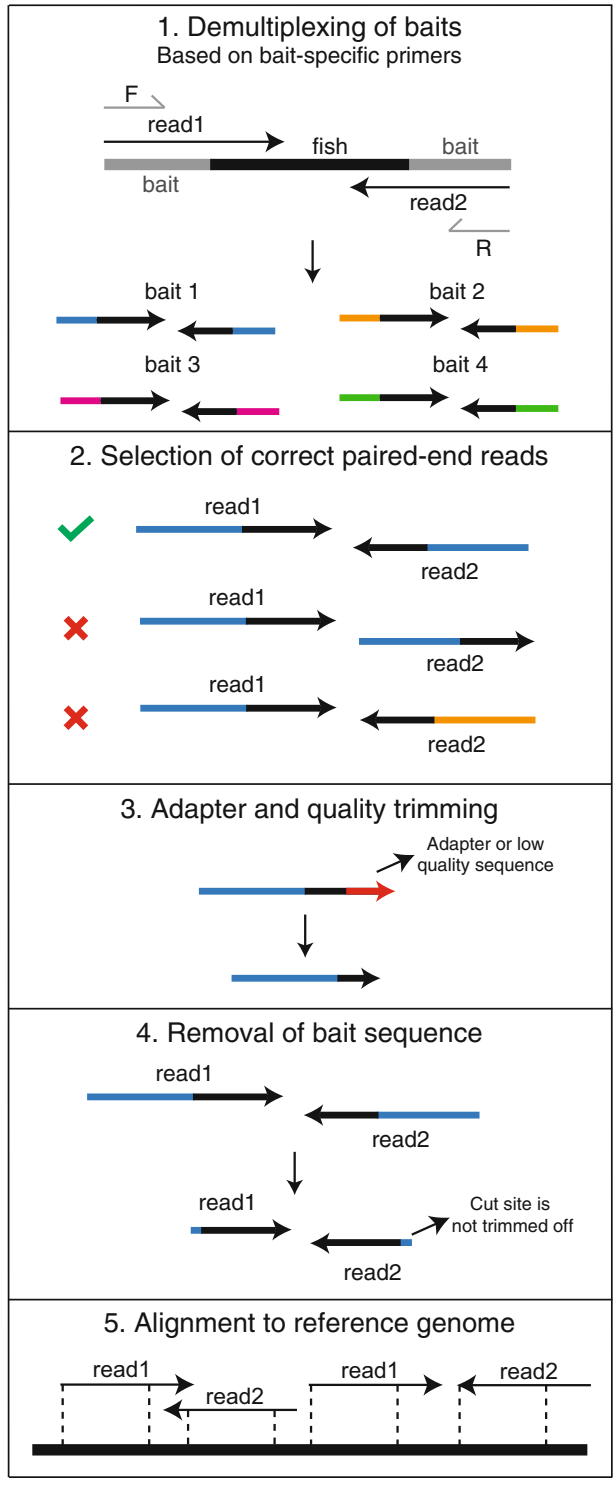


Fig. 2 Outline of the processing of 4C-seq reads. (1) Demultiplex reads based on bait sequences consisting of the primer sequences up to and including the restriction enzyme digestion site. (2) Select correctly oriented read1 and read2 pairs. (3) Trim off adapter sequences and bases with a Phred quality score of <20. (4) Trim off bait sequences excluding the digestion site. (5) Map the reads to the reference genome

19. Absolute ethanol (99–100%).
20. 5 mg/mL linear polyacrylamide (LPA) or 20 mg/mL glycogen.
21. High-fidelity PCR polymerase (*see* **Note 4**).
22. AMPure XP beads or a column-based PCR purification kit.
23. DNA library preparation kit, including indexed sequencing adapters (*see* **Note 5**).
24. Qubit high-sensitivity and broad-range dsDNA assay kit.
25. Bioanalyzer high-sensitivity DNA kit.
26. Eppendorf® safe-lock tubes and Parafilm.
27. LoBind® DNA Eppendorf tubes.
28. Reagents for in-house MiSeq sequencing (or outsource this to a sequencing facility): MiSeq reagent kit (e.g., V2 Nano 300 bp), PhiX Control V3 library, and 1 N NaOH.

2.2 Equipment

1. Dounce homogenizer.
2. Shaking or nonshaking incubator.
3. Qubit fluorometer.
4. NanoDrop spectrophotometer.
5. PCR thermocycler (ideally a gradient thermocycler).
6. DNA agarose gel equipment.
7. Agilent Bioanalyzer 2100 or TapeStation system.
8. In-house MiSeq system, or outsource this to a sequencing facility.

3 Methods

3.1 4C-seq Design

1. Choose the primary and secondary restriction enzymes. First, determine the bait region of interest. Identify the digestion sites of 4-bp (base pair) restriction enzymes around this region and choose an appropriate primary restriction enzyme (*see* **Notes 6–8**). Preferably, the restriction fragment chosen contains your DNA region of interest, or should be in close proximity of this (<1 kb). The choice of primary restriction enzyme is limited, as not all restriction enzymes digest crosslinked DNA efficiently (*see* **Note 7**). For the secondary restriction enzyme, a wide range of restriction enzymes can be used, as at this stage the DNA is reverse–crosslinked. The bait fragment size after the primary and secondary digestion should ideally be >200 bp to allow for efficient circularization, although shorter bait fragments may work fine too [6, 7]. The primary bait fragment does not have to contain secondary restriction enzyme cut sites, although it is preferable if they do.

2. Design inverse 4C-seq PCR primers that orient outward from the restriction fragment. Standard primer design parameters and software can be used for this, such as Primer3. Design the primers as close to the restriction sites as possible, ideally overlapping the restriction site. The maximum distance of the primers from the digestion site is dependent on the deep-sequencing read length chosen. After trimming of bait sequences excluding the digestion site, it is recommended to have a minimum of 30 bp left for mapping. Only one of the primers should be a certain distance away from the restriction site, so that 30-bp of the captured fragment is left for mapping. For example, when using a 75-bp read length, one of the primers should be within 45 bp of the digestion site, so that after trimming of the bait sequence, 30 bp is left for mapping. The other primer could be further away, as it is sufficient to use one of the reads out of a read pair (e.g., read 1) for mapping (when using paired-end sequencing).

3.2 Crosslinking of Cells

1. Harvest and count cells with your standard lab protocol, and spin down 5×10^6 cells in a 15-mL Falcon tube (*see* **Note 9**).
2. Resuspend the cell pellet in 5 mL PBS/10% FCS.
3. Add 5 mL of 4% formaldehyde and incubate while tumbling (gentle inversion on a rotating wheel or by hand) for 10 min at room temperature (RT).
4. Add 500 μL of 2.5 M glycine for a final concentration of 125 mM and incubate for 5 min on ice, mix a few times by inversion.
5. Pellet cells by centrifuging at 250 × *g* for 8 min at 4 °C. Discard the supernatant.
6. Resuspend cells in 10 mL ice-cold 1× PBS and centrifuge at 250 × *g* for 8 min at 4 °C. Discard the supernatant.
7. Directly proceed with the next step, or snap-freeze cell pellets (using liquid nitrogen, dry ice, or methanol/ethanol stored at −80 °C) and store at −80 °C.

3.3 Lysis and Digestion

1. Resuspend nuclei in 1 mL lysis buffer and add another 9 mL of lysis buffer. Incubate on ice for 30 min.
2. Dounce homogenize 10× up and down.
3. Spin cells at 650 × *g* for 5 min at 4 °C.
4. Resuspend the pellet in 1 mL 1× ice-cold PBS and transfer to 2 mL Eppendorf Safe-Lock tubes.
5. Spin at 650 × *g* for 5 min at 4 °C. Remove the supernatant and snap-freeze for storage at −80 °C (*see* **step 7** of Subheading 3.2), or proceed with the protocol.
6. Resuspend nuclei pellet in 95 μL MQ and 5 μL 10% SDS.

7. Incubate at 60 °C for 10 min.
8. Add 306.6 μL MQ and 33.4 μL 15% Triton X-100 (for a final concentration of Triton X-100 of 1.14%). Incubate at 37 °C for 20 min. If aggregates are formed, break them up every 5 min using a p200 pipet (*see* **Note 10**).
9. Add a restriction enzyme mix containing 100 μL restriction enzyme buffer, 800 U of restriction enzyme and MQ up to 1000 μL. Seal the lids of the Eppendorf tubes with Parafilm to prevent evaporation (*see* **Note 11**). Take out 8 μL as undigested control, and add 39.5 μL of 10 mM Tris–HCl pH 7.5, then store at RT.
10. Digest with the tubes placed horizontally at 37 °C (or at the temperature indicated in the restriction enzyme manual) for 4 h without shaking (*see* **Note 10**). Mix the samples gently by inverting the tube every 10 min for the first hour and every 30 min after that. Take out 16 μL of digested sample, and add 29 μL Tris–HCl.
11. Determine the digestion efficiency: add 2.5 μL of 10 mg/mL ProtK to the undigested sample, and 5 μL of 10 mg/mL ProtK to the digested sample, and incubate for 1 h at 65 °C. Run 20 μL of each on a 0.6% agarose gel, together with a DNA ladder. If the digestion is sufficient (Fig. 3), proceed with the next step. If the digestion is insufficient, repeat **steps 8–10** by adding a total of 500 μL extra volume containing 400–800 units of restriction enzyme, restriction enzyme buffer to 1× and MQ (samples can also be incubated overnight at 37 °C).
12. Inactivate the restriction enzyme by incubating at 65 °C for 20 min.

3.4 Ligation

1. Transfer the sample to a 15-mL Falcon tube and add 500 μL of 10× ligase buffer, 3350 U of T4 ligase (*see* **Note 12**), and MQ up to 5 mL. Ligate overnight (O/N) at 16 °C at 30 rpm, followed by 30 min at RT.
2. Check the ligation efficiency: take out 90 μL of ligated sample, add 5 μL of 10 mg/mL ProtK and incubate at 65 °C for 1 h. Run 20 μL of digested sample and 40 μL of ligated sample on a 0.6% gel, along with a DNA ladder. If the ligation is sufficient, proceed with the next step (Fig. 3). If the ligation is insufficient, add a total volume of 100 μL containing 1675–3350 U ligase, 89 μL of 100 mM ATP and 10 μL ligase buffer and repeat **steps 1–2**.

1. Add 150 μg of 10 mg/mL ProtK to the sample, and incubate for 4 h or O/N at 65 °C.

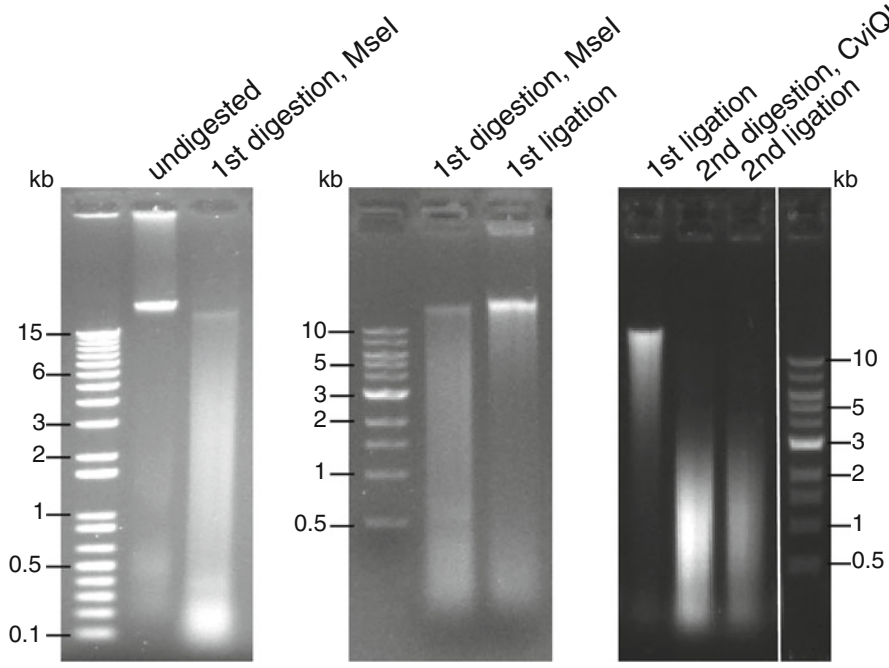


Fig. 3 Example of digestion and ligation checks. Aliquots of digested and ligated 4C samples were taken out, processed and run on a DNA gel, as described in the protocol. The primary restriction enzyme used was MseI, and the secondary CviQI

3.5 Reverse–Crosslinking, RNase A Treatment, and DNA Extraction

2. Add 150 μg of 10 mg/mL RNase A to the sample, and incubate for 30–45 min at 37 °C.
3. Add an equal volume of P/C. Vortex for 30 s.
4. Centrifuge at 5000 × *g* (or the maximum speed of your centrifuge) for 15 min at RT.
5. Transfer the water phase to a new 15-mL Falcon tube.
6. Repeat **steps 3–5**.
7. Precipitate the DNA by adding 3 M NaOAC pH 5.6 to a concentration of 240 mM, and 2.5 volumes of absolute ethanol. Vortex and incubate at −80 °C for around 1.5 h.
8. Centrifuge at 5000 × *g* at 4 °C for 1 h.
9. Remove the supernatant, add 5 mL 70% ethanol (do not resuspend) and centrifuge at 5000 × *g* at RT for 15 min.
10. Remove the supernatant, spin briefly, remove the rest of the supernatant using a p200 pipette, and air-dry for 10–15 min at RT.
11. Dissolve the pellet in 75 μL of MQ (*see* **Note 13**). Remeasure the volume as it will increase to around 100 μL. Take out 1 μL of the sample for running on a gel and add to an Eppendorf tube containing 9 μL of 10 mM Tris–HCl pH 7.5. Five

microliters of this can be run on a gel later, together with the second digestion and ligation samples.

3.6 Second Digestion

1. Add 25 μL of 10× RE buffer, 50 U of the secondary restriction enzyme and MQ to a total volume of 250 μL.
2. Incubate O/N at 37 °C (or at the temperature indicated in the restriction enzyme manual) with the tubes oriented horizontally at 30 rpm. Parafilm-seal the tubes.
3. Take out 7.5 μL of digestion control and add up to 75 μL 10 mM Tris–HCl pH 7.5.
4. Check the digestion efficiency: run 5 μL of P/C-extracted ligated sample and 20 μL of digested sample on a 0.6% agarose gel. If the digestion is sufficient (*see* Fig. 3), proceed with the next step. If the digestion is insufficient, repeat **steps 1–4** by adding 25–50 U additional restriction enzyme (incubation can also be performed for 2–4 h instead of O/N).
5. Inactivate the restriction enzyme by incubation at 65 °C for 20 min. If the restriction enzyme cannot be heat-inactivated, perform a P/C extraction and ethanol precipitation to remove the enzyme. For this, add up to 400 μL MQ. Add an equal volume of P/C and vortex for 30 s. Centrifuge at 15,000 × *g* (or maximum speed) for 10 min at RT. Transfer the water phase to a new tube. Ethanol precipitation is performed by adding 3 M NaOAC pH 5.6 to a concentration of 240 mM and 2.5 volumes absolute ethanol. Vortex and incubate at −80 °C for around 1.5 h. Centrifuge at 15,000 × *g* (or maximum speed) for 20 min at 4 °C. Remove the supernatant, add 1 mL of 70% ethanol and centrifuge at 15,000 × *g* (or maximum speed) for 5 min at RT. Remove the supernatant, spin briefly, remove the rest of the supernatant using a p200 pipette, and air-dry for 10–15 min. Resuspend the DNA pellet in 100 μL MQ. Take out 5 μL as the digestion control. 2.5 μL of this can directly be run on a gel.

3.7 Second Ligation and Clean-up of DNA

1. Transfer the sample to a 15-mL Falcon tube. Add 700 μL 10× ligation buffer, 3350 U ligase and MQ water to 7 mL. Ligate for 4 h or O/N at 16 °C with the tubes oriented horizontally, shaking at 30 rpm. 80 μL of this could directly be run on a DNA gel, however, an upward shift in size is not always seen (*see* Fig. 3). Therefore, the ligation efficiency cannot be reliably checked by agarose gel electrophoresis (*see* **Note 14**).
2. If using a heat-inactivatable secondary restriction enzyme, perform a P/C extraction according to **step 5** in Subheading 3.6, but perform the ethanol precipitation according to the next step in this section. If using a restriction enzyme that cannot be

heat-inactivated, and a P/C extraction was already performed, directly proceed with the ethanol precipitation in the next step.

3. Ethanol-precipitate the DNA with glycogen or LPA. Add 70 µL of 5 mg/mL LPA or 3.5 µL of 20 mg/mL glycogen, 3 M NaOAC pH 5.6 to a final concentration of 240 mM, and 2 volumes of absolute EtOH. Incubate at −80 °C for around 1.5 h. Spin at 5000 × *g* (or maximum speed) at 4 °C for 1 h. Remove the supernatant, add 5 mL of 70% ethanol, and centrifuge at 5000 × *g* at RT for 15 min. Remove the supernatant, spin briefly, remove the rest of the supernatant using a p200 pipette, and air-dry for 10–15 min at RT. Dissolve the pellet in 50 µL 10 mM Tris–HCl pH 7.5 (*see* **Note 13**).
4. Measure the concentration by Qubit using a broad-range dsDNA assay kit according to the manufacturer's instructions. Store 4C samples at −20 °C (*see* **Note 15**).

3.8 4C-seq PCR

Perform PCR reactions according to the manufacturer's guidelines for the Taq Polymerase you are using. Use an extension time that amplifies fragments of up to 1 kb (*see* **Note 16**).

1. Optimize the 4C-PCR reactions by testing different reaction conditions, and running 15 µL of PCR products on a 1.5% agarose gel. First, optimize the annealing temperature using 100 ng of 4C input DNA, by running an annealing temperature gradient using a range of ~10° (±5°) around the predicted annealing temperature of your primers (*see* **Note 17**). The optimal condition is the highest annealing temperature that produces a wide variety of PCR products, in addition to the most prevalent bands derived from self-ligated and undigested bait fragments (Fig. 4). We advise to use an annealing temperature of one degree lower than the optimal annealing temperature, to account for PCR run variations (the optimal annealing temperature in the first PCR run, may not be optimal in other PCR runs). Second, test different DNA input amounts (25, 50, 100 and 200 ng), and use an input that amplifies DNA fragments efficiently. Primer concentrations could also be optimized; for this use the recommended amount and double or triple of that amount.
2. For the final 4C-PCRs, amplify a total of 1 µg starting material by performing multiple PCR reactions (e.g., 10 reactions with 100 ng of input DNA each). At least 1 µg of 4C input DNA is amplified to ensure a wide variety of interacting fragments is represented in the library.
3. Pool all 4C-PCR reactions and purify the amplicons using a PCR purification kit or AMPure beads according to the manufacturer's instructions. If using a PCR purification kit,

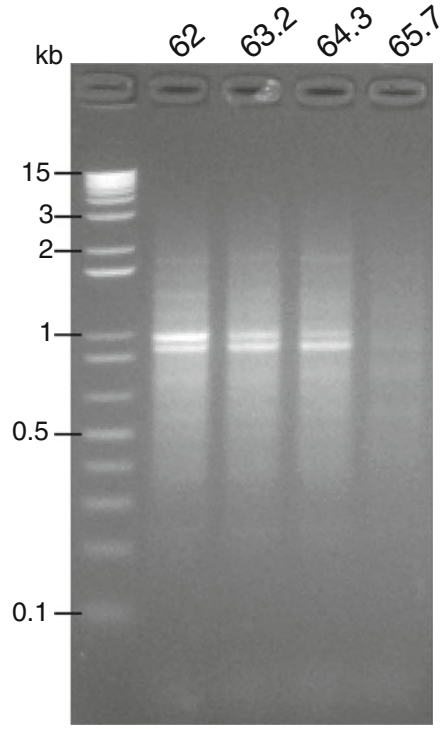


Fig. 4 Example of 4C PCR annealing temperature optimization. 4C PCRs were performed using Q5 Hot Start High-Fidelity 2× master mix, using bait-specific primers described in Marsman et al. [5]. A gradient PCR was run using annealing temperatures of 62, 63.2, 64.3, and 65.7 °C. The optimal annealing temperature based on this PCR is 64.3 °C. Based on this, an annealing temperature of 63.3 °C is best for future PCRs, allowing for run-to-run variations in PCR efficiency

depending on the capacity of the columns, use 1–2 columns per 4C-PCR sample.

4. Measure the concentration by Qubit using a broad-range or high-sensitivity dsDNA assay kit according to the manufacturer's instructions. Also check the purity of the DNA by Nanodrop (*see* **Note 18**).

3.9 Library Preparation

1. When performing 4C-seq on multiple baits: per sample (e.g., per replicate), pool the PCR fragments of multiple baits in equal amounts (*see* **Note 19**). These can later be demultiplexed based on bait-specific primer sequences.
2. Perform the library preparation according to the manufacturer's instructions (*see* **Note 5**). Use different Illumina indexed sequencing adapters per sample (e.g., per replicate); take care to select combinations of indexes that have sufficient sequence variation, as per Illumina's index adapters pooling guide. For the adapter PCR, perform 7 cycles, using an

extension time of 30 s to minimize the formation of large products not suitable for Illumina flow cells.

3. Quantify the libraries by Qubit using a broad-range or high-sensitivity dsDNA assay kit according to the manufacturer's instructions.
4. Library quality check: dilute each library to 1 ng/μL with nuclease-free water. Analyse 1 μL of the diluted library using a High Sensitivity DNA chip and an Agilent BioAnalyzer 2100 system, as per the manufacturer's instructions. Good quality libraries have a peak of DNA with the majority of the mass between 200 and 1000 bp, and no evidence of sharp peaks smaller than this that likely indicate adapter/primer contamination.

3.10 Deep-Sequencing

Prepared libraries are first sequenced on a small scale using a MiSeq to check if each bait and each library is approximately represented equally (*see* **Note 20**). Obtaining several tens of thousands of reads per bait is sufficient for this. This part can also be outsourced to a sequencing facility. When sequencing, it is important that enough sequence diversity is present in the beginning of reads, which is necessary for accurate base calling. This can be achieved by (1) combining amplicons derived from several different baits, (2) by running the 4C-seq libraries together with other libraries in the same flow cell (these can be from any type of experiment, e.g., ChIP-seq), or (3) by mixing in a PhiX control library.

1. Dilute libraries to 4 nM (*see* **Note 21**) and combine equivalent amounts to create an equimolar library mixture. The molarity calculations are based on the average size of the library trace taken from the BioAnalyzer and the quantification by Qubit. Libraries can be mixed with other sequencing libraries in the same MiSeq run (*see* **Note 22**).
2. If using MiSeq, the MiSeq (V2 Nano 300 bp) reagent cartridge and the tube of HS buffer should be taken out from the −20 °C freezer at least 1 h prior to diluting and denaturing the libraries. The HS buffer can stand at RT on your bench. Place the cartridge into a container containing tap water such that the bottom section of the cartridge is submerged to the level marked on the white plastic bottom section of the cartridge.
3. Add 4.0 μL of nuclease-free water and 1.0 μL of 1 N NaOH to a 1.5 mL LoBind Eppendorf tube (total = 5 μL of 0.2 N NaOH). To this, add 5 μL of the 4 nM combined library and mix by pipetting. Incubate at RT for 5 min to denature the library, before adding 990 μL of HS buffer (total = 1 mL of 20 pM denatured library). This can be stored in a −20 °C freezer.
4. Repeat **step 3** for the PhiX Control V3 library (or use an already denatured PhiX library).

5. Combine 30 μL of the denatured PhiX library, 270 μL of the denatured library and 300 μL of HS buffer in a 1.5 mL LoBind Eppendorf tube to make a final volume of 600 μL, and pulse on a vortex mixer. The final preparation is now at a concentration of 8 pM (*see* **Note 23**).
6. Remove the MiSeq cartridge from the water bath and dry well. Puncture the foil, covering the position marked "Sample" using a sterile 1-mL pipette tip. Using a new 1-mL pipette tip, transfer 600 μL of denatured library/PhiX mix into position 18.
7. Follow the manufacturer's instructions and visual guides to load the cartridge and flow cell and begin the run (*see* **Note 24**).
8. After obtaining the data, calculate the percentage of reads derived from each bait (when using multiple baits) by demultiplexing the reads based on bait-specific sequences (*see* Subheading 3.11). When each bait is not represented equally, bait inputs for library preparation have to be adjusted accordingly, and libraries have to be prepared again (Subheading 3.9). Also assess the percentage of the individual libraries in the mixture via the percentage barcode representation. The mixture can be adjusted to ensure even representation before sending it off for high-depth sequencing.
9. When baits and libraries are mixed equally based on the MiSeq run, send the samples for high-depth sequencing. The platform can be chosen based on the number of reads required (e.g., MiSeq or HiSeq). We recommend a sequencing depth of 1–-10 million reads per bait, per sample.

3.11 Data Processing and Analysis Options

The data is processed using the 'fastq.gz' files obtained from the Illumina deep-sequencing run. These files are already demultiplexed per library index by the sequencing center or the MiSeq software (using bcl2fastq). Complete analysis packages that include processing of the reads and data analysis have previously been described, which could be used instead of the data processing steps described here [7, 14].

1. Perform a quality check of the sequencing data using FastQC [15]. Check if the quality of base calling is good. Note that the standard quality control for sequencing is expected to fail in several parts due to the amplicon nature of the 4C-seq experiment. Per base sequence content, sequence duplication levels and overrepresented sequences are likely to not pass the quality criteria.
2. Demultiplex the reads derived from each bait (*see* Fig. 2). Demultiplexing is done using the bait-specific primer sequences up to and including the restriction enzyme site,

allowing 0 mismatches (*see* **Note 25**). Depending on the experimental design and sequencing strategy, primers can be present only in read 1 (R1) for single-end sequencing, or in both R1 and R2 for paired-end sequencing. R1 and R2 can start with either the forward or reverse primer sequences. Demultiplexing can be achieved using the "grep" command from a Linux/Unix terminal. To keep this in a fastq format, it is crucial to use "--no-group-separator -A 2 -B 1" and to use the "^" sign to indicate that the reads should start with the primer sequence. See the example below:

```
grep ^FWDPRIMERSEQ --no-group-separator -A 2 -B 1 sample_R1.
fastq > demultiplexed_R1.fastq
grep ^REVPRIMERSEQ --no-group-separator -A 2 -B 1 sample_R2.
fastq > demultiplexed_R2.fastq
```

We have deposited this script in a GitHub repository: https://github.com/gregomics/MiMB4Cseq. Other bioinformatics tools such as fastq-multx (from the ea-utils toolset) or Cutadapt can also be used [16, 17]. At the end of this step reads are demultiplexed, but the bait sequence is not trimmed off yet (*see* **Note 26**).

3. Select read pairs that have the forward and reverse primers correctly oriented (only when using paired-end sequencing) (*see* Fig. 2). This step selects reads that start with one of the primer pairs in R1, and the other pair in R2, or the other way around (e.g., R1 starting with the forward primer, and R2 with the reverse primer, or the other way around). Odd read pairs are removed. For this, we provide a script named "keep_only_paired_seq.plx" which is available on GitHub (https://github.com/gregomics/MiMB4Cseq). This script can be run as below:

```
./scripts/get_reads_in_paired.sh -h
Usage: ./scripts/get_reads_in_paired.sh
-f R1 file
-r R2 file
-o filtered R1 file
-p filtered R2 file
-h shows this help
```

4. Trim off sequencing adapters and low-quality sequences (*see* Fig. 2). Sequencing adapters and low-quality sequences (Phred quality score of <20) should be removed. Several tools are available such as Cutadapt, fastq-mcf (from the ea-utils toolset), or trimomatic [16–18].

5. Trim off bait sequences up to, but excluding, the digestion site (*see* Fig. 2). For this, seqtk (https://github.com/lh3/seqtk) or fastq-multx (from the ea-utils toolset) can be used.
6. Map the reads to the reference genome (*see* Fig. 2). Align cleaned reads with a minimum length of 30 bp to the reference genome using Burrows-Wheeler Aligner (BWA) MEM algorithm [19, 20]. Mark split read mapping using the –M option. Alignment files in SAM format can be converted to BAM files, removing spurious low quality mapping ($q > 10$) and multi-mapped reads, and finally reads can be sorted by read location using SAMtools [21].
7. From this point onward, several R packages can be used to (1) calculate the read count per restriction fragment, (2) calculate the running mean of multiple adjacent restriction fragments, and (3) perform statistical analysis. Examples of packages that could be used for this are FourSig, 4C-ker, r3Cseq, FourCSeq, and peakC [9–13].

4 Notes

1. 4% buffered formaldehyde solution can be commercially purchased, or can be made from paraformaldehyde (PFA) powder. Home-made 4% buffered formaldehyde solution can be aliquoted and stored at −20 °C for up to 2 years. If using frozen 4% formaldehyde, before use, incubate frozen aliquots for 15 min at 65 °C while vortexing every 5 min (to dissolve any formed paraformaldehyde to formaldehyde), then cool to RT before use.
2. Add amount of solid or solubilized protease inhibitors according to the manufacturer's instructions.
3. Always make the ATP stock separately prior to adding it to the lysis buffer, and use within several months (ATP degrades over time). Active ATP is crucial for efficient ligation.
4. We recommend using a ready-made master mix as they perform well and consistent, and it makes optimization faster (we used the Q5 Hot Start High-Fidelity 2× master mix).
5. A low-input DNA library preparation kit is recommended that does not include a shearing step, or for which the shearing step can be skipped. We have used the ThruPLEX® DNA-seq Kit (Rubicon Genomics) using 10 ng input DNA and TruSeq indexed adapters. These are compatible with Illumina sequencing.

6. Preferably use enzymes insensitive to CpG methylation, that create sticky ends and recognize sites with ~50% GC content, to ensure a homogeneous digestion across the genome.
7. Not all restriction enzymes digest crosslinked DNA efficiently. The following 4-bp restriction enzymes have been used successfully for 4C-seq experiments: DpnII (top choice if the design allows it, as it digest crosslinked DNA very efficiently and adheres to the guidelines described in **Note 6**), MseI, MspI, AluI, Csp6I, NlaIII, MboI, and HhaI. For 6-bp cutters: HindIII, EcoRI, BglII, BamHI, KpnI, XhoI, PstI, DraI, SacI, AseI, NcoI.
8. Six-base pair cutters can also be used, but this will result in a lower resolution of chromatin interactions, not only because of larger primary digestion fragment lengths, but also because larger fragments are not efficiently sequenced.
9. Although it is preferable to use 5–10 million cells, fewer cells can be used. For this, reduce the volumes proportionally from Subheading 3.2 onward. To reproducibly capture the large variety of DNA contacts present in a population of cells, a minimum of a few hundred thousand cells is recommended [7].
10. Aggregates appear as white, stringy clumps. In our experience, the more aggregates are formed, the less efficient the digestion is. Counter intuitively, it is better to treat these aggregates gently, instead of rigorously shaking them. Shaking results in more aggregate formation, which reduces the digestion efficiency.
11. Evaporation can lead to increased salt concentrations, which can induce Star activity (unspecific digestion) of the enzyme.
12. Ligase units can be denoted in cohesive-end or Weiss units; make sure you check which ligase unit denotation is used for the ligase you use. In this protocol we denote the amount of ligase to add in cohesive-end units. One cohesive-end unit equals 0.015 Weiss units.
13. If the DNA does not dissolve easily, follow this method: add MQ to the DNA pellet (do not resuspend), incubate at 37 °C for 0.5–1 h and resuspend to dissolve the DNA.
14. If no shift in size is seen on a DNA gel after the second ligation, another way of checking if the ligation worked is by the following: an aliquot of the sample could be taken out before ligation and mixed with a prior digested plasmid. After the ligation reaction, an aliquot of the sample before and after ligation could be run on a DNA gel, and a shift in plasmid bands should be visible if the ligation worked.

15. Nanodrop is highly inaccurate for measuring 4C DNA concentrations, as the samples are impure. So always use the Qubit for this.
16. For most Taq polymerases, 30 s of extension time suffices. Fragments of >1 kb will not be sequenced efficiently in a flow-cell, so amplifying larger fragments is not necessary.
17. Predicted annealing temperatures depend on the type of Taq Polymerase reagents you are using. Most producers provide an online annealing temperature calculator tool.
18. Contaminants in PCR samples could affect library preparation. Therefore, nanodrop is performed to check the purity of the samples. An $A_{260/280}$ of ~1.8 and an $A_{260/230}$ of ~2–2.2 indicates pure DNA. Our purified 4C products often have a good $A_{260/230}$ ratio, but a lower $A_{260/280}$ of ± 1.5. The latter is acceptable and does not seem to affect library preparation.
19. While it is possible to mix baits in equivalent amounts based on the average size of the library trace taken from the BioAnalyzer and the quantification by Qubit, this often does not help in obtaining an equal representation of each bait. Adding an equal amount (ng) of each bait is often the best approach.
20. The MiSeq run can be performed on part of the samples, e.g., on 1 replicate per condition, to see which baits produce more reads. Often this is very unequal. After demultiplexing, bait input amounts can be adjusted for the final library preparation, based on the number of reads obtained per bait.
21. In which volume the dilutions are made does not matter, as long as the total volume after all libraries are combined is at least 5 μL.
22. As only several tens of thousands of reads per bait is sufficient to determine the representation of each bait in a library, 4C-seq libraries can be run on the same flow-cell together with other sequencing libraries. This makes it cost-effective.
23. Whilst metrics will vary by individual MiSeq machine, loading 8 pM final library/PhiX mix onto our local machine results in approximately 700–800 cluster density, 7–10% aligned reads to PhiX and 800,000–1,000,000 reads pass filter.
24. Most of this process is shown to the user by the MiSeq software. The sequencing setup software includes visual guides to identify the locations for reagent loading.
25. It is good to check if common single nucleotide polymorphisms or mutations are present in the bait. If so, one or more mismatches can be allowed during demultiplexing. If there are homozygous or heterozygous single nucleotide polymorphisms or mutations present in your bait sequence, these may reveal themselves during demultiplexing, as few (for homozygous variants) or halve of the expected number (for

heterozygous variants) of reads from a particular bait would be obtained after demultiplexing.

26. We find that the adapter trimming works best on demultiplexed reads that still contain the bait sequence. FastQC can be used to check whether adapter sequences are removed completely after this step.

Acknowledgments

The authors were supported in this work by grants from the Heart Foundation of New Zealand (post-doctoral research fellowship [1691] and small project grant [1804]), the Dunedin School of Medicine Dean's Bequest fund, and Health Research Council of New Zealand Grants [14-155, 17-402].

References

1. Grob S, Cavalli G (2018) Technical review: a Hitchhiker's guide to chromosome conformation capture. In: Bemer M, Baroux C (eds) Plant chromatin dynamics: methods and protocols. Springer, New York, NY, pp 233–246
2. Simonis M, Klous P, Splinter E et al (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38:1348–1354
3. Zhao Z, Tavoosidana G, Sjölinder M et al (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 38:1341–1347
4. Marsman J, Thomas A, Osato M et al (2017) A DNA contact map for the mouse Runx1 gene identifies novel Haematopoietic enhancers. Sci Rep 7:13347
5. Marsman J, Gimenez G, Day RC et al (2020) A non-coding genetic variant associated with abdominal aortic aneurysm alters ERG gene regulation. Hum Mol Genet 29:554–565
6. van de Werken HJG, de Vree PJP, Splinter E et al (2012) 4C technology: protocols and data analysis. Methods Enzymol 513:89–112
7. Krijger PHL, Geeven G, Bianchi V et al (2020) 4C-seq from beginning to end: a detailed protocol for sample preparation and data analysis. Methods 170:17–32
8. R Core Team (2013) R: A language and environment for statistical computing. R Foundation for StatisticalComputing, Vienna, Austria. http://www.R-project.org/
9. Williams RL, Starmer J, Mugford JW et al (2014) fourSig: a method for determining chromosomal interactions in 4C-Seq data. Nucleic Acids Res 42:e68
10. Klein FA, Pakozdi T, Anders S et al (2015) FourCSeq: analysis of 4C sequencing data. Bioinformatics 31:3085–3091
11. Raviram R, Rocha PP, Müller CL et al (2016) 4C-ker: a method to reproducibly identify genome-wide interactions captured by 4C-Seq experiments. PLoS Comput Biol 12:e1004780
12. Geeven G, Teunissen H, de Laat W et al (2018) peakC: a flexible, non-parametric peak calling package for 4C and capture-C data. Nucleic Acids Res 46:e91
13. Thongjuea S, Stadhouders R, Grosveld FG et al (2013) r3Cseq: an R/Bioconductor package for the discovery of long-range genomic interactions from chromosome conformation capture and next-generation sequencing data. Nucleic Acids Res 41:e132
14. Brouwer RWW, van den Hout MCGN, van IJcken WFJ et al (2017) Unbiased interrogation of 3D genome topology using chromosome conformation capture coupled to high-throughput sequencing (4C-Seq). In: Wajapeyee N, Gupta R (eds) Eukaryotic transcriptional and post-transcriptional gene expression regulation. Springer, New York, NY, pp 199–220
15. Andrews, Simon FastQC: A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

16. Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J 17:10–12
17. Aronesty E Ea-utils: command-line tools for processing biological sequencing data. ExpressionAnalysis. https://github.com/ExpressionAnalysis/ea-utils
18. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120
19. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM
20. Li H, Durbin R (2009) Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 25:1754–1760
21. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Chapter 17

Mammalian Micro-C-XL

Nils Krietenstein and Oliver J. Rando

Abstract

Chromosome Conformation Capture (3C) methods are a family of sequencing-based assays to measure the three-dimensional structure of genomes, with Hi-C as the most prominent method in widespread use. The Micro-C-XL protocol is technical variant that improves the resolution and signal-to-noise ratio of the Hi-C protocol and therefore offers enhanced detection of chromatin features such as chromosome loops and fine-grained resolution of topologically associated domains. Here we describe a detailed step-by-step protocol for Micro-C-XL in mammalian cells.

Key words Chromosome conformation capture (3C), Hi-C, Genome architecture, Nucleosome, Chromatin fiber, MNase, Chromosome loops, TADs

1 Introduction

The development of chromosome conformation capture (3C) assays, in particular the genome-wide variant Hi-C, allowed for a novel perspective onto three-dimensional genome organization [1–3]. Increasing resolution and sequencing depth have successively revealed multiple layers of regulatory structures, including chromosome compartments, topologically associated domains (TADs), and chromosome loops [1, 2, 4, 5]. The three-dimensional organization of chromosomes is captured using three essential steps [2, 3]. In the first step, spatial information of chromosomes within the nucleus is captured by covalently cross-linking proximal protein–DNA complexes. In the second step, the genome is fragmented to allow for detection of long-range interactions for a given chromatin fragment free of the spatial constrains of the one-dimensional genome organization; typically, this is achieved with restriction enzymes that have defined DNA recognition sequences. In the third step, the resulting free DNA ends are ligated to each other. Here, DNA ends that co-occur within covalently cross-linked complexes are preferentially ligated because they remain in physical proximity. The resulting chimeric DNA

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_17,

molecules contain information about the three-dimensional organization of the genome, which can be read out by paired-end deep-sequencing.

The resolution of 3C assays is largely driven by the first two steps, cross-linking and chromatin fragmentation. The level of cross-linking defines the size and number of covalently cross-linked, spatially informative protein–DNA complexes. Next, the fragmentation size is presumably the most resolution defining step in 3C experiments; the finer the genome is fragmented (and the more even the spacing of resulting fragments), the higher is the theoretical resolution. In Micro-C XL, two modifications have been made to typical Hi-C protocols to improve these two steps. First, while typical Hi-C protocols rely on the zero-length crosslinker formaldehyde, Micro-C-XL utilizes additional, longer crosslinkers, such as DSG and EGS, to improve the effective cross-linking rate and radius [6–8]. This additional cross-linking step has now been shown to generally improve various Hi-C protocols [9]. Second, Micro-C uses Micrococcal Nuclease to fragment chromatin to mononucleosomal fragments [10]. Nucleosomes are the fundamental unit of chromatin that are distributed throughout the genome and occupy most of the genomic DNA. At proper digestion degrees, MNase digests most of all DNA except that which is protected by its association with nucleosomes, and MNase is therefore widely used in chromatin biology to profile nucleosome footprints [11–14]. Micro-C-XL leverages the ubiquitous abundance of nucleosomes (approx. 200 bp) to improve the signal-to-noise ration of 3C measurements that results in enhanced detection of high-resolution features, such as chromosome loops and TADs.

2 Materials

Prepare all solutions using ultrapure water and analytical grade reagents.

2.1 Prepare Cross-Linked Chromatin from Cell Culture

1. Dulbecco's phosphate buffered saline (PBS).
2. 37% Formaldehyde.
3. 300 mM DSG stock solution: 0.3 M disuccinimidyl glutarate (DSG) in DMSO. Make fresh before use, equilibrate DSG to room temperature before weighing.
4. 2.5 M Glycine: 187.7 g/L Glycine in H_2O.

2.2 MNase Titration and MNase Digestion of Chromatin

1. MB#1: 50 mM NaCl, 10mM Tris–HCl pH 7.5, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.2% NP-40, 1× Roche cOmplete EDTA-free.

2. Micrococcal Nuclease: resuspended from lyophilized MNase at 20 U/μL in Tris–HCl pH 7.4. Aliquot into tubes upon first use and freeze at −80 °C.
3. 0.5 M EGTA solution.
4. Proteinase K: reconstitute to a concentration of 20 mg/mL in TE and 50% glycerol. Store at −20 °C.
5. 10% SDS.
6. Phenol–chloroform–isoamyl alcohol (PCI).
7. Zymo DNA Clean & Concentrator-5.

2.3 MNase Digestion of Chromatin (See Note 1)

1. Dephosphorylation Mix (1x): 40 μL H_2O, 5 μL NEBuffer 2.1 (NEB), 5 μL Shrimp Alkaline Phosphatase rSAP (1 U/μL, NEB).
2. End-Chewing Mix (1×): 30 μL H_2O, 5 μL NEBuffer 2.1, 2 μL 100 mM ATP, 3 μL 100 mM DTT, 8 μL DNA Polymerase I Large (Klenow) Fragment (5 U/μL, NEB), 2 μL T4 Polynucleotide Kinase (10 U/μL, NEB).
3. End-Labeling Mix (1×):): 38.5 μL H_2O, 25 μL 0.4 mM Biotin-14-dATP (Invitrogen™), 25 μL 0.4 mM Biotin-14-dCTP (Invitrogen™), 1 μL 10 mM dTTP + dGTP (each), 10 μL 10× T4 DNA Ligase Reaction Buffer (NEB), 0.5 μL 20 mg/mL bovine serum albumin (BSA, NEB).
4. Proximity-Ligation Mix (1×): 2225 μL H_2O, 250 μL 10× T4 DNA Ligase Reaction Buffer (NEB), 12.5 μL 20 mg/mL BSA (NEB), 12.5 μL T4 DNA Ligase (400 U/μL, NEB).
5. Biotin-end Removal Mix (1×): 179 μL H_2O, 20 μL 10× NEBuffer 1, 2 μL Exonuclease III (100 U/μL, NEB).

2.4 Dinucleosomal DNA Purification

1. Zymo DNA Clean & Concentrator-5.
2. Zymoclean Gel DNA Recovery Kit.

2.5 Streptavidin Pull-down and on-Bead Library Preparation

1. 1× TBW: 5 mM Tris–HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20.
2. 2× B&W buffer: 10 mM Tris–HCl pH 7.5, 1 mM EDTA, 2 M NaCl.
3. Prewashed streptavidin beads: transfer 5 μL Dynabeads® MyOne Streptavidin C1 to a fresh tube, place into a suitable magnetic rack and wait until solution clears. Discard supernatant, resuspend in 500 μL 1× TBW, mix by pipetting up and down, place into a suitable magnetic rack and wait until solution clears. Repeat the 1× TBW wash and resuspend the beads in 150 μL 2× B&W buffer.

2.6 Library Amplification

1. NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB).
2. 0.1× TE: 1 mM Tris–HCl pH 8.0, 0.1 mM EDTA pH 8.0.
3. NEBNext®Multiplex Oligos for Illumina® (NEB).
4. PCR Amplification Mix (1×, *see* **Note 1**): 65 µL H_2O, 100 µL 2× KAPA HiFi HotStart Readymix, 8 µL NEBNext Universal PCR Primer, 8 µL NEBNext Index Primer for Illumina (individual for each sample).
5. AMPure XP Beads (Beckman Coulter).

3 Methods

3.1 Prepare Cross-Linked Chromatin from Cell Culture

Start this protocol with a suspension of cells grown under your favored experimental condition. Use low binding tubes for all steps and keep the sample on ice if not indicated otherwise.

1. Pellet the cells by centrifugation (1000 × *g*, 5 min, RT) and discard supernatant.
2. Resuspend the pellet in 1 mL DPBS per one million cells.
3. Repeat **steps 1** and **2**.
4. FA cross-linking: add 27 µL 37% formaldehyde per 1 mL cell suspension (1% final concentration) and incubate at room temperature for 10 min with rotation.
5. Quenching: add 0.05 mL 2.5 M Glycine per 1 mL cell suspension (0.125 M final concentration) and incubate at room temperature for 5 min with rotation.
6. Pellet the cells by centrifugation (1000 × *g*, 5 min, RT) and discard supernatant.
7. Resuspend the pellet in 1 mL DPBS per four million cells (*see* **Note 2**).
8. Repeat **steps 6** and **7**.
9. DSG cross-linking: add 10 µL 300 mM DSG stock solution per 1 mL cell suspension (3 mM final concentration) and incubate at room temperature for 40 min with rotation.
10. Quenching: add 0.05 mL 2.5 M Glycine per 1 mL cell suspension (0.125 M final concentration) and incubate at room temperature for 5 min with rotation.
11. Pellet the cells by centrifugation (1000 × *g*, 5 min, RT) and discard supernatant.
12. Resuspend the pellet in 1 mL DPBS per five million cells.
13. Make two aliquots of one million cells per tube to titrate the MNase digestion conditions (*see* Subheading 3.2).

14. Aliquot the remaining cells to five million cells per tube (preparative libraries; *see* Subheading 3.3).
15. Pellet the cells by centrifugation (1000 × *g*, 5 min, RT) and discard supernatant.
16. Snap freeze cell pellets in liquid nitrogen and store at −80 °C until further processing.

3.2 MNase Titration (See *Note 3*)

1. Thaw one cell pellet of one million cells on ice for 10 min (*see* **Note 4**).
2. Resuspend the cell pellet in 500 μL DPBS.
3. Incubate on ice for 20 min (add 1× NEB BSA if cells stick to the tube wall).
4. Pellet the cells by centrifugation (10,000 × *g*, 5 min, RT) and discard supernatant.
5. Resuspend the pellets in 500 μL MB#1 buffer each.
6. Pellet sample by centrifugation (10,000 × *g*, 5 min, RT) and discard supernatant.
7. Resuspend the pellets in 200 μL MB#1 buffer.
8. Split the sample to 4 tubes (50 μL each).
9. Make a 1:4 serial dilution of your MNase stock in 10 mM Tris, pH 7.4 buffer (20 U, 5 U, 1.25 U, and 0.3125 U per μL).
10. MNase digestion: with 10 s intervals, add 1 μL of one of your MNase dilutions to one of your samples, vortex and spin briefly, and incubate for 10 min at 37 °C on shaking Thermomixer or equivalent (850 rpm).
11. With 10 s intervals, stop MNase digestion by addition of 1.6 μL of 0.5 M EGTA with the same order the MNase was added.
12. Add 150 μL 0.1× TE, 25 μL of 10% SDS, and 25 μL of 20 mg/mL Proteinase K.
13. Incubate the sample overnight at 65 °C.
14. Add 500 μL PCI to your sample and mix by vortexing for 2 min.
15. Separate phases by centrifugation (19,800 × *g*, 5 min, RT) and transfer the aqueous phase to a fresh tube.
16. Repeat **steps 14** and **15**.
17. Concentrate the DNA with the ZymoClean DNA purification kit and elute in 15 μL volume.
18. Analyze the digestion degree by gel electrophoresis on 1.5% agarose gel (Fig. 1) (*see* **Note 5**).

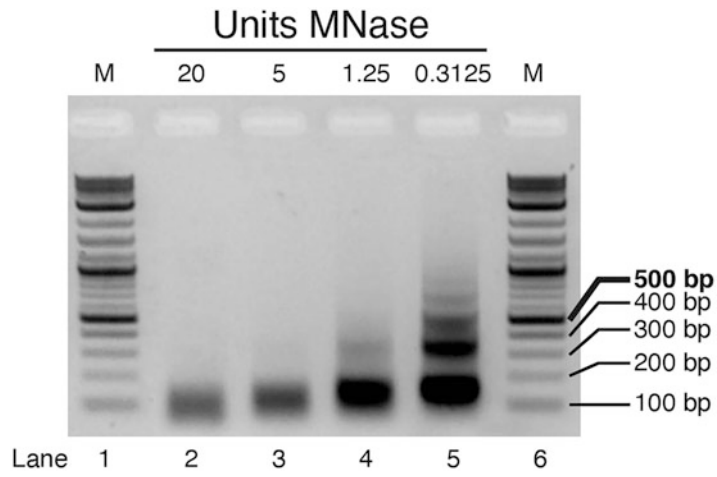


Fig. 1 MNase titration. Lanes 1/6: NEB 2-log DNA Ladder. **Lane 2–5**: Digestion of chromatin from 250,000 cells with various amounts of MNase. **Lane 4**: optimal digestion degree for Micro-C experiments. Note: A high fraction of mononucleosomes is achieved (compare 150 to 300 bp bands). Also, at higher digestion degrees, the size of the mononucleosomal band (150 bp) slightly decreases (particularly visible in lane 2), indicating an overdigestion

3.3 Preparative MNase Digestion of Chromatin

1. Thaw the cell pellet of 5 million cells on ice for 10 min.
2. Resuspend the cell pellet in 1 mL DPBS per five million cells and make aliquots of one million cells per tube (200 µL cell suspension per tube) (*see* **Note 6**).
3. Incubate on ice for 20 min (add 1× NEB BSA if cells stick to the tube wall).
4. Pellet sample by centrifugation (10,000 × *g*, 5 min, RT) and discard supernatant.
5. Resuspend the pellets in 500 µL MB#1 buffer each.
6. Pellet sample by centrifugation (10,000 × *g*, 5 min, RT) and discard supernatant.
7. Resuspend the pellets in 200 µL MB#1 buffer each.
8. MNase digestion: digest chromatin by appropriate amounts of MNase (5–20 U), mix well (vortex and spin quickly), and incubate for 10 min, 37 °C.
9. Stop the MNase digestion by addition of 1.6 µL of 0.5 M EGTA and incubate at 65 °C for 10 min.
10. Pellet the sample by centrifugation (10,000 × *g*, 5 min, 4 °C) and discard supernatant.
11. Resuspend the pellet in 500 µL NEBuffer 2.1. At this point, the digestion aliquots can be pooled for further processing. Do not pool more than the equivalent of five million cells input (*see* **Note 7**).

12. Transfer an aliquot as digestion/input control to a fresh tube (the equivalent of 250–500 thousand cells is sufficient). Adjust volume to 200 μL TE buffer. Purify the DNA alongside the Micro-C proximity ligation library (**step 17** of Subheading 3.3—**step 5** of Subheading 3.4—*see* Fig. 1 lane 1).

3.4 DNA and Processing and Proximity Ligation

1. Pellet the sample by centrifugation (10,000 × *g*, 5 min, 4 °C) and discard supernatant.
2. Resuspend the pellet in 50 μL "De-Phosphorylation Mix" and mix by gently pipetting up and down.
3. Incubate the sample for 45 min at 37 °C on a shaking Thermomixer or equivalent (850 rpm).
4. Incubate the sample for 5 min at 65 °C to deactivate rSAP.
5. Add 50 μL "End-Chewing Mix" to the sample and mix by gently pipetting up and down.
6. Incubate the sample for 15 min at 37 °C on shaking Thermomixer or equivalent (850 rpm).
7. Add 100 μL "End-Labeling Mix" to the sample and mix by gently pipetting up and down.
8. Incubate the sample for 45 min at 25 °C on shaking Thermomixer or equivalent (850 rpm).
9. Add 12 μL 0.5 M EDTA to the sample.
10. Incubate the sample for 20 min at 65 °C to deactivate enzymes.
11. Pellet sample by centrifugation (10,000 × *g*, 5 min, RT) and discard supernatant.
12. Resuspend sample in 2500 μL "Proximity-Ligation Mix" (preferably used 5 mL low binding Eppendorf tubes).
13. Rotate the sample for 2.5–3 h at RT.
14. Split samples into two 1.5 mL tubes, pellet by centrifugation (10,000 × *g*, 5 min, RT) and discard supernatant.
15. Pool both pellets by resuspension in 100 μL "Biotin-end Removal Mix".
16. Incubate the sample for 5 min at 37 °C on shaking Thermomixer or equivalent (850 rpm).
17. Add 25 μL Proteinase K and 25 μL 10% SDS to the sample.
18. Incubate the sample overnight at 65 °C.

3.5 Dinucleosomal DNA Purification

1. Add 500 μL PCI to your sample and mix by vortexing for 2 min.
2. Separate phase by centrifugation (19,800 × *g*, 5 min, RT) and transfer the aqueous, top phase to a fresh tube.
3. Repeat **steps 1** and **2**.

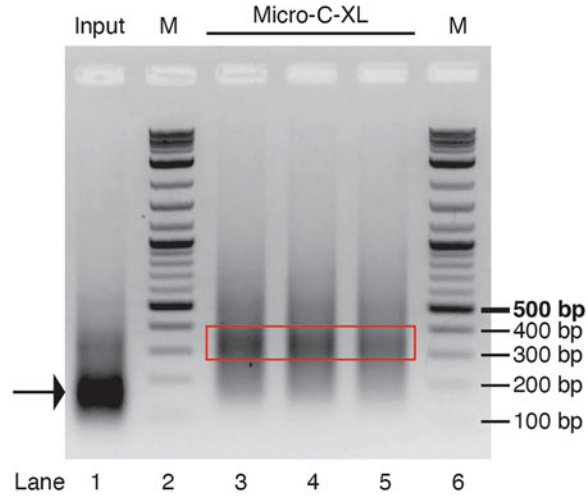


Fig. 2 Micro-C proximity ligation product in 1.5% agarose gel electrophoresis. Lane 1: MNase digestion input control (**step 12** of Subheading 3.211). **Lanes 2/6**: NEB 2-log marker. **Lanes 3–5**: Proximity ligation products. The processed DNA from chromatin of 2.5 million cells was loaded per lane. The proximity ligation product has the size of 2 nucleosomes (~300 bp). A successful Micro-C experiment is indicated by shift from mono-nucleosomal size fragments (~150 bp, indicated by arrow) to dinucleosomal sizes (~300 bp red box). **Red Box**: Indicated target region that was excised from the gel and further processed

4. Concentrate the DNA with the Zymo DNA Clean & Concentrator-5 kit and elute in 30 μL volume.
5. Separate the DNA Fragments by gel electrophoresis on a 1.5% agarose gel (Fig. 2).
6. Excise the DNA fragments that have a dinucleosomal size (~300 bp).
7. Extract the DNA from the gel peace with the Zymoclean Gel DNA Recovery Kit and elute in 50 μL.

3.6 Streptavidin Pull-down and On-bead Library Preparation

1. Add 100 μL of Elution buffer to you sample (total volume of 150 μL).
2. Add 150 μL of prewashed streptavidin beads to your sample.
3. Rotate for 20 min at RT.
4. Place the tubes into an appropriate magnet and wait until the solution clears.
5. Discard the supernatant and resuspend the beads in 300 μL 1× TBW.
6. Repeat **steps 4** and **5**.
7. Place the tubes into an appropriate magnet and wait until the solution clears.

8. Discard the supernatant and resuspend the beads in 100 μL 0.1× TE.
9. Place the tubes into an appropriate magnet and wait until the solution clears.
10. Resuspend the beads in 50 μL 0.1× TE and transfer to PCR reaction tubes.
11. Follow the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® protocol **step 1** (NEBNext End Prep) (*see* **Note 8**).
12. Follow the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® protocol **step 2** (Adaptor Ligation).
13. After USER® Enzyme treatment, place the tubes into an appropriate magnet and wait until the solution clears.
14. Discard the supernatant and resuspend the beads in 300 μL 1x TBW.
15. Place the tubes into an appropriate magnet and wait until the solution clears.
16. Discard the supernatant and resuspend the beads in 100 μL 0.1× TE.
17. Place the tubes into an appropriate magnet and wait until the solution clears.
18. Discard the supernatant and resuspend the beads in 20 μL 0.1× TE.

3.7 Library Amplification

1. Add 180 μL "PCR Amplification Mix" containing the preferred NEB index primers.
2. PCR amplify the library in a thermocycler according to the primer specific protocol (*see* **Note 9**).
3. Purify the DNA with AMPure XP beads at a ratio of 1× according to manufactures protocol and elute in 20 μL 0.1× TE (*see* **Note 10**).
4. Determine the concentration with Qubit Fluorometer and Agilent Bioanalyzer (or similar).
5. Measure your sample on Illumina sequencing platforms in paired-end mode according to your sequencing provider's requirements (*see* **Note 11**).

4 Notes

1. Prepare Master mixes shortly before use and keep on ice until use. If a master mix for multiple reactions is prepared, use an additional 10% of each component to account for pipetting variability.

2. Cells of some particular cell types tend stick to the tube walls. This motivates the addition of BSA at a concentration of 0.5% to the DPBS buffers in **step 3** of Subheading 3.2 and **step 3** of Subheading 3.3.
3. The degree of chromatin digestion is the most crucial parameter in Micro-C experiments. A careful titration of the MNase digestion degree is essential.
4. If no aliquots of smaller cell numbers are available for the MNase titration, larger cell aliquots can be thawed on ice, resuspended in DPBS, aliquoted as convenient and snap-frozen for a second time.
5. The MNase digestion degree is crucial for successful experiments. It is important to avoid overdigestion of nucleosomes as overdigested nucleosomes are inefficiently ligated during the proximity ligation step. Overdigested nucleosomes appear slightly shifted in MNase titrations (*see* Fig. 1 lane 2). Underdigested chromatin contains a larger fraction of dinucleosomal fragments (~300 bp). These fragments have a similar size as Micro-C-XL ligation products and may reduce the number of 3D informative reads in deep-sequencing experiments. For an example of underdigested chromatin, *see* Fig. 1 (lane 5).
6. The digestion degree is sensitive to the digestion reaction volume for larger volumes. Routinely, MNase titration is performed in a volume of 50 μL and the batch preparation in a volume of 200 μL. Scaling up to larger volumes is not recommended and multiple digestions at volumes of 200 μL are preferred.
7. The reaction conditions have successfully been applied to material from up to five million cells starting material. If you want to process more cells, we recommend processing them in aliquots of five million cells in multiple reactions.
8. Follow the NEBNext End Prep and Adaptor Ligation steps (NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® protocol **steps 1** and **2**). Do not proceed with the bead-based DNA size selection and purification step as DNA is still bound to streptavidin beads.
9. Depending on the amount of your starting material, 8–-12 cycles are sufficient to amplify your libraries for sequencing. However, an approximation of how many cycles are required can be obtained by including a minimal-PCR step. Amplify 1 μL of your sample in a 10 μL PCR volume using 16 cycles. Determine the DNA concentration via Qubit (1 μL) and run the remaining 9 μL on a 1% agarose gel. This can be done without additional DNA purification. Compute the cycles for the preparative library assuming an exponential amplification and 20× the amount of starting material. The final DNA

amount should be at least 60 ng. The DNA agarose gel should show a clear band with a size of roughly 420 bp. There should be no 120 bp band visible, which would be indicative for adapter-dimers. A slight shadow might be visible at lower molecular weights that is caused by remaining PCR primers.

10. Because of the Streptavidin wash **steps 14–18** of Subheading 3.5 after adapter ligation, adapter dimers are minimal and no size selection step should be required after library amplification.
11. Before proceeding to very deep sequencing that is often required for 3C experiment analysis, it is recommended to sequence the sample to 5–10 million reads. The sample can be processed with dedicated 3C pipelines, such as distiller (https://github.com/open2c/distiller-nf) or Hi-C Pro [15]. Successful libraries should meet the following criteria: a low duplication rate and an even distribution of sequencing read orientations. The duplication rates for shallow sequenced samples should be below 5%. The sequencing read orientations are informative about the ligation efficiency and the purification of proximity ligated products. Mono- and dinucleosomes that are not products of proximity ligation will always yield sequencing read pairs with one read in forward and one read in reverse orientation. The distance between these two mapped reads will be equal to the footprint of the sequenced particle, that is typically smaller than 500 bp. In contrast, proximity ligation products can result in all possible mapping orientations, that is forward-forward, reverse-reverse, forward-reverse, reverse-forward. Good Micro-C-XL libraries will yield a distribution of 25% of each of these combinations, with a slight overrepresentation of forward-reverse reads, which indicate contaminating, noninformative dinucleosomes. Typically, the abundance of forward-reverse reads is less than 30% (approximately 7% dimer contamination) and for excellent libraries close to 25%. Importantly, contaminating dimers can be filtered in silico and are ignored by default in many Hi-C pipelines (as diagonal proximal bins). Therefore, dinucleosome contamination affects the sequencing cost more than the experimental measurement.

References

1. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558):1306–1311. https://doi.org/10.1126/science.1067799
2. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science

326(5950):289–293. https://doi.org/10.1126/science.1181369

3. Denker A, de Laat W (2016) The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev 30(12):1357–1382. https://doi.org/10.1101/gad.281964.116
4. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159(7):1665–1680. https://doi.org/10.1016/j.cell.2014.11.021
5. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485(7398):376–380. https://doi.org/10.1038/nature11082
6. Hsieh TS, Cattoglio C, Slobodyanyuk E, Hansen AS, Rando OJ, Tjian R, Darzacq X (2020) Resolving the 3D landscape of transcription-linked mammalian chromatin folding. Mol Cell 78(3):539–553 e538. https://doi.org/10.1016/j.molcel.2020.03.002
7. Hsieh TS, Fudenberg G, Goloborodko A, Rando OJ (2016) Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. Nat Methods 13(12):1009–1011. https://doi.org/10.1038/nmeth.4025
8. Krietenstein N, Abraham S, Venev SV, Abdennur N, Gibcus J, Hsieh TS, Parsi KM, Yang L, Maehr R, Mirny LA, Dekker J, Rando OJ (2020) Ultrastructural details of mammalian chromosome architecture. Mol Cell 78(3):554–565 e557. https://doi.org/10.1016/j.molcel.2020.03.003
9. Oksuz BA, Yang L, Abraham S, Venev SV, Krietenstein N, Parsi KM, Ozadam H, Oomen ME, Nand A, Mao H, Genga RM, Maehr R, Rando OJ, Mirny LA, Gibcus JH, Dekker J (2020) Systematic evaluation of chromosome conformation capture assays. bioRxiv:2020.2012.2026.424448. https://doi.org/10.1101/2020.12.26.424448
10. Hsieh TH, Weiner A, Lajoie B, Dekker J, Friedman N, Rando OJ (2015) Mapping nucleosome resolution chromosome folding in yeast by micro-C. Cell 162(1):108–119. https://doi.org/10.1016/j.cell.2015.05.048
11. Axel R (1975) Cleavage of DNA in nuclei and chromatin with staphylococcal nuclease. Biochemistry 14(13):2921–2925. https://doi.org/10.1021/bi00684a020
12. Lieleg C, Krietenstein N, Walker M, Korber P (2015) Nucleosome positioning in yeasts: methods, maps, and mechanisms. Chromosoma 124(2):131–151. https://doi.org/10.1007/s00412-014-0501-x
13. Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ (2005) Genome-scale identification of nucleosome positions in S. cerevisiae. Science 309(5734):626–630. https://doi.org/10.1126/science.1112178
14. Hansen AS, Hsieh T-HS, Cattoglio C, Pustova I, Saldaña-Meyer R, Reinberg D, Darzacq X, Tjian R (2019) Distinct classes of chromatin loops revealed by deletion of an RNA-binding region in CTCF. Mol Cell 76(3):395–411.e313. https://doi.org/10.1016/j.molcel.2019.07.039
15. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E (2015) HiC-pro: an optimized and flexible pipeline for hi-C data processing. Genome Biol 16(1):259. https://doi.org/10.1186/s13059-015-0831-x

Chapter 18

In Situ HiC

Timothy M. Johanson and Rhys S. Allan

Abstract

In situ HiC uses the relative frequency of DNA–DNA ligation events to reconstruct the three-dimensional architecture of a genome. As such, restriction enzyme digested ends of genomic DNA within fixed nuclei are tagged with biotinylated dNTPs. DNA–DNA ligation events generated via proximity ligation are then captured, amplified and next generation sequenced to determine their linear genomic position, but also their three-dimensional relationship. Here, we describe these steps in detail.

Key words In situ HiC, Chromosome conformation capture, Genome organization, Chromatin architecture

1 Introduction

All complex genomes are intricately and hierarchically organized in three dimensions. These three-dimensional DNA structures vary vastly in size, from the compaction of entire chromosomes to elegant loops of DNA joining gene promoters to distant regulatory regions [1]. While three-dimensionally organized genomes were assumed for many years, the definitive demonstration of gene-regulatory genome architecture was made possible by the invention of chromosome conformation capture (3C) technology [2].

Chromosome conformation capture in all its forms, from the original 3C [2] to 4C [3], 5C [4], in situ HiC [5], and beyond [6–11], use the relative frequency of DNA–DNA ligation events to quantify the proximity of regions of DNA to others. While groundbreaking, early iterations could only reveal the proximity of a handful of regions to others. In situ HiC was revolutionary in being able to reveal the physical relationship between all regions and all others, enabling exploration of the three-dimensional architecture of entire genomes.

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_18,

While today there are numerous versions of in situ HiC, from low input [6, 7, 9, 11], further enriched [8, 10] to industrially produced kits, the original protocol (outlined below) is the bed rock of these protocols, the gold standard for many laboratories and an accessible starting point for those unfamiliar with chromosome conformation capture technologies. The procedure involves formaldehyde fixation, lysis, restriction enzyme digestion, DNA end tagging, proximity ligation, cross-link reversal, DNA shearing and size selection, biotin-enrichment, adaptor ligation and PCR.

As with most next generation sequencing based technologies, much of the challenge in deriving meaning from in situ HiC lies in the bioinformatic alignment and analysis of the resultant DNA reads. The paired and potentially genomically distant nature of the reads makes in situ HiC libraries particularly challenging. Thus, the various bioinformatic methods for examining the data should be explored prior to performing the protocol [12].

2 Materials

Prepare all solutions using ultrapure water and analytical grade reagents.

1. 1× Phosphate Buffered Saline (PBS): 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4.
2. 4% buffered formaldehyde, pH 6.9.
3. 2.5 M glycine.
4. HiC lysis buffer: 10 mM Tris–HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630, and 50 μL of protease inhibitor cocktail.
5. 10% sodium dodecyl sulfate (SDS) in water.
6. 10% Triton X-100 in water (v/v).
7. MboI restriction enzyme and 10× MboI buffer.
8. Individual reagents for fill-in master mix: 0.4 mM biotin-14-dATP, 10 mM dCTP, 10 mM dGTP, 10 mM dTTP, 5 U/μL DNA polymerase I, Large (Klenow) Fragment and 10× Klenow buffer.
9. 10× T4 DNA ligase buffer and 400 U/μL T4 DNA Ligase.
10. 20 mg/mL bovine serum albumin.
11. 20 mg/mL Proteinase K.
12. 5 M sodium chloride.
13. 100% ethanol.
14. 70% ethanol in water: make from 100% ethanol.
15. 3 M sodium acetate, pH 5.2 with glacial acetic acid.
16. 1× Tris–HCl buffer: 10 mM Tris–HCl pH 8.0.

17. AMPure XP beads.
18. Low-bind DNA tubes.
19. Dynabeads MyOne Streptavidin T1 beads, 10 mg/mL.
20. 2× Binding Buffer: 10 mM Tris–HCl pH 7.5, 1 mM EDTA, 2 M NaCl.
21. 1× Tween Washing Buffer (TWB): 5 mM Tris–HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl, and 0.05% Tween-20.
22. 25 mM dNTP mix.
23. 10× T4 DNA ligase buffer with 10 U/μL T4 Polynucleotide Kinase (PNK).
24. 3 U/μL T4 DNA polymerase I.
25. 10 mM dATP.
26. 5 U/μL Klenow exo minus.
27. 10× Quick ligation reaction buffer and DNA Quick ligase.
28. Illumina indexed adapters.
29. Phusion Taq polymerase and 5× Phusion HF buffer.
30. Instrument: Covaris S220 (Covaris, Woburn, MA).
31. Magnet rig.
32. PCR machine.
33. Rotator.
34. Heat block.

3 Method

3.1 Cross-Linking

1. Pellet 2×10^5 to 5×10^6 cells by centrifugation at $300 \times g$ for 5 min.
2. Resuspend cells in fresh medium (cultured cells) or 1× PBS (primary cells) at a concentration of 1×10^6 cells per 1 mL.
3. In a fume hood, add fresh formaldehyde to a final concentration of 1% in media.
4. Mix at room temperature for 10 min (*see* **Note 1**).
5. Add 2.5 M glycine solution to a final concentration of 0.2 M (*see* **Note 2**).
6. Mix at room temperature for 5 min.
7. Pellet by centrifugation for 5 min at $300 \times g$ at 4 °C.
8. Discard all supernatant into an appropriate formaldehyde waste container.
9. Resuspend cells in 1 mL of ice-cold 1× PBS.
10. Pellet by centrifugation for 5 min at $300 \times g$ at 4 °C.

11. Discard supernatant and flash-freeze cell pellets in liquid nitrogen or dry ice/ethanol (*see* **Note 3**).
12. Cell pellets can be stored at −80 °C if required.

3.2 Lysis and Restriction Digestion

1. Add 250 μL of ice-cold HiC lysis buffer to each crosslinked cell pellet.
2. Incubate on ice for 20 min (*see* **Note 4**).
3. Pellet by centrifugation at 2500 × *g* for 5 min.
4. Discard the supernatant.
5. Resuspend cell pellet in 500 μL of ice-cold HiC lysis buffer.
6. Pellet by centrifugation at 2500 × *g* for 5 min.
7. Gently resuspend pellet in 50 μL of 0.5% SDS in water (*see* **Note 5**).
8. Incubate at 62 °C for 10 min.
9. Add 145 μL of water, 25 μL of 10% Triton X-100 (*see* **Note 6**).
10. Mix well and gently.
11. Incubate at 37 °C for 15 min (*see* **Note 7**).
12. Add 25 μL of 10× MboI restriction enzyme buffer and 100 U of MboI restriction enzyme.
13. Digest chromatin overnight or for at least 2 h at 37 °C with rotation.

3.3 Tagging of DNA Ends, Proximity Ligation, and Cross-Link Reversal

1. Incubate cell suspensions at 62 °C for 20 min to inactivate MboI (*see* **Note 4**).
2. Cool samples to room temperature.
3. Add 50 μL of fill-in master mix (*see* **Note 8**): 37.5 μL of 0.4 mM biotin-14-dATP, 1.5 μL of 10 mM dCTP, 1.5 μL of 10 mM dGTP, 1.5 μL of 10 mM dTTP, 8 μL of 5 U/μL DNA Polymerase I, Large (Klenow) Fragment.
4. Incubate at 37 °C for 90 min with rotation.
5. Add 900 μL of ligation master mix: 669 μL of water, 120 μL of 10× T4 DNA ligase buffer, 100 μL of 10% Triton X-100, 6 μL of 20 mg/mL Bovine Serum Albumin, 5 μL of 400 U/μL T4 DNA Ligase.
6. Mix by inverting and incubate at room temperature for 4 h with slow rotation (*see* **Note 9**).
7. Pellet nuclei at 2500 × *g* for 5 min and discard supernatant (*see* **Note 10**).
8. Resuspend in 1210 μL of the following solution (ligation master mix without T4 DNA Ligase): 984 μL of water, 120 μL of 10x T4 DNA ligase buffer, 100 μL of 10% Triton X-100, 6 μL of 20 mg/mL Bovine Serum Albumin.

9. Add: 50 μL of 20 mg/mL proteinase K, 120 μL of 10% SDS in water.
10. Incubate at 55 °C for 30 min (*see* **Note 11**).
11. Add 130 μL of 5 M sodium chloride (*see* **Note 12**).
12. Incubate at 68 °C overnight or for at least 90 min.

3.4 DNA Shearing and Size Selection

1. Cool samples to room temperature (*see* **Notes 13** and **14**).
2. Split each sample into three 500-μL aliquots.
3. Mix with 1.6× volumes of pure ethanol.
4. Add 0.1× volumes of 3 M sodium acetate, pH 5.2.
5. Mix by inverting and incubate at −80 °C for 15 min (*see* **Note 15**).
6. Centrifuge at 17,000 × *g* at 4 °C for 15 min.
7. Carefully remove the supernatant by pipetting (*see* **Note 16**).
8. Resuspend and combine the three aliquots in 800 μL of 70% ethanol in water (*see* **Note 17**).
9. Centrifuge at 17,000 × *g* for 5 min.
10. Remove all supernatant.
11. Wash the pellet once more with 800 μL of 70% ethanol in water (*see* **Note 18**).
12. Dissolve the DNA pellet in 130 μL of 1× Tris–HCl buffer.
13. Incubate at 37 °C for 15 min.
14. Load the DNA solution into a Covaris microtube (*see* **Note 19**).
15. Shear DNA into 300–500 bp fragments (*see* **Note 20**) using a Covaris S220 and the following parameters: Fill Level: 10, Duty Cycle: 10, PIP: 175, Cycles/Burst: 200, Time: 58 s.
16. Transfer sheared DNA to a fresh 1.5 mL tube.
17. Wash the Covaris vial with 70 μL of water and add this to the sample. The total sample volume should be 200 μL.
18. For libraries containing fewer than 2×10^6 cells, the size selection using AMPure XP beads described in **steps 19–31** should be performed on the final amplicons rather than before biotin pull-down. If so, bring sample volumes to 300 μL each with water and skip to Subheading 3.5.
19. Verify successful shearing using either traditional agarose gel electrophoresis or a TapeStation automated electrophoresis machine.
20. Add exactly 0.55× volumes of AMPure XP beads to the reaction (110 μL in 200 μL) (*see* **Note 21**).
21. Mix well by pipetting and incubate at room temperature for 5 min.

22. Use a magnet rig to transfer the solution to a fresh tube, avoiding any beads (*see* **Note 22**).
23. Add exactly 30 μL of fresh AMPure XP beads to the solution.
24. Mix by pipetting and incubate at room temperature for 5 min.
25. Use a magnet rig to isolate the beads. Discard the clear solution (*see* **Note 23**).
26. Wash the in-place beads twice with 700 μL of 70% ethanol in water without mixing.
27. Allow 5 min for ethanol to evaporate from the in-place beads.
28. Resuspend beads in 300 μL of 1× Tris–HCl buffer.
29. Incubate at room temperature for 5 min.
30. Isolate solution using a magnet rig and transfer to a fresh 1.5 mL tube. Discard beads.
31. Verify successful size selection using either traditional agarose gel electrophoresis or a TapeStation automated electrophoresis machine.

3.5 Biotin Pull-down and Preparation for Illumina Sequencing

Perform all of the following steps in ***low-bind tubes****.*

1. Wash 150 μL of Dynabeads MyOne Streptavidin T1 beads with 400 μL of 1× TWB (*see* **Note 24**).
2. Use a magnet rig to discard the solution.
3. Resuspend the beads in 300 μL of 2× Binding Buffer.
4. Add washed beads to the DNA solution from either **step 18** or **30** from Subheading 3.4, depending on starting cell number.
5. Incubate at room temperature for 15 min with rotation.
6. Isolate the beads using a magnet rig. Discard the solution.
7. Wash the beads by adding 600 μL of 1× TWB and transferring the mixture to a new tube.
8. Heat the tubes at 55 °C for 2 min with 800 rpm on a benchtop shaker/incubator.
9. Isolate the beads using a magnet rig.
10. Repeat **steps 7–9**. These two washes will be hereafter referred to as the TWB washes.
11. Resuspend beads in 100 μL 1× T4 DNA ligase buffer (*see* **Note 25**) and transfer to a new tube.
12. Isolate the beads using a magnet rig.
13. Resuspend beads in 100 μL of master mix (*see* **Note 26**): 88 μL of 1× T4 DNA ligase buffer with 10 mM ATP, 2 μL of dNTP mix (25 mM), 5 μL of 10 U/μL T4 PNK, 4 μL of 3 U/μL T4 DNA polymerase I, 1 μL of 5 U/μL DNA polymerase I, Large (Klenow) Fragment.

14. Incubate at room temperature for 30 min.
15. Isolate the beads using a magnet rig.
16. Perform 2× TWB washes (**steps 7–9**).
17. Resuspend beads in 100 μL 1× Klenow buffer (we used NEB Buffer 2) and transfer to a new tube.
18. Isolate the beads using a magnet rig.
19. Resuspend beads in 100 μL of dATP attachment master mix (*see* **Note 27**): 90 μL of 1× Klenow buffer, 5 μL of 10 mM dATP, 5 μL of 5 U/μL NEB Klenow exo minus.
20. Incubate at 37 °C for 30 min.
21. Isolate the beads using a magnet rig.
22. Perform 2× TWB washes (**steps 7–9**).
23. Resuspend beads in 100 μL 1×Quick ligation reaction buffer and transfer to a new tube.
24. Isolate the beads using a magnet rig.
25. Resuspend in 50 μL of 1× Quick ligation reaction buffer.
26. Add 2 μL of DNA Quick ligase, 3 μL of an Illumina indexed adapter. Record the sample-index combination.
27. Mix thoroughly and incubate at room temperature for 15 min.
28. Isolate the beads using a magnet rig.
29. Perform 2×TWB washes (**steps 7–9**).
30. Resuspend beads in 100 μL 1×Tris–HCl buffer and transfer to a new tube.
31. Isolate the beads using a magnet rig.
32. Resuspend in 50 μL of 1× Tris–HCl buffer.
33. Heat samples at 98 °C for 10 min.
34. Quickly spin the tubes to collect precipitate, then isolate the solution using a magnet rig. Discard beads (*see* **Note 28**).

3.6 Final Amplification and Purification

1. Amplify the HiC library with 4–12 cycles of PCR, using Illumina primers and the protocol shown in Table 1.
2. After the amplification is complete, bring the total library volume to 250 μL with water.
3. Add 175 μL of room temperature homogenous AMPure XP beads to the PCR reaction (0.7× volumes).
4. Mix by pipetting and incubate at room temperature for 5 min.
5. Isolate the beads using a magnet rig.
6. Wash in-place beads once with 700 μL of 70% ethanol in water.
7. Remove ethanol completely.
8. Resuspend beads in 100 μL of 1×Tris–HCl buffer.

Table 1
PCR protocol for the amplification of the HiC library

PCR recipe	*Program*	
34.1 μL DNA	98 °C, 30 s	
10 μL 5× Phusion HF buffer 0.4 μL dNTPs (25 mM) 5 μL PCR primer cocktail (Illumina)	98 °C, 10 s 60 °C, 30 s 72 °C, 30 s	×8 (9 total cycles)
0.5 μL Phusion Taq polymerase	72 °C, 5 min	

9. Add another 70 μL of AMPure XP beads.
10. Mix by pipetting and incubate at room temperature for 5 min.
11. Isolate the beads using a magnet rig.
12. Wash in-place beads twice with 700 μL of 70% ethanol in water.
13. Leave the beads on the magnet for 5 min to allow the remaining ethanol to evaporate.
14. Resuspend beads in 25 μL of 1×Tris–HCl buffer.
15. Incubate at room temperature for 5 min.
16. Separate on a magnet and transfer the solution to a new labeled tube.

4 Notes

1. Unless otherwise stated, mixing is performed at 6 full revolutions a minute. We use an Intelli-Mixer RM-2 (Daigger).
2. Glycine is an amino acid used in high concentrations to quench the formaldehyde fixation.
3. Some protocols [6] perform the snap freeze after lysis.
4. A small aliquot of each sample can be taken after this incubation step to check the integrity of nuclei on a bench-top light microscope. The cell solution is pelleted by centrifugation at 2500 × *g* for 5 min, resuspended in 10 μL PBS, and examined under a light microscope. Nuclei should be intact and rounded.
5. SDS removes any non–cross-linked proteins, and partly denatures chromatin. This dramatically increases the accessibility of DNA to the subsequent restriction enzyme digestion.
6. Triton X-100 sequesters SDS to allow restriction enzymes to function.
7. A small aliquot can be taken at this stage as a predigestion control. For this, take up to 10% volume of the lysed cell solution and incubate for 30 min at 65 °C with 5× volume of

1× NEBuffer 2 and 0.5× volume of Proteinase K (20 mg/mL). Add 200 μL of 0.5 M NaCl and 200 μL of phenol–chloroform. Mix by vortexing, then centrifuge at maximum speed (e.g., >13,000 × *g*) for 5 min. Carefully transfer the aqueous phase to a new tube before incubating for 15 min at 37 °C with 2.5 μL of RNase A (20 mg/mL). Check the quality of the sample by running it on a 0.7% agarose gel or on a TapeStation. Good quality DNA will run as a single high molecular weight band (>23 kb). Fragmentation suggests poor quality DNA and restarting with a fresh sample should be considered.

8. Fill-in master mix fills in the overhangs left by Mbo1 digestion with dNTPs including biotinylated d-ATP using Klenow 3′ to 5′ exonuclease activity. The reaction creates blunt-ended DNA by filling in 5′ overhangs and degrading 3′ overhangs. Of further note, if the initial cell number is below 5×10^5, half the amount of dNTPs (including biotinylated dATP) can be used. 8 μL of Klenow enzyme is still used.
9. Use this time to preheat an incubator to 68 °C.
10. This centrifugation step removes or reduces the number of captured ligation events that occurred outside of nuclei, and are therefore unlikely to reflect biologically relevant DNA–DNA interactions.
11. Proteinase K removes all remaining proteins and disrupts the nuclear membrane.
12. High concentrations of sodium chloride inhibit DNA ligase.
13. Bring a bottle of AMPure XP beads to room temperature in anticipation of **step 20**, Subheading 3.4.
14. Some more recent iterations of the in situ HiC protocol use a phenol–chloroform clean up in place of the ethanol precipitation outlined here.
15. This process will precipitate the DNA allowing it to be separated from solution by centrifugation.
16. Keep the tubes of ice after centrifugation.
17. For samples with $<5 \times 10^5$ initial cells, 400 μL should be used to resuspend and combine the split sample, while the other 400 μL can be used to wash out residual material from each of the three tubes.
18. The DNA pellet will often be less visible, if not invisible, to the naked eye after the 70% ethanol wash. Remove the ethanol carefully.
19. Covaris microtubes have a slitted cap and an internal glass column. The best way to load them is to tilt them and pierce the slit above the glass column with your sample-loaded pipette

tip. They should be loaded from the bottom upward to avoid low lying bubbles. Bubbles within the tube body will impair shearing. Bubbles near the cap are not problematic.

20. 300–500 bp fragments are optimal for subsequent Illumina sequencing.
21. The liquid phase of AMPure XP mixture precipitates DNA onto the beads. The size of the DNA molecules precipitated depends on the ratio of the volume of the AMPure XP liquid phase to the volume of the sample. Increasing the proportion of the AMPure XP beads decreases the cutoff size of the DNA.
22. The supernatant will contain fragments shorter than 500 bp.
23. Fragments in the range of 300–500 bp will be retained on the beads.
24. Make sufficient TWB (5.2 mL) for each sample for the entire protocol. Of further note, some recent iterations of the protocol use Dynabeads MyOne Streptavidin C1 beads and the NEBNext Ultra DNA Library Prep Kit [6] in place of **steps 1–32** in Subheading 3.5. In our experience this greatly improves library yield.
25. Stock NEB T4 DNA ligase buffer is 10×. Dilute prior to use here and in subsequent steps.
26. T4 Polynucleotide Kinase adds 5′ phosphates, facilitating subsequent ligation, while T4 DNA polymerase I degrades 3′ overhangs (~200× more efficiently than Klenow) and also fills in 5′ overhangs. The addition of both varieties of polymerase is an old technique and not wholly explored or understood.
27. dATP attachment master mix adds an adenosine to the 3′ end of the blunted DNA fragments to allow Illumina adaptor ligation.
28. While it is possible to perform amplification with the beads present, there are some reports that the beads can interfere with the subsequent PCR; thus, we routinely remove them.

References

1. Furlan-Magaril M, Varnai C, Nagano T, Fraser P (2015) 3D genome architecture from populations to single cells. Curr Opin Genet Dev 31:36–41. https://doi.org/10.1016/j.gde.2015.04.004
2. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558):1306–1311. https://doi.org/10.1126/science.1067799
3. Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38(11):1348–1354. https://doi.org/10.1038/ng1896
4. Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J (2006) Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res

16(10):1299–1309. https://doi.org/10.1101/gr.5571506

5. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159(7):1665–1680. https://doi.org/10.1016/j.cell.2014.11.021
6. Diaz N, Kruse K, Erdmann T, Staiger AM, Ott G, Lenz G, Vaquerizas JM (2018) Chromatin conformation analysis of primary patient tissue using a low input Hi-C method. Nat Commun 9(1):4938. https://doi.org/10.1038/s41467-018-06961-0
7. Du Z, Zheng H, Huang B, Ma R, Wu J, Zhang X, He J, Xiang Y, Wang Q, Li Y, Ma J, Zhang X, Zhang K, Wang Y, Zhang MQ, Gao J, Dixon JR, Wang X, Zeng J, Xie W (2017) Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature 547(7662):232–235. https://doi.org/10.1038/nature23263
8. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Varnai C, Thiecke MJ, Burden F, Farrow S, Cutler AJ, Rehnstrom K, Downes K, Grassi L, Kostadima M, Freire-Pritchett P, Wang F, Consortium B, Stunnenberg HG, Todd JA, Zerbino DR, Stegle O, Ouwehand WH, Frontini M, Wallace C, Spivakov M, Fraser P (2016) Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell 167(5):1369–1384 e1319. https://doi.org/10.1016/j.cell.2016.09.037
9. Ke Y, Xu Y, Chen X, Feng S, Liu Z, Sun Y, Yao X, Li F, Zhu W, Gao L, Chen H, Du Z, Xie W, Xu X, Huang X, Liu J (2017) 3D chromatin structures of mature gametes and structural reprogramming during mammalian embryogenesis. Cell 170(2):367–381 e320. https://doi.org/10.1016/j.cell.2017.06.029
10. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, Herman B, Happe S, Higgs A, LeProust E, Follows GA, Fraser P, Luscombe NM, Osborne CS (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet 47(6):598–606. https://doi.org/10.1038/ng.3286
11. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature 502(7469):59–64. https://doi.org/10.1038/nature12593
12. Nicoletti C, Forcato M, Bicciato S (2018) Computational methods for analyzing genome-wide chromosome conformation capture data. Curr Opin Biotechnol 54:98–105. https://doi.org/10.1016/j.copbio.2018.01.023

Chapter 19

LncRNA–Chromatin Pull-Down Using Biotin-Conjugated DNA Probes

Debina Sarkar and Sarah D. Diermeier

Abstract

Long noncoding RNAs (lncRNAs) are a class of RNA molecules that have been associated with several important biological processes and linked to numerous diseases. Due to their cell type- and tissue specific expression, lncRNAs are involved in a wide range of molecular pathways. To fully understand how a lncRNA is linked to a biological process, its mechanism of action needs to be uncovered. Nuclear retained lncRNAs have been described to modulate gene expression directly or indirectly by interacting with chromatin and associated factors. Described here is an RNA pull-down strategy, which enables the identification of chromatin regions directly bound by a lncRNA of interest. This method is an important step toward investigating how lncRNAs regulate gene expression and/or chromatin states.

Key words lncRNA, Pull-down assay, Chromatin

1 Introduction

Long noncoding RNAs (lncRNAs) are a class of RNA molecules that are longer than 200 nucleotides and contain no or very short open reading frames [1]. The majority of lncRNAs is also characterized by cell type– and tissue-specific expression [1]. A growing number of lncRNAs have been linked to important biological and cellular processes [2] in development and diseases [3–5]. Thus far, detailed molecular mechanisms have been deciphered for only a few of the 17,948 currently annotated human lncRNAs (according to GENCODE v37) [1, 6]. LncRNAs have been linked to a wide range of processes associated with yeast and animal reproduction including sex determination [7, 8], meiosis [9], spermatogenesis [10] and imprinting [11]. LncRNAs also appear to play key roles in plants, which include sexual reproduction including floral transition [12], meiosis progression and anther and pollen development [13–15].

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_19,

Studies into the molecular mechanisms of these lncRNAs require robust methods, which can help identify interaction partners depending on the localization of a particular lncRNA [16]. LncRNAs are known to regulate gene expression directly and/or indirectly and the modes of action depend on whether they are present in the nucleus, the cytoplasm, or both [17]. While initial studies reported that the majority of lncRNAs is localized to the nucleus [18], recent publications indicate that the number of cytoplasmic lncRNAs is higher than previously estimated [19]. Bouvrette et al. reported that around 75% of lncRNAs are present in the cytoplasmic fraction of human and *Drosophila* cells [20]. Nuclear lncRNAs have been linked to crucial roles in regulating chromatin architecture. For instance, the lncRNA *XIST* induces a cascade of events, which involves chromatin remodeling leading to silencing of one of the two X-chromosomes during early embryonic development in female mammals [21]. LncRNAs can also promote or prevent the recruitment of chromatin modifiers. The lncRNA *Kcnq1ot1* is associated with imprinting, interacts with chromatin and with the H3K9- and H3K27-specific histone methyltransferase G9a and the PRC2 complex. This interaction results in lineage-specific transcriptional silencing in the *Kcnq1* domain [22]. On the other hand, the lncRNA *lncPRESS1* promotes pluripotency by interacting with SIRT6, a class III HDAC that removes the acetyl group from H3K56 and H3K9 [23]. *LncPRESS1* prevents the binding of SIRT6 at pluripotency gene promoters, resulting in high levels of H3K56 and H3K9 acetylation and active transcription [23]. Some nuclear lncRNAs are known to regulate mRNA transcription by formation of R-loops. Formation of R-loops tether lncRNA in *cis* and enable recruitment of transcription cofactors to the promoter regions [24]. The lncRNA *KHPS1* binds to *SPHK1* upstream of its transcription start site (TSS) and forms an R-loop that anchors *KHPS1*-interacting histone acetyltransferase p300/CBP to the *SPHK1* promoter [24]. This interaction increases local chromatin accessibility, which facilitates E2F1 binding to *SPHK1* [24]. In addition, nuclear lncRNAs can also influence gene expression indirectly by interfering with splicing. An example of a lncRNA that uses this mode of action is *MALAT1*, which regulates alternative splicing by modulating the levels of SR proteins [25].

In light of the various modes by which lncRNAs in the nucleus can operate, it is essential to identify directly bound target genes and chromatin regions. Two methodologies have been developed that enables the purification of chromatin regions directly interacting with an lncRNA, CHART-seq [26] and ChIRP-seq [27], which are conceptually similar. CHART-seq uses a target-specific 25-mer desthiobiotin-conjugated DNA oligonucleotide [26] and ChIRP-

seq uses tiling pools of 20-mer biotin-conjugated DNA oligonucleotide probes [27] to bind the target lncRNA. Described below is an RNA pull-down method based on the protocol by Chu et al. [27], using biotinylated probes targeting an lncRNA, which can be coupled to qRT-PCR or DNA-sequencing. Analysis of the identified genes and chromatin regions represents a first step toward elucidating the molecular role of the lncRNA.

2 Materials

Prepare all solutions using ultrapure or RNase-free water and analytical grade reagents.

2.1 Buffers and Reagents

1. 1× Phosphate-buffered saline (1× PBS, pH 7.4): dissolve 8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, and 0.24 g KH_2PO_4 in 800 mL of RNase-free water at room temperature (RT). Adjust the pH to 7.4 with HCl and make up the volume to 1 L.
2. 1% Glutaraldehyde (for crosslinking, prepare fresh): prepare 20 mL of 1% glutaraldehyde per 20 million cells at RT (add 0.8 mL of 25% glutaraldehyde stock solution with 19.2 mL of 1× PBS).
3. 1.25 M Glycine.
4. Cell lysis buffer: 150 mM NaCl, 25 mM Tris-Cl (pH 7.4), 1% Igepal 630, 1 mM EDTA, 5% glycerol. Supplement with 1× complete protease inhibitor cocktail (1 tablet in 10 mL of lysis buffer, Roche), 100 U/mL SUPERaseIN (Invitrogen), 100 mM PMSF (make 100× stock in isopropanol, Sigma-Aldrich) fresh before use.
5. Antisense DNA biotinylated probes with a biotin modification at the 3′ terminus of an oligonucleotide. ChIRP Probe Designer software can be used to design DNA probes/oligos available online at Biosearch.com (and ordered from Biosearch Technologies).
6. Streptavidin C1 Dynabeads (Invitrogen).
7. Proteinase K buffer (for DNA and RNA): 100 mM NaCl, 50 mM Tris–HCl (for RNA use pH 7.0 and for DNA use pH 8.0), 1 mM EDTA, 0.5% SDS. Add 5% v/v Proteinase K before use.
8. DNA elution buffer: 50 mM $NaHCO_3$, 1% SDS.
9. PCR purification kit.
10. RNase A.
11. RNase H.

12. TRIzol.
13. Phenol–chloroform–isoamyl alcohol.
14. Chloroform.
15. GlycoBlue.
16. Qubit high sensitivity (HS) RNA and Qubit HS DNA assay kits.
17. Buffer kit (Invitrogen).
18. 3 M Sodium Acetate: dissolve 123.05 g of sodium acetate in 400 mL of water. Adjust pH to 5.2 using glacial acetic acid and make up the volume to 500 mL.
19. Agarose.
20. 50X TAE (Tris-Acetate-EDTA) buffer: dissolve 242 gm Tris, 100 mL of 0.5 M EDTA, and 57.1 mL of glacial acetic acid. Make up the volume to 1 L using water. Make a working stock of 1× TAE before use.
21. Ethidium Bromide.
22. BlueJuice™ Gel Loading Buffer (10X).
23. 1 kb Plus DNA ladder.
24. SuperScript IV VILO Master Mix with ezDNase enzyme (Invitrogen).

2.2 Cell Culture

1. Cell culture hoods to culture and passage cell lines.
2. Incubator to maintain cells at 37 °C and 5% CO_2.
3. Cell culture media, fetal bovine serum (FBS) and antibiotics (penicillin and streptomycin) as required, 0.05% trypsin and 1× PBS (all sterile).

2.3 Other Equipment

1. Bioruptor® Pico and 0.2 mL Bioruptor tubes (Diagenode).
2. DynaMag-2 magnet (Invitrogen).
3. Pipette controller and sterile Stripette pipettes.
4. Micropipettes and filter tips. Sharp and long 10 μL filter pipette tip.
5. 1.7 mL microcentrifuge and 0.2 mL PCR tubes.
6. Refrigerated centrifuge (4 °C) with max capacity of 13,000 × *g*.
7. End to end rotating shaker.
8. Phase-lock gel tubes (Quantabio).
9. Benchtop mini centrifuge.
10. PCR thermocycler.

3 Methods

3.1 Harvesting Mammalian Cells

1. Grow cells in an appropriate culture medium to 80% confluency in 150 mm^3 plates (cell yield is approximately 20 million cells per plate for MDA-MB-231).
2. Rinse cells using half the volume of media (10 mL) with 1× PBS once.
3. Add 2–3 mL of 0.05% trypsin to the cells and incubate at 37 °C for 2–3 mins or until cells detach. Observe cells using a microscope to ensure their complete detachment from the surface of the plate. Gently tap the side of the plate to dislodge cells if necessary.
4. Add 4–6 mL of culture media (2× the volume of trypsin added) to quench the trypsin and resuspend cells into a single cell suspension by pipetting up and down gently a few times.
5. Transfer the cell suspension to a 50 mL falcon tube and count cells using a hemocytometer or an automated cell counter. Typically, 20 million cells are used per pull-down.
6. Spin cells down at 800 × *g* for 5 min at RT. Aspirate medium using a 10 mL pipette and resuspend 20 million cells in 20 mL of 1× PBS.
7. Spin again at 800 × *g* for 5 min at RT. Remove PBS and aspirate any residual PBS using a long and sharp 10 μL pipette tip, leaving just the cell pellet.

3.2 Crosslinking Using 1% Glutaraldehyde to Conserve Chromatin and RNA Interactions

1. Prepare 20 mL of 1% glutaraldehyde per 20 million cells at RT. Add 5 mL of 1% glutaraldehyde initially and tap the bottom of the 50 mL falcon tube to carefully dislodge the cell pellet. Top up with the remaining 15 mL of 1% glutaraldehyde, gently pipette up and down to resuspend cells and invert tubes further to mix.
2. Crosslink for 10 min at RT using an end-to-end shaker or rotor.
3. Quench the crosslinking reaction by adding 2 mL of 1.25 M glycine (1/10th volume of 1% glutaraldehyde added in **step 1**) at RT for 5 min on the shaker. The colour of the reaction will turn from clear to yellow.
4. Spin cells at 2000 × *g* for 5 min at RT.
5. Aspirate the supernatant and wash the cell pellet with 20 mL of 1× PBS (prechilled at 4 °C).
6. Spin cells at 2000 × *g* for 5 min at RT.
7. Aspirate the supernatant and resuspend the pellet in 1 mL of prechilled 1× PBS and transfer the resuspended cells to a microcentrifuge tube.

8. Spin cells at 2000 × *g* for 5 min at 4 °C. Remove PBS carefully with a long and sharp pipette tip without disturbing the pellet.
9. Snap-freeze cell pellets in liquid nitrogen and store at −80 °C indefinitely or continue on to cell lysis and sonication immediately.

3.3 Cell Lysis and Sonication to Shear Chromatin

1. If using a frozen cell pellet, thaw on ice.
2. Tap the bottom of the microcentrifuge tube carefully to dislodge the cell pellet.
3. Spin at 2000 × *g* for 5 min at 4 °C.
4. Aspirate any residual PBS using a long sharp pipette tip.
5. Tare the weight of an empty microcentrifuge tube and record the weight of the cell pellet.
6. Supplement cell lysis buffer by adding protease inhibitor (50× stock), PMSF (100× stock) and SUPERaseIn (100 U/mL).
7. Add 10× the volume of cell lysis buffer to each cell pellet. For example, add 500 μL cell lysis buffer to a cell pellet that weighs 50 mg. Mix well to get a smooth suspension by pipetting up and down, then proceed immediately to sonication.
8. Sonication conditions need to be optimized to obtain fragmented chromatin of 100–500 bp in length. Sonication conditions for this protocol have been optimized using a Bioruptor® Pico. Settings for other instruments may differ and have to be determined experimentally (*see* **Note 1**).
9. Sonicate cell lysate using optimized sonication settings. Transfer and combine lysates in a new microcentrifuge tube with up to 1 mL in each tube.
10. Centrifuge sonicated samples at 16,000 × *g* for 10 min at 4 °C. Transfer supernatant (= chromatin) to a new tube and continue on to the pull-down step. If stopping at this step, snap-freeze in liquid nitrogen to store at −80 °C indefinitely.

3.4 Chromatin Pull-Down Using Biotinylated DNA Probes

1. If using frozen samples, thaw tubes of chromatin at RT. Use 500 μL of chromatin for each pull-down.
2. Take out 5 μL (1%) for DNA and RNA input each and set aside on ice.
3. Thaw lncRNA-specific probes (see at RT and add 100 pmol of probe per reaction. Mix well and incubate at RT for 1.5 h with shaking (*see* **Notes 2** and **3**).
4. With 20 min remaining for hybridization, prepare 100 μL of streptavidin C1 magnetic beads per pull-down. Let the beads separate on a magnetic stand for 1 min and remove the supernatant carefully without disturbing the beads, with tubes still on the magnetic stand.

5. Take the tube(s) containing magnetic beads off the magnetic stand and add 800 μL of un-supplemented lysis buffer to the beads (800 μL to 100 μL of beads).
6. To wash the beads, resuspend and gently pipette 10× up and down to resuspend beads followed by a quick spin using a mini benchtop centrifuge and place the tubes back on the magnetic stand.
7. Let the beads separate for 1 min and aspirate the supernatant.
8. Repeat **steps 5**–7 for a total of three washes.
9. Add RNA pull-down reaction to the beads and incubate samples for 30 min at RT with shaking.
10. Separate beads on a magnetic stand for 1 min. Remove the supernatant without touching or disturbing the beads.
11. Add 800 μL of un-supplemented lysis buffer to each pull-down reaction and wash the beads three times as done previously (**steps 5**–7) using a magnetic stand.
12. After the last wash, resuspend beads in 500 μL of supplemented lysis buffer. Transfer 50 μL to a new microcentrifuge tube for RNA isolation and keep the remaining 450 μL for DNA isolation.
13. Place all tubes (includes samples for both DNA and RNA isolation) on a magnetic stand, separate beads for 1 min and aspirate the supernatant using a long and sharp 10 μL pipette tip to remove any residual liquid.

3.5 RNA Isolation for Quantification Using qRT-PCR to Check RNA Enrichment

1. Take the tubes kept aside containing 50 μL of RNA pull-down reaction beads (from **step 12** in Subheading 3.4) and the 5 μL RNA input (from **step 2** in Subheading 3.4). Add 42.5 μL of RNA PK buffer (pH 7.0) to the RNA input and resuspend the beads (from **step 12** in Subheading 3.4) in 47.5 μL RNA PK buffer. Add 2.5 μL of Proteinase K to both tubes. The total volume in the RNA input and RNA pull-down samples should each be 50 μL.
2. Incubate at 50 °C for 45 min for reverse crosslinking.
3. Spin briefly and boil samples for 10 min at 95 °C.
4. Chill on ice immediately and add 500 μL of TRIzol reagent, vortex and incubate at RT for 10 min. Store at −80 °C for up to a month or proceed straight to RNA isolation.
5. Add 100 μL of chloroform and vortex for 10 s. If starting with frozen samples, thaw samples at RT and then add 100 μL of chloroform. Centrifuge at 12,000 × *g* for 15 min at 4 °C.
6. Carefully remove and transfer the aqueous phase to a new microcentrifuge tube by holding the tubes at a 75° angle without touching or disturbing the interphase and the phenol phase.

7. Add 250 μL of isopropanol and 1 μL of GlycoBlue and incubate for 10 min at RT.
8. Centrifuge at 12,000 × *g* for 30 min at 4 °C. Aspirate the supernatant carefully.
9. Wash RNA pellet with 1 mL of freshly prepared 75% ethanol.
10. Centrifuge at 7500 × *g* for 5 min at 4 °C. Carefully remove the supernatant. Repeat this step again for a total of two washes.
11. Air dry tubes for 10 min or until all the ethanol evaporates (*see* **Note 4**).
12. Resuspend RNA in 10 μL of RNase-free water and keep on ice.
13. Use 1 μL to check the RNA concentration using a Qubit HS RNA kit (RNA concentrations typically range between 20–30 ng). Store RNA at −80 °C indefinitely or proceed to cDNA synthesis and qRT-PCR directly (follow-on from **step 1** in Subheading 3.7).

3.6 DNA Isolation from Pull-Down Samples from Beads for Downstream Analysis

1. Prepare DNA elution buffer fresh, 75 μL per sample, and supplement by adding 0.75 μL of RNase A (10 mg/mL) and 1.5 μL of RNase H (5 U/μL) to the DNA elution buffer. Resuspend beads in 75 μL of supplemented DNA elution buffer per RNA pull-down reaction and add 70 μL of supplemented DNA elution buffer to the DNA input sample.
2. Incubate for 30 min at 37 °C to elute DNA from the beads.
3. Spin briefly using a mini benchtop centrifuge and place tubes on a magnetic stand to separate the beads from the supernatant. Transfer the supernatant containing the eluted DNA to newly labelled microcentrifuge tubes.
4. To ensure complete elution of DNA from the beads prepare and add another 75 μL of DNA elution buffer containing the same amounts of RNase A and RNase H (*see* **step 1**) to the beads. Incubate for 30 min at 37 °C and repeat **steps 2** and **3**. Transfer the supernatant containing eluted DNA to the tubes containing the first elution (from **step 3**).
5. Add another 75 μL of supplemented DNA elution buffer to the DNA input sample to make up the total volume to 150 μL. The total volume in the new tubes for all samples should be 150 μL.
6. Add 7.5 μL proteinase K to each eluted sample and the DNA input.
7. Incubate for 45 min at 50 °C with shaking. Allow samples to equilibrate to RT.
8. Prespin empty phase-lock gel tubes at 12,000–16,000 × *g* for 20–30 s at RT.

9. Transfer each RNA pull-down sample and the DNA input to a phase-lock tube and add 150 μL of phenol–chloroform–isoamyl alcohol per sample.
10. Shake vigorously for 10 min and centrifuge at 16,000 × *g* for 5 min at 4 °C. Transfer the upper aqueous phase to a new microcentrifuge tube.
11. Add 2 μL GlycoBlue, 15 μL of 3 M sodium acetate, and 450 μL of prechilled AR grade absolute ethanol (95%, kept at 4 °C). Store tubes at −20 °C overnight to precipitate DNA.
12. Centrifuge at 16,000 × *g* for 30 min at 4 °C and remove the supernatant without disturbing the DNA pellet.
13. Wash pellet with 1 mL of freshly prepared 70% ethanol and vortex to mix. Spin at 16,000 × *g* for 5 min at 4 °C.
14. Remove the supernatant completely and allow pellets to air dry for 10 min or until the ethanol evaporates (avoid overdrying).
15. Resuspend each DNA pellet in 30 μL of nuclease-free water.
16. Measure the DNA concentrations using 1 μL of the DNA sample and a Qubit DNA HS assay kit (DNA concentrations typically range from 100–200 ng).
17. Samples are ready for preparation of DNA sequencing libraries or analysis by qPCR.

3.7 RNA Pull-Down Validation Using qRT-PCR and DNA qPCR Analysis

1. Synthesize cDNA using Superscript IV VILO cDNA kit, 0.2 PCR tubes and a PCR thermocycler. Use the entire RNA yield from the pull-down for cDNA synthesis. Remove DNA from the samples by adding 1 μL of ezDNase buffer (10×) and 1 μL of ezDNase (make up the volume to 10 μL if required with RNase-free water), gently mix and incubate at 37 °C for 2 min. Chill samples on ice immediately.
2. Proceed immediately with cDNA synthesis by adding 6 μL of RNase-free water and 4 μL of Superscript VILO master mix while tubes are placed on ice (total volume in the tube is now 20 μL). Gently mix followed by brief centrifugation.
3. Programme the following on a PCR thermocycler to synthesize cDNA: 10 min at 25 °C (anneal primers), 10 min at 50 °C (reverse-transcribe RNA), 5 min at 85 °C (inactivate enzyme), and hold at 4 °C.
4. Following cDNA synthesis, use primers against the lncRNA of interest and, if available, a technical control (e.g., *TERC*) to assess pull-down efficiency (*see* **Note 3**).
5. Adjust the Ct (threshold cycle, or the number of cycles needed to detect a signal of the desired product) of the input by factoring in the dilution factor.

6. Calculate ΔCt of pull-down samples by subtracting the Ct of the adjusted input and analyze the enrichment levels (log2) of your target of interest comparing the input and pull-down samples. In order to demonstrate the results as percentage of input, multiply log2 enrichment levels by 100 (*see* **Note 5**).
7. DNA qPCR and analysis can be done in a similar way or DNA-seq libraries can be prepared using a ChIP-seq library kit.

4 Notes

1. For sonication optimization using the Bioruptor Pico, aliquot 100 μL of cell lysate each per Bioruptor tube. Prepare up to 12 tubes as the Pico rotor has the capacity to hold a maximum of 12 tubes. Use sonication settings 30 s ON: 45 s OFF starting from 20× cycles (take a tube out at 20× cycles) and continue sonicating by increasing the number of sonication cycles (can range from 20–150×) until the lysate turns clear. Sonication cycle numbers need to be optimized for each cell type (e.g., 70× cycles were required for MDA-MB-231 and HCT-116 cells). Next, take 5 μL of cell lysate out of each tube and transfer to a fresh microcentrifuge tube. Add 90 μL of Proteinase K buffer (for DNA) and 5 μL of Proteinase K. Vortex and spin down briefly. Incubate tubes for 45 min at 50 °C to reverse the crosslink. Purify DNA using a PCR purification kit. Load the entire sample on a 1% agarose gel and check DNA fragmentation. If the bulk of the fragmented DNA is within 100–500 bp, sonication is in the optimal size range for this pull-down protocol (Fig. 1). Should none of the tested settings result in sufficient chromatin shearing, increase the cycle numbers.
2. The protocol can be optimized to be used with single probes, or pools of multiple tiling probes. The pull-down efficiency of each individual probe correlates with the accessibility to the targeted lncRNA regions; therefore, the pull-down efficiency of each probe can be tested separately. The advantage of using one probe is that the chromatin regions pulled down can be more consistent between biological replicates. Combining multiple probes may be required if the pull-down efficiency of individual probes do not result in sufficient RNA enrichment (>10% is desirable).
3. Positive and negative controls should be included in each pull-down experiment. For example, a well-studied, relatively abundant lncRNA such as *TERC* can be used as a positive and technical control. To control for unspecific binding, pull-downs should be repeated independently with different probes or probe pools. In addition, loss-of-function models such as genetic knockout (KO) cell lines of the lncRNA of interest

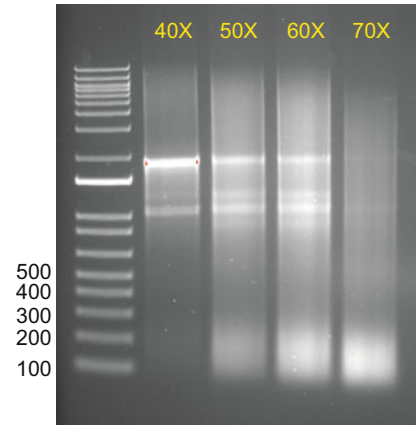


Fig. 1 Sonication optimization cycles for MDA-MB-231 cells. 70× cycles of sonication resulted in desired chromatin shear range of 100–500 bp

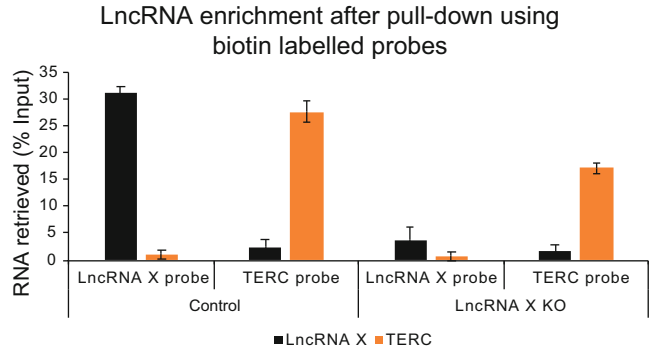


Fig. 2 RNA-pull-down efficiency: three biotin labelled oligonucleotides binding *LncRNA X* transcripts were used for chromatin isolation by RNA pull-down. qRT-PCR showing *LncRNA X* RNA retrieved in the MDA-MB-231 control cells and MDA-MB-231 *LncRNA X* knockout (KO) cells. *TERC* probes were used as positive technical controls in the same cell lines. *TERC* RNA levels were undetectable in the *LncRNA X* enriched pull-down and vice versa. $n = 2$ and error bars represent SD

should be included to examine the specificity of each pull-down (Fig. 2). Cytoplasmic mRNA (β-actin, *GAPDH*, *HPRT*, etc.) can also be used as negative controls.

4. Avoid to over dry the tubes after removing ethanol for RNA or DNA isolation as it will dehydrate the nucleic acids and reduce yields.
5. Pull-down efficiencies typically range between 10% and 40% (Fig. 2).

References

1. Kopp F, Mendell JT (2018) Functional classification and experimental dissection of long noncoding RNAs. Cell 172(3):393–407. https://doi.org/10.1016/j.cell.2018.01.011
2. Statello L, Guo C-J, Chen L-L, Huarte M (2021) Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol 22(2):96–118. https://doi.org/10.1038/s41580-020-00315-9
3. Sparber P, Filatova A, Khantemirova M, Skoblov M (2019) The role of long non-coding RNAs in the pathogenesis of hereditary diseases. BMC Med Genet 12(2):42. https://doi.org/10.1186/s12920-019-0487-6
4. Slack FJ, Chinnaiyan AM (2019) The role of non-coding RNAs in oncology. Cell 179(5):1033–1055. https://doi.org/10.1016/j.cell.2019.10.017
5. Hobuß L, Bär C, Thum T (2019) Long non-coding RNAs: at the heart of cardiac dysfunction? Front Physiol 10:30. https://doi.org/10.3389/fphys.2019.00030
6. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J, Barnes I, Berry A, Bignell A, Carbonell Sala S, Chrast J, Cunningham F, Di Domenico T, Donaldson S, Fiddes IT, García Girón C, Gonzalez JM, Grego T, Hardy M, Hourlier T, Hunt T, Izuogu OG, Lagarde J, Martin FJ, Martínez L, Mohanan S, Muir P, Navarro FCP, Parker A, Pei B, Pozo F, Ruffier M, Schmitt BM, Stapleton E, Suner MM, Sycheva I, Uszczynska-Ratajczak B, Xu J, Yates A, Zerbino D, Zhang Y, Aken B, Choudhary JS, Gerstein M, Guigó R, Hubbard TJP, Kellis M, Paten B, Reymond A, Tress ML, Flicek P (2019) GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res 47(D1):D766–d773. https://doi.org/10.1093/nar/gky955
7. Mulvey BB, Olcese U, Cabrera JR, Horabin JI (2014) An interactive network of long non-coding RNAs facilitates the Drosophila sex determination decision. Biochim Biophys Acta 1839(9):773–784. https://doi.org/10.1016/j.bbagrm.2014.06.007
8. Hansen TB, Jensen TI, Clausen BH, Bramsen JB, Finsen B, Damgaard CK, Kjems J (2013) Natural RNA circles function as efficient microRNA sponges. Nature 495(7441):384–388. https://doi.org/10.1038/nature11993
9. van Werven FJ, Neuert G, Hendrick N, Lardenois A, Buratowski S, van Oudenaarden A, Primig M, Amon A (2012) Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. Cell 150(6):1170–1181. https://doi.org/10.1016/j.cell.2012.06.049
10. Nolasco S, Bellido J, Gonçalves J, Tavares A, Zabala JC, Soares H (2012) The expression of tubulin cofactor A (TBCA) is regulated by a noncoding antisense Tbca RNA during testis maturation. PLoS One 7(8):e42536. https://doi.org/10.1371/journal.pone.0042536
11. Latos PA, Pauler FM, Koerner MV, Şenergin HB, Hudson QJ, Stocsits RR, Allhoff W, Stricker SH, Klement RM, Warczok KE, Aumayr K, Pasierbek P, Barlow DP (2012) Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science 338(6113):1469–1472. https://doi.org/10.1126/science.1228110
12. Csorba T, Questa JI, Sun Q, Dean C (2014) Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. Proc Natl Acad Sci U S A 111(45):16160–16165. https://doi.org/10.1073/pnas.1419030111
13. Ding J, Lu Q, Ouyang Y, Mao H, Zhang P, Yao J, Xu C, Li X, Xiao J, Zhang Q (2012) A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. Proc Natl Acad Sci U S A 109(7):2654–2659. https://doi.org/10.1073/pnas.1121374109
14. Wang M, Wu H-J, Fang J, Chu C, Wang X-J (2017) A long noncoding RNA involved in rice reproductive development by negatively regulating Osa-miR160. Sci Bull 62(7):470–475. https://doi.org/10.1016/j.scib.2017.03.013
15. Ma J, Yan B, Qu Y, Qin F, Yang Y, Hao X, Yu J, Zhao Q, Zhu D, Ao G (2008) Zm401, a short-open reading-frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the floret formation in maize. J Cell Biochem 105(1):136–146. https://doi.org/10.1002/jcb.21807
16. Marchese FP, Raimondi I, Huarte M (2017) The multidimensional mechanisms of long noncoding RNA function. Genome Biol 18(1):206. https://doi.org/10.1186/s13059-017-1348-2
17. Yao R-W, Wang Y, Chen L-L (2019) Cellular functions of long noncoding RNAs. Nat Cell Biol 21(5):542–551. https://doi.org/10.1038/s41556-019-0311-8
18. Ransohoff JD, Wei Y, Khavari PA (2018) The functions and unique features of long

intergenic non-coding RNA. Nat Rev Mol Cell Biol 19(3):143–157. https://doi.org/10.1038/nrm.2017.104

19. Fazal FM, Han S, Parker KR, Kaewsapsak P, Xu J, Boettiger AN, Chang HY, Ting AY (2019) Atlas of subcellular RNA localization revealed by APEX-Seq. Cell 178(2):473–490.e426. https://doi.org/10.1016/j.cell.2019.05.027
20. Benoit Bouvrette LP, Cody NAL, Bergalet J, Lefebvre FA, Diot C, Wang X, Blanchette M, Lécuyer E (2018) CeFra-seq reveals broad asymmetric mRNA and noncoding RNA distribution profiles in Drosophila and human cells. RNA 24(1):98–113. https://doi.org/10.1261/rna.063172.117
21. Chen C-K, Blanco M, Jackson C, Aznauryan E, Ollikainen N, Surka C, Chow A, Cerase A, McDonel P, Guttman M (2016) Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. Science 354(6311):468–472. https://doi.org/10.1126/science.aae0047
22. Pandey RR, Mondal T, Mohammad F, Enroth S, Redrup L, Komorowski J, Nagano T, Mancini-DiNardo D, Kanduri C (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol Cell 32(2):232–246. https://doi.org/10.1016/j.molcel.2008.08.022
23. Jain AK, Xi Y, McCarthy R, Allton K, Akdemir KC, Patel LR, Aronow B, Lin C, Li W, Yang L, Barton MC (2016) LncPRESS1 is a p53-regulated LncRNA that safeguards pluripotency by disrupting SIRT6-mediated de-acetylation of histone H3K56. Mol Cell 64(5):967–981. https://doi.org/10.1016/j.molcel.2016.10.039
24. Postepska-Igielska A, Giwojna A, Gasri-Plotnitsky L, Schmitt N, Dold A, Ginsberg D, Grummt I (2015) LncRNA Khps1 regulates expression of the proto-oncogene SPHK1 via triplex-mediated changes in chromatin structure. Mol Cell 60(4):626–636. https://doi.org/10.1016/j.molcel.2015.10.001
25. Tripathi V, Ellis JD, Shen Z, Song DY, Pan Q, Watt AT, Freier SM, Bennett CF, Sharma A, Bubulya PA, Blencowe BJ, Prasanth SG, Prasanth KV (2010) The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. Mol Cell 39(6):925–938. https://doi.org/10.1016/j.molcel.2010.08.011
26. Simon MD, Wang CI, Kharchenko PV, West JA, Chapman BA, Alekseyenko AA, Borowsky ML, Kuroda MI, Kingston RE (2011) The genomic binding sites of a noncoding RNA. Proc Natl Acad Sci U S A 108(51):20497–20502. https://doi.org/10.1073/pnas.1113536108
27. Chu C, Quinn J, Chang HY (2012) Chromatin isolation by RNA purification (ChIRP). J Vis Exp (61):3912. https://doi.org/10.3791/3912

Chapter 20

Superresolution Microscopy for Visualization of Physical Contacts Between Chromosomes at Nanoscale Resolution

Zulin Yu and Tamara A. Potapova

Abstract

This protocol describes the fluorescence in situ hybridization (FISH) of DNA probes on mitotic chromosome spreads optimized for two super-resolution microscopy approaches—structured illumination microscopy (SIM) and stimulated emission depletion (STED). It is based on traditional DNA FISH methods that can be combined with immunofluorescence labeling (Immuno-FISH). This technique previously allowed us to visualize ribosomal DNA linkages between human acrocentric chromosomes and provided information about the activity status of linked rDNA loci. Compared to the conventional wide-field and confocal microscopy, the quality of SIM and STED data depends a lot more on the optimal specimen preparation, choice of fluorophores, and quality of the fluorescent labeling. This protocol highlights details that make specimens suitable for super-resolution microscopy and tips for good imaging practices.

Key words Structured illumination microscopy (SIM), Stimulated emission depletion (STED), Chromosome spreads, Fluorescence in situ hybridization (FISH), Immuno-FISH, Ribosomal DNA (rDNA)

1 Introduction

DNA FISH has been a standard cytogenetics tool for decades and more recently has been employed for studying genome organization and other applications [1–4]. In our recent study, we used FISH and Immuno-FISH combined with super-resolution microscopy to show that ribosomal DNA can form physical linkages between chromosomes [5]. rDNA associations can be detected on wide-field images (Fig. 1a). However, SIM revealed that these associations were comprised of thin rDNA filaments connecting rDNA loci from different chromosomes (Fig. 1b). These structures were validated by using SIMcheck (Fig. 1c). To confirm this finding, rDNA linkages were imaged using STED. Again, rDNA associations can be detected in the confocal mode (Fig. 2a), but

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*,
Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0_20,

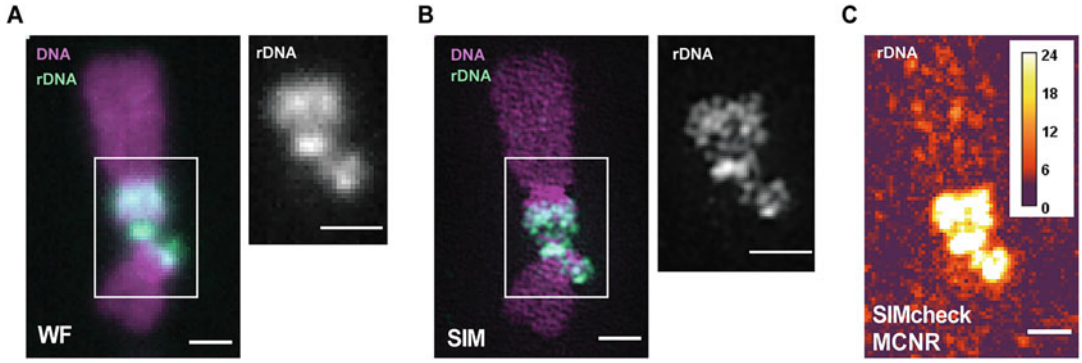


Fig. 1 Wide-field and (WF) and structural illumination microscopy (SIM) images of rDNA linkage between human acrocentric chromosomes. (**a**) Image of rDNA-linked acrocentric chromosomes from the human mammary epithelial cell (HMEC) acquired in wide-field mode. Mitotic chromosomes were labeled by FISH with Fluorescein-conjugated rDNA probe (green) and DAPI (purple). Magnified insert shows linked rDNA loci. (**b**) Image of the same chromosomes acquired in SIM mode. Thin fibers of rDNA are visible within the compact linkage structure. (**c**) Heatmap of local modulation contrast-to-noise ratio (MCNR) values obtained by SIMcheck analysis of the image shown in figure (**b**). Color of a lookup table (LUT) is indicative of the MCNR values: 0–4 (purple)- inadequate, to 8 (red)—acceptable, to 12 (orange) —good, to 18 (yellow)—very good, to 24 (white)—excellent. Bar, 1 μm

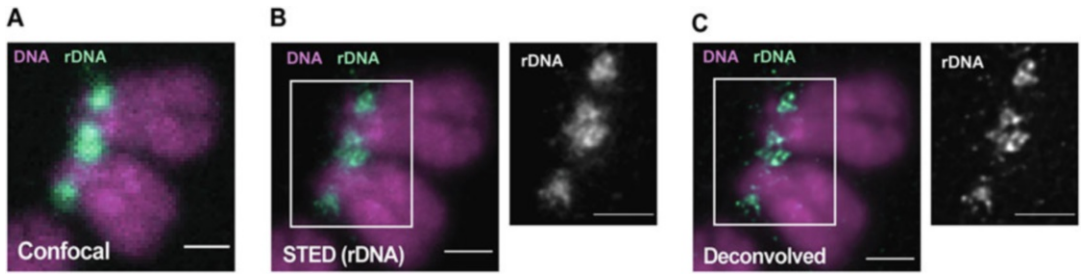


Fig. 2 Confocal and STED images of rDNA linkage between human acrocentric chromosomes. (**a**) Confocal image of rDNA-linked acrocentric chromosomes from hTERT-RPE1 cell line labeled by FISH with TAMRA-conjugated rDNA probe (green) and DAPI (purple). Magnified insert depicts linked rDNA loci. (**b**) Image of the same rDNA linkage acquired in STED mode. Fibers of rDNA are detectable within the linkage structure. (**c**) Deconvolved STED image of rDNA linkage shown in A and B. Bar, 1 μm

STED imaging exposed rDNA linkage structures analogous to those detected by SIM (Fig. 2b). Furthermore, we immunolabeled the rDNA transcription factor UBF and used SIM to show that rDNA linkages also contained UBF (Fig. 3).

General techniques for mitotic chromosome spreads preparation and hybridization of DNA probes are well-established [6–9]. Mitotic chromosomes can be isolated from many sources including tissue culture cells, donor lymphocytes, embryonic tissue, and other systems where cells are actively dividing. Mitotic spreads preparation involves arresting cells in mitosis by microtubule poisons followed by hypotonic treatment and fixation in the solution of methanol and acetic acid.

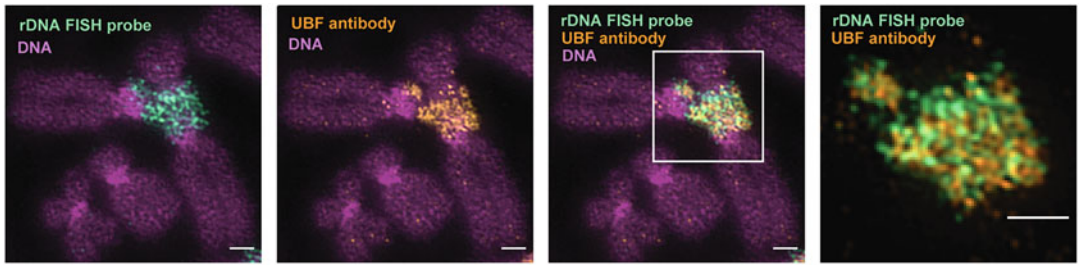


Fig. 3 SIM images of rDNA-linked mitotic chromosomes labeled by immuno-FISH. SIM images of rDNA-linked mitotic chromosomes from hTERT-RPE1 cell line labeled with Fluorescein-conjugated rDNA probe (green), and UBF antibody detected with secondary antibody conjugated to Alexa Fluor 561. DNA was counterstained with DAPI (purple). Both rDNA and UBF form filamentous connections between chromosomes. UBF and rDNA signals do not show a complete overlap because UBF binds to the coding region of rDNA repeats, but the rDNA probe also covers the entire intergenic spacer region. Magnified insert shows rDNA and UBF composing the linkage. Bar, 1 μm

Super-resolution imaging requires the specimen to be as close to the coverslip as possible. For that reason, fixed cells are dropped directly on #1.5 glass coverslips and then utilized for FISH or immuno-FISH. Fixed cells then are subjected to denaturation in hot formamide solution followed by hybridization to the fluorescently labeled DNA probe. In this protocol, spreads and fluorescent probes are denatured separately.

For super-resolution microscopy, it is essential to use brightly labeled fluorescent probes to achieve a high signal-to-noise ratio, and the choice of fluorophores is also important. For SIM, dyes with shorter emission wavelengths result in higher resolution, similar to conventional microscopy [10]. For STED, the photophysical properties of the dyes are most important [11]. DNA probe generation and fluorescent label conjugation strategies depend largely on biological questions. We are not going to discuss here the DNA probe generation, because, at present, there are numerous strategies for making fluorescent probes and many sources of commercially available ones.

Posthybridization steps include washing out the unbound probe and counter-staining of chromosomes with dyes whose excitation and emission spectra do no overlap with fluorescent channels of interest. FISH specimens can be further immunolabeled for chromatin-bound proteins. The efficiency of labeling depends on the degree of antigen preservation and the ability of the antibody to detect it. In our studies, immunofluorescence labeling was performed after the FISH. However, not all antigens will be well-preserved after exposure to harsh denaturation conditions. In this case, immunostaining may be attempted before the denaturation [12–14]. After the final labeling steps, the specimen is mounted on a glass slide and is ready for imaging.

The resolution of conventional optical microscopy is diffraction-limited to ~200 nm [15]. In reality, many conventional microscopes do not even approach this physical resolution limit due to the limitations of their optical and detection equipment. Fluorescently labeled objects that are separated by a distance lower than the resolution limit appear as one object, and the fine details within that object are not resolved. To bypass the ~200 nm resolution limit in light microscopy, several super-resolution microscopy technologies have been invented recently, including SIM and STED [16]. SIM utilizes a high-frequency sinusoidal striped pattern to excite samples. After processing raw data with the SIM reconstruction algorithm, the lateral resolution of the reconstructed image exceeds the conventional resolution limit by about twofold, to 100–130 nm [17]. SIM has been utilized broadly due to the relatively straightforward operation, fast imaging speed, and not many restrictions on fluorescent dye or protein selection.

The STED imaging system has optical configurations similar to the conventional confocal microscope. It utilizes an additional doughnut-shaped STED beam with a central zero node in the excitation pathway that silences (depletes) fluorophores. Because of a much smaller effective point spread function (PSF), the resolution of STED images can be, theoretically, improved down to molecular size [18]. In reality, depending on the experimental system, it approaches 30–90 nm. Unlike SIM, STED does not require complex imaging processing; but may be aided by deconvolution [19].

2 Materials

2.1 Preparation of Mitotic Chromosome Spreads

1. Appropriate cell culture medium and disassociating reagent (Trypsin or TrypLE), prewarmed to 37 °C for mammalian cells.
2. Colcemid solution: 10 μg/mL (100×) of N-desacetyl-N-methylocolchicine in HBSS or PBS, sterile, available from Gibco under the brand name KaryoMAX or prepared in house.
3. 50 mL conical centrifuge tubes.
4. Hypotonic solution: 0.075 M KCl/H_2O, sterile, available from Gibco or prepared in-house.
5. Methanol–acetic acid (3:1) fixative solution: 30 mL of methanol and 10 mL acetic acid. Must be prepared fresh.
6. #1.5 glass coverslips size 22 mm × 30 mm or larger, precleaned.
7. Precleaned Superfrost microscope glass slides.

8. Phase-contrast microscope equipped with 20× or 40× objective.
9. 65 °C heating block for slides.

2.2 Fluorescent In Situ Hybridization

1. RNAse A (any source, but must be high quality and DNase-free).
2. 37 °C incubator.
3. 20× SSC stock solution: 175.3 g NaCl, 88.2 g sodium citrate, make up to 1 L with H_2O, adjust to pH 7, sterilize. This solution can be further diluted with sterile water to prepare 2xSSC and 1xSSC solutions.
4. Water bath preheated to 72 °C.
5. 65 °C heating block for slides.
6. 80 °C heating source for probe denaturation (can be a heating block, PCR machine, or water bath).
7. Coplin staining jars.
8. Ethanol series: Prepare 2 sets of 70%, 80%, and 100% ethanol. One set is kept at room temperature, and the other is prechilled at −20 °C.
9. Denaturation solution (70% deionized formamide/2× SSC): 35 mL deionized formamide, 5 mL 20× SSC, 10 mL distilled H_2O. Preheat in 72 °C water bath for no longer than 20 min. Must be prepared fresh. Mix the solution right before use by pipetting up and down a few times.
10. Fluorescently labeled probe, commercial or in-house generated. The source and the generation of the probe depend on the scientific question.
11. Hybridization buffer: this reagent can be purchased or made in-house. For example, probes from Empire genomics are supplied with the proprietary hybridization buffer, and probes from Cytocell are supplied prediluted in hybridization buffer. Recipes for in-house hybridization buffers vary: a typical composition contains 50% formamide, 2× SSC, and 10% dextran sulfate.
12. 22 mm × 22 mm glass or HybriSlip hybridization covers.
13. Rubber cement or CytoBond removable coverslip sealant.
14. Light-protecting humidified slide-staining chamber.

2.3 Washes and Mounting

1. Water bath preheated to 45 °C.
2. Washing solution I (50% formamide/2× SSC): 15 mL 20× SSC, 60 mL distilled H_2O, 75 mL formamide, adjust the pH to 7.0 using HCl, and preheat to 45 °C for longer than 20 min. Must be prepared fresh.

3. Washing solution II (1× SSC): Prepare by diluting 20xSSC with sterile H_2O and preheat to 45 °C. Also, have some solution at room temperature (enough for one wash).
4. DAPI (4,6-diamidino-2-phenylindole) stock solution: dissolve at 1 mg/mL in H_2O and store at −20 °C.
5. Mounting medium: ProLong Gold or ProLong Glass can be used for both SIM and STED. Vectashield can be used for SIM.
6. Molecular biology grade H_2O.
7. Clear nail polish.

2.4 Immunofluorescence Labeling

1. Washing solution III: 4× SSC, 0.1% Tween 20.
2. Antigen unmasking solution (citrate buffer): 82 mL 0.1 M sodium citrate, 18 mL 0.1 M citric acid, 900 mL distilled H_2O, pH 6.0.
3. Immunofluorescence (IF) wash buffer: 0.1% Triton X-100 in 1× PBS.
4. Blocking Buffer: 5% normal goat serum or BSA, 0.1% Triton X-100 in 1× PBS.
5. Antibody Dilution Buffer: 1% BSA, 0.1% Triton X-100 in 1× PBS.
6. Primary antibody of interest.
7. Fluorochrome-conjugated secondary antibody.

2.5 Microcopy Hardware and Software Package

1. GE Healthcare OMX structured illumination microscope (V4) with three scientific complementary metal–oxide semiconductor (sCMOS) cameras, or a similar microscope (e.g., Nikon N-SIM, Zeiss Elyria).
2. Olympus PlanApo N 60×, 1.42 NA oil immersion objective.
3. Leica Inverted TCS SP8 STED scope equipped with White Light Laser (470–670 nm range) and STED laser lines 592 nm, 660 nm, and 775 nm.
4. Leica 100×, 1.4-NA HC PL APO OIL CS2 Objective.
5. OMX acquisition computer and software v3.70 (GE Healthcare).
6. OMX image-processing station with SoftWoRx v6.52 (GE Healthcare) software installed.
7. Leica Application Suite (LAS X) software with Hyvolution Module.
8. ImageJ or FIJI (Java software for image-processing analysis).
9. SIMcheck, the ImageJ/Fiji plugin for SIM data evaluation.

3 Methods

3.1 Preparation of Mitotic Chromosome Spreads

For optimal preparation of chromosomal spreads, the cell culture needs to be healthy and in an exponential growth phase. Cultures with a high percentage of senescent, quiescent, or differentiated cells, or cultures growing in depleted or acidified medium may not yield a sufficient number of mitotic cells. Each prep typically requires one T-75 flask or one 100 mm plate of adherent cells that are 70–90% confluent. For cells growing in suspension, start with 1–2 million cells.

1. Accumulation of mitotic cells (mitotic arrest).
2. Add 0.1 μg/mL of Colcemid solution to the cell culture for 4–10 h, depending on the proliferation rate of a cell line. Colcemid is a microtubule poison that causes depolymerization of mitotic spindle microtubules and metaphase arrest mediated by mitotic spindle checkpoint [20]. Other microtubule depolymerizing agents can be used to induce mitotic arrest such as sterile 0.1–1 μg/mL nocodazole [21]. Cells that proliferate faster typically require shorter mitotic blocks than cells with a longer cell cycle. Anticipate the mitotic index of 5–25% at the end of the mitotic block. Note that prolonged mitotic arrest causes hyper-condensation of chromosomes and loss of arm cohesion [22]. Therefore, it may be better to have fewer mitotic spreads with good chromosomal morphology than lots of mitotic spreads with problematic morphology.
3. Trypsinize adherent cells and collect cell suspension in a 50 mL sterile conical centrifuge tube containing 10 mL of cell culture medium/Colcemid solution. Spin cells down for 3–5 min at 200 × *g* and aspirate supernatant avoiding disturbing the cell pellet.
4. Gently resuspend the pellet in 0.5 mL of media by tapping the tube.
5. Slowly add 10 mL of prewarmed 0.075 M KCl hypotonic solution. Mix well by tapping the tube.
6. Allow cells to swell for 10–12 min at 37 °C (*see* **Note 1**). At this step, cells uptake a large amount of water, becoming fragile. Mix cells periodically, avoiding vigorous shakes and vortexing.
7. Prefix cells by adding 1% of total volume (i.e., 100 μL/10 mL) methanol–acetic acid (3:1) fixative solution. Mix well by gently tapping the tube.
8. Collect cells by centrifuging for 5 min at 200 × *g*.
9. Discard supernatant leaving a small amount at the bottom of the tube and gently resuspend the pellet in the remaining supernatant. Avoid vigorous mixing yet ensure that cells are fully resuspended. If clumps of cells are left at this stage, they will persist after the fixation.

10. Fix cells by quickly adding 10 mL methanol–acetic acid (3:1) fixative solution. Ensure that cells are well resuspended.
11. Incubate at room temperature for at least 10 min. At this point, cells can be stored in the fixative solution at −20 °C.
12. Centrifuge fixed cells for 5 min at 200 × *g*, discard the supernatant, and resuspended again in smaller volume (0.5–1.5 mL) of fresh methanol–acetic acid (3:1) fixative solution. The cell suspension can be moved to a smaller conical tube for easier handling. The optimal volume of the fixative solution depends on the cell density and can be adjusted if needed. A good cell concentration range is around 0.2–1 million cells/mL.
13. Now cells are ready to be dropped on the precleaned glass slides or coverslips. For SIM and STED, the distance between sample and coverslip is critical and optimal results are achieved when spreads are deposited directly on precleaned #1.5 glass coverslips. For regular wide-field or confocal microscopy, chromosomal spreads can be deposited on a precleaned glass slide (*see* **Note 2**). Pipet up 20–25 μL of the fixed cell suspension and drop them from arm's length on the precleaned coverslip (*see* **Note 3**).
14. Examine spreading under a phase-contrast microscope with 20× or 40× objective, find mitotic cells, and observe the spreading. If cells appear too dense or too sparse, adjust the density by adding or removing some fixative. In low humidity conditions, holding the slide with spreads before they are completely dry under the water vapor for a moment improves the spreading. Placing slides on the 65 °C warmer immediately stops the spreading, which can be helpful if chromosomes spread too much and start floating away.
15. Allow fixative to dry and label the top with a diamond pen or other nonfluorescent alcohol-resistant label. It also helps to demarcate the area where the spreads are.
16. For easier handling, coverslips can be temporarily mounted face-up on a glass slide with very small drops of nail polish applied to the corners of the coverslip, which can be carefully removed after all staining procedures are complete. This protocol refers to coverslip slides as "slides."
17. Place coverslip slides with spreads on a heating block prewarmed to 65 °C for at least 1 h. For FISH, they can be left on the heating block overnight and processed the next day.
18. Leftovers of the fixed cell suspension can be stored in fixative solution at −20 °C for months. For prolonged storage, it is good to seal the top of the tightly closed tube with Parafilm because methanol is very volatile and can evaporate over time.

3.2 Fluorescent In Situ Hybridization (FISH)

1. In a Coplin jar, treat slides with mounted coverslips with 0.1 mg/mL RNAse A in 2xSSC solution for 30–60 min at 37 °C.
2. Rinse slides in 2× SSC solution three times, 5 min each, at room temperature.
3. Dehydrate slides, consecutively, in 70%, 80%, and 100% ethanol series, 2 min in each solution, at room temperature.
4. Place slides back on 65 °C warmer until ready for denaturation (*see* **Note 4**).
5. Denature chromosomes in a Coplin jar containing Denaturation solution preheated to 72 °C. Immerse slides in the solution for 1–1.5 min (*see* **Note 5**). The timing of this step should be very precise. Always wear gloves and exercise caution when handling hot formamide.
6. Rapidly stop the denaturation by immersing slides in 70% ethanol precooled to −20 °C for 2 min. Continue with 80%, and 100% -20 °C ethanol series for 2 min each. Shake off the ethanol and place slides on 65 °C heating block for a short period of time, until ready to apply the probe (*see* **Note 6**).
7. Prepare the working solution of the fluorescently labeled probe (*see* **Note 7**) by diluting an appropriate amount of labeled probe in the hybridization buffer. Vortex well and place 12 μL of prepared probe per each drop of spreads into a microcentrifuge tube. Avoid exposing the fluorescently labeled probe to direct light as much as possible and use brown Eppendorf tubes if available. Denature the probe by incubation at 80 °C for 10 min. Vortex again and briefly spin the solution down to the bottom of the tube.
8. Pipette 12 μL of denatured probe mixture on the coverslip slide with denatured chromosomes, on the marked area where the spreads were deposited. Avoid the formation of bubbles.
9. Place a 22 × 22 mm square glass coverslip or HybriSlip over the probe mix. Avoid air bubbles. Seal the edges with rubber cement or CytoBond.
10. Transfer slides to a light-protected humidified chamber and place in 37 °C incubator for 24–72 h.

3.3 Posthybridization Washes and Mounting or Preparation for Immunostaining

During all posthybridization procedures, slides should be protected from direct light as much as possible.

1. Remove slides from the humidified chamber and carefully peel off the rubber cement.
2. Carefully remove coverslip or HybriSlip covers and discard. If they are a little stuck on the coverslip slides, float them with washing solution I and gently move them off the slides.

3. Transfer slides to Coplin jar containing washing solution I; wash slides 3 times for 3–5 min each at 45 °C. The total wash time should not exceed 15 min.
4. Wash slides twice in washing solution II at 45 °C for 5 min each.
5. Wash slides once in washing solution II at room temperature for 5 min.
6. Optional DAPI labeling. Incubate slides in 0.5 μg/mL DAPI solution in 1xSSC for 30 min. If immunostaining is planned, this step can be done at the end of the procedure.
7. If the immunostaining is not planned, briefly rinse slides in molecular biology grade water. If the coverslip was temporarily mounted on a glass slide, carefully remove it from the slide without breaking. Air-dry in the dark.
8. For SIM, mount in coverslips on precleaned glass slides in a small amount of Vectashield, ProlongGold, or ProlongGlass and seal with nail polish. For STED, mount in ProlongGold or ProlongGlass. Allow ProlongGlass to cure overnight in the dark.

3.4 Immunofluorescence Labeling

The key to success is finding the antibody that works on chromosome spreads prepared as above. Antigen unmasking in hot Citrate buffer helps uncover antigenicity and does not reduce the FISH signal by a lot. Protect slides from the direct light at all steps as much as possible to preserve the FISH signal.

1. After washing slides in washing solution II, wash slides in washing solution III for 5 min at room temperature.
2. Antigen unmasking/rehydration: Immerse slides in 65 °C citrate buffer in a Coplin jar and incubate at 65 °C for 1 h (*see* **Note 8**).
3. Wash slides in IF wash buffer 2 times for 5 min at room temperature.
4. Block slides in Blocking Buffer for at least 60 min at room temperature.
5. Primary antibody staining (*see* **Note 9**). Incubate specimen in the primary antibody diluted in antibody dilution buffer. 30–50 μL of antibody solution can be applied per slide under Parafilm or HybriSlip covers. Incubate in humidified chamber overnight at 4 °C.
6. Gently remove hybridization covers from the slides. Use wash buffer to dislodge them if they are a little stuck.
7. Wash slides three times in IF wash buffer for 5 min each.
8. Secondary antibody staining: incubate specimen in an appropriate fluorochrome-conjugated secondary anti-species

antibody (*see* **Note 10**) diluted in antibody dilution buffer for 2–4 h at room temperature. Use 30–50 μL of secondary antibody solution per slide under Parafilm or HybriSlip covers as in the primary incubation step. For counterstaining, DAPI can be added to the secondary antibody solution at 0.5 μg/mL, or it can be applied separately as in subheading 3.3, **step 6**.

9. Wash slides three times in IF wash buffer for 5 min each.
10. Briefly rinse slides in molecular biology grade water. If the coverslip was temporarily mounted on a glass slide, carefully remove it from the slide without breaking. Air-dry in the dark, mount on a glass slide, and seal with nail polish.

3.5 Super-Resolution Microscopy (SIM and STED)

For successful SIM and STED imaging, it is important to strive for a high signal-to-noise ratio.

3.5.1 SIM Image Acquisition and Validation

SIM works best with samples that have brightly labeled discrete structures with low background. Samples with a low signal-to-noise ratio may not benefit from SIM. Diffuse or dim structures or structures labeled with easily photobleached fluorophores can be plagued with reconstruction artifacts [23].

It is important to select the immersion oil with an optimal refractive index and ensure that the entire structure is imaged from top to bottom with minimal photobleaching. Photobleaching, spherical aberration, and clipping structures off in *Z* can result in data reconstruction artifacts due to an asymmetrical point spread function (PSF). The reconstruction algorithm for SIM images assumes symmetrical PSF, and the out-of-focus light from asymmetrical PSF can be reconstructed as a real signal, resulting in reconstruction artifacts.

Selecting an immersion oil with an appropriate refractive index (RI) is very important for achieving symmetrical PSF. The OMX manufacturer provides an oil kit with refractive indices ranging from 1.500 to 1.534. The optimal RI of the oil depends on many factors, including the sample RI, the distance between the sample and coverslip, the RI of the mounting medium, the emission wavelength of the fluorophore, and the temperature of the environment. For multicolored specimens, the oil should be matched for the channel of most interest (*see* **Note 11**). For instance, for images shown in Fig. 1 the channel of most interest was 488 (Fluorescein). Consequently, for reference channels (DAPI in this case) RI of the oil is slightly suboptimal and the reconstruction may suffer from artificial patterns. In this case, increasing the Wiener Filter Constant in reconstruction parameters for the 405 channel (DAPI) can help to reduce artificial patterns at a cost of a decreased resolution.

To assess the quality of SIM data and reconstruction, it is recommended to utilize SIMcheck, a collection of ImageJ plugins available for ImageJ or FIJI [24] (*see* **Note 12**). SIMcheck detects problems, or lack of thereof, in SIM data acquisition and processing. For example, Fig. 1c shows a heatmap of local modulation contrast-to-noise ratio (MCNR) values from SIMcheck. Structures that have high MCNR values are likely to be real. Using SIMcheck is a good practice for producing interpretable SIM images and making informed decisions about the quality of data.

1. Apply the optimal RI immersion oil from the oil kit to the objective lens; for fluorophores with green emission spectra, we recommend starting at RI of 1.514 or 1.516 at the first round.
2. Set up multichannel image acquisition on Omxworx software: chose correct laser lines and emission filters based on the excitation/emission spectra of the dyes, and adjust laser intensities and camera exposure to achieve a good signal-noise ratio. Do keep in mind photobleaching when setting up the laser power and the exposure.
3. Set up the *Z*-stack: use the recommended optimal distance between *Z* slices, ensure that the entire structure is imaged and the best focal planes are in the mid-range of the stack.
4. Image the channel or channels of interest first, followed by the reference channel.
5. Generate wide-field images from raw SIM images, which can be done by clicking the "Generate wide-field image" option in SoftWoRx processing software. Find small bright spots in the widefield image and examine them in orthogonal views (XZ and YZ). PFS from these spots should be symmetrical if the RI of oil was chosen correctly. If the PFS is asymmetric and the asymmetry (flare) is towards the coverslip, the RI of the oil may be too high, and if it is away from the coverslip, it may be too low. Remove the oil from the slide and the objective and try the different RI oil (one step up or down at a time).
6. Set up a processing queue in SoftWoRx. First, reconstruct raw images with the measured PSF and optimized parameters (i.e., Wiener filter), and then align multichannel images following the manufacturer's alignment protocol.
7. Load raw and reconstructed SIM image files into ImageJ or FIJI to run SIMcheck plugin. For large fields of view, it is recommended to crop them to a region containing structures of interest, but including areas of representative background. SIMcheck options also allow cropping the stack in *Z*. For large multichannel datasets fluorescent channels may be split and analyzed separately.

3.5.2 STED Image Acquisition and Processing

The important considerations for successful STED imaging include fluorescent labeling strategies, specimen preparations, and acquisition settings.

The selection of fluorophores for labeling structures of interest is critical for successful STED imaging. As with SIM, it is important to strive for the best possible signal-to-noise ratio. Not all fluorophores are suitable for STED, as they require specific spectral and photostability properties and also have to be compatible with the lasers and optical configurations of the imaging system (*see* **Note 13**). Even generally recommended fluorophores or combinations of fluorophores should be tested experimentally on a given microscope.

STED microscopy relies on combining the excitation laser with the depletion laser whose beam shape is modified into a central zero node "doughnut" that drives the fluorescence at the periphery of the focal point of the excitation laser to a ground state. Therefore, the excitation laser and the depleting STED laser must be perfectly aligned.

Because of the necessity for depleting a fraction of the signal, the overall loss in the fluorescent signal is unavoidable. However, it affects the signal and the background equally, therefore the signal-to-noise ratio of STED images will still be good if the specimen labeling had a high fluorescent signal and low background. STED is not the most sensitive method of imaging and may not be the best for weakly labeled structures. It is important to find a balance between resolution and sensitivity when setting the STED laser power for a given specimen.

For counter-staining of reference structures that can be imaged in regular confocal mode, other dyes could be used, as long as the emission spectra of these dyes do not overlap with the range of the STED detection. For images shown in Fig. 2, rDNA for STED was labeled with 5-TAMRA dUTP, and chromosomes were labeled with DAPI. DAPI is not an appropriate dye for STED but also does not overlap with TAMRA emission. Since bleaching by powerful STED laser will also affect the reference channel dyes, it is best to acquire these images in the confocal mode first, before the STED acquisition.

Unlike SIM, STED directly enhances the resolution and generally does not require postacquisition reconstruction based on mathematical algorithms. However, deconvolution can be used as a postacquisition processing method to enhance image visualization and improve resolution further. Deconvolution algorithms reassign out-of-focus signal based on estimated or real PSF, thereby improving the signal-to-noise ratio. Raw STED images are informative if the STED was successful, but deconvolution can improve the resolution and the contrast for better visualization.

1. Acquire the reference (DAPI) channel in regular confocal mode with 405 nm excitation and 426–468 nm as emission.
2. Set up a sequential mode with 2 color images in LAS X. Set up the channel of interest as the first channel following by the reference channel and choose "between frame" as the default setting for channel switching.
3. Acquire the channel of interest in STED mode. For the data in Fig. 2b, TAMRA (the rDNA channel) was excited with a pulsed white light (80 MHz) tuned to 555 nm and was depleted with a pulsed 660 nm laser with 85–90% maximum output.
4. For the TAMRA channel, the time gate was set from 0.6–6 ns at the internal Leica HyD detector, and the emission spectrum was collected in the range of 568–614 nm. STED image was acquired in 2D mode with recommended pixel size to maximize lateral resolution, and each image was averaged 8 times in line averaging mode.
5. For optional deconvolution, load both raw STED and confocal images into the Huygens software module. For the deconvolution of the image shown in Fig. 2c, theoretical estimated PSF was used. Most other used parameters were default values. The background was measured from raw image and signal to noise was set at 15 for STED images and 20 for confocal images.

4 Notes

1. Optimal swelling time in hypotonic may need to be determined experimentally for each cell line. The excessive swelling time can result in cell bursting, while insufficient swelling time can result in poor separation of chromosomes. For instance, human iPS cells generally require a shorter swelling time (8–9 min), while human mammary epithelial cells (HMEC) can be incubated in the hypotonic solution for up to 15 min.
2. To preclean coverslips or glass slides, they are put in a beaker filled with 70% ethanol and sonicated in an ultrasonic bath sonicator for 60 min continuous operation. Coverslips can then be stored in the same beaker covered with foil or Parafilm. Wipe the ethanol off before using; coverslips or slides must be completely dry.
3. Proper chromosome spreading is essential. We recommend a 30–50 cm distance from the pipette to the slide pr coverslip for good dispersal. For consistency, the pipette can be mounted on a chemical stand with the slide placed directly underneath it.
4. In this protocol, spreads and the fluorescent probes are denatured separately because the optimal timing and temperature is different for spreads and for the probe.

5. Optimal denaturation time varies and may need to be experimentally determined. Over- and under-denaturation can result in reduced fluorescent signal. The longer the chromosome spreads have been stored, the longer it takes for them to denature.
6. It helps to inspect slides under the phase-contrast microscope after denaturation and dehydration in cold ethanol series. Properly denatured chromosomes should appear lighter-colored than they did before denaturation.
7. For super-resolution microscopy, it is essential to use brightly fluorescently labeled probes to achieve a good signal-to-noise ratio. There are many ways of generating fluorescently labeled FISH probes in-house that are not discussed in this protocol, and multiple sources of commercially available probes. The effectiveness and optimal concentration of each probe may need to be determined experimentally. We have been using fluorescently labeled DNA probes from Empire Genomics, PNA Bio, and Cytocell at concentrations recommended by the manufacturer, or in-house generated probes.
8. The antigen unmasking step may need to be piloted for every new antigen and antibody. Not all chromatin-bound and chromatin-associated proteins are preserved sufficiently after methanol fixation followed by multiple ethanol dehydration steps and harsh hybridization conditions. Some antigens require a shorter unmasking period and some longer but do not exceed 60 min to minimize loss of the FISH signal. If immunostaining of mitotic chromosomes is the primary goal, it may be better to resort to protocols that utilize paraformaldehyde fixation [25].
9. Having a primary antibody that works in these conditions is the key to success. Not all antibodies that recognize the antigen well in paraformaldehyde-fixed specimens will work on denatured chromosomal spreads. Generally, primary antibody dilution needs to be 2–4 fold higher than for regular immunofluorescence. Importantly, immunolabeling is most effective on freshly prepared spreads that have not been stored in the fixative for very long and have not been backed at 65 °C overnight.
10. When choosing the fluorochrome-conjugated secondary antibody, make sure its emission spectrum does not overlap with the FISH signal.
11. GE Oil Immersion Calculator, an online tool for estimating RI of Oil, can be found here: https://www.cytivalifesciences.com/en/us/support/online-tools/cell-imaging-and-microscopy/immersion-oil-calculator

12. SIMcheck toolbox for ImageJ or FIJI can be downloaded here: https://github.com/MicronOxford/SIMcheck
    SIMcheck user guide is available here: https://www.micron.ox.ac.uk/software/SIMcheck/index.html
13. Leica provides a list of fluorophores recommended for single- and dual-color STED available here:
    https://www.leicabiosystems.com/fileadmin/academy/2012/STED_sample_preparation_QUICK_GUIDE_V02.pdf
    Many other dyes have been reported in the literature to be suitable for STED, but we recommend testing them in a given experimental system before assuming that they will work.

Acknowledgments

We are thankful to the Microscopy core facility at the Stowers Institute for enabling super-resolution experiments. We thank Patrina Pellett, Jay Unruh, and Sean McKinney for conceptual help, and Brian Slaughter for the critical review of the manuscript. We thank Jennifer Gerton for mentorship and members of the Gerton lab for discussions. We are grateful to Scott Rider for assistance with labeled probes and to Martha Stampfer for HMECs. We thank Christophe Leterrier for making his collection of colorblind-friendly LUTs publicly available. This study was supported by funding from Stowers Institute for Medical Research.

References

1. Hu Q, Maurais EG, Ly P (2020) Cellular and genomic approaches for exploring structural chromosomal rearrangements. Chromosom Res 28(1):19–30. https://doi.org/10.1007/s10577-020-09626-1
2. Cui C, Shu W, Li P (2016) Fluorescence in situ hybridization: cell-based genetic diagnostic and research applications. Front Cell Dev Biol 4:89. https://doi.org/10.3389/fcell.2016.00089
3. Speicher MR, Carter NP (2005) The new cytogenetics: blurring the boundaries with molecular biology. Nat Rev Genet 6(10):782–792. https://doi.org/10.1038/nrg1692
4. Trask BJ (2002) Human cytogenetics: 46 chromosomes, 46 years and counting. Nat Rev Genet 3(10):769–778. https://doi.org/10.1038/nrg905
5. Potapova TA, Unruh JR, Yu Z, Rancati G, Li H, Stampfer MR, Gerton JL (2019) Super-resolution microscopy reveals linkages between ribosomal DNA on heterologous chromosomes. J Cell Biol 218(8):2492–2513. https://doi.org/10.1083/jcb.201810166
6. Dal Cin P (2003) Metaphase harvest and cytogenetic analysis of malignant hematological specimens. Curr Protoc Hum Genet Chapter 10:Unit 10.12. https://doi.org/10.1002/0471142905.hg1002s36
7. Bangs CD, Donlon TA (2005) Metaphase chromosome preparation from cultured peripheral blood cells. Curr Protoc Hum Genet Chapter 4:Unit 4.1. https://doi.org/10.1002/0471142905.hg0401s45
8. Schuck PL, Stewart JA (2019) FISHing for damage on metaphase chromosomes. Methods Mol Biol 1999:335–347. https://doi.org/10.1007/978-1-4939-9500-4_24
9. Landegent JE, Jansen in de Wal N, van Ommen GJ, Baas F, de Vijlder JJ, van Duijn P, Van der Ploeg M (1985) Chromosomal localization of a unique gene by non-autoradiographic in situ hybridization. Nature 317(6033):175–177. https://doi.org/10.1038/317175a0
10. Gustafsson MG, Shao L, Carlton PM, Wang CJ, Golubovskaya IN, Cande WZ, Agard DA, Sedat JW (2008) Three-dimensional resolution

doubling in wide-field fluorescence microscopy by structured illumination. Biophys J 94(12):4957–4970. https://doi.org/10.1529/biophysj.107.120345

11. Jahr W, Velicky P, Danzl JG (2020) Strategies to maximize performance in STimulated emission depletion (STED) nanoscopy of biological specimens. Methods 174:27–41. https://doi.org/10.1016/j.ymeth.2019.07.019
12. Solovei I, Cremer M (2010) 3D-FISH on cultured cells combined with immunostaining. Methods Mol Biol 659:117–126. https://doi.org/10.1007/978-1-60761-789-1_8
13. Perea-Resa C, Bury L, Cheeseman IM, Blower MD (2020) Cohesin removal reprograms gene expression upon mitotic entry. Mol Cell 78(1):127–140.e7. https://doi.org/10.1016/j.molcel.2020.01.023
14. Chan FL, Marshall OJ, Saffery R, Kim BW, Earle E, Choo KH, Wong LH (2012) Active transcription and essential role of RNA polymerase II at the centromere during mitosis. Proc Natl Acad Sci U S A 109(6):1979–1984. https://doi.org/10.1073/pnas.1108705109
15. Wegel E, Gohler A, Lagerholm BC, Wainman A, Uphoff S, Kaufmann R, Dobbie IM (2016) Imaging cellular structures in super-resolution with SIM, STED and localisation microscopy: a practical comparison. Sci Rep 6:27290. https://doi.org/10.1038/srep27290
16. Schermelleh L, Heintzmann R, Leonhardt H (2010) A guide to super-resolution fluorescence microscopy. J Cell Biol 190(2):165–175. https://doi.org/10.1083/jcb.201002018
17. Gustafsson MG (2000) Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J Microsc 198(Pt 2):82–87. https://doi.org/10.1046/j.1365-2818.2000.00710.x
18. Hell SW, Wichmann J (1994) Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. Opt Lett 19(11):780–782. https://doi.org/10.1364/ol.19.000780
19. Zanella R, Zanghirati G, Cavicchioli R, Zanni L, Boccacci P, Bertero M, Vicidomini G (2013) Towards real-time image deconvolution: application to confocal and STED microscopy. Sci Rep 3:2523. https://doi.org/10.1038/srep02523
20. Rieder CL, Palazzo RE (1992) Colcemid and the mitotic cycle. J Cell Sci 102(Pt 3):387–392
21. Gorbsky GJ (1997) Cell cycle checkpoints: arresting progress in mitosis. BioEssays 19(3):193–197. https://doi.org/10.1002/bies.950190303
22. Lai SK, Wong CH, Lee YP, Li HY (2011) Caspase-3-mediated degradation of condensin cap-H regulates mitotic cell death. Cell Death Differ 18(6):996–1004. https://doi.org/10.1038/cdd.2010.165
23. Demmerle J, Innocent C, North AJ, Ball G, Muller M, Miron E, Matsuda A, Dobbie IM, Markaki Y, Schermelleh L (2017) Strategic and practical guidelines for successful structured illumination microscopy. Nat Protoc 12(5):988–1010. https://doi.org/10.1038/nprot.2017.019
24. Ball G, Demmerle J, Kaufmann R, Davis I, Dobbie IM, Schermelleh L (2015) SIMcheck: a toolbox for successful super-resolution structured illumination microscopy. Sci Rep 5:15915. https://doi.org/10.1038/srep15915
25. Hirota T, Gerlich D, Koch B, Ellenberg J, Peters JM (2004) Distinct functions of condensin I and II in mitotic chromosome assembly. J Cell Sci 117(Pt 26):6435–6445. https://doi.org/10.1242/jcs.01604

INDEX

Julia Horsfield and Judith Marsman (eds.), *Chromatin: Methods and Protocols*, Methods in Molecular Biology, vol. 2458, https://doi.org/10.1007/978-1-0716-2140-0,

Q

R

S

T

U

V

Y

Z