Methods in Molecular Biology 791

Springer Protocols

Trygve O. Tollefsbol *Editor*

Epigenetics Protocols

Second Edition

Humana Press

METHODS IN MOLECULAR BIOLOGY™

Series Editor
John M. Walker
School of Life Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Epigenetics Protocols

Second Edition

Edited by

Trygve O. Tollefsbol

Department of Biology, University of Alabama at Birmingham, Birmingham, AL, USA

Humana Press

Editor
Trygve O. Tollefsbol
Department of Biology
University of Alabama at Birmingham
Birmingham, AL 35294-1170, USA
trygve@uab.edu

ISSN 1064-3745 e-ISSN 1940-6029
ISBN 978-1-61779-315-8 e-ISBN 978-1-61779-316-5
DOI 10.1007/978-1-61779-316-5
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011935043

Printed on acid-free paper

Humana Press is part of Springer Science+Business Media (www.springer.com)

Preface

Since the first edition of *Epigenetics Protocols* was published in 2004, the field of epigenetics has continued to have a major role at the forefront not only of molecular biology, but also of medical genetics and clinical medicine. Few disciplines have experienced growth comparable to that we have witnessed for epigenetics in the past decade. I believe that much of the innovative force driving the remarkable development of epigenetics can be attributed to a steady flow of novel techniques in this field. Due to the rapid pace of development of new methods in epigenetics and the exponentially increasing interest in this area, a second edition of *Epigenetics Protocols* seems timely. Moreover, for the same reasons, the second edition differs significantly from the first edition as new methods have been devised and new areas of focus have evolved.

It would not be realistic to attempt to amass all of the epigenetic techniques that have been invented in one book. Rather, the goal of this volume has been to highlight select techniques that have been mainstays in the field and to also cover methods that are especially relevant to extant discoveries in epigenetics. Leading the way in the renaissance this field is currently enjoying is the advent of numerous tools for studying the epigenome, both in terms of deriving experimental findings and in analyzing these data to unravel the power of epigenetic processes to influence phenotypic expression of the genome.

The two broad areas of epigenetics that receive the most attention in this book are DNA methylation and chromatin modifications. These major epigenetic processes can be further subdivided into topics related to gene- or region-specific analyses, genome-wide studies, and analyses of modulation or measurement of the mediators of DNA methylation and chromatin modifications. Lastly, this volume covers the complex topic of computational methods for epigenetic analyses which is essential to a complete understanding of the vast body of information that is being derived with the use of these newly developed tools.

The protocols detailed in this book are intended to provide advanced students, basic scientists, and clinical researchers as well as clinicians and biotechnology investigators with a contemporary set of tools that can be applied to understanding epigenetics. The methods covered in *Epigenetics Protocols II* will continue to drive the exciting field of epigenetics and facilitate even more fascinating discoveries of the many important roles of epigenetics in basic molecular biology, medical sciences, and clinical applications.

Birmingham, AL, USA ***Trygve O. Tollefsbol***

Contents

Contributors

Paola B. Arimondo • *Muséum National d'Histoire Naturelle, Paris, France*
Lars Bolund • *Institute of Human Genetics, Aarhus University, Aarhus C, Denmark*
Kenneth R. Bondioli • *Louisiana State University Agricultural Center, School of Animal Sciences, Baton Rouge, LA, USA*
Ida L. M. Candiloro • *Molecular Pathology Research and Development Laboratory, Department of Pathology, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia*
Alexandre Ceccaldi • *Muséum National d'Histoire Naturelle, Paris Cedex, France*
Hsueh-Wei Chang • *Department of Biomedical Science and Environmental Biology, Center of Excellence for Environmental Medicine, Cancer Center, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan*
Yu-Huei Cheng • *Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan*
Li-Yeh Chuang • *Department of Chemical Engineering, I-Shou University, Kaohsiung, Taiwan*
Pierre Close • *Unit of Medical Chemistry, GIGA-Signal Transduction, University of Liege, Liege, Belgium*
Philippe Collas • *University of Oslo, Institute of Basic Medical Sciences & Norwegian Center for Stem Cell Research, Oslo, Norway*
Céline Cosseau • *Ecologie et Evolution des Interactions, 2EI, Université de Perpignan, Perpignan Cedex, France*
Nancy E. Davidson • *University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA*
J. Tyson DeAngelis • *Department of Biology, University of Alabama at Birmingham, Birmingham, AL, USA*
Alexander Dobrovic • *Molecular Pathology Research and Development Laboratory, Department of Pathology, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia*
Lorigail Echipare • *Department of Pharmacology and the Genome Center, University of California-Davis, Davis, CA, USA*
Tomas J. Ekström • *Clinical Neuroscience, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden*
Osman El-Maarri • *Institute of Experimental Haematology and Transfusion Medicine, University of Bonn, Germany*
Peggy J. Farnham • *Department of Biochemistry and Molecular Biology, USC/Norris Comprehensive Cancer Center, Los Angeles, CA, USA*
Jörg Felsberg • *Department of Neuropathology, Heinrich-Heine-University, Düsseldorf, Germany*

Jorge Ferrer • *Genomic Programming of Beta Cells, Institut d'Investigacions Biomediques August Pi i Sunyer, Barcelona, Spain*
Angelica M. Giraldo • *Revivicor, Inc., Blacksburg, VA, USA*
Christoph Grunau • *Ecologie et Evolution des Interactions, 2EI, Université de Perpignan, Perpignan Cedex, France*
Yoshihide Hayashizaki • *Omics Science Center, Riken, Yokohama, Kanagawa, Japan*
Chelsee A. Hewitt • *Molecular Pathology Research and Development Laboratory, Department of Pathology, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia*
Chih-Lin Hsieh • *Department of Urology, University of Southern California, Los Angeles, CA, USA; Department of Biochemistry and Molecular Biology, University of Southern California, Los Angeles, CA, USA*
Yi Huang • *University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA*
Ryan A. Irvine • *Department of Urology, University of Southern California, Los Angeles, CA, USA; Department of Biochemistry and Molecular Biology, University of Southern California, Los Angeles, CA, USA*
Albert Jeltsch • *Biochemistry Lab, School of Engineering and Science, Jacobs University Bremen, Bremen, Germany*
You-Kyung Jeon • *Laboratory of Cell Biology, Korean Cell Line Bank, Cancer Research Institute, Seoul National University College of Medicine, Seoul, South Korea*
Renata Z. Jurkowska • *Biochemistry Lab, School of Engineering and Science, Jacobs University Bremen, Bremen, Germany*
Mohsen Karimi • *Department of Medicine, Karolinska University Hospital, Huddinge, Stockholm, Sweden*
Ja-Lok Ku • *Laboratory of Cell Biology, Korean Cell Line Bank, Cancer Research Institute, Seoul National University College of Medicine, Seoul, South Korea*
Jian Li • *Institute of Human Genetics, Aarhus University, Aarhus C, Denmark*
Yuanyuan Li • *Department of Biology, University of Alabama at Birmingham, Birmingham, AL, USA*
Karin Luttropp • *Molecular Medicine and Surgery, Karolinska University Hospital, Stockholm, Sweden*
Ricardo B. de Medeiros • *Department of Antibody Applications and Stem Cells, R&D Systems, Inc., Minneapolis, MN, USA*
Thomas Mikeska • *Molecular Pathology Research and Development Laboratory, Department of Pathology, Peter MacCallum Cancer Centre, Melbourne, VIC, Australia*
Natalie Mitchell • *Department of Biology, University of Alabama at Birmingham, Birmingham, AL, USA*
Annette Molinaro • *Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT, USA*
Takao Nammo • *Genomic Programming of Beta Cells, Institut d'Investigacions Biomediques August Pi i Sunyer, Barcelona, Spain*
Nicole Nüsgen • *Institute of Experimental Haematology and Transfusion Medicine, University of Bonn, Germany*

Henriette O'Geen • *Department of Pharmacology and the Genome Center, University of California-Davis, Davis, CA, USA*
Yasushi Okazaki • *Research Center for Genomic Medicine, Saitama Medical University, Hidaka, Saitama, Japan*
Cindy Okitsu • *Department of Urology, University of Southern California, Los Angeles, CA, USA; Department of Biochemistry and Molecular Biology, University of Southern California, Los Angeles, CA, USA*
Hisato Okuizumi • *Genetic Resources Center, National Institute of Agrobiological Sciences (NIAS), Tsukuba, Ibaraki, Japan*
Jae-Gahb Park • *Laboratory of Cell Biology, Korean Cell Line Bank, Cancer Research Institute, Seoul National University College of Medicine, Seoul, South Korea*
Mattia Pelizzola • *Genomic Analysis Lab, Salk Institute for Biological Studies, La Jolla, CA, USA*
Santiago A. Rodríguez-Seguí • *Genomic Programming of Beta Cells, Institut d'Investigacions Biomediques August Pi i Sunyer, Barcelona, Spain*
Luke A. Selth • *Dame Roma Mitchell Cancer Research Laboratories, The University of Adelaide, Adelaide, SA, Australia*
Patrick G. Shaw • *University of Pittsburgh Cancer Institute, Pittsburgh, PA, USA*
Heike Singer • *Institute of Experimental Haematology and Transfusion Medicine, University of Bonn, Germany*
Jesper Q. Svejstrup • *Mechanisms of Gene Transcription Laboratory, Clare Hall Laboratories, Cancer Research UK London Research Institute, Hertfordshire, UK*
Tomoko Takamiya • *School of Pharmacy, Nihon University, Chiba, Japan*
Trygve O. Tollefsbol • *Department of Biology, University of Alabama at Birmingham, Birmingham, AL, USA*
Cheng-Hong Yang • *Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan*
Yingying Zhang • *Biochemistry Lab, School of Engineering and Science, Jacobs University Bremen, Bremen, Germany*
Qian Zhao • *School of Life Science and Technology, Tongji University, Shanghai, China*

Chapter 1

Advances in Epigenetic Technology

Trygve O. Tollefsbol

Abstract

Epigenetics refers to the collective heritable changes in phenotype that arise independent of genotype. Two broad areas of epigenetics are DNA methylation and histone modifications and numerous techniques have been invented to analyze epigenetic processes not only at the level of specific genes, but also to analyze epigenetic changes that occur in defined regions of the genome as well as genome-wide. Advances have also been made in techniques devised to assess the enzymes that mediate epigenetic processes. These methods that are currently driving the field of epigenetics will greatly facilitate continued expansion of this exponentially growing discipline of genetics.

Key words: Epigenetics, DNA methylation, Chromatin, Methods, Histone, Technique

1. Introduction

Although the term "epigenetics" has several variants of its definition, taken together they can be summarized as the collective heritable changes in phenotype due to processes that arise independent of primary DNA sequence. DNA methylation is the most studied of the subfields of epigenetics and in most eukaryotes is characterized by the enzymatic addition of a methyl group to the cytosine-5 position. This usually occurs in CpG dinucleotides and the DNA methyltransferases (DNMTs), of which there are three major types in mammals, are responsible for most of the methylation that occurs in the genome. DNMT1 is the enzyme primarily involved in the maintenance of methylation patterns with each cell replication and it preferentially methylates hemimethylated DNA (1). Also very important in shaping the methylome are the enzymes DNMT3a and DNMT3b, which have relatively high de novo methylating activity (2). The most significant aspect of DNA methylation is its role in gene expression and the more methylated a

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_1,

gene regulatory region, the more likely it is that the gene activity will become downregulated and vice versa although there are exceptions to this general rule (3).

Like DNA methylation, chromatin modification is another central epigenetic process that impacts gene expression as well as many other biological processes. For example, histone acetylation has been associated with an increase in gene activity whereas histone deacetylation often silences transcription (4). However, methylation of histones has variable effects on gene activity where lysine 4 (K4) methylation of histone H3 is frequently associated with increasing gene activity while methylation of lysine 9 (K9) of histone H3 may lead to transcriptional repression. Even more significant perhaps is the fact that there is often cooperation between DNA methylation and histone modifications (5) such that cytosine methylation may contribute to the methylation of H3-K9. Likewise, H3-K9 methylation may promote cytosine methylation to lead to gene downregulation. Therefore, it is often the case that both DNA methylation and histone modifications act interdependently to contribute to the overall chromatin state and its epigenetic control of numerous cellular processes.

2. DNA Methylation Gene- or Region-Specific Techniques

A major breakthrough in the analysis of DNA methylation occurred with the development of bisulfite methylation sequencing. This method (Chapter 2) is a gold-standard for the detection of DNA methylation largely because it allows identification of 5-methylcytosine (5mC) at single base-pair resolution. In this technique, bisulfite treatment of denatured DNA converts cytosines to uracil while 5mC is not converted. After PCR amplification and DNA sequencing, nonmethylated cytosines are recognized as thymines and methylated cytosines (largely in CpG dinucleotides) remain as cytosines. This leading method of DNA methylation analysis has lead to numerous subsequent methods such as methylation-specific PCR (MSP), combined bisulfite restriction analysis (COBRA) and methylation-sensitive single nucleotide primer extension (MS-SNuPE) which are also covered in this volume.

The real power of fundamental bisulfite methylation sequencing is in its application to a variety of methods and notable among these techniques is MSP (Chapter 3) that allows assessment of methylation patterns in CpG islands. Advantages of MSP are that it requires no cloning or methylation-sensitive restriction enzymes and it can be performed on very small quantities of DNA as well as DNA from paraffin-embedded samples. For discrimination of methylated or nonmethylated DNA, two primers are synthesized with one pair specific for methylated DNA and the other specific

for nonmethylated DNA. Amplification using PCR of a methylated primer pair indicates methylated DNA and amplification from a nonmethylated primer pair occurs if the DNA does not contain methylation within the site covered by the primer. MSP is especially useful for screening DNA samples and has found numerous applications in clinical medicine.

A strong feature of bisulfite pyrosequencing (Chapter 4) is its quantitative nature. This biotin-based method also amplifies PCR products from bisulfite-treated DNA and incorporated nucleotides allow for quantification based on emission of light. Although methylation patterns at a single allele are not derived with this method, it does allow the identification of heterogeneous DNA methylation patterns. A drawback of pyrosequencing can be its more limited covered regions of DNA (typically about a hundred base pairs) but serial bisulfite pyrosequencing allows longer PCR products to be assessed. This method can also be applied to formalin-fixed paraffin-embedded tissue which greatly extends its useful applications.

In certain locus-specific DNA methylation analyses, the need for removing PCR products can be eliminated in that amplification and analysis of bisulfite-modified DNA is performed in one tube. This is referred to in Chapter 5 as a closed-tube method and some examples of this convenient approach are real-time analysis – methylation-specific PCR (SMART-MSP), methylation-sensitive high-resolution melting (MS-HRM), and MethyLight. The MethyLight approach utilizes a fluorescent hydrolysis probe while the SMART-MSP method uses a double-stranded DNA binding fluorescent dye and melting curve analysis. If the methylation level or patterns of the DNA of interest is not known, MS-HRM can be used as a pre-screening technique. Detection and quantification of methylation is better achieved for DNA that has less than 5% methylation using the MethyLight and SMART-MSP approaches (Chapter 5).

COBRA is commonly used for analyzing specific regions of DNA for methylation and is also dependent on bisulfite treatment of DNA and subsequent restriction enzyme treatments. Chapter 6 introduces a novel visualization software referred to as Methyl-Typing. In this new approach, comprehensive restriction enzymes for sequences containing 5mC after bisulfite conversion are provided. This method, which makes COBRA analysis much easier, provides computation and visualization of the essential information for performing COBRA and provides all possible methylation sites of restriction enzymes.

A very promising and useful region-specific DNA methylation analysis method is the SIRPH technique (Chapter 7). SIRPH refers to *s*ingle nucleotide primer extension (SNuPE) that is combined with *i*on paired *r*everse *p*hase *H*PLC. This method utilizes one universal HPLC gradient and one SNuPE annealing temperature for all primers. This quantitative approach to analyzing methylation in

specific regions of DNA requires only small quantities of PCR product. Another advantage is that it is highly accurate and reproducible.

Thus, there are numerous approaches for determining the DNA methylation status of specific genes or regions of DNA of interest. Many of these methods have bisulfite conversion of DNA as their fundamental basis. Any of the techniques covered in Chapters 2–7 can be incorporated into most laboratories depending on the specific needs of each laboratory and the desired resolution and applications that are needed. Most notably, each of these methods is reliable in achieving accurate assessments of DNA methylation in defined DNA sequences.

3. Methods of Epigenomic Analysis

In some cases, one may wish to achieve a broader view of DNA methylation status within the genome and a number of exciting techniques have been developed that allow analysis of DNA methylation of very large segments of DNA if not the entire genome. One of the early developments in this regard is referred to as restriction landmark genome scanning (RLGS). This approach described in Chapter 8 allows methylation status determination of the whole genome based on two-dimensional electrophoresis. An improvement of this method is described in Chapter 8 in that the original method utilized *NotI* to detect methylated sites and this has now been refined to the use of the methylation-sensitive isoschizomers *MspI* and *HpaII*. A widely used method for genome-wide methylation analysis is the methylated DNA immunoprecipitation (MeDIP) technique. This method is based on the uses of antibody specific for 5mC and allows quantification of enriched methylated DNA fragments. However, analysis of MeDIP results can be complex. In Chapter 9, Pelizzola and Molinaro describe newly developed analyses for MeDIP results referred to as MEDME (Modeling Enrichment Derived from MeDIP Experiments) and BATMAN (Bayesian Tool for Methylation Analysis) which are designed to assist with evaluation of MeDIP enrichment measures to estimate the absolute or relative number of 5mCs in a genomic sample.

Two additional approaches to assessing the methylome are the methylated-CpG island recovery assay (MIRA) originally developed by Rauch and Pfeifer (6) and the luminometric methylation assay (LUMA). An advantage of the MIRA method (Chapter 10) is that it is not dependent upon bisulfite conversion of DNA and, unlike MeDIP, does not require specific antibody recognition. The fundamental basis of MIRA is that the high affinity of methylated DNA binding proteins, MBD2b and MBD3L1, is able to recover methylated DNA at single nucleotide resolution and can also be

used with microarray analysis or next-generation sequencing for methylome determinations. In Chapter 11, another exciting genome-wide approach to assessing DNA methylation status is described that is dependent upon methylation-specific restriction enzymes followed by pyrosequencing. This LUMA technique allows for quantification of the number of restriction enzyme cuts in a genomic sample based on methylation sensitivity relative to an internal standard. Notable advantages of the LUMA approach are its quantitative abilities and the high-throughput potential of this method.

The advent of techniques to analyze the methylome has been one of the most remarkable advances in epigenetics and these methods described in Chapters 8–11 represent some of the promising developments in this area. However, advances in epigenomics are occurring at a remarkable pace and it is likely that this area will herald the most breakthroughs over the next decade.

4. Techniques to Inhibit DNA Methylation and Assess DNA Methylation Activity

Since analyses of DNA methylation are a central component of epigenetic processes, and DNMTs are the mediating enzymes of DNA methylation, methods to inhibit the DNMTs as well as to measure their activity are important components of epigenetics. Inhibition of DNA methylation in somatic cells often allows determination of the role of DNA methylation in gene expression as well as other processes such as cellular differentiation and nuclear reprogramming (Chapter 12). The most commonly used approach for inhibiting DNA methylation is the employment of pharmacological agents such as 5-azacytidine, 5-aza-2′deoxycytidine, and nonnucleoside inhibitors such as procainamide, which inhibit the activity of DNMTs. DNMT heterozygous mouse models have also been used although mice null in DNMT1 result in embryonic lethality. The employment of small interfering RNA (siRNA) has also been applied to reducing DNA methylation in cells and this approach has the advantage of specificity in terms of which the three major mammalian DNMTs (1, 3a or 3b) are selectively knocked down.

Another basic method of DNA methylation analysis involves assessment of the enzymatic activity of the DNMTs. Techniques to accurately determine the activity of the DNMTs have many uses such as monitoring the effectiveness of knockdown of the DNMTs as well as measuring changes in DNMT enzymatic activity during various biological processes such as cellular differentiation and tumorigenesis. In Chapter 13, three reliable approaches to determine DNMT activity are described. These assays consist of use of radioactively labeled S-adenosyl-l-methionine, use of bisulfite

conversion of in vitro methylated DNA, and a novel fluorescence-coupled array using restriction enzymes that has high-throughput applications for screening of DNMT inhibitors. Therefore, inhibition of DNA methylation as well as methods to determine the activity of the DNMT mediators of DNA methylation are fundamental approaches to understanding the role of DNA in biological processes.

5. Chromatin Immunoprecipitation-Based Protocols

Chromatin Immunoprecipitation (ChIP) assays are widely used for a number of different assessments such as analysis of DNA–protein interactions and determining posttranslational modifications of histones. The most common use of ChIP assays in epigenetic analyses involve the determination of histone modifications which, along with DNA methylation, shares the role as a mainstay of epigenetic processes. However, a drawback to ChIP assays is it often requires a large number of cells for reliable measurements. Recent advances have overcome many of these limitations. For example, in Chapter 14, a micro-ChIP protocol is detailed that can be used for multiple parallel ChIPs using only a thousand cells which is remarkably less than the typical number of about 10^6–10^7 cells for conventional ChIP analyses. This development should greatly enhance the utility of ChIP-based measurements and expand its practical applications. However, conventional ChIP requires cross-linking of DNA and protein while the development of native ChIP (nChIP) bypasses the cross-linking process (Chapter 15). Some advantages of foregoing cross-linking and using nChIP are that the native form of the proteins are preserved, more sensitivity is generally possible and special equipment (e.g., sonicator) is not required.

Two very useful modifications of the ChIP method are Q-PCR in combination with ChIP and sequential ChIP or SeqChIP. Q-PCR is a highly sensitive and reproducible technique and when combined with conventional ChIP, these characteristics of Q-PCR are maintained which greatly enhances traditional ChIP assays. For example, Q-PCR can be combined with ChIP to assess the presence of acetylated histones H3 and H4 on different regions of a target locus as described in Chapter 16. Often it is desirable to assess the binding of more than one protein to a particular DNA sequence. For instance, epigenetic proteins such as the DNMTs can have many interactions with other proteins and it is often useful to assess how these proteins are interacting at a particular site on the DNA. SeqChIP is a relatively newer technique that provides this advantage. In SeqChIP, one can assess the interactions of two or more proteins or various histone modifications at a specific site in the genome (Chapter 17). The sequential nature of this

technique refers to the use of different antibodies to specific proteins that are provided sequentially during ChIP reactions. The significant advantage of SeqChIP is that many epigenetic modifications do not occur in a single-protein manner but rather, often involve interactions. Likewise, more than one histone modification often is involved in many epigenetic processes. Therefore, SeqChIP allows a more accurate determination of the key-interdependent processes that occur in vivo. Both Q-PCR in combination with ChIP and SeqChIP are extremely useful techniques that are greatly expanding the potential of epigenetic analyses.

6. Combined ChIP and DNA Methylation Analysis

Most of the chapters describe techniques that are valuable in assessing either DNA methylation or chromatin modifications. However, these two major mediators of epigenetic changes often cross-talk and can modulate each other. We therefore developed a novel technique that is a combination of conventional ChIP with *b*isulfite *m*ethylation *s*equencing which we refer to as ChIP-BMS (Chapter 18). This method renders DNA methylation information at the single-nucleotide level of defined DNA fragments precipitated by specific antibodies to histones or transcription factors of interest. This allows investigation of the interaction patterns between histone modification and DNA methylation as well as transcription factor binding and methylation of recognition sites. It can also serve to provide valuable information about multiple interactions between genetic and epigenetic factors Additional applications of this new technique besides analyses of interactions between histone modifications and DNA methylation are identification of methylation-sensitive transcription factors as well as simultaneous assessment of epigenetic regulation of gene expression by the two major epigenetic mediators, DNA methylation and the vast array of chromatin modifications. It also seems possible that ChIP-BMS could be combined with SeqChiP to render more meaningful results of the myriad of epigenetic and genetic factors that can contribute to transcriptional regulation.

7. In Vivo RNA–Protein Interaction Assessment

RNA also has important roles in epigenetics and many proteins interact with RNA to modulate RNA-based epigenetic processes. It is therefore important to have the tools available to detect direct and indirect interactions between specific proteins and RNA in vivo. This is best achieved through the RNA immunoprecipitation

technique (RIP) that is not too unlike ChIP assays. The key difference, however, between ChIP and RIP is that DNA is enzymatically removed and RNA that is bound to the proteins captured by the antibody are immunoprecipitated (Chapter 19). The uses of the RIP technique are vast and may be applied to epigenetics to help unravel the increasingly appreciated role of RNA in epigenetic processes.

8. Epigenomic Chromatin Methods

Genome-wide assessments of epigenetic modifications have definitely come of age and many investigators have a strong interest in methods that can achieve global information about chromatin modifications such as histone acetylation or histone methylation. ChIP-chip, or the application of purified ChIP products to a microarray (chip), was for many years the leading method of choice for epigenomic changes in histone modifications. However, multiple DNA microarrays are required for global analyses using the ChIP-chip approach and in part due to this reason ChIP-seq (Chapter 20) is rapidly becoming more popular than ChIP-chip. ChIP-seq can interrogate the whole genome with only one lane of sequencing, it renders high resolution and it requires only very small amounts of chromatin for analyses.

One of the major roles of epigenetic processes is the control of gene expression and it is often advantageous to determine regions of open chromatin associated with gene activity on a genome-wide scale. Chapter 21 describes the mapping of open chromatin using formaldehyde-assisted isolation of regulatory elements (FAIRE). This method is also based on cross-linkages and in this case it identifies open regulatory regions due to their relatively lower levels of association with nucleosomes. FAIRE has also been combined with high-throughput sequencing (FAIRE-seq) as described in Chapter 21 and is relatively easy to perform. FAIRE can be used for small amounts of tissue and has numerous applications such as facilitating understanding of how sequence variation affects open chromatin structures. When applied to a global analysis, FAIRE can be a very powerful tool in epigenetics.

9. Knockdown of Histone Deacetylases

Similar to the case with DNA methylation where inhibitors can render information about the role of DNA methylation in various contexts, knockdown of histone deacetylases (HDACs) can also be a very useful tool in epigenetic studies. Also reminiscent of DNA methylation inhibition, HDAC inhibition has a number of applications perhaps most notably in the anticancer field since

HDAC inhibitors have been found to have utility in a number of different cancer treatments. A major mechanism through which HDAC inhibitors affect cancer is through re-activation of aberrantly silenced tumor suppressor genes and Chapter 22 details protocols for HDAC inhibition in human breast cancer cells as one important application of this approach. HDAC inhibitors can be subdivided into hydroxamic acids, short-chain fatty acids, benzamindes, and cyclic tetrapeptides groups and assessment of their effectiveness can be determined through processes such as Q-PCR for gene-specific expression changes or microarray expression analyses to determine global gene expression changes after HDAC inhibition. Given the increasing clinical significance of HDAC inhibition, advances in technology development in this field could have a major impact.

10. Bioinformatics Applied to Epigenomics

The development of sophisticated methods to determine changes in the epigenome under various conditions has been a major advancement in the field of epigenetics. For example, massively parallel sequencing allows mapping of epigenetic variants but also renders vast amounts of data that can be daunting to understand and analyze. Computational methods for epigenetic analysis have therefore developed rapidly to answer the call for efficient techniques that can manage the volumes of data that are derived using epigenomic methods. For instance, a protocol has been developed consisting of computational analysis for modified methylation-specific digital karyotyping (MMSDK) based on massively parallel sequencing (Chapter 23). A protocol is described for a mapping process based on the in silico simulation of combined enzyme cutting and tag extraction of the reference genome. The tags are mapped to the simulated library using Mapping and Assembly with Qualities (MAQ) which is suitable for any tag profiling-based method. It can also be applied to other epigenetic analyses based on massively parallel sequencing. The endonuclease digestion-based MMSDK method allows for mapping of DNA methylation in the human genome and is an accurate and fast computational tool that is being developed for epigenomic studies.

11. Conclusion

Numerous methods have been developed that provide means to accurately assess many epigenetic processes, most notably, DNA methylation and chromatin modifications as the most prevalent epigenetic mediators. Two of these methods, bisulfite sequencing

for DNA methylation and ChIP for chromatin modifications and DNA–protein interactions, have given rise to numerous additional techniques that improve the speed, accuracy, and applications of these fundamental methods. New protocols have also been developed, such as ChIP-BMS, that combine these two basic techniques not only to render information about the interdependence of DNA methylation and chromatin modifications, but also to provide a more comprehensive understanding of the overall dynamics of epigenetic processes in modulating the phenotypic expression of the genome. Along with significant advances in methods to unravel the epigenome have come new challenges in managing massive amounts of information that is derived with these advanced techniques. Further developments in computational epigenetics will be needed for deciphering the many data that are accumulating due to the exciting advances in experimental epigenomics.

References

1. Robertson, K.D. (2001) DNA methylation, methyltransferases and cancer. *Oncogene*. 20, 3139–3155.
2. Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999) Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell*. 99, 247–257.
3. Lai, S.R., Phipps, S.M., Liu, L., Andrews, L.G., and Tollefsbol, T.O. (2005) Epigenetic control of telomerase and modes of telomere maintenance in aging and abnormal systems. *Front Biosci*. 10, 1779–1796.
4. Lehrmann, H., Pritchard, L.L., and Harel-Bellan, A. (2002) Histone acetyltransferases and deacetylases in the control of cell proliferation and differentiation. *Adv Cancer Res*. 86, 41–65.
5. Fuks, F., Burgers, W. A., Brehm, A., Hughes-Davies, L., and Kouzarides, T. (2000) DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat Genet*. 24, 88–91.
6. Rauch, T., and Pfeifer, G. (2005) Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Lab Invest*. 85, 1172–1180.

Chapter 2

DNA Methylation Detection: Bisulfite Genomic Sequencing Analysis

Yuanyuan Li and Trygve O. Tollefsbol

Abstract

DNA methylation, which most commonly occurs at the C5 position of cytosines within CpG dinucleotides, plays a pivotal role in many biological procedures such as gene expression, embryonic development, cellular proliferation, differentiation, and chromosome stability. Aberrant DNA methylation is often associated with loss of DNA homeostasis and genomic instability leading to the development of human diseases such as cancer. The importance of DNA methylation creates an urgent demand for effective methods with high sensitivity and reliability to explore innovative diagnostic and therapeutic strategies. Bisulfite genomic sequencing developed by Frommer and colleagues was recognized as a revolution in DNA methylation analysis based on conversion of genomic DNA by using sodium bisulfite. Besides various merits of the bisulfite genomic sequencing method such as being highly qualitative and quantitative, it serves as a fundamental principle to many derived methods to better interpret the mystery of DNA methylation. Here, we present a protocol currently frequently used in our laboratory that has proven to yield optimal outcomes. We also discuss the potential technical problems and troubleshooting notes for a variety of applications in this field.

Key words: DNA methylation, Epigenetics, Bisulfite genomic sequencing

1. Introduction

Extensive studies have demonstrated that DNA methylation plays a major role in regulating various physiological and pathological processes in mammals. DNA methylation is an important epigenetic event in modulating embryonic development, genomic imprinting, X inactivation, cellular differentiation, and proliferation (1, 2). However, abnormal patterns of DNA methylation are correlated with DNA instability which may ultimately trigger diseases such as cancer (3). DNA methylation, primarily occurring at C5 of the cytosine ring within cytosine–guanine (CpG) dinucleotides, is frequently

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_2,

found clustered at gene regulatory sites such as promoter regions (4). Dense methylation of CpGs in the gene promoter region is associated with a compacted chromatin structure resulting in transcriptional silencing of the affiliated gene. If DNA hypermethylation occurs at the promoter regions of certain critical cancer-related genes, it could lead to tumor suppressor gene silencing and ultimately tumorigenesis (5). Therefore, the importance of DNA methylation in a variety of biological processes represents an attractive diagnostic and therapeutic target. A precise and efficient method is required to determine the exact DNA methylation status to further elucidate the essential roles of DNA methylation in biological procedures.

Bisulfite genomic sequencing is regarded as a gold-standard technology for the detection of DNA methylation because it provides a qualitative, quantitative, and efficient approach to identify 5-methylcytosine (5mC) at single base-pair resolution. This method was first introduced by Frommer et al. and it is based on the finding that the amination reactions of cytosine and 5mC proceed with very different consequences after the treatment of sodium bisulfite (6). In this regard, cytosines in single-stranded DNA will be converted into uracil residues and recognized as thymine in subsequent PCR amplification and sequencing; however, 5mCs are immune to this conversion and remain as cytosines allowing 5mCs to be distinguished from unmethylated cytosines. A subsequent PCR process is necessary to determine the methylation status in the loci of interest by using specific methylation primers after the bisulfite treatment. The actual methylation status can be determined either through direct PCR product sequencing (detection of average methylation status) or subcloning sequencing (detection of single molecule distribution of methylation patterns) (Fig. 1). Moreover, bisulfite sequencing analysis can not only indentify DNA methylation status along the DNA single strand, but also detect the DNA methylation patterns of DNA double strands since the converted DNA strands are no longer self-complementary and the amplification products can be measured individually.

Over the past few years, several techniques have arisen based on the working basis of bisulfite including methylation-specific PCR (MSP), combined bisulfite restriction analysis (COBRA), methylation-sensitive single nucleotide primer extension (Ms-SNuPE), and several other techniques depending on different applications (7–10). Compared with other DNA methylation approaches based on the sensitivity of restriction enzymes that can specifically recognize methylated cytosine within their cleavage recognition site (11), bisulfite-based DNA methylation analysis has more quantitative accuracy, detection sensitivity, high efficiency, and a wide spectrum for sample analysis.

Bisulfite genomic sequencing, as a fundamental method of DNA methylation analysis, has been widely used in various research and clinical settings. To optimize the final results of the bisulfite

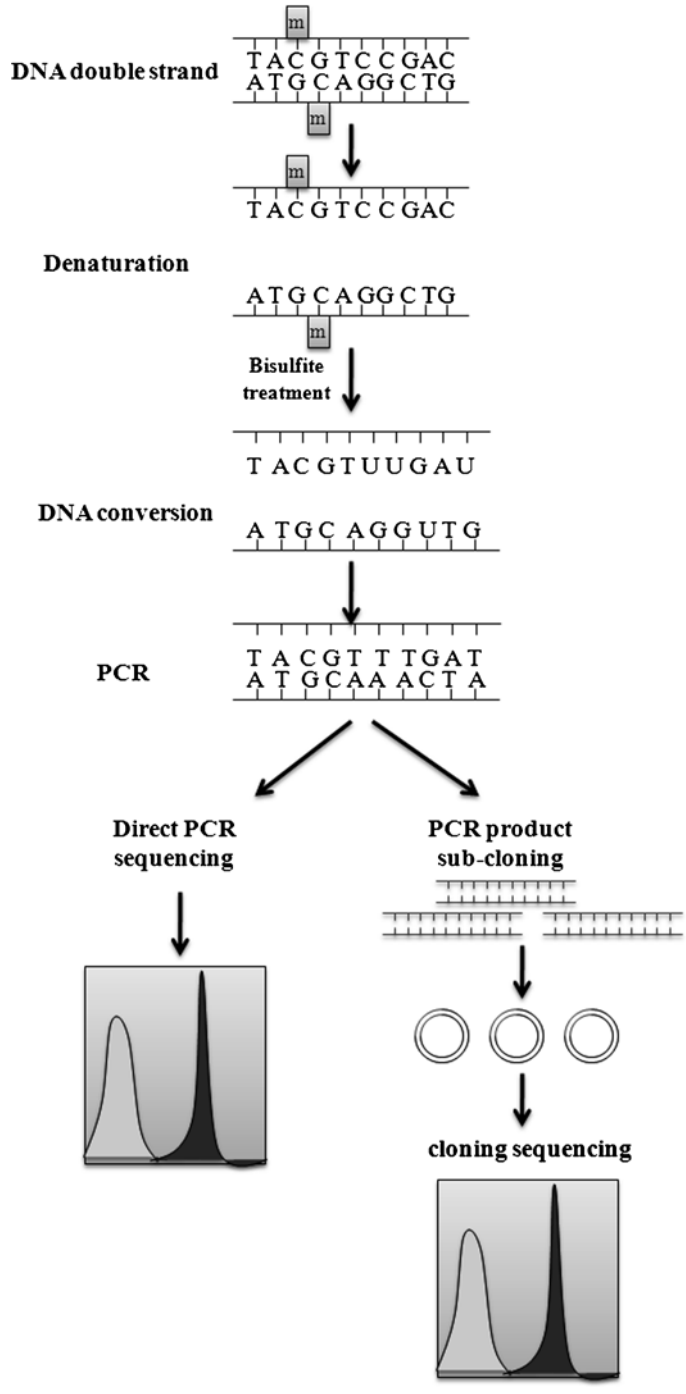


Fig. 1. Principles of methylation analysis using bisulfite genomic sequencing. After treatment with sodium bisulfite, unmethylated cytosine residues are converted to uracil whereas 5-methylcytosine (5mC) remains unaffected. After PCR amplification, uracil residues are converted to thymine. DNA methylation status can be determined by direct PCR sequencing or cloning sequencing.

genomic sequencing protocol, numerous modifications have been explored and have significantly improved the sensitivity and accuracy in this procedure (12–16). In this chapter, we introduce a modified bisulfite sequencing protocol, which is currently used in our laboratory and consistently working well. Detailed discussion of technical problems, solutions, and troubleshooting is also included to enhance this technique.

2. Materials

2.1. Genomic DNA Extraction

1. Wizard Genomic DNA purification kit (Promega).
2. 70% of ethanol.
3. Microcentrifuge.
4. Incubator at 65°C.

2.2. Bisulfite Reaction

1. Commercially available bisulfite reaction kit, EpiTect Bisulfite Kit (Qiagen).
2. For conventional bisulfite genomic treatment, 3 M NaOH solution, 5 M sodium bisulfite solution (with 125 mM hydroquinone, pH 5.0), 5 M ammonium acetate, isopropanol, ethanol, and mineral oil are needed. The mixtures of bisulfite reaction solutions are illuminated in Table 1. Bisulfite and hydroquinone solutions are light-sensitive and should be protected from light in all steps.
3. Wizard DNA clean-up system (Promega) for purification of bisulfite-treated DNA.
4. Disposable 5-ml lure-lock syringes.
5. Deionized water or TE buffer.

2.3. PCR Purification, Cloning, and Sequencing Analysis of Target DNA Fragment

1. Regular 2 × PCR Mastermix.
2. Agarose gel, ethidium bromide, and electrophoresis apparatus.
3. QIAquick PCR Purification Kit (Qiagen) for purification of PCR product.
4. QIAquick Gel Extraction Kit (Qiagen) for purification of target PCR fragment from multiple nonspecific PCR products.
5. pGEM-T Easy vector system II (Promega).
6. For bacterial culturing and positive cloning selection, bactotryptone (BD), yeast extract, sodium chloride, ampicillin solution, isopropyl-β-D-thiogalactoside (IPTG), X-Gal (Bio-Rad), and bacterial shaker incubator at 37°C are required.
7. QIAprep Spin Miniprep Kit (Qiagen).
8. ABI 3730 DNA Analyzer.

Table 1
Reagents of 5 M sodium bisulfite solution

Reagents/volumes	1 ml	2 ml	3 ml	4 ml	5 ml
Sodium bisulfite	0.475 g	0.95 g	1.425 g	1.9 g	2.375 g
Deionized water	0.625 ml	1.25 ml	1.875 ml	2.5 ml	3.125 ml
2 M NaOH (80 mg/ml)	175 µl	350 µl	525 µl	700 µl	875 µl
1 M hydroquinone (110 mg/ml)	125 µl	250 µl	375 µl	500 µl	625 µl

3. Methods

3.1. Genomic DNA Preparation (see Notes 1–5)

Genomic DNA from cultured cells, cultured bacteria, animal tissues, and paraffin-embedded tissue sections can be isolated by using a number of commercially available DNA Extraction Kits followed by the corresponding manufacturer's protocols. Genomic DNA (1–10 µg) is dissolved in deionized water with 18 µl final volume. Proceed to Subheading 3.2, step 1.

3.2. Bisulfite Modification

1. Predenature the DNA from Subheading 3.1 by boiling in a water bath for 20 min (see Note 6).
2. Denature the DNA by adding 2 µl of 3 M freshly made NaOH and 380 µl 5 M sodium bisulfite solution (Table 1) and mix well (see Note 7).
3. Add 500 µl of heavy mineral oil on the top of 400 µl DNA solution from step 2 to diminish evaporation and incubate the solution in the dark at 50°C for 12–16 h (see Note 8).
4. Purification of the bisulfite treated-DNA: We routinely use the Wizard DNA clean-up kit from Promega to purify the bisulfite-treated DNA. After carefully removing the heavy mineral oil from the reaction solution from step 3, the procedure involving this step is followed according to the manufacturer's protocol.
5. The bisulfite-modified DNA is eluted in 50 µl deionized water and 11 µl 3 M NaOH is added. Incubate at 37°C for 15 min to desulfonate the DNA.
6. Add 166 µl 5 M ammonium acetate, 750 µl of absolute ethanol, and 200 µl isopropanol to precipitate the DNA at –20°C for 2–4 h.
7. Centrifuge the DNA at maximum speed for 10 min and discharge the supernatant.
8. Wash the DNA with 200 µl 70% ethanol and centrifuge as in step 7.

9. Carefully remove the ethanol and dry the DNA pellet at room temperature for 10 min. Resuspend the DNA in 10–20 μl of TE or deionized water (see Note 9).

3.3. Bisulfite PCR Amplification

1. Bisulfite PCR primer design is critical for successful implementation of subsequent bisulfite sequencing analysis. The detailed guidelines for primer design of bisulfite-treated DNA templates are discussed in Note 10.
2. Bisulfite PCR amplification can be performed as a regular PCR reaction. However, the PCR conditions for amplifying bisulfite-treated material should be carefully optimized (see Note 11). The PCR results will be verified by gel-based electrophoresis and a single, bright, and specific band will be considered as a successful PCR amplification.

3.4. Direct PCR Sequencing and Cloning Sequencing (see Note 12)

1. Prior to the direct PCR sequencing, purification of PCR products is necessary to remove the residue of the PCR reaction that might interfere with the outcome of sequencing results. Commercially available kits such as QIAquick PCR Purification Kit (Qiagen) can be used for specific PCR fragments, whereas QIAquick Gel Extraction Kit (Qiagen) can help purify the target PCR product from multiple nonspecific PCR bands. The purified PCR products can be directly sequenced.
2. Cloning sequencing is necessary to observe the distribution of methylation patterns in single molecules. We prefer to use pGEM-T Easy vector system II (Promega) for cloning purposes which provides the T4 DNA ligase system, a pGEM-T Easy vector, and competent JM109 cells as well. By using this kit, purified PCR products can be ligated to the pGEM-T Easy vector and transformed into competent JM109 cells. The JM109 cells that carry the ligated vectors can be selected on agar plates containing ampicillin/X-gal/IPTG by color change where blue colonies represent empty vector, and white colonies represent vectors inserted with target PCR product. The white colonies can then be selected and grown in LB medium. Plasmids containing the target DNA are extracted by using the QIAprep Spin Miniprep Kit (Qiagen) and subjected to standard sequencing analysis. All the procedures follow the manufacturer's protocol.

3.5. Data Interpretation

After successful bisulfite PCR amplification or subcloning procedures, DNA methylation status can be interpreted by subsequent sequencing analysis. Direct sequencing of PCR products may be easily accessible, however, a series of problems limit its application such as failing to read the entire target region and high background interference. Cloning sequencing can provide useful methylation information on a molecular basis. To obtain high confidence in the results, a large number of clones (minimum 5, ideally 10) need to be sequenced, which can be time- and labor-intensive.

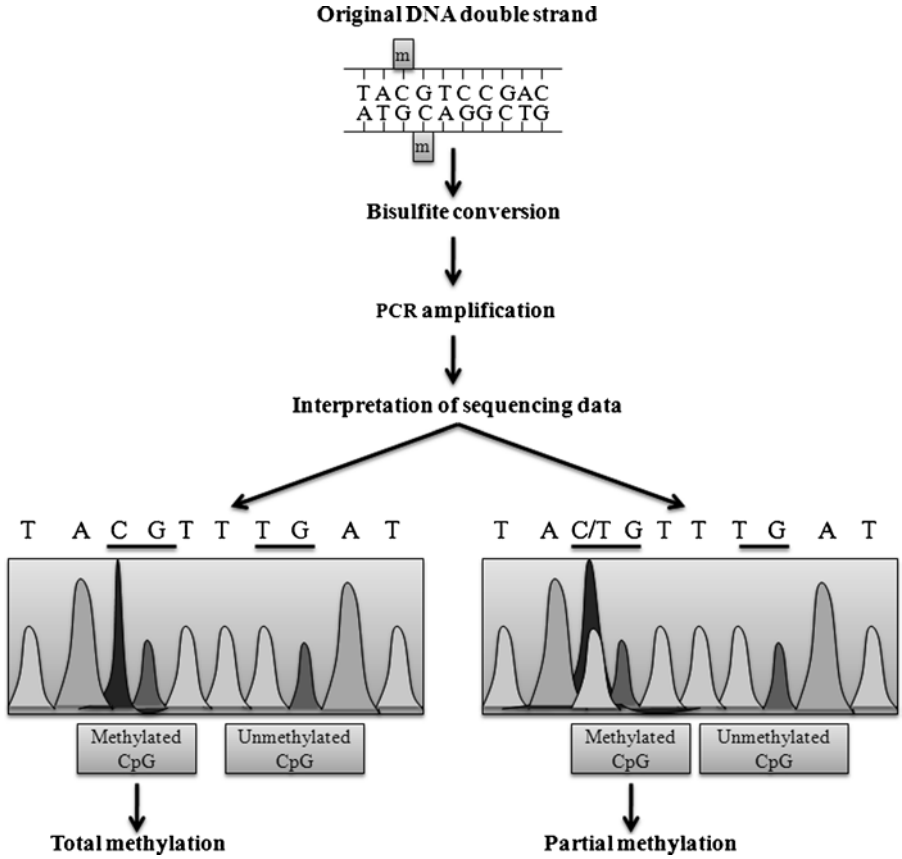


Fig. 2. Interpretation of methylation sequencing results. After bisulfite treatment, all unmethylated cytosines (C) convert to thymine (T) and the presence of a C peak indicates the presence of 5mC in the genome. Total methylation or complete conversion of a single residue shows a single peak. The presence of both C and T peaks indicates partial methylation or potentially incomplete bisulfite conversion.

Both procedures have the potential of artifacts due to incomplete conversion which may be prevented by strictly following Notes 1–9.

DNA methylation status can be interpreted by comparing the sequencing results and the original DNA sequence. Basically, all unmethylated cytosines (C) convert to thymine (T) and the presence of a C peak indicates the presence of 5mC in the genome. If both C and T peaks appear, this indicates partial methylation or potentially incomplete bisulfite conversion has occurred. The proportion of 5mC to C can be interpreted by analyzing the relative square area of these two bands (Fig. 2).

4. Notes

1. The quality and quantity of DNA are important in the bisulfite reaction. Typically, most protocols recommend >1 μg of high-quality DNA extracted from cultured cells or fresh tissue

samples to obtain reliable results. Therefore, samples from formalin-fixed, paraffin-embedded tissue sections may yield unfavorable results due to the limited quantity of initial DNA and subsequent DNA degradation during the bisulfite treatment. Several modifications have integrated into the conventional bisulfite protocol to help optimize analysis of low-quantity DNA such as samples from paraffin-embedded tissue blocks (17).

2. Alternatively, some have recommended an extended proteinase K treatment involving DNA isolation to increase accessibility for subsequent sodium bisulfite to DNA by removing residual protein (18). In this procedure, an overnight proteinase K (2 mg/ml) incubation at 37°C can be employed before proceeding to Subheading 3.2, step 1.
3. An additional DNA digestion by a restriction methylation-insensitive endonuclease can be incorporated prior to bisulfite modification which helps to reduce DNA strand annealing following denaturation (20).
4. A further process that reduces DNA losses during bisulfite modification is to embed the DNA in low-melting-point agarose blocks (14, 16). This modification will allow the subsequent bisulfite reaction to be performed in the agaroses where the DNA is physically captured. This will greatly reduce DNA loss during the procedures, especially when a small amount of DNA sample is applied.
5. Carriers such as salmon sperm DNA and glycogen can be used to increase bisulfite conversion and DNA precipitation, respectively, which also reduce the loss of DNA throughout the whole procedure (19).
6. The most critical step for the bisulfite reaction is DNA denaturation since sodium bisulfite can only react with cytosine in single-stranded DNA. Therefore, complete DNA denaturation is an essential prerequisite for successful DNA conversion by bisulfite treatment. According to the modification protocol originally developed by Frommer et al. (6), genomic DNA is denatured in high sodium bisulfite salt at high temperature and low pH. However, these harsh conditions can cause the DNA double-strand to form an unfavorable conformation leading to partial DNA renaturation, thus increasing the risk of an incomplete conversion reaction (20). Various modifications have been attempted to reduce strand reannealing that are listed below.
   (a) DNA can be fragmented by the use of proteinase K and appropriate restriction enzymes as mentioned in Notes 2 and 3.
   (b) DNA can be imbedded in low-melting point agarose block to prevent DNA reannealing as mentioned above in Note 4.

(c) DNA can be preboiled prior to bisulfite treatment to improve the denaturing step as described in Subheading 3.2, step 1.

(d) A high concentration of urea (6 M) may be added to the bisulfite solution to destabilize base-pairing in the DNA (21).

7. The bisulfite reaction solution in Table 1 and the NaOH solution must be freshly prepared each time prior to the conversion reaction.

8. A standard overnight incubation at 50°C ensures a complete reaction. However, the prolonged incubation was found to increase deaminated 5mC in the genome (12, 22) therefore leading to an underrepresentation of 5mC in subsequent PCR analysis. Although several groups have claimed that 4- to 5-h incubations with bisulfite are sufficient for complete conversion (13–15, 20, 23), we prefer to incubate DNA for more than 10 h of bisulfite reaction, which can yield complete conversion without further damage of DNA template. If further PCR reactions give poor results due to extensive DNA damage by a long incubation, a reduced incubation may be employed.

9. Owing to a noncomplementary DNA conformation after bisulfite treatment, the converted DNA is not stable and repeated freezing–thawing should be avoided. Freshly made bisulfite-modified DNA is recommended to yield optimal results. Using our protocol, the bisulfite-modified DNA can be used for up to a year posttreatment with good quality if stored at −80°C or in liquid nitrogen.

10. The principles for designing bisulfite PCR primers vary to meet different research purposes and protocols following the conventional bisulfite reaction. The primer guidelines listed below are used for bisulfite genomic sequencing analysis. A more detailed description of methylation primer design is provided in refs. 24, 25.

(a) After bisulfite treatment, the unmethylated cytosines convert to thymine and methylated cytosines remain cytosines. Since the methylation status of CpG dinucleotides is unknown, the bisulfite primer sequences should strictly avoid CpG dinucleotides. Therefore, primers should be generated to replace all cytosines to thymines according to the original DNA sequence. Primer designing software can also be used to avoid potential hairpin structures and possible primer dimers based on this modified sequence.

(b) The length of the primers should be around 25–30 nucleotides.

(c) The length of the PCR product should not exceed 400 bp due to potential DNA degradation during the bisulfite modification that might influence the PCR amplification.

11. The bisulfite PCR conditions should be carefully optimized. Since the bisulfite treatment reduces the specificity of DNA double-strands, the processes for determining the optimal PCR conditions with bisulfite-modified DNA template can be more laborious than regular PCR.
    (a) Annealing temperature: A gradient PCR thermocycler can help to determine the appropriate annealing temperature. If there is no access to a gradient PCR thermocycler, a touchdown PCR can be applied to increase the annealing sensitivity.
    (b) PCR reaction system: The commercially available PCR MasterMix which mixes *Taq* DNA polymerase and dNTP with optimal salt concentration can be easily used for bisulfite PCR. If this common PCR reaction system cannot produce a clean band, it is advisable to try a different PCR reaction system. In our laboratory, we normally use JumpStart (Sigma) or SureStart PCR system (Startagene) to improve the bisulfite PCR results.
    (c) Nested PCR reaction: A nested or a seminested PCR approach is recommended to obtain a sufficient PCR product especially when a limited amount of DNA is used.
12. pGEM-T Easy vector is a T-A cloning system and requires an extra adenosine added to the 5′ end of both strands during the PCR reaction. Therefore, it is important to choose the appropriate *Taq* DNA polymerase before cloning. Generally, sequencing requires more than five colonies to determine the accurate methylation patterns on a target region.

Acknowledgments

This work was supported in part by grants from the National Cancer Institute (R01 CA 129415), the Susan G. Komen for the Cure and a Postdoctoral Award (PDA) sponsored by the American Institute for Cancer Research (AICR).

References

1. Bird, A. (2007) Perceptions of epigenetics. *Nature* **447**, 396–398.
2. Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nature Rev. Genet.* **3**, 662–673.
3. Robertson, K. D. and Wolffe, A. P. (2000) DNA methylation in health and disease. *Nature Rev. Genet.* **1**, 11–19.
4. Bird, A. P. (1986) CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209–213.
5. Jones, P. A. and Baylin, S. B. (2002) The fundamental role of epigenetic events in cancer. *Nature Rev. Genet.* **3**, 415–428.
6. Frommer, M., McDonald, L. E., Millar, D. S., Collis, C.M., Watt, F., Grigg, G. W., Molloy, P. L.,

and Paul, C. L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* **89**, 1827–1831.

7. Rand, K., Qu, W., Ho, T., Clark, S. J. and Molloy, P. (2002) Conversion-specific detection of DNA methylation using real-time polymerase chain reaction (ConLight-MSP) to avoid false positives. *Methods* **27**, 114–120.
8. Xiong, Z. and Laird, P. W. (1997) COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* **25**, 2532–2534.
9. Gonzalgo, M. L. and Jones, P. A. (1997) Rapid quantitation of methylation differences at specific sites using methylationsensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.* **25**, 2529–2531.
10. Suzuki, M. M., and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9**, 465–476.
11. Bird, A. P. (1978) Use of restriction enzymes to study eukaryotic DNA methylation. II: the symmetry of methylated sites supports semiconservative copying of the methylation pattern. *J. Mol. Biol.* **118**, 48–60.
12. Clark, S. J., Harrison, J., Paul, C. L., and Frommer, M. (1994) High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* **22**, 2990–2997.
13. Raizis, A. M., Schmitt, F., and Jost, J. P. (1994) A bisulphite method of 5-methylcytosine mapping that minimises template degradation. *Anal. Biochem.* **226**, 161–166.
14. Olek, A., Oswald, J., and Walter, J. (1996) A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res.* **24**, 5064–5066.
15. Paulin, R., Grigg, G. W., Davey, M. W., and Piper, A. A. (1998) Urea improves efficiency of bisulphite-mediated sequencing of 5-methylcytosine in genomic DNA. *Nucleic Acids Res.* **26**, 5009–5010.
16. Hajkova, P., El-Maarri, O., Engemann, S., Oswald, J., Olek, A., and Walter, J., (2002) DNA-methylation analysis by the bisulfite-assisted genomic sequencing method. In: Mills, K.I., and Ramsahoye, B.H.(ed) DNA Methylation Protocols, Humana Press Inc., New Jersey.
17. Tan, L. W., and Dobrovic, A. (2001) Methylation analysis of formalin-fixed, paraffin-embedded sections using a nontoxic DNA extraction protocol. *Biotechniques* **31**, 1354, 1356–1357.
18. Warnecke, P. M., Stirzaker, C., Song, J., Grunau, C., Melki, J. R., and Clark, S. J. (2002) Identification and resolution of artifacts in bisulfite sequencing. *Methods* **27**, 101–107.
19. Herman, J. G., Graff, J. R., Myohanen S., Nelkin, B. D., and Baylin, S. B. (1996) Methylationspecific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. USA* **93**, 9821–9826.
20. Rein, T., Zorbas, H., and DePamphilis, M. L. (1997). Active mammalian replication origins are associated with a high-density cluster of mCpG dinucleotides. *Mol. Cell Biol.* **17**, 416–426.
21. Paulin, R., Grigg, G. W., Davey, M. W., and Piper, A. A. (1998) Urea improves efficiency of bisulphite-mediated sequencing of 59-methylcytosine in genomic DNA. *Nucleic Acids Res.* **26**, 5009–5010.
22. Wang, R. Y., Gehrke, C. W., and Ehrlich, M. (1980). Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res.* **8**, 4777–4790.
23. Shapiro, R., Servis, R. E., and Welcher, M. (1970b). Reactions of uracil and cytosine derivatives with sodium bisulphite. A specific deaminationmethod. *J. Am. Chem. Soc.* **92**, 422–424.
24. Li, L. C. (2007) Designing PCR primer for DNA methylation mapping. *Methods Mol. Biol.* **402**, 371–384.
25. Li, L. C. (2002) Dahiya R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **18**, 1427–1431.

Chapter 3

Methylation-Specific PCR

Ja-Lok Ku, You-Kyung Jeon, and Jae-Gahb Park

Abstract

DNA methylation patterns in CpG-rich regions of promoter, CpG islands, are concerned in regulation of gene expression in mammalian cells. Excessive methylation of CpG dinucleotides in promoter represses the gene expression. In cancer, especially, gene silencing is occurred through aberrant methylation in promoter of tumor suppressor genes. Methylation-specific PCR (MSP) is a method for analysis of DNA methylation patterns in CpG islands. For performing MSP, DNA is modified by and PCR performed with two primer pairs, which are detectable methylated and unmethylated DNA, respectively. MSP is a rapid measure for assession of the methylation status in CpG island.

Key words: Methylation-specific PCR, CpG islands, Bisulfite conversion, Methylation and unmethylation primer, MSP primer design

1. Introduction

DNA methylation is an important epigenetic modification that occurs at the carbon 5 of pyrimidine ring of cytosine in CpG dinucleotides. Methyl group on C-5 of cytosine in promoter is maintained and added by family of DNA methyltransferases (DNMTs) (1). Extensive methylation in CpG-rich areas of promoters, known as CpG islands, interrupts gene expression and leads gene silencing. The suppression might be caused that transcription factors cannot bind to methylated promoters (1, 2).

Regulation of gene expression by DNA methylation is associated with various fields of mammalian genetic control such as embryonic development, differentiation, regulation of imprinted gene expression, X chromosome inactivation, and tumor suppressor gene silencing in human cancer (1–3). In cancer, many tumor suppressor genes, such as RUNX3, p16, APC, were shown hypermethylated state of the promoter and repressed the gene expression.

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_3,

This suppression due to DNA methylation assists tumor initiation and development (4, 5).

Methylation-specific PCR (MSP) is a PCR-based method for the analysis of methylation patterns in CpG islands. To perform the MSP assay, DNA is purified and undergoes modification by using sodium bisulfate, in a sulphonation process (Fig. 1) that converts all unmethylated cytosine residues to uracil (6, 7). Methylated cytosine, in which a methyl group is attached to carbon 5, cannot be converted since the sulphonation reaction cannot occur (8).

For a MSP experiment, two pairs of primers are needed. One pair is specific for methylated DNA (M) and the other is specific for unmethylated DNA (U) (Fig. 2). For discrimination of methylated and unmethylated DNA, one or more CpG sites are included in each primer (or at least one of the pair) sequence (9). PCR reactions are performed using each primer pair, M and U primer pair. Successful amplification from the M primer pair is indicative of

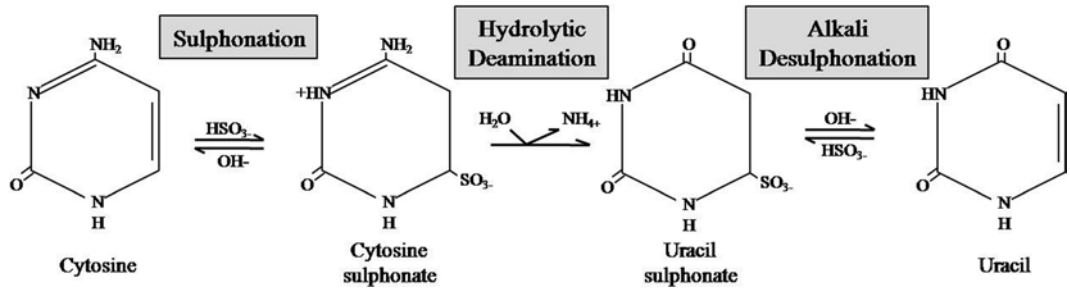


Fig. 1. A mechanism of nucleotide conversion from cytosine to uracil with sodium sulfite treatment (Clark et al., Nucleic Acids Res., 1994, 22:2290–2997).

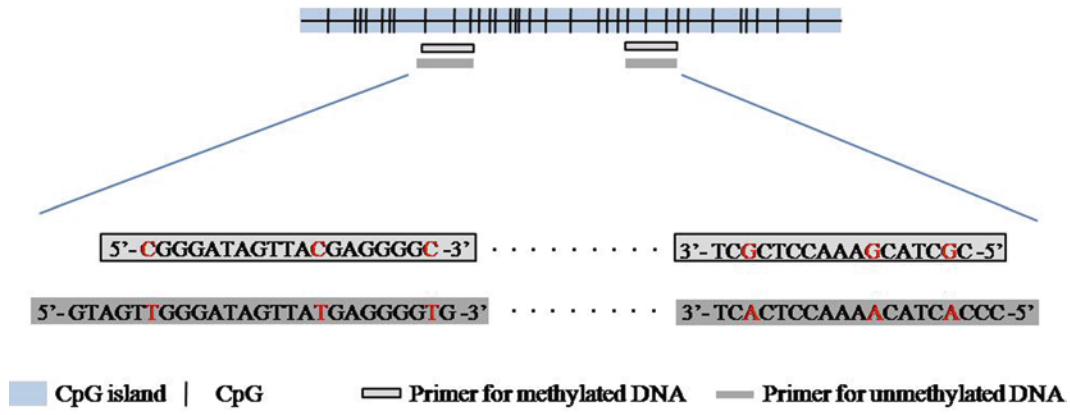


Fig. 2. Primers selection for methylation-specific PCR. Primers for MSP are required in pairs which detect only methylated DNA (M primers) and only unmethylated DNA (U primers). Primers contain at least more than one CpG site and two pairs of primers include same CpG sites. However, two sets of primers including same CpC sites could not be of same length and start point. Origin sequences are from the promoter region of RUNX3 gene (accession number, NM_004350).

methylated DNA and PCR product by U primer pair is reflective of unmethylated DNA.

MSP can rapidly assess the methylation status of almost any cytosine of CpG sites within a CpG island, apart from the use of cloning or methylation-sensitive restriction enzymes. MSP requires very small quantities of DNA, is sensitive to 0.1% methylated alleles of a given CpG island locus, and can be performed in DNA extracted paraffin-embedded samples (6).

2. Materials

2.1. Sample Preparation and Lysis

1. Culture medium supplemented with 10% fetal bovine serum (FBS).
2. Centrifuge tubes, 15 ml.
3. Culture flasks, 75 cm^2.
4. Sterile phosphate-buffered saline (PBS), pH 7.0.
5. Trypsin–EDTA, 0.05%.
6. Xylene.
7. Ethanol, 75 and 100%.

2.2. Cell Lysis and DNA Isolation

1. Proteinase K lysis buffer: proteinase K 20 mg/ml, 50 µl; 1 M Tris–HCl solution, 10 µl; 0.5 M EDTA, 2 µl; 10% sodium dodecyl sulfate (SDS), 100 µl; distilled water, 838 µl.
2. Phenol solution, phenol:chloroform:isopropanol in a ratio of 25:24:1.
3. Chloroform.
4. Sodium acetate, 3 M.
5. Isopropanol.
6. Cold 75% ethanol.
7. Distilled water or TE buffer.
8. Liquid nitrogen.
9. Mortar and pestle.
10. QIAamp DNA mini kit (Qiagen, Valencia, CA, USA).

2.3. DNA Modification

1. NaOH, 3 M.
2. Hydroquinone, 10 mM.
3. Sodium bisulfate, 3 M.
4. Wizard DNA purification resin (Promega, Madison, WI, USA) or another DNA purification kit as preferred.
5. Sodium acetate, 3 M.
6. Isopropanol.

7. Cold 75% ethanol.
8. Distilled water or TE buffer.
9. CpGenome DNA modification kit (Intergen, Edinburgh, UK) or Ez DNA methylation kit (Zymo Research, Orange, CA, USA) or CpGenome Fast DNA modification kit (Millipore, Billerica, MA, USA).
10. CpGenome™ Universal Methylated DNA (Chemincon, Temecula, CA, USA).

3. Methods

3.1. Preparation of Samples

3.1.1. Sample Preparation from Cell Lines

1. Remove the culture medium from the culture flask (or dish) and rinse cultured cells with sterile PBS.
2. Add 5 ml of trypsin–EDTA (usually 0.05%) in PBS and incubate at 37°C for approximately 3 min.
3. Add 10 ml of culture medium supplemented with 10% FBS and transfer the cells into centrifuge tubes.
4. Centrifuge at 1,500 × *g* for 3 min at RT and decant the supernatant from the pellets.

3.1.2. Sample Preparation from Frozen Tissues

1. Approximately 2 g of the surgically removed tissues are frozen immediately and stored in liquid nitrogen.
2. The remaining sections of the samples are fixed with formalin and used for histological examination to confirm the diagnosis postoperatively.
3. Pour liquid nitrogen into the precooled mortar and place the frozen tissues in the mortar.
4. Finely grind the frozen tissues with the precooled pestle, adding fresh liquid nitrogen to replace evaporated liquid nitrogen.
5. Collect the finely grinded tissues into a precooled 15 ml tube.

3.1.3. Sample Preparation from Paraffin-Embedded Tissues

1. Cancer (or normal) cells are precisely obtained from H&E stained slides using a 30-gauge needle or laser capture microdissection devices.
2. Dissected cells are collected in a 1.5-ml Eppendorf tube.
3. Add 1 ml of xylene to the tube to remove paraffin during incubation at RT for 30 min.
4. Centrifuge at 18,000 × *g* for 5 min at RT and discard the supernatant. Repeat this step twice.
5. Add 1 ml of 100% ethanol and incubate at RT for 30 min.
6. Centrifuge at 18,000 × *g* for 5 min at RT and discard the supernatant.

7. Add 1 ml of 75% ethanol and incubate at RT for 30 min.
8. Centrifuge at 18,000 × *g* for 5 min at RT and discard the supernatant.

3.2. DNA Isolation from Collected Samples

3.2.1. DNA Isolation from Frozen Tissues and Cell Pellets

1. Add 500 μl of proteinase K lysis buffer.
2. Incubate at 50°C with shaking until all cells and tissue fragments are completely dissolved.
3. Add 500 μl of the phenol solution to the tube containing the cell lysate and mix by vortexing.
4. Centrifuge for 5 min at 18,000 × *g* at RT.
5. Transfer the supernatant fluid to a new 1.5 ml Eppendorf tube using a pipette.
6. Add 1 volume of chloroform to the supernatant and mix by vortexing.
7. Centrifuge at 18,000 × *g* for 5 min at RT.
8. Transfer the supernatant fluid to a fresh 1.5 ml Eppendorf tube using a pipette.
9. Add 0.1 volume of 3 M sodium acetate and mix by vortexing.
10. Add 1 volume of isopropanol and incubate at –20°C overnight.
11. Centrifuge at 18,000 × *g* for 30 min at 4°C and discard the supernatant.
12. Add 500 μl of cold 75% ethanol and wash the DNA pellet.
13. Centrifuge at 18,000 × *g* for 10 min at 4°C and discard the supernatant.
14. Air-dry the DNA pellet and dissolve in 50 μl of distilled water or TE buffer.

3.2.2. DNA Isolation from Paraffin-Embedded Tissues

1. Place the collected cells from paraffin-embedded tissues into a 1.5-ml microcentrifuge tube, and add 50 μl of proteinase K lysis buffer and incubate at 50°C overnight (see Note 1).
2. DNA can be isolated through the aforementioned proteinase K lysis protocol.

3.3. DNA Modification Using Sodium Bisulfite

1. Denature 1–2 μg of DNA in a volume of 50 μl by adding 3 M NaOH to final concentration of 0.3 M.
2. Incubate at 37°C for 10 min.
3. Add 30 μl of 10 mM hydroquinone and 520 μl of 3 M sodium bisulfate at pH 5.0.
4. Mix gently and incubate under mineral oil at 50°C for 16 h.
5. Purify modified DNA using Wizard DNA purification resin according to the manufacturer (Promega) or another preferred kit.

6. Elute into 50 μl of distilled water.
7. Neutralize by adding 0.1 volume of 3 M sodium acetate and mix by vortexing.
8. Add 1 volume of isopropanol and mix gently.
9. Centrifuge at 18,000 × *g* for 30 min at 4°C and discard the supernatant.
10. Add 500 μl cold 75% ethanol and wash the DNA pellet.
11. Centrifuge at 18,000 × *g* for 10 min at 4°C and discard the supernatant.
12. Air-dry the DNA pellet and dissolve in 50 μl of distilled water or TE buffer.

Sodium bisulfite modification kits are also commercially available such as CpGenome DNA modification kit (Intergen), Ez DNA methylation kit (Zymo Research), and CpGenome Fast DNA modification kit (Millipore).

CpGenome Universal Methylated DNA (Chemincon) is often used as a control DNA.

3.4. Primer Design

3.4.1. Prediction of CpG Islands

For primer design of MSP, CpG islands have to be predicted with the promoter sequence of the target gene. The CpG island is a sequence at least 200 bp length in the promoter, which has a CG content exceeding 50% and an observed CpG/expected CpG ratio exceeding 0.6 (10) (see Note 2). It is also needed for sequence conversion using bisulfate-modified DNA. The conversion can be performed in two different versions (Fig. 3). (1) Methylation DNA; methylated cytosine of CpG dinucleotides remains intact, but cytosine of non-CpGs is converted to thymine. (2) Unmethylated DNA; all cytosines including CpG dinucleotides

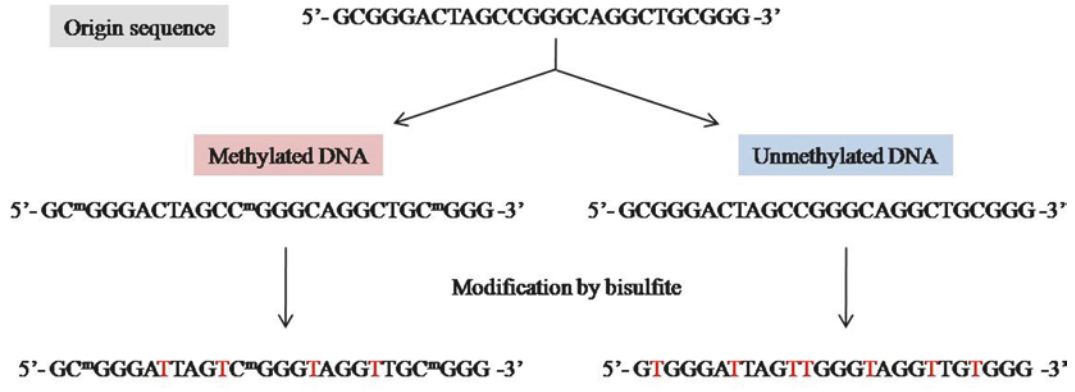


Fig. 3. DNA modification by sodium bisulfate. All cytosines in CpGs and non-CpGs except the methylated cytosines in CpGs were converted to thymine (replacement of uracil in DNA). Origin sequences are from the promoter region of RUNX3 gene (accession number, NM_004350).

are converted to thymine. Two pairs of primers can be determined for methylated or unmethylated DNA based on the converted sequences.

The above procedures can be performed by online software with the promoter sequence of the target gene. The websites are

- MethPrimer: http://www.urogene.org/methprimer/index1.html (Fig. 4).
- Primo: http://www.changbioscience.com/primo/primom.html.
- BiSearch: http://bisearch.enzim.hu/.

Also, Methylprimer Express Software 1.0 is available for free from Applied Biosystems web site.

3.4.2. Primer Selection

CpG Island Prediction and Primer Selection

Concerning the predicted CpG islands, several rules govern the choice of appropriative primers as target region.

1. If more than one island is predicted, any of them can be the target region for amplification.
2. If a CpG island size is smaller than the minimal product size, the primer pair should span the whole island.
3. If a CpG island size is greater than the maximal product size, the primer pair should be within the island.
4. If a CpG island size is between the minimal and maximal product size (see Note 3), primer pair should involve at least two thirds of the island.

Primer Selection for MSP

For performing MSP, two pairs of primers are needed. One of the pairs amplifies modified and methylated DNA and the other amplifies modified and unmethylated DNA. The following restrictions are applied to primer selection for MSP:

1. Primers should include at least one CpG dinucleotide, and at least one of the last three bases in the primer should be a cytosine of CpG for maximal discrimination between methylated and unmethylated DNA.
2. More CpG sites in the primer are preferred including CpG dinucleotide(s) at the most 3′ end.
3. Primers for methylated DNA and for unmethylated DNA should contain the same CpG dinucleotides within their sequence. However, they may vary in length or at their start point for matching of temperature value between two sets of primers (see Note 4).
4. Two pairs of primers having similar annealing temperature are preferred and annealing temperatures equal or higher than 55°C is preferred.


```
Sequence Name:
Sequence Length: 420

CpG island prediction results
Criteria used: Island size > 100, GC Percent > 50.0, Obs/Exp > 0.60
 1 CpG island(s) were found in your sequence
            Size         (Start - End)
 Island 1   319 bp        (47 - 365)

Primer picking results for methylation specific PCR (MSP)

    Primer             Start Size  Tm      GC%    'C's  Sequence
1   Left  M primer      133   19   61.73   78.95    3   CGGGATAGTTACGAGGGGC
    Right M primer      236   17   57.66   76.47    3   CGCTACGAAACCTCGCT
    Product size: 104, Tm: 78.6
    Left  U primer      128   25   61.14   80.00    5   GTAGTTGGGATAGTTATGAGGGGTG
    Right U primer      238   19   52.47   78.95    3   CCCACTACAAAACCTCACT
    Product size: 111, Tm: 73.7


  121 GGCCCAGGCAGCCGGGACAGCCACGAGGGGCGGCCGCACGCGGGGCCGCGCGCCGAGGAT
      +|:::|||:||:++|||:||::|++|||||++|:++:|++++|||:++++++:++|||||
  121 GGTTTAGGTAGTCGGGATAGTTACGAGGGGCGGTCGTACGCGGGGTCGCGCGTCGAGGAT
                M>>>>>>>>>>>>>>>>>>>
              U>>>>>>>>>>>>>>>>>>>>>>>>>

  181 GCGGGACTAGCCGGGCAGGCTGCGGGCGGCCGTCGGGCCAGCGAGGCCTCGCAGCGGGCG
      |++|||:|||:++||:|||:||++||++|:++|++||::||++|||::|++:||++||++
  181 GCGGGATTAGTCGGGTAGGTTGCGGGCGGTCGTCGGGTTAGCGAGGTTTCGTAGCGGGCG
                                            M<<<<<<<<<<<<<<<<<<
                                            U<<<<<<<<<<<<<<<<<<<<<<
```

Fig. 4. Result of CpG island prediction and primer selection using MethPrimer software (online). Input sequence is a part of promoter region of *RUNX3* gene (accession number, NM_004350).

3.5. Methylation-Specific PCR

3.5.1. Direct MSP

After primer design and selection, MSP is performed with general PCR conditions. For example, PCR conditions consist of 5 min at 94°C for initial denaturation, followed by 40 cycles of 94°C (30 s), annealing temperature of primers (30 s), and 72°C (60 s) and a final elongation of 7 min at 72°C. Finally, amplified products are loaded onto a 2% agarose gel and electrophoresed for 20 min at 100 V.

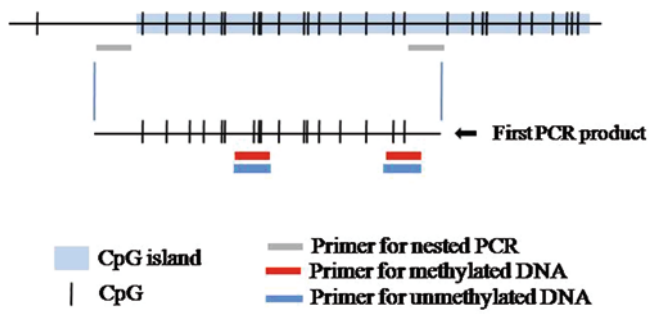


Fig. 5. Diagram of nested PCR. When the products of direct MSP primers is not amplified definitely, it needs another primer sets surrounding the products of direct MSP primers.

3.5.2. Nested MSP

If an experiment could not abundantly amplify the product for analysis by direct MSP, nested MSP can be performed. Nested MSP requires one more primer set, which covers the sequence of the amplified product with selected two pairs of primers (Fig. 5). After the first PCR with primers of nested MSP, a second PCR with two pairs of primers (each primer set for different states of methylation) is performed using the amplified products from the first PCR. They should not contain "C" of CpG dinucleotides because both methylated and unmethylated DNA are amplified with them. PCR conditions consist of 5 min at 94°C for initial denaturation, followed by 20 cycles for first or 35 cycles for second of 94°C (30 s), annealing Tm of primers (30 s), and 72°C (60 s) and a final elongation of 7 min at 72°C. Final products electrophoresed for 20 min at 100 V in a 2% agarose gel.

4. Notes

1. In the case of paraffin-embedded tissues, most of the DNA is low molecular weight, and so the majority can be lost through the aforementioned standard proteinase K DNA isolation protocol (Subheading 3.2.1). To minimize this loss, boiling for approximately 10 min will inactivate proteinase K. The remaining solution containing DNA in the tube can be directly used for further analysis. In this approach, the phenol solution is not necessary. High-quality DNA of the collected cells from paraffin-embedded tissues can be also isolated using a commercial DNA isolation kit such as QIAamp DNA mini kit (Qiagen).
2. The CpG count is the number of CG dinucleotides in island. The percentage CpG is the ratio of CpG nucleotide bases

(twice the CpG count) to the length. The ratio of observed to expected CpG is calculated according to the formula (10):

$$\frac{\text{Obs}}{\text{Exp CpG}} = \frac{\text{number of CpG} \times N\left(\text{length of sequence}\right)}{(\text{number of C} \times \text{number of G})}.$$

3. When MSP performed, PCR product size could be predicted optimal length according to specifications including minimal and maximal product size.
4. Bisulfite-modified DNA is a different sequence between methylated and unmethylated DNA. If two primer sets are made in same length and from same start point, the pair for methylated DNA, which remains cytosine with methyl group, has higher annealing temperature than the pair for unmethylated DNA, which is converted to thymine from cytosine.

References

1. Bird AP, Wolffe AP. (1999) Methylation-induced repression-belts, braces, and chromatin. *Cell* 99(5), 451–454.
2. En Li. (2002) Chromation modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* 3, 662–673.
3. Song F, Smith JF, Kimura MT, Morrow AD, Matsuyama T, Nagase H, et al. (2005) Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proc Natl Acad Sci USA* 102, 3336–3341.
4. Herman JG, Jen J, Merlo A, Baylin SB. (1996) Hypermethylation-associated inactivation indicates a tumor suppressor role for p15INK4B. *Cancer Res* 56(4), 722–727.
5. Suzuki M, Yoshino I. (2010) Aberrant methylation in non small cell lung cancer. *Surg Today* 40(7), 602–627.
6. Herman JG, Graff JR, Myöhänen S, Nelkin BD, Baylin SB. (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci USA* 93, 9821–9826.
7. Oakeley EJ.(1999) DNA methylation analysis: a review of current methodologies. *Pharmacol Ther* 84, 389–400.
8. Frommer M, McDonald L E, Millar D S, Collis C M, Watt F, Grigg G W, Molloy P L, and Paul C L (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci US.* 89, 1827–1831.
9. Li LC, Dahiya R. (2002) MethPrimer: designing primers for methylation PCRs. *Bioinformatics* 18, 1427–1431.
10. Gardiner-Carden M, Frommer M. (1987) CpG islands in vertebrate genomes. *J Mol Biol* 196(2), 261–282.

Chapter 4

Analysing DNA Methylation Using Bisulphite Pyrosequencing

Thomas Mikeska, Jörg Felsberg, Chelsee A. Hewitt, and Alexander Dobrovic

Abstract

Bisulphite pyrosequencing is a quantitative methodology for the investigation of DNA methylation of sequences up to 100-bp in length. Biotin-labelled, single-stranded PCR products generated from bisulphite-treated DNA are used as a template with an internal primer to perform the pyrosequencing reaction. Nucleotides are added in a predetermined order in each pyrosequencing cycle and the amount of incorporated nucleotide results in a proportional emission of light. DNA methylation ratios are calculated from the levels of light emitted from each nucleotide incorporated at individual CpG positions in a strand-dependent manner. The methylation detection limit at individual CpG sites is approximately 5% and the results are displayed as an average methylation level for each CpG position assayed across all amplification products generated during a PCR reaction. As a consequence, bisulphite pyrosequencing allows the identification of heterogeneous DNA methylation patterns but does not provide information at a single allele resolution. This methodology is suited to analyse short DNA sequences such as those typically extracted from formalin-fixed paraffin-embedded specimens. Nevertheless, longer PCR products can be sequenced by serial bisulphite pyrosequencing, which utilises tandem assays along the amplicon. The general information provided is applicable for all formats of current pyrosequencing instruments, however, a specific protocol for the PyroMark Q24 instrument is provided.

Key words: PyroMark Q24, Quantitative DNA methylation analysis, Heterogeneous methylation, Bisulphite primer design, *MGMT*

1. Introduction

Pyrosequencing is a sequencing-by-synthesis technique (1). It can be used for *de novo* sequencing, but it is most commonly used to determine the base composition of a single or a small number of variable positions within a region of known sequence. The technique is relatively cost effective, delivers the results quickly and most importantly in a quantitative manner. These features have opened

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_4, © Springer Science+Business Media, LLC 2011

pyrosequencing to a broad range of applications encompassing the analysis of genetic variation, genotyping, determination of gene copy number (reviewed in refs. 2–4) and the analysis of DNA methylation (5–7).

The analysis of DNA methylation patterns by bisulphite pyrosequencing has been applied to the investigation of single gene loci (e.g., (8–12)) and where applicable in an allele-specific manner (13).

Bisulphite pyrosequencing has also been used to investigate genome-wide DNA methylation levels present in repetitive elements, such as ALU repeats (e.g., (14–16)), LINE-1 repeats (e.g., (15, 17–19)), and tandem repeats (20). However, in this chapter, we focus on the analysis of DNA methylation of single gene loci following bisulphite treatment.

It is important to note that determination of the DNA methylation level at a certain CpG position is based on parallel sequencing of all possible templates present in a single PCR reaction and is displayed as an average. As a consequence, while pyrosequencing can accurately determine the amount of methylation above the threshold level (approximately 5%) at each CpG position, it will underestimate the number of methylated alleles in heterogeneously methylated templates (for further details see ref. 21).

The accuracy of pyrosequencing to deliver quantitative results is mainly determined by the reaction kinetics of the enzyme cascade in each cycle. We describe here an overview (Fig. 1) of the different steps of each pyrosequencing cycle. Further details of the sophisticated biochemistry are reviewed in refs. 22, 23.

PCR products amplified from bisulphite-treated DNA are used as the template for bisulphite pyrosequencing. The PCR reaction is performed in the presence of either a biotinylated forward or reverse PCR primer, depending on the strand that one wishes to pyrosequence. The positions of the bisulphite-specific PCR primers are generally chosen so that sequences are amplified regardless of their DNA methylation status. The biotinylated PCR products are then captured by binding to streptavidin-coated Sepharose beads and the unlabelled strand is washed away following strand separation under denaturing conditions. An internal pyrosequencing primer, proximal to the region of interest, is then annealed to the single-stranded PCR product. During the assembly step, the primed and single-stranded DNA template is extended in a stepwise fashion in each pyrosequencing cycle by the incorporation of the nucleotide dATP-α-S (deoxyadenosine alpha-thio triphosphate) (see Note 1), dCTP, dGTP, or dTTP according to the nucleotide dispensation order, which is determined by the sequence being analysed. The nucleotide is incorporated by a mutated Klenow fragment of DNA polymerase I, which lacks proofreading 3′→5′ exonuclease activity to avoid out of phase sequencing.

The successful addition of a nucleotide to the polynucleotide results in the release of inorganic pyrophosphate (PPi) which triggers

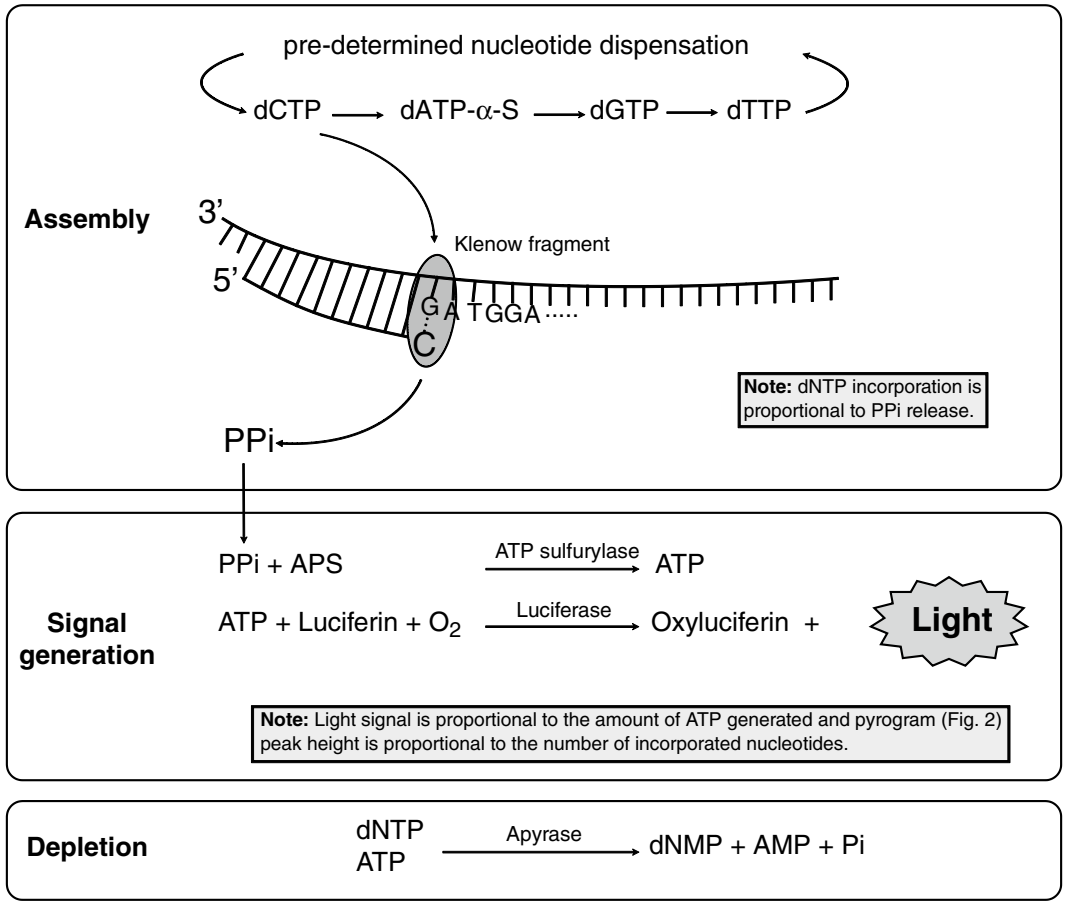


Fig. 1. A general overview of the pyrosequencing cascade reaction. *dATP-α-S* deoxyadenosine alpha-thio triphosphate; *dCTP* deoxycytidine-5′-triphosphate; *dGTP* deoxyguanosine-5′-triphosphate; *dTTP* thymidine-5′-triphosphate; *PPi* inorganic pyrophosphate; *APS* adenosine-5′-phosphosulfate; *ATP* adenosine-5′-triphosphate; *dNTP* deoxyribonucleotide-5′-triphosphate; *dNMP* deoxyribonucleotide-5′-monophosphate; *AMP* adenosine-5′-monophosphate; *Pi* inorganic phosphate. For a more extensive description see Subheading 1.

a reaction cascade and light signal generation. The generated PPi is a substrate for ATP sulfurylase, which synthesises adenosine-5′-triphosphate (ATP) from adenosine-5′-phosphosulfate (APS).

The newly synthesised ATP is then used by the luciferase enzyme in the presence of D-luciferin and oxygen to generate oxyluciferin and light (see Note 2). The amount of emitted bioluminescence is equivalent to the amount of incorporated nucleotides and is detected by a charge-coupled device (CCD) camera. A linear correlation is observed for the incorporation of up to three dATP-α-S residues or five dCTP, dGTP, or dTTP residues in a row.

In the final depletion step, apyrase cleans up the reaction mixture prior to the next cycle of nucleotide addition by degrading the unincorporated nucleotides and ATP into dNMPs, AMP, and two inorganic phosphate (Pi) residues (see Note 3).

The dispensation order of nucleotides depends on the downstream application of the pyrosequencing and on which of the strands of DNA has been biotinylated. When the sequence is unknown, for example in *de novo* sequencing, the dispensation order is dATP-α-S, dCTP, dGTP, dTTP during every cycle. When the majority of sequence is known, such as for quantification of methylation at CpG islands a single and fixed nucleotide is dispensed during each cycle except at the variable positions, which are then essentially treated as C/T SNPs (G/A SNPs in the reverse direction). At these positions, the combination of known possible nucleotides is dispensed.

If the reverse PCR primer is biotinylated and the reverse strand is thus used as the template (as is most common), the bisulphite pyrosequencing primer will be extended according to the base composition of the forward strand during the pyrosequencing reaction. Therefore, the relative amount of methylation at a CpG position is determined by the ratio of incorporated dCTP (methylated CpG site) and dTTP (unmethylated CpG site). Conversely, when it is preferable to label the forward strand with biotin (often called a "reverse assay"), the bisulphite pyrosequencing primer is extended according to the base composition of the reverse strand and the amount of DNA methylation at an individual CpG site is calculated from the ratio of incorporated dGTP (methylated CpG site) and dATP-α-S (unmethylated CpG site).

The major drawback of the pyrosequencing reaction compared to Sanger sequencing is the limited read length. However, this is assay dependent and read lengths of up to 100 bp can be achieved (24, 25). The read length is limited by out of phase DNA extension and the accumulation of by-products during each pyrosequencing cycle.

To investigate the DNA methylation status of longer amplicons, serial bisulphite pyrosequencing of tandem assays can be used (26). This approach investigates PCR products with more than one bisulphite pyrosequencing primer in independent and consecutive bisulphite pyrosequencing reactions using templates recovered from the previous bisulphite pyrosequencing run.

Pyrosequencing does not suffer from loss of sequence data adjacent to the primer encountered by Sanger sequencing. Depending on the assay, nucleotides closer than 10–25 bp downstream of Sanger sequencing primers show poor peak quality and result in loss of information. Bisulphite pyrosequencing generates reliable sequence data from the nucleotide immediately 3′ of the pyrosequencing primer. Thus, bisulphite pyrosequencing is suited for the analysis of PCR products generated from degraded genomic DNA extracted from sources such as formalin-fixed paraffin-embedded (FFPE) tissue (e.g., (27–29)). (Genomic DNA extracted from FFPE tissue is usually fragmented and amplifying a shorter PCR product results in more target sequences being available).

Bisulphite pyrosequencing is also superior to direct Sanger sequencing of bisulphite PCR products showing heterogeneous DNA methylation patterns (reviewed in ref. 21). The determination of DNA methylation levels at a given CpG position calculated from dye peaks is influenced by several factors and may need the assistance of appropriate software packages for normalisation (30). Bisulphite pyrosequencing, on the other hand calculates the DNA methylation level directly from the ratio of emitted light following the incorporation of nucleotides at the given CpG position.

Pyrosequencing instruments are currently available in 24- and 96-well formats. The following protocol is for the PyroMark Q24 instrument but most of the information provided is also valid for the 96-well formats. Specific protocols for 96-well formats have been published elsewhere (e.g., (31, 32)).

2. Materials

2.1. Primer Design

1. MethPrimer: http://www.urogene.org/methprimer/index.html.
2. BiSearch Primer Design and Search Tool: http://bisearch.enzim.hu/.
3. Methyl Primer Express (Life Technologies): http://www.appliedbiosystems.com.
4. PyroMark Assay Design Software (Qiagen).
5. RepeatMasker: http://www.repeatmasker.org/.
6. Ensembl database: http://www.ensembl.org/index.html.
7. SNP database (dbSNP) BLAST: http://www.ncbi.nlm.nih.gov/projects/SNP/index.html.
8. OligoAnalyzer (Integrated DNA Technologies): http://www.idtdna.com.
9. MethMarker: http://methmarker.mpi-inf.mpg.de/.

2.2. Bisulphite Treatment and Preparation of DNA Methylation Standard Series

1. Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Bio-Sciences, Cat. No. 25-6600-30), store at –80°C.
2. QIAquick PCR Purification Kit (Qiagen, Cat. No. 28104).
3. CpGenome Universal Methylated DNA (Millipore, 10 µg, Cat. No. S7821), store at –20°C.
4. CpG Methyltransferase (M.SssI) (New England BioLabs, Cat. No. M0226S), store at –20°C.
5. EpiTect Bisulfite Kit (Qiagen, Cat. No. 59104), MethylEasy *Xceed* (Human Genetic Signatures, Cat. No. Me002) or the EZ DNA Methylation-Gold Kit (Zymo Research, Cat. No. D5005).

6. HotStarTaq DNA Polymerase (Qiagen, Cat. No. 203203), store at –20°C.
7. SYTO 9 fluorescent dye (5 mM solution in DMSO, Life Technologies, Cat. No. S-34854), store at –20°C.

2.3. Bisulphite-Specific PCR

1. Biotin-labelled primer (5′-modification, HPLC purified [see Notes 4 and 5]; e.g., Sigma-Aldrich).
2. Unlabelled PCR primer (desalted, PCR quality).
3. HotStarTaq DNA Polymerase (Qiagen, Cat. No. 203203), store at –20°C.

2.4. Gel Electrophoresis

1. Agarose, LE, Analytical Grade (Promega, Cat. No. V3125) or equivalent.
2. 1× TBE buffer system (pH 8.3).

2.5. Template Preparation

1. 0.2-mL Thermo-Strip (0.2-mL thin-walled eight tube strips, Thermo Scientific, Cat. No. AB-0264).
2. Ultra Clear Cap Strips (Optically clear flat eight cap strips, Thermo Scientific, Cat. No. AB-0866).
3. 96-well (8 × 12 format) PCR rack with lid (e.g., SSI, Cat. No. 5230–29).
4. Streptavidin Sepharose High Performance; referred to in the text as streptavidin-coated Sepharose beads (5 mL, GE Healthcare Bio-Sciences, Cat. No. 17-5113-01), store at 4°C.
5. Binding buffer (200 mL, 2× concentrate, Qiagen, Cat. No. 979006), store at 4°C (see Note 6).
6. Microplate shaker with 1,400 rpm (e.g., MS3 digital with microtiter attachment MS 3.4, IKA, Cat. No. IKA3319000).
7. PyroMark Q24 Vacuum Prep Workstation (Qiagen).
8. High purity water (Milli-Q 18.2 MΩ × cm) or equivalent.
9. Ethanol absolute (GR for analysis ACS; e.g., Merck, Cat. No. 410230) is used to prepare 70% (v/v) ethanol. Where analytical grade ethanol is not available, laboratory grade ethanol may suffice for the washing step.
10. Denaturing solution (500 mL, Qiagen, Cat. No. 979007), store at 4°C (see Note 7).
11. Washing buffer (200 mL, 10× concentrate, Qiagen, Cat. No. 979008), store at 4°C (see Note 6).
12. Annealing buffer (250 mL, Qiagen, Cat. No. 979009), store at 4°C (see Note 6).
13. Bisulphite pyrosequencing primer(s) (desalted, PCR quality).
14. PyroMark Q24 Plate (Qiagen, Cat. No. 979201).
15. Heating block or thermoblock suitable for the PyroMark Q24 Plate Holder.

2.6. Bisulphite Pyrosequencing

1. PyroMark Q24 instrument (Qiagen).
2. PyroMark Q24 Cartridge (Qiagen, Cat. No. 979202).
3. PyroMark Gold Q24 Reagents (5 × 24) (Qiagen, Cat. No. 970802). Contains enzyme mix (Klenow fragment, ATP sulfurylase, luciferase, apyrase, SSB = single-stranded binding protein), substrate mix (APS = adenosine 5′-phosphosulfate, D-luciferin), and nucleotides (dATP-α-S, dCTP, dGTP, and dTTP), store at 4°C.

2.7. Serial Bisulphite Pyrosequencing (Optional)

1. Given in Subheading 2.5 and 2.6.

2.8. Data Analysis

1. PyroMark Q24 software (Qiagen), (see Note 8).

3. Methods

3.1. Primer Design

3.1.1. Design of Bisulphite-Specific PCR Primers

PCR primer design is one of the most important issues for investigating DNA methylation using bisulphite-treated DNA. Software tools such as MethPrimer (33), the BiSearch Primer Design and Search Tool (34), Methyl Primer Express (Life Technologies) and the PyroMark Assay Design Software (Qiagen) are available to assist with bisulphite-specific PCR primer design. Some general rules and advice are given below:

1. PCR primers should not be placed in repetitive regions. Repetitive regions can be identified using the RepeatMasker software (35).
2. PCR primers should not contain extensive homopolymeric stretches and should avoid polymorphic sites, especially those, which are retained following bisulphite treatment as this may result in a biased PCR amplification. Polymorphic sites can be identified by screening the corresponding genomic sequence in the Ensembl database or by using a BLAST search of the SNP database (dbSNP) at the National Center for Biotechnology Information (NCBI) website.
3. PCR primers should not form hairpin structures and homo or hetero primer-dimers. Online services such as OligoAnalyzer can be used to screen for hairpin structures and primer–dimers (see Note 9).
4. PCR primers should be 20–35 bp in length and their annealing temperature should preferentially be in the range of 55–65°C. The melting temperature difference between a primer pair should be less than 2°C.
5. It is recommended that each PCR primer should contain one or more bisulphite conversion-specific thymines corresponding

to non-CpG cytosines at or close to its 3′-end. The inclusion of bisulphite conversion-specific thymines helps to discriminate bisulphite-treated DNA from genomic DNA and incompletely bisulphite converted DNA during PCR amplification.

6. PCR bias during amplification should be determined. This is most readily done using methylation-sensitive high resolution melting (MS-HRM) of DNA methylation standards using a mixture of methylated and unmethylated templates (see Subheading 3.2 and Chapter 5). A CpG position at or close to the 5′-end in one or both PCR primers (36) can be used to compensate for a PCR amplification bias (37).
7. The length of the PCR product should not exceed 350 bp to avoid secondary structure formation, which may have a negative impact on the pyrosequencing reaction. PCR products of 250–300 bp work well for high-quality genomic DNA, such as extracted from cell lines or fresh-frozen tissues. For PCR amplification of material that is likely to be highly degraded, such as genomic DNA extracted from FFPE tissue, the PCR product should be 100 bp or less.
8. Either the forward or reverse PCR primer contains a 5′-end biotin label.

3.1.2. Design of Bisulphite Pyrosequencing Primers

Pyrosequencing primers can be designed using the PyroMark Assay Design Software (Qiagen) and/or the MethMarker software (38). Steps 1–3 from Subheading 3.1.1 should be taken into account when designing pyrosequencing primers. However, additional specific rules and advice are given below:

1. Pyrosequencing primers need to be complementary to the biotin-labelled strand of the PCR product.
2. Pyrosequencing primers should be 16–24 bp in length.
3. Pyrosequencing primers should not contain any CpG sites as this may result in an overestimation of DNA methylation present. If this cannot be avoided make sure that the CpG position is at or close to the 5′-end of the pyrosequencing primer.
4. Pyrosequencing primers should be unique to one position within the PCR product.
5. Pyrosequencing primers should preferably have a bisulphite conversion-specific thymine corresponding to a non-CpG cytosine at their 3′-end (Concept of “double discrimination”) (see Note 10).

3.1.3. Considerations for the Region Subjected to Bisulphite Pyrosequencing

1. Depending on the biotinylated strand the DNA methylation status of a CpG position is displayed as the ratio of incorporated dCTP (methylated CpG site)/dTTP (unmethylated CpG site) and dGTP (methylated CpG site)/dATP-α-S (unmethylated CpG site) nucleotides, respectively.

2. Accurate pyrosequencing analysis is possible for up to 55 bisulphite pyrosequencing cycles of nucleotide dispensation. However, the actual read length depends on the sequence composition, for example the existence of homopolymers (single nucleotide repeats) enables a read length of more than 55 bp. The nucleotide dispensation order is calculated by the PyroMark Q24 software from the bisulphite-modified genomic input sequence provided by the user.
3. SNP positions which are retained following bisulphite treatment (i.e., non-C/T SNPs) need to be included as variable nucleotide positions in the bisulphite-modified genomic input sequence. The occurrence of SNPs may otherwise lead to failed reference peak patterns during data analysis (see Subheading 3.8).
4. At least one isolated bisulphite conversion-specific thymine corresponding to a non-CpG cytosine should be included in the nucleotide dispensation order to estimate the residual contamination of incompletely bisulphite converted sequences during PCR amplification (see Note 11).
5. If possible, avoid long homopolymeric stretches as well as multiple short repetitive sequences within the amplified region. Repetitive sequences can be identified using the RepeatMasker software (35).

3.2. Bisulphite Treatment and Preparation of DNA Methylation Standard Series

3.2.1. Sources of Unmethylated Template

1. Currently, there is no ideal source of universally unmethylated DNA, which can be used for all types of DNA methylation analysis and an appropriate source should be chosen carefully according to the individual experimental problem being addressed.

   DNA from peripheral blood (either total nucleated cells or mononuclear cells) from normal individuals is generally a good source of unmethylated DNA. However, this DNA is not suitable as an unmethylated standard for the investigation of imprinted genes. In addition, some individuals may show a certain amount of DNA methylation due to age, individual variation, or other circumstances. Cell line DNA with a verified DNA methylation status at the region amplified during bisulphite-specific PCR can be used as an unmethylated DNA standard, but care must be taken because the DNA methylation status may vary depending on the source. Whole-genome amplified (WGA) products can also be used as a source of unmethylated DNA. Whole-genome amplification is performed in two consecutive amplification rounds according to Kristensen et al. (39). However, the WGA product consists of fragments of different size and contains less amplifiable material of a certain concentration compared to high-quality DNA. It is therefore recommended to use at least a threefold higher

concentration of WGA product than high-quality DNA for bisulphite treatment to ensure a similar magnitude of amplifiable fragments.

One nanogram of genomic DNA extracted from peripheral blood in 1.0 μL is subjected to the first round of WGA amplification (2 h at 30°C) utilising the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare) according to the manufacturer's directions. For the second round of WGA amplification, 0.1 μL of the unpurified WGA product of the first round is diluted in PCR grade water to a final volume of 1.0 μL and processed as described above. The WGA product of the second round is purified using the QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions, eluted in 50 μL EB buffer and quantified with a NanoDrop ND-1000 (Thermo Scientific) spectrophotometer.

3.2.2. Sources of Fully Methylated Template

1. Cell line DNA with a verified DNA methylation status at the region under investigation can be used as a fully methylated DNA standard, but care must be taken because the DNA methylation status may vary depending on the source.

   Fully methylated DNA is commercially available from suppliers such as Millipore or Qiagen. Alternatively, CpG Methyltransferase (M.SssI), which methylates genomic DNA in vitro can be used to generate an almost fully methylated standard following several rounds of M.SssI treatment. M.SssI treatment is performed in a 200-μL PCR tube containing 500 ng of genomic DNA (100 ng/μL) in PCR grade water in a volume of 17 μL. Two U (0.5 μL) M.SssI and 1 μL fresh SAM (20×) in 1× NEBuffer 2 are incubated for 3 h at 37°C. The tube is then placed on ice and 2 U (0.5 μL) M.SssI and 1 μL SAM (20×) are added and the reaction is incubated for 2 h at 37°C. The tube is placed on ice and 2 U (0.5 μL) M.SssI and 1 μL SAM (20×) are added, reaching a final volume of 20 μL and incubated for 15 h at 37°C followed by a final incubation at 65°C for 20 min (see Note 12).

3.2.3. Bisulphite Treatment

1. 100–500 ng of genomic DNA is subjected to bisulphite conversion using commercially available kits such as the EpiTect Bisulfite Kit (Qiagen), MethylEasy *Xceed* (Human Genetic Signatures), or the EZ DNA Methylation-Gold Kit (Zymo Research) following the manufacturer's recommendations and instructions.
2. Home-made protocols can be used for bisulphite treatment (e.g., http://www.epigenome-noe.net/WWW/researchtools/protocol.php?protid=35). However, unless the experimenter is very experienced with these protocols, the use of commercial available bisulphite conversion kits is advised.

3.2.4. Normalisation of the Amplifiable Amounts of Starting Material

1. To prepare a DNA methylation standard series, it is necessary to make sure that the concentrations of amplifiable unmethylated and fully methylated DNA are equal. Estimate the amplifiable DNA concentration of each using the C_q values obtained from the *COL2A1* control assay (40). The primers used for amplification of bisulphite-treated human DNA are COL2A1-F1: 5′-GTAATGTTAGGAGTATTTTGTGGGTA-3′ and COL2A1-R1: 5′-CTACCCCAAAAAAACCCAATCCTA-3′ (39). PCR cycling is performed on the Rotor-Gene Q (Qiagen, formerly the Rotor-Gene 6000 from Corbett Research) but can also be performed on other suitable real-time PCR thermocyclers. PCR is performed in a 100-µL PCR tube (Qiagen) with a final volume of 20 µL, containing 200 nM of each primer, 200 µM of each dNTP, 0.5 U of HotStarTaq DNA Polymerase (Qiagen) in 1× of the supplied PCR buffer containing 2.5 mM $MgCl_2$, 5 µM SYTO 9 (Life Technologies), and 1 µL of bisulphite-treated DNA (10 ng, theoretical amount) as template. The initial denaturation (95°C, 15 min) is followed by 45 cycles of 10 s at 95°C, 20 s at 64°C, and 20 s at 72°C.

   The take-off value (see Note 13) of each sample is given by comparative quantification (a feature of the Rotor-Gene Q software) and used as the C_q value to calculate the DNA concentration of the sample. The least concentrated sample is used as a calibrator, whereas the sample with the highest DNA concentration is diluted with an appropriate volume of buffer (elution buffer for bisulphite treated DNA) to match the concentration of the calibrator. Following dilution, the *COL2A1* control assay is performed again to confirm that the correct adjustment has been made.

3.2.5. Preparation of DNA Methylation Standard Series

1. Prepare a DNA methylation standard series by diluting fully methylated DNA in unmethylated DNA following concentration normalisation (as described in Subheading 3.2.4). For most bisulphite pyrosequencing assays, a dilution series of 100% (methylated), 50, 25, 10, and 0% (equals 100% unmethylated) standards is sufficient.

3.3. Bisulphite-Specific PCR

1. The PCR conditions are strongly dependent on the DNA polymerase used and the length of the PCR product in addition to the base composition of the region amplified. The following PCR reaction conditions include a range of reagent concentrations with a suggested starting point in parentheses:

   Perform PCR reactions in a final volume of 20–30 µL (20 µL), containing 100–400 nM of each primer (200 nM), 200 µM of each dNTP, 0.5–2.5 U of DNA polymerase (0.5 U) in the supplied PCR buffer containing 1.5–2.5 mM $MgCl_2$ (2.5 mM) and 1–3 µL of bisulphite-treated DNA (10–25 ng,

theoretical amount) (1 μL; 10 ng, theoretical amount) as template (see Note 14). The initial denaturation/activation step (95–97°C, 5–15 min depending on the DNA polymerase used) is followed by 40–50 cycles (45 cycles) of 20–60 s (30 s) at 95°C, 30–60 s (30 s) at 55–65°C, 30–60 s (30 s) at 72°C, and a final extension step at 72°C for 10 min.

3.4. Gel Electrophoresis

1. Evaluate the quality of the PCR products by agarose gel electrophoresis. Run the samples on a 2.0–2.5% (w/v) agarose gel in a 1× TBE buffer system and stain with ethidium bromide. Load the wells with 2–5 μL of the PCR products and mix with 1.5 μL 5× loading dye. Run an appropriate DNA Molecular Weight Marker alongside the PCR products to determine their size (see Note 15).

3.5. Template Preparation

1. Bring all reagents to room temperature before use.
2. Mix 40 μL binding buffer (2×) and 2 μL Streptavidin Sepharose (well shaken; do not vortex as this may damage the streptavidin-coated Sepharose beads) per bisulphite pyrosequencing reaction. The Sepharose beads will sink quickly to the bottom of the bottle and should be transferred within 2 min of shaking.
3. Pipette 10–20 μL of each PCR product into a 200-μL tube. Add PCR grade water to 38 μL. Add 42 μL of Streptavidin Sepharose and binding buffer mixture (shaken gently from time to time to disperse the Sepharose beads) to a final volume of 80 μL (see Notes 16 and 17).
4. Seal tubes and shake them at room temperature for at least 5 min on a microplate shaker at 1,400 rpm.
5. Switch the vacuum on. Wash the Vacuum Prep Tool filter probes for 20 s in high-purity water. Switch the vacuum off.
6. Stop shaking the tubes and remove the lids carefully to avoid dispersal of the PCR product and to avoid cross contamination. Place the tubes in the appropriate frame on the Vacuum Prep Workstation (*Important note*: The Sepharose beads will sink quickly to the bottom of the tube and should therefore be captured by the Vacuum Prep Tool filter probes within 2 min of discontinuing shaking).
7. Switch the vacuum on and sink the Vacuum Prep Tool filter probes carefully and slowly into the tubes until the filter probes nearly reach the bottom of the tubes. Make sure that the liquid of every tube is completely removed. Remove the Vacuum Prep Tool filter probes carefully from the tubes and avoid touching any surfaces.
8. Place the Vacuum Prep Tool filter probes for 5–7 s successively into the chambers containing 50 mL of 70% (v/v) ethanol (see Note 18), 40 mL denaturation solution, and 50 mL

washing buffer (1×, prepared from 10× stock solution) (see Notes 19 and 20).

9. Invert the Vacuum Prep Tool to 90° (vertically) for 10–15 s to remove the liquid from the filter probes.
10. Place the Vacuum Prep Tool horizontally over the PyroMark Q24 plate and switch the vacuum off. Place the Vacuum Prep Tool filter probes carefully into the appropriate wells loaded with 0.3 μM [0.25 μL pyrosequencing primer working solution (30 μM) in PCR grade water + 24.75 μL annealing buffer] pyrosequencing primer in a final volume of 25 μL annealing buffer (*Important note*: Make sure that the vacuum is switched off!). Let the Vacuum Prep Tool filter probes rest for 2 min in each well and shake gently from time to time to release the Sepharose beads. A proper release of the Sepharose beads is recognised by the solution becoming opaque. Remove the Vacuum Prep Tool carefully from the PyroMark Q24 plate and place it in high-purity water for 30 s and shake gently. Place the Vacuum Prep Tool in another trough containing high-purity water, switch the vacuum on, and wash the Vacuum Prep Tool filter probes for 1 min. Switch the vacuum off and place the Vacuum Prep Tool in the resting spot.

3.6. Bisulphite Pyrosequencing

1. Place the PyroMark Q24 plate on the provided plate holder on a prewarmed heating block at 80°C for 2 min. This allows annealing of the bisulphite pyrosequencing primer and disruption of any secondary structures that the PCR product may have formed. Avoid a longer exposure period to prevent excessive evaporation from the wells. After 2 min incubation, cool the PyroMark Q24 plate to room temperature.
2. Fill the appropriate slots of the reagent cartridge with the precalculated amounts (by the PyroMark Q24 software) of the enzyme mixture (E), the substrate mixture (S), and the four nucleotides dATP-α-S (A), dCTP (C), dGTP (G), and dTTP (T) (*Important note*: Pipette the liquids carefully into the cartridge to avoid the formation of air bubbles. Tap the filled cartridge on the bench to bring the liquids to the bottom of each chamber). E and S are delivered freeze dried and need to be dissolved in 620 μL of PCR grade water before use (see Note 21). The nucleotides are kept at 4°C and are brought to room temperature before filling the reagent cartridge. Put the reagent cartridge into the PyroMark Q24 instrument and close the cartridge support bar. The label of the reagent cartridge should face the experimenter.
3. Place the PyroMark Q24 plate into the PyroMark Q24 instrument and close the frame holding the plate with care.
4. Start the bisulphite pyrosequencing run (see Note 22).

5. Once the run has finished, remove the PyroMark Q24 plate from the PyroMark Q24 instrument. Discard the bisulphite pyrosequencing reaction unless you wish to use it for serial bisulphite pyrosequencing (see Subheading 3.7).
6. The reagent cartridge should be cleaned immediately after use by washing the chambers at least three times with high-purity water (see Note 23).

3.7. Serial Bisulphite Pyrosequencing (Optional)

1. Take the PyroMark Q24 plate out of the instrument. Remove the bisulphite pyrosequencing reaction mixture from each well and place it into a new 200-µL tube. Pipette the liquid up and down three times to disperse the Sepharose beads before adding them to the 200-µL tubes. Add 15 µL PCR grade water to each tube (see Note 24). Add 20 µL binding buffer (2×) to each occupied well of the PyroMark Q24 plate and pipette the liquid up and down three times before transferring it to the tube containing the pyrosequencing reaction mixture. Repeat the last step to make a final volume of 80 µL (see Note 25).
2. Continue with Subheading 3.5 from step 4.

3.8. Data Analysis

The PyroMark Q24 software creates a theoretical histogram based on the nucleotide input sequence entered by the user. A histogram for a bisulphite pyrosequencing *MGMT* assay is shown in Fig. 2a. The *x*-axis represents the nucleotide dispensation order (each nucleotide is numbered but only labelled every 5 bp), whilst the *y*-axis represents the number of nucleotides expected to be incorporated. The non-variable positions in the sequence are represented by filled bars (dark blue on the machine's display), whereas the variable nucleotide positions are represented by clear bars containing a vertical double arrow. The arrows indicate the potential theoretical number of incorporated nucleotides.

Using the theoretical histogram as an internal control, the PyroMark Q24 software generates a pyrogram from the collected data points (Fig. 2b). The bisulphite converted nucleotide input sequence is shown above the pyrogram. The variable positions at the CpG sites are given in the appropriate International Union of Pure and Applied Chemistry (IUPAC) code (Y stands for the bases C or T). Similar to the histogram, the *x*-axis represents the nucleotide dispensation order, whereas the *y*-axis on the pyrogram represents the relative peak intensity in arbitrary units. Incorporated nucleotides are shown as peaks.

The variable positions highlighted by the darker shading (light blue on the machine's display) indicate the peaks which are used to determine the methylation ratio at a given CpG position (e.g., nucleotide positions 2 and 3). Above these boxes, the calculated percentage of DNA methylation at each CpG position is given. On the machine's display, these boxes are colour coded in blue, yellow,

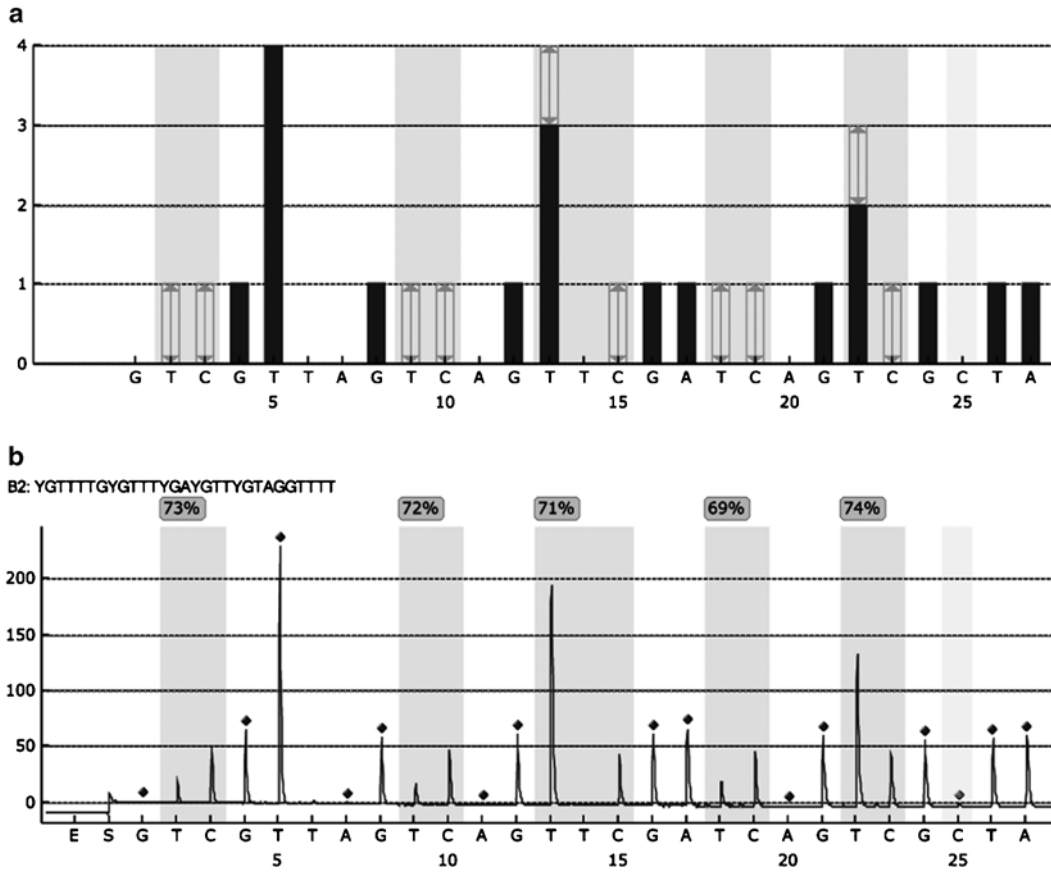


Fig. 2. Results of a bisulphite pyrosequencing assay to determine the DNA methylation status of the human *MGMT* promoter. (**a**) The theoretical histogram calculated by the PyroMark Q24 software. It shows the expected peak pattern for the *MGMT* assay based on the provided nucleotide input sequence. (**b**) The pyrogram obtained from a bisulphite pyrosequencing *MGMT* assay. For a more extensive description see Subheading 3.8.

and red and serve as quality controls for the reliability of the calculated DNA methylation values according to the quality control settings in the software (see Note 26). A blue code indicates a pass, a yellow code indicates that a manual check is required, and a red code indicates a failure.

The diamond symbols identify peaks used as references. Reference peaks are either non-variable nucleotide positions or blank nucleotide dispensations and serve as references for peak height as well as for internal quality control. Blank nucleotide dispensations (e.g., nucleotide position 7) are expected to show no peak. Nucleotide 25 is highlighted by light shading (light yellow on the machine's display). Such shading indicates a control peak for the bisulphite conversion-specific thymine corresponding to a non-CpG cytosine that is used to estimate the residual contamination of

incompletely bisulphite converted sequences. These non-CpG cytosines are not expected to incorporate nucleotides in the theoretical histogram, indicating that the bisulphite conversion rate is 100%. Nucleotide positions 5 and 6 in the pyrogram show the addition of dTTP in consecutive pyrosequencing cycles. The second cycle allows the incorporation of additional thymine nucleotides to complete the extension of the homopolymeric thymine stretch, thus minimising out of phase sequencing.

It should be noted that the peak heights of the A nucleotides are slightly higher than the peak heights derived from the other three nucleotides. This increase is due to the differences in the reaction kinetics of the Klenow fragment and apyrase when they are processing the alternative dATP-α-S substrate (see Note 27). The increase in A-peak height is corrected by the A-peak reduction factor utilised by the PyroMark Q24 software.

For most assays, reliable results are observed for up to 55 bisulphite pyrosequencing cycles. The limited read length is mainly caused by out of phase DNA extensions. The recent introduction of added SSB protein to the enzyme mix has enabled longer read lengths by preventing secondary structures and reducing the effects of plus and minus frame shifts (23). A plus frame shift is caused by insufficient apyrase activity and results in incomplete degradation of nucleotides. This type of frame shift is observed in later cycles of pyrosequencing and recognised by the occurrence of out of phase DNA extensions, which are at least one nucleotide ahead. A minus frame shift is caused by insufficient Klenow activity. This type of frame shift is caused by the incomplete extension of homopolymeric stretches (three to four nucleotides) and recognised by the occurrence of out of phase DNA extensions, which are at least one nucleotide behind.

The DNA methylation detection limit varies according to the particular bisulphite pyrosequencing assay. DNA methylation levels are usually reliably detected down to 5% for each CpG position, which determines the lower and upper quantification limits of 5 and 95%, respectively. The deviation between replicates is usually small and is generally less than 5%.

A poorly performing bisulphite pyrosequencing assay (e.g., several of the CpG positions do not pass the quality control or a very short read length is captured) can be improved through the following optimisation steps. A good starting point is to vary the concentration of the bisulphite pyrosequencing primer and/or to adjust the amount of PCR product added to the bisulphite pyrosequencing reaction. In some circumstances, it may be necessary to redesign the bisulphite pyrosequencing primer, the PCR amplification primers or even designing a new assay based on the other strand.

4. Notes

1. Luciferin is capable of using dATP as an alternative substrate to ATP. Therefore, dATP-α-S is used instead of dATP, which serves as a substrate for the Klenow fragment but not for luciferin.
2. During this two-step process, a luciferyl adenylate intermediate is first formed, which is then oxidised to oxyluciferin in an electronically excited state. The loss of energy during the relaxation of excited oxyluciferin to its ground state results in the emission of light.
3. The enzyme cascade described above is based on a four-enzyme system. A three-enzyme system is used in 454 sequencing which does not use the apyrase enzyme in the final step. Here, the apyrase is substituted by a washing step, which cleans up the pyrosequencing reaction before the next cycle.
4. If the free biotin is not removed properly, it will compete with the biotinylated PCR product for streptavidin binding and may result in decreased peak intensity during the pyrosequencing reaction.
5. Prepare a 50 or 100 μM stock solution of the biotinylated primer in PCR grade water and store it at –20°C. In our experience, biotinylated primers can be stored at this temperature for at least 1 year without any loss in performance. In addition, limit the freeze–thaw cycles to a minimum. Prepare a 5 or 10 μM PCR working solution and store the tubes in aluminium foil at 4°C. Under these conditions, the biotinylated primer can be stored for at least 6 weeks without any loss in performance.
6. A recipe for a home-made solution is given in ref. 31.
7. 0.2 M sodium hydroxide (NaOH) solution.
8. PyroMark software is continually improved and updated.
9. Primer–dimers may include the biotin label and therefore compete with the biotinylated PCR products for the streptavidin-coated Sepharose beads and subsequently block the positions on the Vacuum Prep Tool filter probes. Primer–dimers should also be avoided as they may cause increased background noise during the pyrosequencing step.
10. The bisulphite-specific PCR primers contain bisulphite conversion-specific thymines corresponding to non-CpG cytosines. These positions help to prevent PCR amplification from DNA templates with a poor bisulphite conversion rate. A pyrosequencing primer with a bisulphite conversion-specific thymine

at its 3′-end adds stringency to the DNA methylation analysis by minimising the contribution of PCR amplification products without a proper bisulphite conversion rate to the calculation of DNA methylation levels.

11. A poor bisulphite conversion rate results in the overestimation of DNA methylation.
12. The contents of the tube can be used without further purification for bisulphite treatment.
13. The take-off point is defined as the amplification cycle at which the second derivative is at 20% of the maximum level, which indicates the end of the background noise.
14. A new PCR set-up includes genomic DNA (2 ng/μL) as a template to ensure that the bisulphite-specific PCR primers specifically amplify bisulphite-treated templates only.
15. The agarose gel should show a single band of strong intensity PCR product at the expected size with no excessive formation of primer–dimers. The no template PCR control (NTC) should not show any evidence of contamination. If contamination is present the PCR products should be discarded and the experiment repeated.
16. Always include the no template PCR control (NTC). Peaks observed in the NTC pyrogram may represent primer–dimers and will help to determine whether they interfere with peaks from the targeted bisulphite sequence.
17. For every new bisulphite pyrosequencing assay, additional controls are recommended. Process the PCR product in the absence of the pyrosequencing primer to screen for secondary structures resulting in extension during the pyrosequencing reaction. In addition, perform the pyrosequencing reaction with solutions that contain the biotinylated PCR primer and the pyrosequencing primer separately and in combination to screen for the formation of primer–dimers also resulting in extension. Furthermore, pyrosequence at least three replicates of an appropriate control sample to determine the variability and reproducibility of the pyrosequencing assay.
18. Fill the chamber with 70% ethanol just before using the Vacuum Prep Workstation to avoid excessive evaporation.
19. 50 mL of each solution used can be prepared in 50-mL tubes and should be stored at room temperature.
20. Washing buffer can be used at least twice.
21. Swirl the bottles gently and let them rest at room temperature for 10 min to ensure that the mixture is fully dissolved (*Important note*: do not vortex!). Aliquot 5 × 124 μL of each

enzyme mixture (E) and substrate mixture (S) into 1.5-mL tubes. The aliquots can be freeze–thawed up to three times and should be stored at –20°C.

22. The pyrosequencing reaction is performed at 28°C due to the thermo-instable luciferase and the assay takes 1 min per nucleotide dispensation.
23. Fill the chambers to the top with high-purity water and press gently with the finger tips on each chamber. The pressure releases a thin stream of liquid from the nozzle of each chamber. The stream should be straight, otherwise replace the reagent cartridge.
24. Binding buffer can be used instead.
25. This approach is dependent on and limited by the amount of recovered biotinylated PCR product. The loss of PCR product will affect the peak intensity during the bisulphite pyrosequencing reaction and is therefore the most limiting parameter.
26. The default settings should not be altered except by experienced users.
27. The DNA polymerase activity for incorporation of dATP-α-S is reduced and to compensate for this a higher concentration of dATP-α-S is added to each of the appropriate pyrosequencing cycles. Additionally, the ability of the apyrase to breakdown the dATP-α-S is also reduced, which results in ATP being available for a longer time period compared to the other dNTPs. The combination of a higher dATP-α-S concentration accompanied by a reduced breakdown of ATP leads to an increase in ATP availability to the luciferase, resulting in an increase in the amount of light generated per time frame and therefore a slight increase in A-peak height.

Acknowledgements

The authors thank Ida Candiloro and Phillip Hudson (Qiagen) for critical reading of this manuscript and helpful discussions. We acknowledge funding from the National Breast Cancer Foundation (Australia), the Cancer Council of Victoria, the Victorian Cancer Agency, the CLL Global Research Foundation, and the Susan G. Komen for the Cure Foundation.

References

1. Ronaghi, M., Uhlen, M., and Nyren, P. (1998) A sequencing method based on real-time pyrophosphate *Science* **281**: 363–5.
2. Fakhrai-Rad, H., Pourmand, N., and Ronaghi, M. (2002) Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms *Hum. Mutat.* **19**: 479–85.
3. Ronaghi, M., Shokralla, S., and Gharizadeh, B. (2007) Pyrosequencing for discovery and analysis of DNA sequence variations *Pharmacogenomics* **8**: 1437–41.
4. Lee, J. H., and Jeon, J. T. (2008) Methods to detect and analyze copy number variations at the genome-wide and locus-specific levels *Cytogenet. Genome Res.* **123**: 333–42.
5. Uhlmann, K., Brinckmann, A., Toliat, M. R., Ritter, H., and Nurnberg, P. (2002) Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis *Electrophoresis* **23**: 4072–9.
6. Colella, S., Shen, L., Baggerly, K. A., Issa, J. P., and Krahe, R. (2003) Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites *Biotechniques* **35**: 146–50.
7. Tost, J., Dunker, J., and Gut, I. G. (2003) Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing *Biotechniques* **35**: 152–6.
8. Shaw, R. J., Liloglou, T., Rogers, S. N., et al. (2006) Promoter methylation of P16, RARbeta, E-cadherin, cyclin A1 and cytoglobin in oral cancer: quantitative evaluation using pyrosequencing *Br. J. Cancer* **94**: 561–8.
9. White, H. E., Durston, V. J., Harvey, J. F., and Cross, N. C. P. (2006) Quantitative analysis of SNRPN(correction of SRNPN) gene methylation by pyrosequencing as a diagnostic test for Prader-Willi syndrome and Angelman syndrome *Clin. Chem.* **52**: 1005–13.
10. Song, M. A., Park, J. H., Jeong, K. S., Park, D. S., Kang, M. S., and Lee, S. (2007) Quantification of CpG methylation at the 5′-region of XIST by pyrosequencing from human serum *Electrophoresis* **28**: 2379–84.
11. Mikeska, T., Bock, C., El-Maarri, O., et al. (2007) Optimization of Quantitative MGMT Promoter Methylation Analysis Using Pyrosequencing and Combined Bisulfite Restriction Analysis *J. Mol. Diagn.* **9**: 368–81.
12. Lof-Ohlin, Z. M., and Nilsson, T. K. (2009) Pyrosequencing assays to study promoter CpG site methylation of the O6-MGMT, hMLH1, p14ARF, p16INK4a, RASSF1A, and APC1A genes *Oncol. Rep.* **21**: 721–9.
13. Wong, H.-L., Byun, H.-M., Kwan, J. M., et al. (2006) Rapid and quantitative method of allele-specific DNA methylation analysis *Biotechniques* **41**: 734–9.
14. Yang, A. S., Estecio, M. R. H., Doshi, K., Kondo, Y., Tajara, E. H., and Issa, J. P. (2004) A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements *Nucl. Acids Res.* **32**: e38.
15. Richards, K. L., Zhang, B., Baggerly, K. A., et al. (2009) Genome-wide hypomethylation in head and neck cancer is more pronounced in HPV-negative tumors and is associated with genomic instability *PLoS One* **4**: e4941.
16. Xie, H., Wang, M., Bonaldo, M. de F., et al. (2009) High-throughput sequence-based epigenomic analysis of Alu repeats in human cerebellum *Nucleic Acids Res.* **37**: 4331–40.
17. Yamamoto, E., Toyota, M., Suzuki, H., et al. (2008) LINE-1 hypomethylation is associated with increased CpG island methylation in Helicobacter pylori-related enlarged-fold gastritis *Cancer Epidemiol. Biomarkers Prev.* **17**: 2555–64.
18. Aparicio, A., North, B., Barske, L., et al. (2009) LINE-1 methylation in plasma DNA as a biomarker of activity of DNA methylation inhibitors in patients with solid tumors *Epigenetics* **4**: 176–84.
19. Irahara, N., Nosho, K., Baba, Y., et al. (2010) Precision of pyrosequencing assay to measure LINE-1 methylation in colon cancer, normal colonic mucosa, and peripheral blood cells *J. Mol. Diagn.* **12**: 177–83.
20. Choi, S. H., Worswick, S., Byun, H.-M., et al. (2009) Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer *Int. J. Cancer* **125**: 723–9.
21. Mikeska, T., Candiloro, I. L., and Dobrovic, A. (2010) The implications of heterogeneous DNA methylation for the accurate quantification of methylation *Epigenomics* **2**: 561–73.
22. Ronaghi, M. (2001) Pyrosequencing sheds light on DNA sequencing *Genome Res.* **11**: 3–11.
23. Ahmadian, A., Ehn, M., and Hober, S. (2006) Pyrosequencing: history, biochemistry and future *Clin. Chim. Acta* **363**: 83–94.
24. Gharizadeh, B., Nordstrom, T., Ahmadian, A., Ronaghi, M., and Nyren, P. (2002) Long-read pyrosequencing using pure 2′-deoxyadenosine-5′-O′-(1-thiotriphosphate) Sp-isomer *Anal. Biochem.* **301**: 82–90.

25. Brakensiek, K., Wingen, L. U., Länger, F., Kreipe, H., and Lehmann, U. (2007) Quantitative high-resolution CpG island mapping with Pyrosequencing reveals disease-specific methylation patterns of the CDKN2B gene in myelodysplastic syndrome and myeloid leukemia *Clin. Chem.* **53**: 17–23.
26. Tost, J., El abdalaoui, H., and Gut, I. G. (2006) Serial pyrosequencing for quantitative DNA methylation analysis *Biotechniques* **40**: 721–2, 724, 726.
27. Pasquali, L., Bedeir, A., Ringquist, S., Styche, A., Bhargava, R., and Trucco, G. (2007) Quantification of CpG island methylation in progressive breast lesions from normal to invasive carcinoma *Cancer Lett.* **257**: 136–44.
28. Dunn, J., Baborie, A., Alam, F., et al. (2009) Extent of MGMT promoter methylation correlates with outcome in glioblastomas given temozolomide and radiotherapy *Br. J. Cancer* **101**: 124–31.
29. Flanagan, J. M., Cocciardi, S., Waddell, N., et al. (2010) DNA methylome of familial breast cancer identifies distinct profiles defined by mutation status *Am. J. Hum. Genet.* **86**: 420–33.
30. Lewin, J., Schmitt, A. O., Adorján, P., Hildmann, T., and Piepenbrock, C. (2004) Quantitative DNA methylation analysis based on four-dye trace data from direct sequencing of PCR amplificates *Bioinformatics* **20**: 3005–12.
31. Tost, J., and Gut, I. G. (2007) DNA methylation analysis by pyrosequencing *Nat. Protoc.* **2**: 2265–75.
32. Dejeux, E., El abdalaoui, H., Gut, I. G., and Tost, J. (2009) Identification and quantification of differentially methylated loci by the pyrosequencing technology *Methods Mol. Biol.* **507**: 189–205.
33. Li, L. C., and Dahiya, R. (2002) MethPrimer: designing primers for methylation PCRs *Bioinformatics* **18**: 1427–31.
34. Tusnády, G. E., Simon, I., Váradi, A., and Arányi, T. (2005) BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes *Nucl. Acids Res.* **33**: e9.
35. Smit, A. F. A., Hubley, R., and Green, P. (1996–2004) *RepeatMasker Open-3.0* <http://www.repeatmasker.org>.
36. Wojdacz, T. K., Hansen, L. L., and Dobrovic, A. (2008) A new approach to primer design for the control of PCR bias in methylation studies *BMC Res. Notes* **1**: 54.
37. Warnecke, P. M., Stirzaker, C., Melki, J. R., Millar, D. S., Paul, C. L., and Clark, S. J. (1997) Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA *Nucl. Acids Res.* **25**: 4422–6.
38. Schuffler, P., Mikeska, T., Waha, A., Lengauer, T., and Bock, C. (2009) MethMarker: User-friendly design and optimization of gene-specific DNA methylation assays *Genome Biol.* **10**: R105.
39. Kristensen, L. S., Mikeska, T., Krypuy, M., and Dobrovic, A. (2008) Sensitive Melting Analysis after Real Time- Methylation Specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection *Nucleic Acids Res.* **36**: e42.
40. Virmani, A. K., Tsou, J. A., Siegmund, K. D., et al. (2002) Hierarchical clustering of lung cancer cell lines using DNA methylation markers *Cancer Epidemiol. Biomarkers Prev.* **11**: 291–7.

Chapter 5

Closed-Tube PCR Methods for Locus-Specific DNA Methylation Analysis

Ida L.M. Candiloro, Thomas Mikeska, and Alexander Dobrovic

Abstract

Closed-tube PCR methods (sometimes referred to as in-tube PCR methods) for locus-specific DNA methylation analysis are methodologies in which the amplification and analysis of bisulphite-modified DNA take place in one tube without the need to remove the PCR products for further analysis. Closed-tube methodologies lend themselves to high-throughput applications and molecular diagnostics but are also applicable as a research tool. We review three closed-tube methodologies, methylation-sensitive high-resolution melting (MS-HRM), MethyLight, and sensitive melting after real-time analysis – methylation-specific PCR (SMART-MSP). Closed-tube detection can be performed by simultaneously amplifying both methylated and unmethylated templates and subsequent melting curve analysis (MS-HRM). Alternatively, methylation-specific primers are used in real-time quantitative PCR and monitored either by a fluorescent hydrolysis probe (MethyLight) or using a double-stranded DNA binding fluorescent dye with a subsequent quality control step by melting curve analysis (SMART-MSP).

Key words: High-resolution melting, Methylation-specific PCR, MSP, Cancer, Low-level methylation

1. Introduction

1.1. Background

As DNA methylation analysis begins to be used in molecular diagnostics, there is a growing need for reliable PCR methodologies that can rapidly assess the methylation status of a given region (1), particularly in cancer. Closed-tube methodologies are the most convenient of the many methodologies for DNA methylation analyses that have been developed as they require no removal of the PCR product from the tube to determine the methylation status. This not only allows rapid and relatively simple DNA methylation analysis and the development of high-throughput automated methods, but also importantly leads to a reduction in the potential for PCR contamination.

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_5,

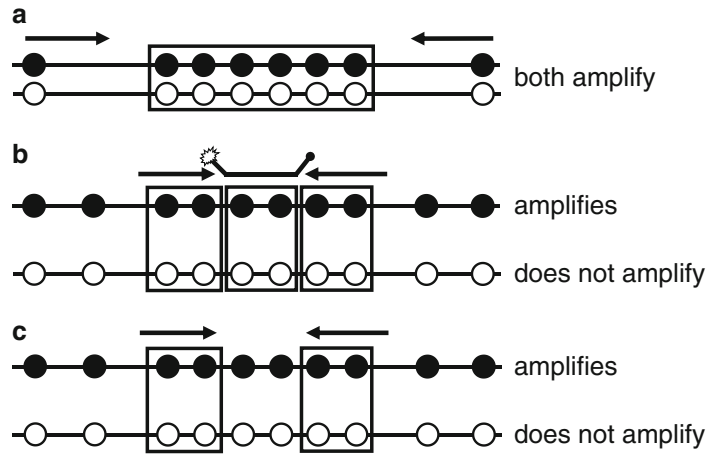


Fig. 1. Schematic representation of the three closed-tube methodologies used to examine DNA methylation is discussed in this chapter. Each *horizontal line* represents a methylated or unmethylated allele. *Open* and *filled circles* are used to indicate unmethylated and methylated CpG sites. The *horizontal arrows* represent the primers and the MethyLight probe has a fluorophore symbol at the 5′-end and a quencher (*solid circle*) at the 3′-end. The *boxes* indicate which CpG sites are analysed by the different methodologies. (**a**) Methylation-sensitive HRM uses methylation-independent primers to amplify bisulphite-modified DNA regardless of methylation status. (**b**, **c**) Methylation-specific PCR (MSP) and its derivative techniques MethyLight (**b**) and SMART-MSP (**c**) normally examine DNA methylation only at the primer (and, where applicable, probe) binding sites. Primers are thus designed to be complementary to methylated DNA.

Methylation information is lost with PCR amplification. The treatment of DNA with sodium bisulphite modifies unmethylated cytosine bases to form uracil (which is read as thymine by the DNA polymerases used in PCR) while leaving 5-methylcytosine intact allows methylation information to be retained (2). The sequence alterations then enable the use of PCR-based methods which take advantage of the fact that the cytosines that remain correspond to 5-methylcytosines in the original template.

This chapter deals with three closed-tube PCR methods (Fig. 1); methylation-sensitive high-resolution melting (MS-HRM) – a (semi)-quantitative methodology that displays both methylated and unmethylated sequences (3), MethyLight – a probe-based quantitative adaptation of methylation-specific PCR (MSP) (4), and sensitive melting after real-time – MSP (SMART-MSP) – a quantitative adaptation of MSP using new generation intercalating dyes which enable quality control analysis of the amplification by high-resolution melting (HRM) (5).

1.2. Methylation-Sensitive High-Resolution Melting

The incorporation of a double-stranded DNA-specific fluorescent dye into PCR not only allows real-time monitoring of the reaction but also enables melting analysis to be performed. Melting curve analysis was first introduced for the quality control of real-time PCR reactions (6). However, in MS-HRM, the melting curve

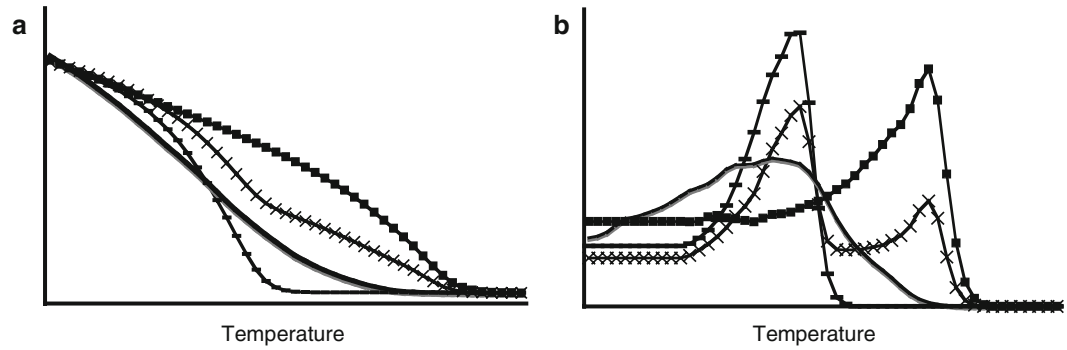


Fig. 2. High-resolution melting (HRM) figures from a hypothetical MS-HRM assay. Samples included are an unmethylated DNA standard (–), a mixture of methylated and unmethylated DNA (×), completely methylated standard DNA (*filled square*), and heterogeneously methylated DNA (*thick solid line*). (**a**) A normalised plot showing that the unmethylated DNA melts first, followed by the mixture of methylated and unmethylated DNA, with the fully methylated DNA melting last. The heterogeneously methylated sample begins melting before the unmethylated sample, and crosses over into the area where the occurrence of methylated samples is expected. (**b**) Negative first derivative of the melting data (T_m curves). This figure is obtained by taking the first derivative of the melt plot (d*F*/d*T*). This plot is usually then inverted (–d*F*/d*T*), so that there are peaks as opposed to troughs in the figure. This representation of the data is often used as the T_m of the products can be easily seen and calculated, as they are the points at which the peaks occur. The unmethylated peak is at a lower temperature than the methylated peak, and the mixture contains peaks at both positions. The heterogeneously methylated sample has a broad peak that encompasses the unmethylated peak, but extends further on both sides.

analysis is the primary analytical technique (Fig. 2), whereas the real-time amplification information is used for quality control to indicate the amount of amplifiable template.

The temperature at which an amplicon will "melt" (denature) is dependent on the base composition. As amplicons generated from methylated templates have cytosines in the positions corresponding to 5-methylcytosine, they have higher melting temperatures than amplicons from unmethylated templates, which contain thymines. When methylation-independent primers are used, this allows the amplification of both methylated and unmethylated sequences which are seen as distinct melting peaks.

The first closed-tube melting curve analysis of DNA methylation was based on SYBR Green fluorescence (7). However, SYBR Green based melting curve analysis failed to gain acceptance as a methodology for methylation analysis. The introduction of HRM using new generation dyes that unlike SYBR Green could be used at saturating conditions, and instruments and software that could generate and analyse the HRM data led to the development of MS-HRM which uses HRM analysis to determine the methylation status of the sample (3).

The melting temperature of an amplicon is dependent on its length and its base composition. Slowly raising the temperature after PCR along a gradient and monitoring fluorescence allows the discrimination of amplicons based on these properties. The separation of methylated and unmethylated peaks is due to the absence of heteroduplex formation because of the multiple sequence

differences between fully methylated and unmethylated amplicons (Fig. 2). Accordingly, MS-HRM has proven particularly valuable in the study of imprinting as it allows ready (semi)-quantitative distinction of the methylated and unmethylated alleles (8–10).

However, when DNA methylation patterns are heterogeneous as they often are in cancer, multiple heteroduplexes can be formed giving rise to a continuous, broad melting profile (Fig. 2). The development of a digital version of MS-HRM (using limiting dilution) enabled the analysis of heterogeneous DNA methylation by amplifying multiple individual alleles whose methylation could be determined by their melting temperature (11, 12).

As MS-HRM amplifies both methylated and unmethylated DNA, it optimally uses primers that are designed to enable compensation for PCR bias by choosing an annealing temperature at which methylated and unmethylated amplicons amplify at the same rate (13). If any CpGs are present in the primers, further increasing the annealing temperature will favour the amplification of methylated sequences which will assist in the detection of low-level DNA methylation in the range of 0.1–1% (3).

1.3. Probe-Based Quantification of Methylated Sequences

MSP was the first rapid DNA methylation analysis technique developed and is still widely used (14). The primers are designed to specifically amplify methylated bisulphite-treated DNA (see Note 1). Following amplification, the PCR products are run out on an agarose gel and methylation is scored as positive in those lanes in which a PCR product is detected. MSP is qualitative and does not allow the comparison of band intensities. However, DNA methylation levels as low as 0.1% may be detected (14). The sensitivity achieved is due to a variety of factors, e.g., appropriate primer design, stringent annealing temperature, and is dependent on a sufficient amount of template being used in the PCR reaction.

As only methylated sequences are amplified, real-time monitoring of amplification transforms MSP into a quantitative methodology. This was first done by the use of a dual-labelled hydrolysis probe which has a fluorophore on the 5′-end and a quencher on the 3′-end. This methodology was first called real-time quantitative MSP (15) but became better known as MethyLight (4).

In MethyLight and its related techniques (see Note 2), the hydrolysis probe binds in a sequence-specific manner to one or the other strand of the amplicon (4, 15). The probe has a higher melting temperature (T_m) than the primers and thus binds earlier to the single-stranded DNA as it is cooled down after the DNA denaturation step. Once primer extension occurs and the newly synthesised DNA strand meets the probe, the DNA polymerase hydrolyses the probe utilising the 5′→3′ exonuclease activity of the DNA polymerase thereby separating the fluorophore from the quencher, resulting in fluorescence. The fluorescence is therefore proportional to the amount of PCR product generated, making the methodology quantitative.

The probe also acts to minimise the false-positive rate, which is common for MSP (especially, when material from formalin-fixed paraffin-embedded (FFPE) specimens are used (16)) as the increased stringency introduced by the CpG sites within the probe means that most of the false-positive amplifications will not be detected. However, this also acts to further suppress the amplification of heterogeneously methylated templates.

1.4. Sensitive Melting After Real-Time Analysis: MSP

In an alternative quantitative adaptation of MSP, fluorescent dyes are used to monitor the amplification in real time (17). SMART-MSP (5) takes advantage of the extra possibilities that HRM analysis allows. As in MS-HRM, the melting profile is obtained by measuring the decrease in fluorescence resulting from denaturing DNA during an increasing temperature gradient (Fig. 3).

The specificity of SMART-MSP is achieved in the HRM analysis that follows PCR amplification. In late amplifying samples, the late amplification may be a consequence of either low amounts of DNA methylation or low levels of incomplete bisulphite conversion. The usual adaptation of SMART-MSP is designed to assist in the identification of such false positives due to incomplete bisulphite conversion. By loading the regions between the primers only

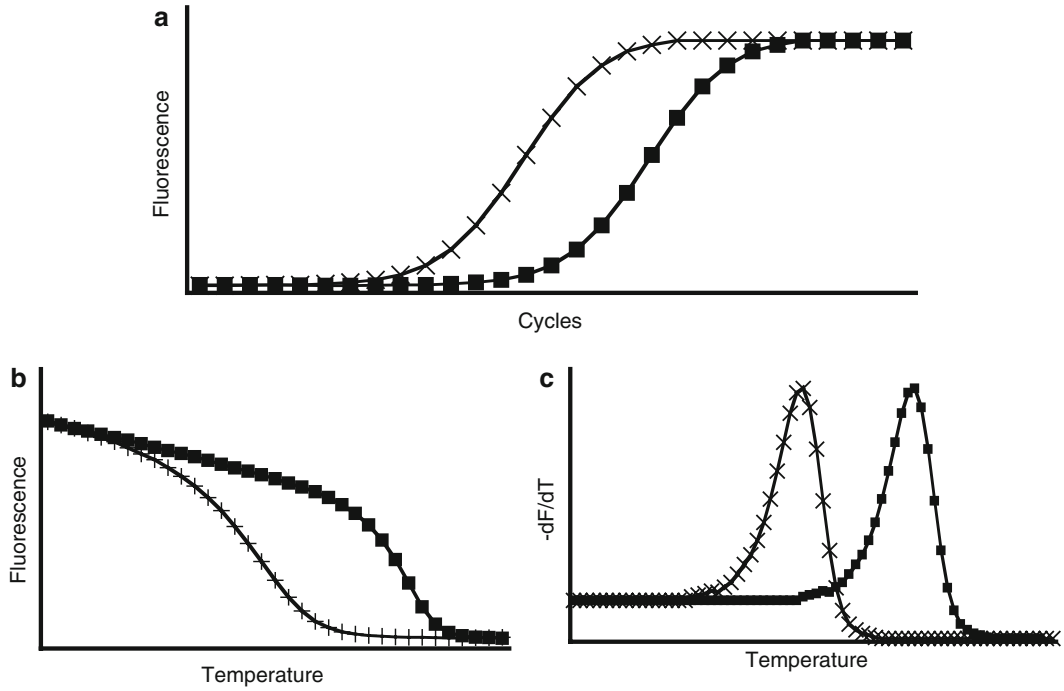


Fig. 3. An example of SMART-MSP data. (**a**) Amplification data for two samples. The sample that amplifies earlier (×) has a higher DNA methylation content than the sample that amplifies later (*filled square*). (**b**) Normalised melting plot. The second sample that melts (*filled square*) has a higher melting temperature than the positive control (×), and is therefore scored as a false positive. (**c**) Negative first derivative of the melting data. This clearly shows that one sample (*filled square*) is a false positive.

with non-CpG cytosines, which should be converted to uracils during bisulphite treatment, the ability to detect incomplete bisulphite conversion is enhanced. The specific melting temperature allows the identification of false-positive results as each incomplete conversion of the intervening non-CpG cytosines results in a detectable increase in T_m (Fig. 3).

SMART-MSP has also been used to analyse allele-specific DNA methylation (18). If the MSP primers frame a polymorphic site, the alleles can be distinguished as in conventional single nucleotide polymorphism genotyping (19). Each genotype will have its own melt profile, and any known heterozygotes that appear homozygous are indicative of allele-specific DNA methylation as only methylated templates are amplified.

1.5. Deciding Which Technique to Use

MS-HRM analysis is based on the comparison of different melting profiles observed for methylated and unmethylated DNA. When analysing homogeneously methylated DNA (mixtures of fully methylated templates in unmethylated templates), the sensitivity achievable can be down to 0.1% (3). Homogeneous methylation can be quantified using a dilution series of DNA methylation standards.

Examination of the melting curves obtained with MS-HRM allows one to readily distinguish homogeneously from heterogeneously methylated samples. However, when analysing heterogeneous methylation, while comparisons can be made between samples, the amount of methylation is difficult to estimate as the melting curve depends on heteroduplex formation between different methylated strands. If necessary, the amount of methylated alleles can be estimated by digital MS-HRM, where each HRM product derives from a single template following limiting dilution of the sample (11) (see Note 3).

Accurate quantification by MethyLight and SMART-MSP is dependent on methylation being homogeneous under the primer binding sites if stringent primer annealing conditions are used. For heterogeneous methylation, positives may only be obtained under conditions of reduced stringency and the value obtained will be an underestimate of the actual methylation and will be dependent on the precise amplification conditions. We discuss quantification of heterogeneous methylation in more detail elsewhere (20).

For samples with unknown DNA methylation levels and patterns, MS-HRM will serve as the most useful pre-screening method. For low-level homogeneous DNA methylation (less than 5%), MethyLight and SMART-MSP may be better at detecting and especially quantifying methylation than MS-HRM. In some instances of low-level heterogeneous methylation, these techniques may also detect methylation levels that MS-HRM is likely to miss if used at an appropriately lowered stringency. SMART-MSP is more cost effective than MethyLight as it does not require a probe.

2. Materials

2.1. Instrumentation

Instruments with combined real-time and HRM capabilities are recommended, such as the LightCycler 480 (Roche), and the RotorGene Q (Qiagen) formerly available as the RotorGene 6000 (Corbett).

2.2. Bisulphite Modification and Control DNA

1. EpiTect Bisulfite Kit (Qiagen, cat. no. 59104), MethylEasy *Xceed* (Human Genetic Signatures, cat. no. Me002), or the EZ DNA Methylation-Gold Kit (Zymo Research, cat. no. D5005).
2. CpGenome Universal Methylated DNA (Millipore, 10 μg, cat. no. S7821), store at –20°C.
3. Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Bio-Sciences, product code 25-6600-30), store at –80°C.
4. QIAquick PCR Purification Kit (Qiagen, cat. no. 28104).

2.3. PCR and High-Resolution Melting

1. PCR primer design software: Amplify 3× (http://engels.genetics.wisc.edu/amplify/) for Macintosh computers or AmplifX (http://ifrjr.nord.univ-mrs.fr/AmplifX-Home-page?lang=en) for PCs.
2. HotStarTaq DNA Polymerase (with 10× buffer and 2.5 mM/L $MgCl_2$) (Qiagen, cat. no. 203203), store at –20°C.
3. SYTO 9 green fluorescent nucleic acid stain (Life Technologies, cat. no. S-34854).
4. PCR primers (desalted, PCR quality).
5. Probes for MethyLight, dual labelled with 5′-FAM and TAMRA probes (Life Technologies), 5′-FAM and 3′-BHQ-1 (Black Hole Quencher), (Biosearch Technologies) or MGB probes (Life Technologies).
6. PCR reaction tubes: Strip Tubes and Caps, 0.1 mL (Qiagen, cat. no. 981103) or equivalent for RotorGene Q, or Multiwell plate-96, white (Roche, cat. no. 04729692001) or equivalent.

3. Methods

3.1. PCR Primer Design

Primer design is a critical step in each of these techniques. A thorough search of the published literature for the optimal regions within a given CpG island should be made to identify the critical methylated residues. The freely available MethMarker (http://methmarker.mpi-inf.mpg.de/) software package might assist in identifying such regions (21).

The DNA strands are no longer complementary after bisulphite modification. Primers are therefore designed only to one strand. The sense strand is generally chosen for convenience, but if it is difficult to find appropriate primers framing the region of interest, the antisense strand should be investigated (see Note 4).

3.1.1. Important Points Regarding Primer Design

- PCR primers should have the same (or as close as possible within 2°C) melting temperature (T_m). Because of the decreased complexity of bisulphite-modified DNA, adequate stringency is an important criterion and the melting temperatures should preferably be within 60°C to 65°C. We find that OligoCalc (http://www.basic.northwestern.edu/biotools/oligocalc.html) (22) is useful as a T_m predictor. Note that the optimal annealing temperature for the PCR reaction is generally about 5°C lower than the primer T_ms. The probe for MethyLight needs to have a melting temperature about 10°C above the melting temperatures calculated for the PCR primers to ensure that it hybridises before the primers bind. Primer-Express (Life Technologies) is useful for the calculation of primer and probe T_ms for MethyLight assays.
- PCR primers should ideally be longer than 20 bp. After bisulphite treatment, some genome complexity is lost; therefore, PCR primers need to be longer to ensure that their binding site is unique within the genome. If working with human, mouse, or rat DNA, their sequence can be checked against the bisulphitome using methBLAST (http://medgen.ugent.be/methBLAST/).
- The inclusion of several thymines resulting from bisulphite converted non-CpG cytosines is highly recommended to ensure only completely bisulphite converted DNA is amplified. One or more of the thymines should be included as close to the 3′-end of the PCR primer as possible, preferably within the last three bases. PCR amplification of incompletely bisulphite converted DNA results in false-positive results and this can be particularly serious with the MSP-based methodologies (23).
- Software tools such as Amplify for Macintosh computers and AmplifX for PCs are useful to check for multiple priming sites and the formation of primer dimers.

For HRM analysis:
Ideally, the analysed amplicon should contain a single melting domain, that is, the PCR amplicon melts entirely over a small temperature interval rather than multiple foci with different melting temperatures. The final output of the online tool POLAND (http://www.biophys.uni-duesseldorf.de/local/POLAND//poland.html) (24, 25) is the predicted negative first derivative of the nucleotide input sequence and is useful in predicting the simplicity or complexity of the melting pattern.

3.1.2. Control Assays for DNA Input

As in RT-qPCR, a control assay for an appropriate reference sequence is needed for the relative quantification of a real-time DNA methylation assay. Assays that control for DNA input post-bisulphite modification (a detailed protocol for a fragment from the human *COL2A1* genomic sequence devoid of CpG dinucleotides is given in Chapter 4) need to be placed in regions devoid of CpGs, as this enables the determination of the amount of template independent of the DNA methylation status. The aim of such assays is to have all templates amplify equally regardless of methylation status. Due to the lack of cytosines in these regions, the DNA code is effectively reduced to three bases, making the region likely to be AT-rich. The inclusion of as many guanosine bases as possible will help to counter this. It is of vital importance to include as many C to T conversion events resulting from the bisulphite treatment as possible in the primers (and probe), particularly at the 3′-ends to ensure that only completely bisulphite converted templates will be amplified (see Note 5).

3.1.3. MethyLight and SMART-MSP

To ensure the greatest specificity towards methylated templates, several CpGs need to be included in the primers (and probe). Usually, three or more CpGs are desirable in each primer. Locating at least one of these close to the 3′-end allows for greater specificity.

3.1.4. MS-HRM

Primers that amplify bisulphite-modified DNA regardless of methylation status are used. When no CpGs are included in these primers as is frequently recommended, there may be a PCR bias towards unmethylated sequences. We find it useful to include one to two CpGs towards the 5′-end of the primers as this introduces a temperature-dependent bias to methylated sequences (26). This also allows user to increase the sensitivity for low-level DNA methylated sequences when desired by increasing the temperature. Incorporation of CpGs also helps to design primers in CpG-rich regions. As these primers will still need to amplify both methylated and unmethylated DNA, more than one to two CpGs per primer should be avoided.

3.2. DNA

3.2.1. DNA Extraction

Most DNA extraction methods and commercially available DNA extraction kits deliver DNA that is suitable for bisulphite treatment (27). For those that do not contain a proteinase K digestion step, the addition of this step is necessary. Incomplete removal of proteins may increase the potential for false-positive results (28).

3.2.2. Bisulphite Modification

Treatment of single-stranded DNA with sodium bisulphite converts cytosine bases to uracils, leaving 5-methylcytosines intact. There are many commercial kits available (e.g., EZ DNA Methylation, Zymo Research; MethylEasy *Xceed*, Human Genetic Signatures; EpiTect Bisulfite Kit, Qiagen). Bisulphite modification kits usually

involve a bisulphite modification step followed by a clean-up protocol. Kits differ mainly in the conditions and length of time used for bisulphite modification.

3.2.3. Control DNA

Fully methylated DNA is commercially available for human and mouse (Millipore, Qiagen, New England Biolabs, Zymo Research). Completely unmethylated human DNA is also commercially available. The generation of unmethylated DNA can be achieved by performing whole genome amplification (WGA) to a suitable sample (preferably one that will already have a low-level of DNA methylation, and repeating the WGA process on the product generated (5)). The product from this second round of WGA should be cleaned up using the QIAquick PCR Purification kit (Qiagen) (Chapter 4). Ideally, the target region should only have low-level DNA methylation (if any), so that two rounds of WGA will sufficiently dilute out any methylation making it undetectable.

Before making a series of DNA methylation standards containing methylated DNA diluted in unmethylated DNA, quantification of the bisulphite-modified DNA is essential to ensure the dilutions are correct. Using a real-time PCR assay as described in Subheading 3.2.2 is recommended over spectrophotometric methods. An extra negative control containing unmodified genomic DNA is required when working with bisulphite-modified DNA to ensure that the primers only amplify successfully bisulphite-modified DNA. Amplification of incompletely bisulphite converted DNA will result in false positives in the cases for MethyLight and SMART-MSP and overestimation of methylation for MS-HRM.

3.3. PCR Setup

A hot start DNA polymerase is preferable to minimise the formation of primer-dimers and non-specific PCR products that may interfere with the fluorescence signals. A typical PCR setup and PCR cycling conditions are usually used, with the addition of a fluorescent double-stranded DNA binding dye added. A suggested PCR setup is shown in Table 1. Two types of dye are available: those that cannot be used at saturating concentrations as they inhibit the PCR (e.g., SYBR Green I (Life Technologies)), and those that can (e.g., SYTO 9 (Life Technologies), LC Green (Idaho) and EvaGREEN (Biotium)).

The number of cycles required during PCR of bisulphite-modified DNA is usually greater than that for genomic DNA, so the addition of extra PCR cycles when performing the first runs during assay development and optimisation is advised. An example of PCR cycling conditions that can be used are 15 min at 95°C, followed by 50 cycles of 10 s at 95°C, 20 s at 60°C, and 20 s at 72°C (see Note 7). The initial heat activation step is dependent on the DNA polymerase used, and the annealing temperature is dependent on the primers used and sensitivity/specificity required for a particular assay. The HRM step should consist of a temperature ramp from 65°C to 95°C at a rate of 0.2°C/s.

Table 1
Final concentrations of all reagents required for PCR and HRM (see Note 6)

Reagent	Final concentration
10× PCR buffer	1×
$MgCl_2$	2.5–4 mM
dNTPs	200 μM each
Forward primer	200 nM
Reverse primer	200 nM
SYTO-9/probe	5 μM/250 nM
Hot Start Taq	0.5 U
Bisulphite-modified DNA	20 ng
PCR grade water	to 20 μL

A HRM-capable, real-time thermocycler is recommended (e.g., LightCycler 480 or RotorGene Q) not only to minimise handling but also for the integration of the real-time and melting analyses. The real-time information is useful as a diagnostic tool to aid in troubleshooting difficult assays. In particular, late amplification is often an indicator to treat the results obtained with caution. If not using such an instrument, melting must be performed as soon as possible. Although re-melting of products at a later stage is possible, this usually results in a greater spread between replicates, making interpretations more difficult.

3.4. Assay Optimisation

3.4.1. MS-HRM

The sensitivity and dynamic range of the MS-HRM assay can be adjusted by altering the stringency of the reaction conditions. Alterations that increase primer specificity such as increasing the annealing temperature increase the sensitivity when there are cytosines/guanines in the primers at the CpG positions. Other PCR additives such as dimethyl sulphoxide (DMSO) (29), betaine (30), DMSO and betaine in combination (31), and formamide (32) will also alter the sensitivity. A trial-and-error empirical approach is usually required to find the exact conditions that best suit the needs of the experimenter at the time. Increasing sensitivity results in a shifted dynamic range, i.e., to detect 0.1% DNA methylation, there may be little discrimination between 25 and 100% DNA methylation.

3.4.2. MethyLight and SMART-MSP

A decision must first be made as to whether heterogeneous or homogeneous DNA methylation is being targeted. For the analysis of homogeneously methylated DNA, the PCR conditions need to be as stringent as possible to minimise the false-positive rate. A series of DNA methylation standards created by diluting fully

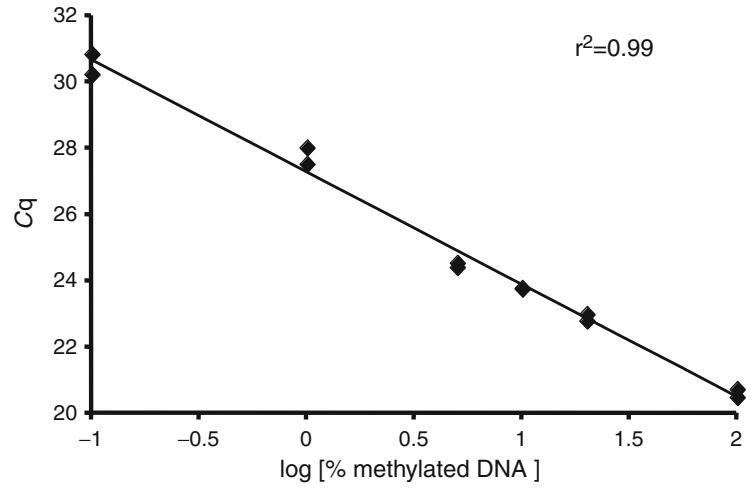


Fig. 4. A plot of log (methylated DNA in %) against C_q values obtained from a series of DNA methylation standards (in duplicate) from a SMART-MSP run. The r^2 value, a measure of fit to the linear plot, approximates 1, indicating that the graph obtained is close to linear across all dilutions.

methylated DNA into unmethylated DNA need to be used and tested during assay development. The completely unmethylated DNA control should never amplify. A plot of log (methylated DNA in %) against C_q values obtained should be linear as shown in Fig. 4 (see Note 8).

If heterogeneous methylation is being targeted, the conditions at which the PCR is performed should be relaxed to allow amplification of incompletely matching sequences due to the incomplete methylation expected, although this also increases the risk of false positives. For this reason, it may be preferable to perform the analysis over a range of annealing temperatures.

3.4.3. DNA Input

The sensitivity of any assay is ultimately limited by the amount of starting template in the PCR tube. To achieve a sensitivity of 0.1%, for example, more than 1,000 copies of the target must be present. Treatment of DNA with sodium bisulphite results in harsh conditions, so some DNA degradation is to be expected (33) The degradation may vary substantially depending on the methodology or kit used.

3.5. Data Analysis

3.5.1. Introduction

Most software packages that are used to operate HRM instruments have analysis modules that allow appropriate evaluation of the melting data obtained. Two different figure types are usually used. The first is a normalised plot of the melting data (normalised melting curves). This adjusts for the different starting and finishing fluorescence levels of the different samples during a HRM step.

Each sample is normalised to an equal starting and finishing point so that comparisons can be readily made (Figs. 2a and 3b). The second type is the negative first derivative (T_m curves) to give melting curves in which the T_m is at the apex of the peak (Figs. 2b and 3c). This is obtained by calculating the first derivative of the fluorescence changes as the samples melt allowing for easy discrimination of products with similar T_ms.

3.5.2. MS-HRM

The real-time data serves as a useful quality control and can be helpful in troubleshooting problematic samples. Replicates of a sample should amplify at similar C_q values with a similar PCR efficiency, and early enough to indicate that a sufficient amount of template was present (see Note 9). Negative controls (no template and unmodified DNA) should not amplify. Unusual melting behaviour can often be traced back to late amplification indicating that there are only a few template copies. In these cases, the experiment should be repeated.

The unmethylated DNA samples should be the first to melt, with the fully methylated control melting last. DNA methylation standards containing mixtures of the fully methylated standard and the unmethylated standard have populations of amplicons that will melt at both positions, resulting in intermediate profiles on a normalised plot and peaks at both positions on a negative first derivative plot (Fig. 2). Estimation of methylation is achieved by comparing against the series of DNA methylation standards used. When amplicons contain four or more CpGs between the primers, heteroduplexes are rarely formed between fully methylated and unmethylated templates.

Melting profiles resulting from heterogeneous methylation result in a melting curve that cannot be compared against DNA methylation standards due to the heteroduplexes formed by the differentially methylated PCR products. This results in melting beginning at a lower temperature (before the unmethylated control), reflecting the less stable nature of heteroduplexes, and finish melting at a higher melting temperature than the unmethylated control due to the presence of methylation (Fig. 2). Since there are no DNA methylation standards possible against which a measurement can be made, MS-HRM in this instance is qualitative.

3.5.3. Melting Analysis for SMART-MSP (and MethyLight)

Unlike MS-HRM, when there are no intervening CpGs in SMART-MSP, only a single PCR product is expected. Any sample that melts differently to the 100% methylated control is presumed to be a false-positive result and should be excluded (Fig. 3). Where there are CpGs between the primers, all samples that melt at higher temperatures than the fully methylated control can be excluded as false positives, however, those that melt at a lower temperature cannot be excluded as there may be incomplete (heterogeneous) methylation.

3.5.4. Analysis of Real-Time Data for SMART-MSP and MethyLight

There are multiple methods for analysis of real-time data. When relative quantification is used, methylation is measured against fully methylated DNA as a reference sample. Using a fully methylated sample as the control and multiplying the value obtained in this formula by 100 will express DNA methylation as a percentage (often referred to as the percentage methylated ratio (PMR) (34)).

The $2^{-\Delta\Delta Ct}$ (delta–delta C_t) method (35) makes the assumption that the PCR efficiencies of both the control fragment and target assays are 100% efficient. PCR efficiency is the fold increase in products per cycle, and can be calculated as $E = 10^{(-1/\text{slope})}$, where E is the efficiency and the slope is measured in the exponential phase.

The method described by Pfaffl (36) is based on similar principles, but takes the actual, measured efficiencies of both assays into account for each sample as shown in the formula below (1) where E is the efficiency of the target assay or the control as appropriate.

$$\text{Ratio} = \frac{(E_{\text{target}})^{\Delta C_{q\text{target}}(\text{methylated control - sample})}}{(E_{\text{control}})^{\Delta C_{q\text{control}}(\text{methylated control - sample})}} \quad (1)$$

The Relative Expression Software Tool (REST) package performs the calculations using this approach (http://www.gene-quantification.de/rest.html) (36).

4. Notes

1. MSP assays that specifically amplify only unmethylated DNA are often used in conjunction with assays that specifically amplify only methylated DNA when conducting MSP experiments. Positive reactions in this case indicate the presence of unmethylated DNA.
2. Strictly, MethyLight is any form of quantitative DNA methylation analysis that uses a hydrolysis probe (4) but its implementation has been almost entirely using MSP primers with a probe that is specific for methylation, i.e., contains one or more CpG sites within its sequence.
3. In many cases, the information obtained from HRM will be sufficient. However, further analysis of the MS-HRM product with other techniques is possible. In particular, digital MS-HRM may be used as an alternative to cloning for cost-effective sequencing analysis of methylation (11).
4. It may be very difficult to find regions where MS-HRM primers can be designed due to the density of CpG dinucleotides, even

when both strands are examined. In these instances, there may not be any option but to use SMART-MSP or MethyLight.

5. Amplicon length using these techniques is largely guided by the template being used. For the HRM methods, one melting domain is preferable. For SMART-MSP, as we normally wish to omit CpG dinucleotides between the primers, these assays are typically less than 100 bp. MS-HRM and MethyLight assays can range from around 70 bp when designed for FFPE-derived material up to 250 bp. The control assay should be around the same length as the methylation assay to control for any effects brought into the assay by amplicon length.
6. SYTO 9 is not required for MethyLight, but can be used in conjunction with a TaqMan-like probe utilising a fluorophore with distinct spectral emission if desired for high-resolution melting analysis (37).
7. It is believed that increasing the number of cycles in the PCR increases the likelihood that non-specific products will be produced during PCR amplification. We have found in our experience that this does not usually occur in well-designed assays. We recommend using 50 cycles in the first trial of an assay. Note that it is not necessary to reach the plateau phase for melting.
8. Threshold cycle (C_t), crossing point (C_p), and take-off point (TOP) are all used in the literature, but it is recommended that quantification cycle (C_q) be used as a standard (38, 39).
9. Variation in C_q values implies stochastic effects due to limiting amounts of the appropriate template but may also be due to random mistakes in the PCR setup.

Acknowledgements

AD has received grant support from the Cancer Council of Victoria, the National Breast Cancer Foundation of Australia, the Victorian Cancer Agency, the National Health and Medical Research Council of Australia, the Susan G. Komen for the Cure Foundation, and the US Department of Defence Breast Cancer Research Program under award numbers W81XWH-05-1-0500 and W81XWH-06-1-0670. Views and opinions of, and endorsements by the authors do not reflect those of the US Army or the Department of Defence.

References

1. Dobrovic, A. (2005) in *Molecular diagnostics for the clinical laboratorian* (Coleman, W. B., and Tsongalis, G. J., Eds.), pp 149–160, Humana Press, Totowa, NJ.
2. Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands, *Proceedings of the National Academy of Sciences of the United States of America* **89**; 1827–1831.
3. Wojdacz, T. K., and Dobrovic, A. (2007) Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation, *Nucleic acids research* **35**; e41.
4. Eads, C. A., Danenberg, K. D., Kawakami, K., Saltz, L. B., Blake, C., Shibata, D., Danenberg, P. V., and Laird, P. W. (2000) MethyLight: a high-throughput assay to measure DNA methylation, *Nucleic acids research* **28**; e32.
5. Kristensen, L. S., Mikeska, T., Krypuy, M., and Dobrovic, A. (2008) Sensitive Melting Analysis after Real Time- Methylation Specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection, *Nucleic acids research* **36**; e42.
6. Ririe, K. M., Rasmussen, R. P., and Wittwer, C. T. (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction, *Analytical biochemistry* **245**; 154–160.
7. Worm, J., Aggerholm, A., and Guldberg, P. (2001) In-tube DNA methylation profiling by fluorescence melting curve analysis, *Clinical chemistry* **47**; 1183–1189.
8. Alders, M., Bliek, J., vd Lip, K., vd Bogaard, R., and Mannens, M. (2009) Determination of KCNQ1OT1 and H19 methylation levels in BWS and SRS patients using methylation-sensitive high-resolution melting analysis, *European Journal of Human Genetics* **17**; 467–473.
9. Wojdacz, T. K., Dobrovic, A., and Algar, E. M. (2008) Rapid detection of methylation change at H19 in human imprinting disorders using methylation-sensitive high-resolution melting, *Human mutation* **29**; 1255–1260.
10. White, H. E., Hall, V. J., and Cross, N. C. (2007) Methylation-sensitive high-resolution melting-curve analysis of the SNRPN gene as a diagnostic screen for Prader-Willi and Angelman syndromes, *Clinical chemistry* **53**; 1960–1962.
11. Candiloro, I. L., Mikeska, T., Hokland, P., and Dobrovic, A. (2008) Rapid analysis of heterogeneously methylated DNA using digital methylation-sensitive high resolution melting: application to the CDKN2B (p15) gene, *Epigenetics & chromatin* **1**; 7.
12. Snell, C., Krypuy, M., Wong, E. M., Loughrey, M. B., and Dobrovic, A. (2008) BRCA1 promoter methylation in peripheral blood DNA of mutation negative familial breast cancer patients with a BRCA1 tumour phenotype, *Breast Cancer Research* **10**; R12.
13. Wojdacz, T. K., and Hansen, L. L. (2006) Reversal of PCR bias for improved sensitivity of the DNA methylation melting curve assay, *BioTechniques* **41**; 274, 276, 278.
14. Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D., and Baylin, S. B. (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands, *Proceedings of the National Academy of Sciences of the United States of America* **93**; 9821–9826.
15. Lo, Y. M., Wong, I. H., Zhang, J., Tein, M. S., Ng, M. H., and Hjelm, N. M. (1999) Quantitative analysis of aberrant p16 methylation using real-time quantitative methylation-specific polymerase chain reaction, *Cancer research* **59**; 3899–3903.
16. Preusser, M., Elezi, L., and Hainfellner, J. A. (2008) Reliability and reproducibility of PCR-based testing of O6-methylguanine-DNA methyltransferase gene (MGMT) promoter methylation status in formalin-fixed and paraffin-embedded neurosurgical biopsy specimens, *Clinical neuropathology* **27**; 388–390.
17. Chan, M. W., Chu, E. S., To, K. F., and Leung, W. K. (2004) Quantitative detection of methylated SOCS-1, a tumor suppressor gene, by a modified protocol of quantitative real time methylation-specific PCR using SYBR green and its use in early gastric cancer detection, *Biotechnology letters* **26**; 1289–1293.
18. Candiloro, I. L., and Dobrovic, A. (2009) Detection of MGMT promoter methylation in normal individuals is strongly associated with the T allele of the rs16906252 MGMT promoter single nucleotide polymorphism, *Cancer prevention research (Philadelphia, Pa)* **2**; 862–867.
19. Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G., and Pryor, R. J. (2003) High-resolution genotyping by amplicon melting analysis using LCGreen, *Clinical chemistry* **49**; 853–860.
20. Mikeska, T., Candiloro, I. L. M., and Dobrovic, A. (2010) The implications of heterogeneous DNA methylation for the accurate quantification of methylation, *Epigenenomics* **2**; 561–573.
21. Schuffler, P., Mikeska, T., Waha, A., Lengauer, T., and Bock, C. (2009) MethMarker: user-friendly

design and optimization of gene-specific DNA methylation assays, *Genome biology* **10**; R105.

22. Kibbe, W. A. (2007) OligoCalc: an online oligonucleotide properties calculator, *Nucleic acids research* **35**; W43–46.
23. Brandes, J. C., Carraway, H., and Herman, J. G. (2007) Optimal primer design using the novel primer design program: MSPprimer provides accurate methylation analysis of the ATM promoter, *Oncogene* **26**; 6229–6237.
24. Poland, D. (1974) Recursion relation generation of probability profiles for specific-sequence macromolecules with long-range correlations, *Biopolymers* **13**; 1859–1871.
25. Steger, G. (1994) Thermal denaturation of double-stranded nucleic acids: prediction of temperatures critical for gradient gel electrophoresis and polymerase chain reaction, *Nucleic acids research* **22**; 2760–2768.
26. Wojdacz, T. K., Hansen, L. L., and Dobrovic, A. (2008) A new approach to primer design for the control of PCR bias in methylation studies, *BMC research notes* **1**; 54.
27. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, 2nd ed., Cold Sprinh Harbor Laboratory Press, Cold Spring Harbor, N.Y.
28. Warnecke, P. M., Stirzaker, C., Song, J., Grunau, C., Melki, J. R., and Clark, S. J. (2002) Identification and resolution of artifacts in bisulfite sequencing, *Methods (San Diego, Calif* **27**; 101–107.
29. Bookstein, R., Lai, C. C., To, H., and Lee, W. H. (1990) PCR-based detection of a polymorphic BamHI site in intron 1 of the human retinoblastoma (RB) gene, *Nucleic acids research* **18**; 1666.
30. Henke, W., Herdel, K., Jung, K., Schnorr, D., and Loening, S. A. (1997) Betaine improves the PCR amplification of GC-rich DNA sequences, *Nucleic acids research* **25**; 3957–3958.
31. Baskaran, N., Kandpal, R. P., Bhargava, A. K., Glynn, M. W., Bale, A., and Weissman, S. M. (1996) Uniform amplification of a mixture of deoxyribonucleic acids with varying GC content, *Genome research* **6**; 633–638.
32. Sarkar, G., Kapelner, S., and Sommer, S. S. (1990) Formamide can dramatically improve the specificity of PCR, *Nucleic acids research* **18**; 7465.
33. Tanaka, K., and Okamoto, A. (2007) Degradation of DNA by bisulfite treatment, *Bioorganic & medicinal chemistry letters* **17**; 1912–1915.
34. Eads, C. A., Lord, R. V., Wickramasinghe, K., Long, T. I., Kurumboor, S. K., Bernstein, L., Peters, J. H., DeMeester, S. R., DeMeester, T. R., Skinner, K. A., and Laird, P. W. (2001) Epigenetic patterns in the progression of esophageal adenocarcinoma, *Cancer research* **61**; 3410–3418.
35. Livak, K. J., and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(–Delta Delta C(T)) Method, *Methods (San Diego, Calif* **25**; 402–408.
36. Pfaffl, M. W., Georgieva, T. M., Georgiev, I. P., Ontsouka, E., Hageleit, M., and Blum, J. W. (2002) Real-time RT-PCR quantification of insulin-like growth factor (IGF)-1, IGF-1 receptor, IGF-2, IGF-2 receptor, insulin receptor, growth hormone receptor, IGF-binding proteins 1, 2 and 3 in the bovine species, *Domestic animal endocrinology* **22**; 91–102.
37. Lind, K., Stahlberg, A., Zoric, N., and Kubista, M. (2006) Combining sequence-specific probes and DNA binding dyes in real-time PCR for specific nucleic acid quantification and melting curve analysis, *BioTechniques* **40**; 315–319.
38. Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., and Wittwer, C. T. (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments, *Clinical chemistry* **55**; 611–622.
39. Lefever, S., Hellemans, J., Pattyn, F., Przybylski, D. R., Taylor, C., Geurts, R., Untergasser, A., and Vandesompele, J. (2009) RDML: structured language and reporting guidelines for real-time quantitative PCR data, *Nucleic acids research* **37**; 2065–2069.

Chapter 6

A Combined Bisulfite Restriction Analysis Bioinformatics Tool: Methyl-Typing

Cheng-Hong Yang, Yu-Huei Cheng, Li-Yeh Chuang, and Hsueh-Wei Chang

Abstract

In this chapter, we introduce our developed freeware tool Methyl-Typing. It provides methylation-related bioinformatics with a special focus on combined bisulfite restriction analysis (COBRA). We give an overview of the implementation and program modules for Methyl-Typing. Various databases and methylation-related functions are integrated into Methyl-Typing and a helpful example is illustrated in detail. Several input protocols and their outputs for COBRA-related information are demonstrated, such as the inputs of multiple gene names in official gene symbols, multiple accession numbers for nucleotide sequence retrieval, multiple template sequences in a free format, primer sequences, and file uploads. The program goal of Methyl-Typing is to provide computation and visualization of the essential information for COBRA assay so that methylation can easily be analyzed by COBRA. It is a fast and efficient tool for providing all possible methylation sites of restriction enzymes.

Key words: Methylation, COBRA, Polymorphism, SNP, Database, BLAST, RFLP, Genotyping, Primer design

1. Introduction

Methylation of DNA is a common epigenetic signaling function that cells use to silence genes and keep them in the "off" position; it is an essential mechanism in many cellular processes, such as development, imprinting, X-chromosome inactivation, and maintenance of chromosome stability (1). DNA hypermethylation had been found to downregulate many genes involved in diverse functions and pathways leading to cancer (2, 3). Therefore, it is important to monitor and determine the DNA methylation status.

Methods for methylation detection rely on bisulfite-independent and-dependent approaches. For bisulfite-independent methods,

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_6,

DNA digestion with methylation-sensitive restriction enzymes (MSRE) (4) and methylated-DNA immunopreciptation (MeDIP) (5) have been reported. They are dependent on the overall CpG content of the corresponding region rather than the individual methylation site. For bisulfite-dependent methods, DNA is treated with bisulfite to convert unmethylated cytosine residues, followed by direct DNA sequencing (6) or methylation-specific PCR (MSP) (7), MethyLight (8), pyrosequencing (9), and COBRA (combined bisulfite restriction analysis) (10).

Among these methods, COBRA is still one of the most common methylation methods for many regular laboratories (11–17). PCR amplification of bisulfite-treated DNA may contain some DNA fragments with newly created or retained restriction sites containing CpG, indicating the relative amounts of DNA sequences with methylated and unmethylated CpG sites (digested and undigested PCR products, respectively) (10). COBRA is thus a quantitative method for determining the methylation levels of particular CpG sites. Therefore, any kind of restriction enzyme used to distinguish between methylated and unmethylated sequences with bisulfite-conversion is regarded as the COBRA method.

In the traditional COBRA approach, only a few restriction enzymes are used, such as BstUI (5′-CG$^{\downarrow}$CG-3′) (11) and Taq$^{\alpha}$I (5′-T$^{\downarrow}$CGA-3′) (10). However, other restriction enzymes available for COBRA, such as HinP1I (5′-G$^{\downarrow}$CGC-3′), HpyCH4IV (5′-A$^{\downarrow}$CGT-3′), and AciI (5′-G$^{\downarrow}$CGG-3′), are less frequently mentioned. Therefore, restriction enzymes with recognition sites containing CG for digestion should be included among COBRA enzymes. Different nucleotides flanking CG may favor different restriction enzymes for COBRA assay. Moreover, the traditional COBRA approach detects the CpG island-containing sequences but it is not restricted to promoter regions. Therefore, it is still challenging to integrate COBRA restriction enzyme mining, CpG island searching, promoter prediction, and other related functions in the development of an improved COBRA bioinformatics tool.

In this chapter, we introduce a novel visualization software called Methyl-Typing (18) (http://www.bio.kuas.edu.tw/methyl-typing), which is able to provide comprehensive restriction enzymes for methyl-cytosine-containing sequences after bisulfite-conversion, i.e., unmethylated cytosine converts to uracil (regarded as thymine for PCR amplification) while 5-methylcytosine remains unchanged. CpG and GpC sites in promoters are available for mining restriction enzymes. Moreover, the CCCTC-binding factor (CTCF)-binding site database (CTCFBSDB) of insulators (19) is embedded in Methyl-Typing. The insulators of chromatin, such as CTCF, can block the activity of a down-stream enhancer and are neutralized by methylation (20), thereby contributing to gene regulation. Possible continuous poly Ts within the promoter sequences from computer-generated bisulfite-converted sequences are detected by

user-defined length. The primer design for bisulfite sequencing PCR (BSP) is also available. Using several examples of inputs, computation, and visualization of essential information for COBRA assay is demonstrated and methylation can easily be analyzed by COBRA.

2. Materials

2.1. Hardware

A standard personal computer platform with an Internet connection.

2.2. Software

A regular Internet browser, such as Internet Explorer, is required. It should support JavaScript 1.1.

3. Methods

Methyl-Typing (http://www.bio.kuas.edu.tw/methyl-typing), an integrated methylation-related and web-based tool, was designed and implemented in Java language. It runs on Apache Tomcat (see Note 1) under JSP (Java Server Pages) and Servlet technologies.

3.1. System Implementation

3.1.1. Integration of DBTSS and CpG Island Searcher

A promoter database for human and mouse genomes and a CpG island searcher tool were downloaded from DBTSS (21) (see Note 2) and CpG island searcher (22) (see Note 3), respectively. They were employed to Methyl-Typing (18) to provide promoter sequences and CpG island prediction.

3.1.2. Integration of GenBank and In Silico PCR

Two well-known public databases, GenBank and In Silico PCR of the University of California, Santa Cruz (UCSC) Genome Browser (23) (see Note 4) are well integrated to Methyl-Typing (18) to retrieve accession number and primer-based sequences, respectively. Sequences are retrieved online when the accession numbers and primer sequences are used for input.

3.1.3. RFLP Analysis

A restriction enzyme database was downloaded from REBASE version 806 (24) (see Note 5) to provide mining of available restriction enzymes for methylated sequences. For example, CpG and GpC sites (see Note 6) within sequences are examined by a restriction fragment length polymorphism (RFLP) program (25) (see Note 7) to mine all available restriction enzymes in REBASE for COBRA assay.

3.1.4. CTCFBSDB Search

The CCCTC-binding factor (CTCF)-binding site database (CTCFBSDB) for vertebrate genomic insulators has been hyperlinked and fed to CTCFBSDB (19) (see Note 8) for online analysis.

3.1.5. Bisulfite T Stretches

For bisulfite T stretches, all C nucleotides in a sequence are converted to T nucleotides by default (computer-generated bisulfite-converted sequences). Subsequently, the continuous T nucleotides are processed to be visualized in red.

3.1.6. Bisulfite Sequencing Primers

Bisulfite sequencing primers (BSPs) were designed in Methyl-Typing (18) according to criteria in the user manual for the freeware-Methyl Primer Express® Software v1.0 (see Note 9). Subsequently, COBRA analysis can be performed using the BSP primers for PCR amplification and the mined COBRA enzymes for digestion.

3.2. Program Modules

The Methyl-Typing (18) system incorporates 11 modules, the (1) Input Module, (2) Promoter Query Module, (3) Nucleotide Query Module, (4) UCSC In Silico PCR Module, (5) CpG Island Search Module, (6) Visualization Module, (7) CpG-RFLP Module, (8) CTCFBS Search Module, (9) Bisulfite T Stretch Module, (10) Primer Design Module, and (11) Output Module. They are described in detail below.

3.2.1. Input Module

Five input types are available in the Input Module including gene name(s), accession#(s), fasta sequences, primer pair information, and file uploads. Gene name(s) input is integrated into the DBTSS database which includes human and mouse organisms. Accession#(s) input uses the comma, space, and enter keys to separate nucleotides for query nucleotide sequences. Input with single or multiple fasta sequences can be used to search for CpG islands in sequence by setting some parameters for CpG islands. Primer pair information input uses the UCSC in silico PCR-based interface to input-related values. File upload format (fasta sequences) is also acceptable for analyzing COBRA.

3.2.2. Promoter Query Module

The Promoter Query Module employs the DBTSS database to query promoter sequences of human or mouse genomes by gene name(s) input from the Input Module.

3.2.3. Nucleotide Query Module

The Nucleotide Query Module integrates the NCBI nucleotide database to query nucleotide sequences by accession#(s) input from the Input Module. It uses online retrieval technology to remote access database servers to request nucleotide information.

3.2.4. UCSC In Silico PCR Module

The UCSC in silico PCR (23, 26) which is a fast and stable bioinformatics tool for locating sequence positions is employed to obtain sequences for PCR products by primer pair sequence input from the Input Module. It also uses online retrieval technology to achieve this purpose.

3.2.5. CpG Island Search Module

The CpG Island Search Module combines with a CpG Island Searcher (22). The command line program cpgi130 is used in the system.

3.2.6. Visualization Module

CpG islands and restriction enzymes can be displayed by the Visualization Module. CpG mapping, CpG distribution, and CpG island sequences are available for visualization of CpG islands. Restriction enzymes gained from the SNP-RFLP Module are displayed and the user can click on the desired selection. The modified core of SNP-RFLPing (25, 27) can identify available restriction enzymes which only recognize sequences containing "CG" or "GC" nucleotides (see Note 6).

3.2.7. CpG-RFLP Module

The CpG-RFLP Module uses the modified core of SNP-RFLPing (25, 27) to analyze the availability of RFLP restriction enzymes for CpG-containing sequences.

3.2.8. CTCFBS Search Module

The CTCFBS Search Module uses CTCFBSDB (19) to search for insulators for a CpG island-containing sequence after use of the CpG-RFLP Module.

3.2.9. Bisulfite T Stretch Module

The Bisulfite T Stretch Module provides a search for repeat "T" nucleotides after the bisulfite process. The corresponding sequence is marked in red and the default length is 10 bp.

3.2.10. Primer Design Module

BSPs are implemented to design feasible primers according to criteria in the Methyl Primer Express® Software v1.0 user manual (see Note 9).

3.2.11. Output Module

The F (forward) and R (reverse) primers are visualized and their sequences are provided in the Output Module. The positions for the primer location and COBRA enzymes are shown, as well as the product size for PCR digestion with COBRA enzymes.

3.3. Examples of Input Protocols and Their Outputs

Modes of input available in the input module include the gene name(s) for the DBTSS input, NCBI nucleotide accession#(s) input, fasta sequence(s) paste input, primer pair information input, and fasta sequence file input (Fig. 1a–e). Detailed information from the user manual is provided below (see Note 10).

3.3.1. Gene Name(s) for DBTSS Input

The user can input multiple gene names using official gene symbols to search for transcriptional start sites from the DBTSS (see Note 2). Two organisms, human and mouse, can be selected. For example, input a gene name "TP53" as shown in Fig. 1a. By clicking the "Query" button, the promoter sequence will be imported to the system. Subsequently, the user can give "CpG islands search" parameter settings as shown in Fig. 2 for CpG island prediction (see Subheading 3.3.6 for the results of a CpG island search).

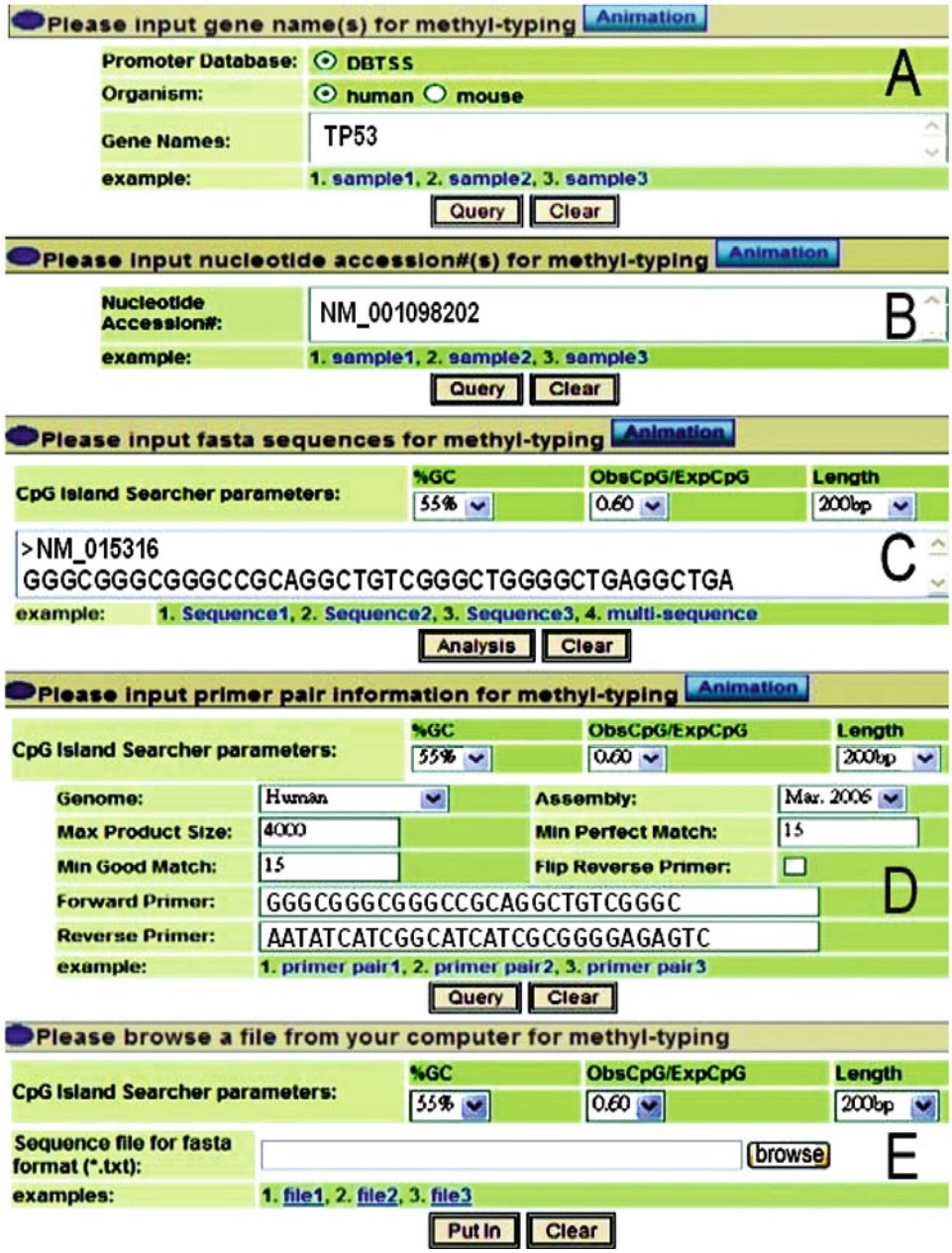


Fig. 1. Five input interfaces in Methyl-Typing. (**a**) Gene name(s) input interface with a gene name "TP53" input. (**b**) Nucleotide accession#(s) input interface with a nucleotide accession# "NM_001098202" input. (**c**) Fasta sequence input interface with a fasta sequence "NM_015316" input by clicking the hyperlink "Sequence1." (**d**) Primer pair information input interface with forward primer: "GGGCGGGCGGGCCGCAGGCTGTCGGGC" and reverse primer: "AATATCATCGGCATCATCGCGGGGAGAGTC" input. (**e**) File upload.

3.3.2. NCBI Nucleotide Accession#(s) Input

NCBI nucleotide accession# input is available in the Methyl-Typing system to provide several sequences. For example, the user can input the nucleotide accession# "NM_001098202" as shown in Fig. 1b, and then click the "Query" button. Subsequently, the nucleotide

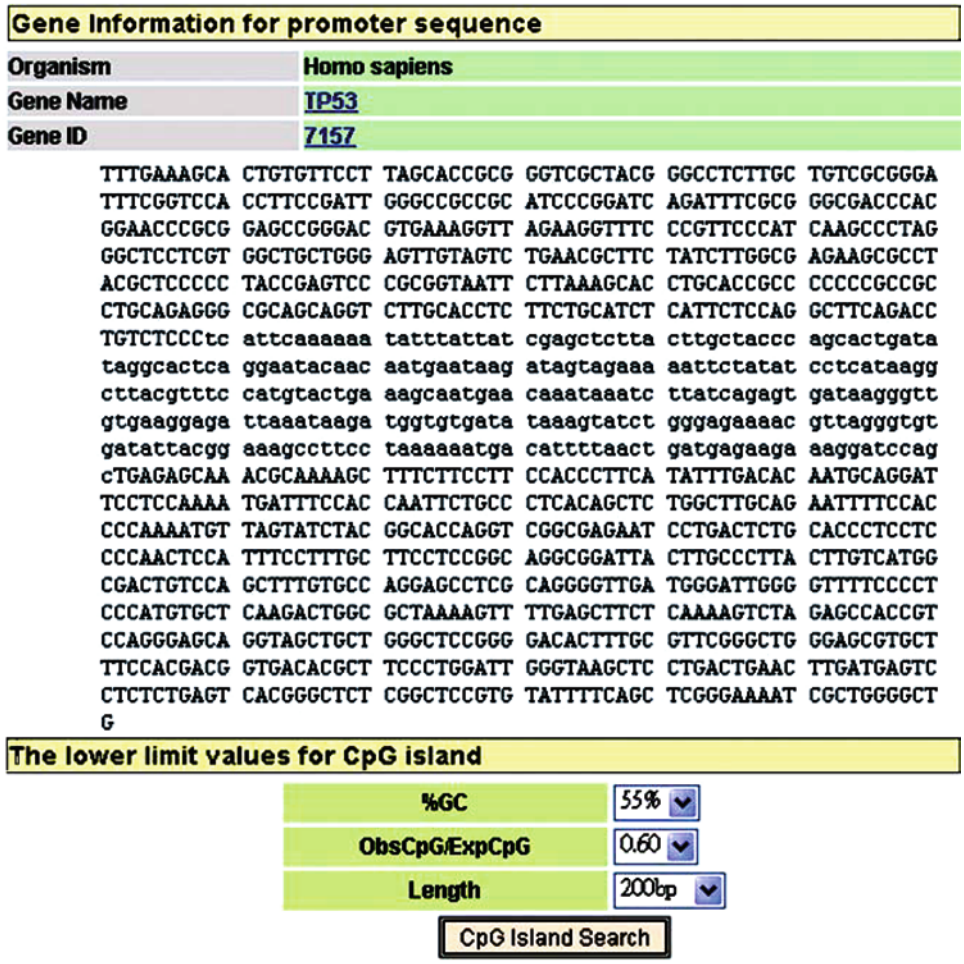


Fig. 2. Promoter sequence output with CpG island parameter settings indicated by gene name "TP53" input. The CpG island search parameters are described in Subheading 3.3.6.

sequences are imported to the system. Subsequently, the user can give "CpG islands search" parameter settings (similar to Fig. 2) for CpG island prediction (see Subheading 3.3.6).

3.3.3. Fasta Sequence(s) Paste Input

Fasta sequences (see Note 11) of templates are accepted for input. Upper and lower case letters are not significant and all other characters, including spaces and digits, are ignored. A fasta format sequence or multiple fasta format sequences can be typed in or pasted into the sequence input text area, and CpG island search parameters can be adjusted (see Subheading 3.3.6). For example, by clicking example 1 (sequence 1) as shown in Fig. 1c, the fasta sequence for "NM_015316" is fed automatically.

3.3.4. Primer Pair Information Input

A primer pair can be typed or pasted into the "Forward Primer" text field and "Reverse Primer" text field as shown in Fig. 1d. For example, forward primer: "GGGCGGGCGGGCCGCAGGCTGTCGGGC" and reverse primer: "AATATCATCGGCATCATCGCGGGGAGAGTC" are used as an example to find the PCR product. The function is embedded in the system from the UCSC in silico PCR database (see Note 4). Several parameters including Genome and Assembly, Forward Primer, Reverse Primer, Max Product Size, Min Perfect Match, Min Good Match, and Flip Reverse Primer are available to adjust the search results. The meanings of these configuration options are described (see Note 12) as shown in Table 1. CpG island search parameters are described below (see Subheading 3.3.6).

3.3.5. Fasta Sequence File Input

A fasta format sequence or multiple fasta format sequences (see Note 13) can be saved to a file in the user's local computer. Users can click the "browse" button to select the file to import to the system. CpG island search parameters are available in Subheading 3.3.6. The following operations are the same as the Subheading 3.3.3.

3.3.6. CpG Islands Search

The default CpG Island Searcher parameters are 55% of the GC, 0.60 of ObsCpG/ExpCpG, and a length of 200 bp to search for CpG islands from the input sequence. For example, for the gene input "TP53," the default CpG Island search result is shown in Fig. 3. Three options "CpG Mapping," "CpG Distribution," and "CpG Island Seq" in each CpG Island # are shown in Fig. 3a, Fig. 3b, and Fig. 3c, respectively. The user then clicks on "*Methyl RFLP analysis*" to access Subheading 3.3.7.

Table 1
Configuration options derived from the UCSC in silico PCR (23, 26)

Configuration option	Description
Genome and Assembly	The sequence database to search.
Forward Primer	Minimum length of 15 bases.
Reverse Primer	Minimum length of 15 bases.
Max Product Size	Maximum size of PCR amplified region.
Min Perfect Match	Number of bases that match exactly on 3′ end of primers. Minimum match size is 15.
Min Good Match	Number of bases on 3′ end of primers where at least 2 out of 3 bases match.
Flip Reverse Primer	Inversion of the sequence order of the reverse primer and its complement.

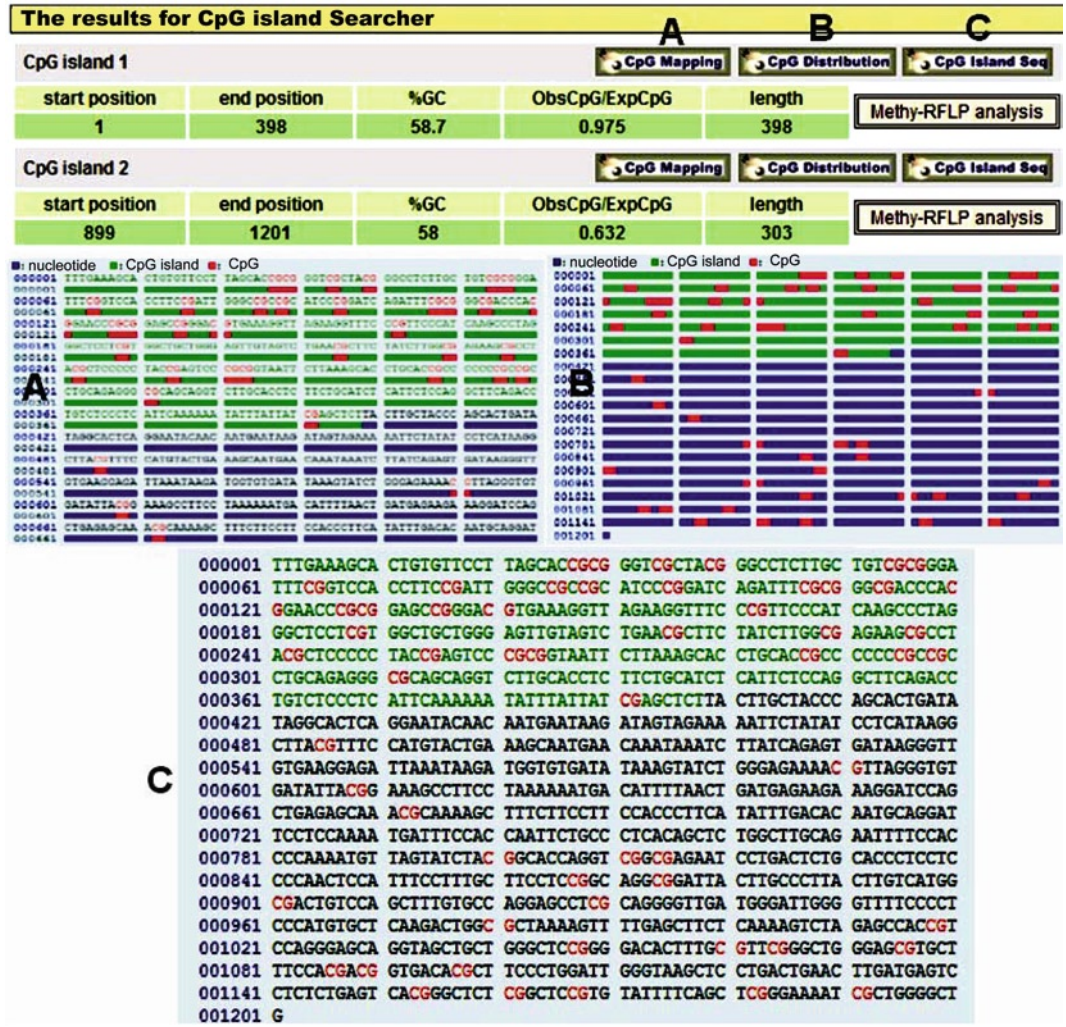


Fig. 3. The results of a CpG island search. CpG island 1 is used as an example. (**a**) "CpG Mapping" shows bar images and nucleotide sequences in undertone, (**b**) "CpG Distribution" shows bar images in undertone, and (**c**) "CpG Island Seq" shows nucleotide sequences in undertone.

3.3.7. Methylation RFLP Analysis

Available restriction enzymes of methylation RFLP are provided for the selected CpG island# by the Subheading 3.3.6. Restriction enzymes are classified as commercial and noncommercial. The recognition sequences contain two types of nomenclature, the International Union of Pure and Applied Chemistry (IUPAC) and non-IUPAC. All positions of the recognition sequences are listed and sorted in alphabetical order as shown in Fig. 4. At the top, four options are available for displaying additional information, "Sequence," "Visualization," "CTCFBS Search," and "Bisulfite T Stretch" with labels A to D, respectively.

By clicking the "Sequence" image (label A) in the top window of Fig. 4, the system shows the original or bisulfite-treated/methylated

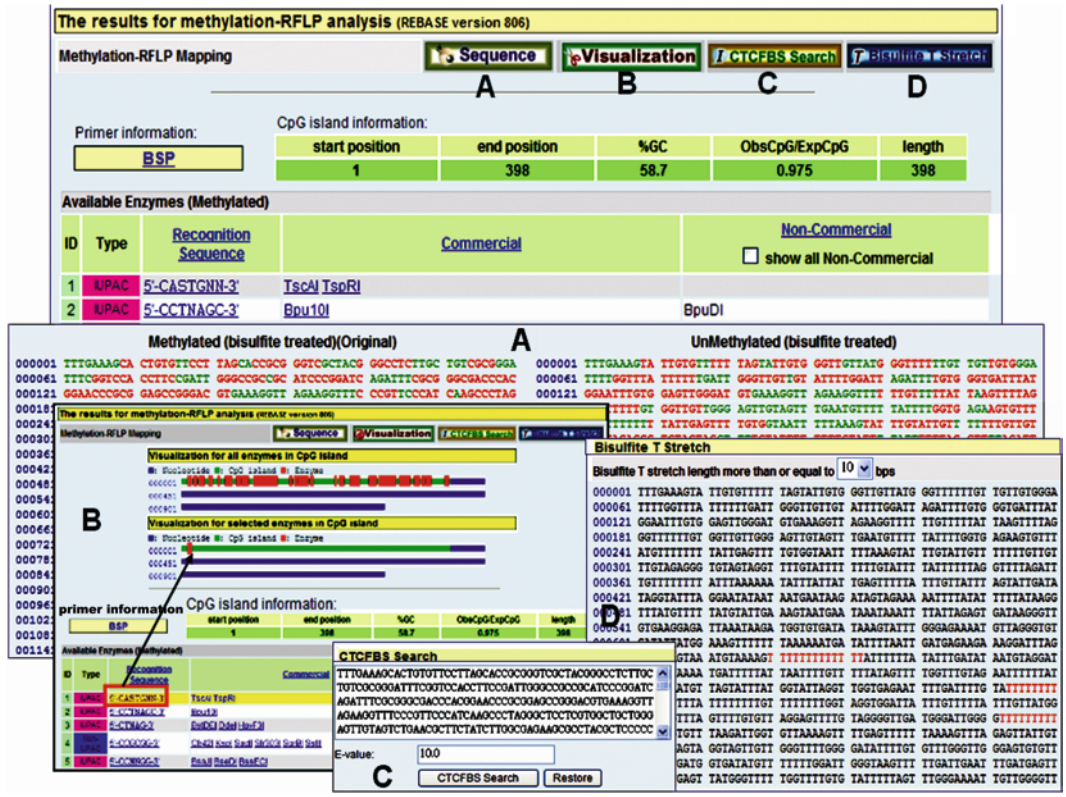


Fig. 4. The results of methylation RFLP analysis. (**a**) Methylated and unmethylated sequences (in undertone) after bisulfite treatment. (**b**) A user-interaction interface for visualizing available restriction enzymes for CpG island-containing methylated sequences (in undertone). (**c**) "CTCFBS Search" provides an insulator search function for CpG islands. (**d**) "Bisulfite T Stretch" provides a prediction function for possible continuous "T" nucleotides (in undertone) after computer-generated bisulfite-converted sequences of the target gene.

sequences with CpG islands marked in green and the mined restriction enzymes marked in red on the left side of Fig. 4a. For comparison, the bisulfite-treated/unmethylated sequences are provided on the right side of Fig. 4a.

By clicking the "Visualization" image (label B) in the top window of Fig. 4, all recognition sequences for the corresponding restriction enzymes are marked in red bars under the CpG region marked in green (Fig. 4b). By clicking a recognition sequence of the restriction enzyme (e.g., 5′-CASTGNN-3′ in Fig. 4b), the selected line is marked in yellow and the visualization for selected enzymes in the CpG island are marked in red bars as indicated by the arrow.

When users click the "CTCFBS Search" image (label C) in the top window of Fig. 4, a new browser page pops up. The input sequence or the promoter sequence is loaded automatically and the default E-value is 10.0 (Fig. 4c).

By clicking the "Bisulfite T Stretch" image (label D) in the top window of Fig. 4, a new browser page pops up. Possible continuous poly Ts (see Note 14) within the promoter sequence from the computer-generated bisulfite-converted sequence of TP53 gene are shown. Under a default length of 10 bp, the corresponding continuous poly T sequence is marked in red (Fig. 4d).

3.3.8. Bisulfite Sequencing Primer Output

By clicking the hyperlink of the "BSP" box in Fig. 4, a bisulfite sequencing primer pair is designed. As shown in Fig. 5, the primer information includes the primer pair (F and R), position, length (bp), GC no., GC%, tm (temperature, °C), tm-diff (temperature difference, °C), and product length (bp). The primers F (forward) and R (reverse) are visualized and marked in the input sequence or the promoter sequence. The available enzymes for the methylated sequences are also provided with information on recognition sequences, commercial COBRA enzymes, and the recognition positions in the sequence, as well as the product size for PCR digestion with COBRA enzymes.

primer pair	position	length (bp)	GC no.	GC %	tm (°C)	tm-diff (°C)	product length (bp)
F: 5'- CCATCAAGCCCTAGGGCT-3'	167	18	11	61.11	64.56	2.48	118
R: 5'- GCAGGTGCTTTAAGAATTACC-3'	264	21	9	42.86	62.07		

```
000001 TTTGAAAGCA CTGTGTTCCT TAGCACCGCG GGTCGCTACG GGCCTCTTGC TGTCGCGGGA
000061 TTTCGGTCCA CCTTCCGATT GGGCCGCCGC ATCCCGGATC AGATTTCGCG GGCGACCCAC
000121 GGAACCCGCG GAGCCGGGAC GTGAAAGGTT AGAAGGTTTC CCGTTCCCAT CAAGCCCTAG
000181 GGCTCCTCGT GGCTGCTGGG AGTTGTAGTC TGAACGCTTC TATCTTGGCG AGAAGCGCCT
000241 ACGCTCCCCC TACCGAGTCC CGCGGTAATT CTTAAAGCAC CTGCACCGCC CCCCCGCCGC
000301 CTGCAGAGGG CGCAGCAGGT CTTGCACCTC TTCTGCATCT CATTCTCCAG GCTTCAGACC
000361 TGTCTCCCTC ATTCAAAAAA TATTTATTAT CGAGCTCTTA CTTGCTACCC AGCACTGATA
000421 TAGGCACTCA GGAATACAAC AATGAATAAG ATAGTAGAAA AATTCTATAT CCTCATAAGG
000481 CTTACGTTTC CATGTACTGA AAGCAATGAA CAAATAAATC TTATCAGAGT GATAAGGGTT
000541 GTGAAGGAGA TTAAATAAGA TGGTGTGATA TAAAGTATCT GGGAGAAAAC GTTAGGGTGT
000601 GATATTACGG AAAGCCTTCC TAAAAAATGA CATTTTAACT GATGAGAAGA AAGGATCCAG
000661 CTGAGAGCAA ACGCAAAAGC TTTCTTCCTT CCACCCTTCA TATTTGACAC AATGCAGGAT
000721 TCCTCCAAAA TGATTTCCAC CAATTCTGCC CTCACAGCTC TGGCTTGCAG AATTTTCCAC
000781 CCCAAAATGT TAGTATCTAC GGCACCAGGT CGGCGAGAAT CCTGACTCTG CACCCTCCTC
000841 CCCAACTCCA TTTCCTTTGC TTCCTCCGGC AGGCGGATTA CTTGCCCTTA CTTGTCATGG
000901 CGACTGTCCA GCTTTGTGCC AGGAGCCTCG CAGGGGTTGA TGGGATTGGG GTTTTCCCCT
000961 CCCATGTGCT CAAGACTGGC GCTAAAAGTT TTGAGCTTCT CAAAAGTCTA GAGCCACCGT
001021 CCAGGGAGCA GGTAGCTGCT GGGCTCCGGG GACACTTTGC GTTCGGGCTG GGAGCGTGCT
001081 TTCCACGACG GTGACACGCT TCCCTGGATT GGGTAAGCTC CTGACTGAAC TTGATGAGTC
001141 CTCTCTGAGT CACGGGCTCT CGGCTCCGTG TATTTTCAGC TCGGGAAAAT CGCTGGGGCT
001201 G
```

Available Enzymes (Methylated)

ID	Type	Recognition Sequence	Commercial	Position	Product size
1	Non-IUPAC	5'-CCGCGG-3'	Cfr42I KspI SacII Sfr303I SqrBI SstII	260	94, 119
2	IUPAC	5'-CCNNGG-3'	BsaJI BseDI BssECI	176, 260	10, 94, 119
3	IUPAC	5'-CCRYGG-3'	BstDSI BtgI	260	94, 119
4	IUPAC	5'-CMGCKG-3'	MspA1I	260	94, 119
5	Non-IUPAC	5'-CCGC-3'	AciI BspACI SsiI	260	94, 119
6	Non-IUPAC	5'-CGCG-3'	AccII Bsh1236I BspFNI BstFNI BstUI MvnI	261	95, 119

Fig. 5. Output for bisulfite sequencing primers. Primer pair information is provided at the *top of window*. Sequences for bisulfite sequencing primers are *underlined*. The available restriction enzymes for the methylated sequences are also provided.

4. Notes

1. Apache Tomcat: http://www.tomcat.apache.org/.
2. DBTSS version 6.0 (21): http://www.dbtss.hgc.jp/.
3. CpG island searcher (22): http://www.uscnorris.com/cpgislands2/cpg.aspx.
4. In Silico PCR of the UCSC Genome Browser: http://www.genome.ucsc.edu/cgi-bin/hgPcr.
5. REBASE version 806: http://www.rebase.neb.com/rebase/rebase.html.
6. Both CpG and GpC sites are recognized by a modified core of SNP-RFLPing (25). About 90% of methylated cytosine residues in mammals are found at CpG dinucleotides (28, 29). Methylated cytosine residues for GpC are not frequently discussed and analyzed. However, methylations in non-CpG sequences have been found in some plants (30–33), animals (34, 35), and human breast cancer (36).
7. SNP-RFLPing 2: http://www.bio.kuas.edu.tw/snp-rflping2/.
8. The CCCTC-binding factor (CTCF)-binding site is an insulator of chromatin which can block the activity of a down-stream enhancer and is neutralized by methylation (20). In Methyl-Typing, the CTCFBSDB (19), located at http://www.insulatordb.utmem.edu/help.php, is employed to provide related information. Similarly, the CTCF paralog BORIS (brother of the regulator of imprinted sites) is an insulator DNA-binding protein and may regulate gene expression. It has been suggested that BORIS is a pro-proliferative factor (37, 38), whereas CTCF acts as an antiproliferative factor. BORIS and CTCF share similar zinc finger DNA-binding domains (39) and seem to bind to identical target sequences. Therefore, prediction sites for CTCFBS implemented in Methyl-Typing may provide possible binding site candidates for BORIS.
9. Methyl Primer Express® Software v1.0: http://www.appliedbiosystems.com/methylprimerexpress.
10. A demonstration and the user manual for Methyl-Typing are available for download at http://www.bio.kuas.edu.tw/methyl-typing and http://www.bio.kuas.edu.tw/methyl-typing/userManual_info.jsp, respectively. Animation is provided on the homepage for tutorial purposes.
11. The fasta format: >name [first line]-sequence [second-line], with A, T, C, and G. For example, the fasta format for NM_015316 (PPP1R13B gene) is shown as follows:

    >NM_015316

GGGCGGGCGGGCCGCAGGCTGTCGGGCTGGGGCTG
AGGCTGAGGCTGAGGTTGAGGCGGCGGCG
GCGGCGGCCGGGTGCCCGGGACAGCGACGCAG
CGCGCCGGCGGCCGCGACAGGGCCAGCGAGAGC
CCCGCAGCCCGCCGCAGCTGCCGCCTCGCCGC
GGCCGGGCCGGAGAGCACGGCGGCGGGAGCG
CGGCCTTAGGAGGCGGCCGGAGCGGTGGGCA
CAGCTCGGCGCGGAGCGTCCTGTCAGGCGG
CGGCCGAGGGCGTCGCGGACTCTCCCCG
CGATGATGCCGATGATATT

12. The meanings of these configuration options in UCSC in silico PCR are listed in http://www.genome.ucsc.edu/cgi-bin/hgPcr.
13. Example of multiple fasta formats:

>NM_015316

GGGCGGGCGGGCCGCAGGCTGTCGGGCTGGGGCT
GAGGCTGAGGCTGAGGTTGAGGCGGCGGCGGC
GGCGGCCGGGTGCCCGGGACAGCGACGCAG
CGCGCCGGCGGCCGCGACAGGGCCAGCGAGAG
CCCCGCAGCCCGCCGCAGCTGCCGCCTCGCCG
CGGCCGGGCCGGAGAGCACGGCGGCGGGAGC
GCGGCCTTAGGAGGCGGCCGGAGCGGTGGGCA
CAGCTCGGCGCGGAGCGTCCTGTCAGGCGGC
GGCCGAGGGCGTCGCGGACTCTCCCCGCGA
TGATGCCGATGATATT

>NM_007499

TTTATAGAGCCGGAAGCGGGAAGGCGTGCGTAC
AGAACCAGCTGCTAGATCCGTGCACGCGGG
AAAAGGCGAAGCGACCTGGGTTTGCATTGGCA
GGCAGAATGCAGCGGTGAGGATGCATGTTC
TGAAATCTTAAACCATGAGTCTAGCACTCA
ATGATCTGCTCATTTGCTGCCGGCAGTTAGA
GCATGACAGAGCTACAGAAAGAAGGAAA
GAAGTGGATAAATTTAAGCGCCTGATTC
AGGATCCTGAAACAGTTCAACATTTAGATA
GGCATTCTGATTCCAAACAAGGAAAATA

>NM_007302

CTTAGCGGTAGCCCCTTGGTTTCCGTGGCAA
CGGAAAAGCGCGGGAATTACAGATAAA
TTAAAACTGCGACTGCGCGGCGTGAGCTC
GCTGAGACTTCCTGGACGGGGGACAG
GCTGTGGGGTTTCTCAGATAACTGGGCC
CCTGCGCTCAGGAGGCCTTCACCCTC
TGCTCTGGGTAAAGTTCATTGGAACAGA
AAGAAATGGATTTATCTGCTCTTCGCGT
TGAAGAAGTACAAAATGTCATTAATGCT
ATGCAGAAAATCTTAGAGTGTCCCATCTG
TCTGGAGTTGATCAAGGAAC

14. In our sequencing experiments (unpublished data), continuous poly Ts always interfered with nearby downstream sequences, i.e., the peaks for poly Ts were high and mixed with nearby sequences leading to computational autoreading in electropherograms. Therefore, prediction of continuous poly Ts within the sequence is helpful to avoid this sequencing problem. In other words, we suggest that users not choose bisulfite sequencing if the predicted sequence is rich in continuous poly Ts. Alternatively, MSP or COBRA assays can avoid sequencing problems with continuous poly Ts.

Acknowledgments

This work was partly supported by the National Science Council in Taiwan under grants NSC98-2622-E-151-024-CC3, NSC98-2622-E-151-001-CC2, NSC98-2221-E-151-040, NSC97-2311-B-037-003-MY3, and by the grants KMU-EM-99-1.4 and DOH 100-TD-C-111-002 Dr. Hsueh-Wei Chang from the Department of Health, Taiwan.

References

1. Phillips, T. (2008) The role of methylation in gene expression. *Nature Education* **1**.
2. Vucic, E. A., Brown, C. J., and Lam, W. L. (2008) Epigenetics of cancer progression. *Pharmacogenomics* **9**, 215–34.
3. Herman, J. G., and Baylin, S. B. (2003) Gene silencing in cancer in association with promoter hypermethylation. *N Engl J Med* **349**, 2042–54.
4. Chu, D. C., Chuang, C. K., Fu, J. B., Huang, H. S., Tseng, C. P., and Sun, C. F. (2002) The use of real-time quantitative polymerase chain reaction to detect hypermethylation of the CpG islands in the promoter region flanking the GSTP1 gene to diagnose prostate carcinoma. *J Urol* **167**, 1854–8.
5. Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L., and Schubeler, D. (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* **37**, 853–62.
6. Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci USA* **89**, 1827–31.
7. Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D., and Baylin, S. B. (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci USA* **93**, 9821–6.
8. Eads, C. A., Danenberg, K. D., Kawakami, K., Saltz, L. B., Blake, C., Shibata, D., Danenberg, P. V., and Laird, P. W. (2000) MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res* **28**, E32.
9. Tost, J., and Gut, I. G. (2007) DNA methylation analysis by pyrosequencing. *Nat Protoc* **2**, 2265–75.
10. Xiong, Z., and Laird, P. W. (1997) COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res* **25**, 2532–4.
11. Rohde, C., Zhang, Y., Jurkowski, T. P., Stamerjohanns, H., Reinhardt, R., and Jeltsch, A. (2008) Bisulfite sequencing Data Presentation and Compilation (BDPC) web server--a useful tool for DNA methylation analysis. *Nucleic Acids Res* **36**, e34.
12. Leakey, T. I., Zielinski, J., Siegfried, R. N., Siegel, E. R., Fan, C. Y., and Cooney, C. A. (2008) A simple algorithm for quantifying DNA methylation levels on multiple independent CpG sites in bisulfite genomic sequencing electropherograms. *Nucleic Acids Res* **36**, e64.
13. Rahner, N., Friedrichs, N., Steinke, V., Aretz, S., Friedl, W., Buettner, R., Mangold, E., Propping, P., and Walldorf, C. (2008) Coexisting somatic promoter hypermethylation and pathogenic MLH1 germline mutation in Lynch syndrome. *J Pathol* **214**, 10–6.

14. Nagasaka, T., Goel, A., Notohara, K., Takahata, T., Sasamoto, H., Uchida, T., Nishida, N., Tanaka, N., Boland, C. R., and Matsubara, N. (2008) Methylation pattern of the O6-methylguanine-DNA methyltransferase gene in colon during progressive colorectal tumorigenesis. *Int J Cancer* **122**, 2429–36.
15. Zhang, F., Pomerantz, J. H., Sen, G., Palermo, A. T., and Blau, H. M. (2007) Active tissue-specific DNA demethylation conferred by somatic cell nuclei in stable heterokaryons. *Proc Natl Acad Sci USA* **104**, 4395–400.
16. Brena, R. M., Auer, H., Kornacker, K., Hackanson, B., Raval, A., Byrd, J. C., and Plass, C. (2006) Accurate quantification of DNA methylation using combined bisulfite restriction analysis coupled with the Agilent 2100 Bioanalyzer platform. *Nucleic Acids Res* **34**, e17.
17. Estecio, M. R., Youssef, E. M., Rahal, P., Fukuyama, E. E., Gois-Filho, J. F., Maniglia, J. V., Goloni-Bertollo, E. M., Issa, J. P., and Tajara, E. H. (2006) LHX6 is a sensitive methylation marker in head and neck carcinomas. *Oncogene* **25**, 5018–26.
18. Yang, C. H., Chuang, L. Y., Cheng, Y. H., Gu, D. L., Chen, C. H., and Chang, H. W. (2010) Methyl-Typing: An improved and visualized COBRA software for epigenomic studies. *FEBS Lett* **584**, 739–44.
19. Bao, L., Zhou, M., and Cui, Y. (2008) CTCFBSDB: a CTCF-binding site database for characterization of vertebrate genomic insulators. *Nucleic Acids Res* **36**, D83–7.
20. Gaszner, M., and Felsenfeld, G. (2006) Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* **7**, 703–13.
21. Wakaguri, H., Yamashita, R., Suzuki, Y., Sugano, S., and Nakai, K. (2008) DBTSS: database of transcription start sites, progress report 2008. *Nucleic Acids Res* **36**, D97–101.
22. Takai, D., and Jones, P. A. (2003) The CpG island searcher: a new WWW resource. *In Silico Biol* **3**, 235–40.
23. Karolchik, D., Kuhn, R. M., Baertsch, R., Barber, G. P., Clawson, H., Diekhans, M., Giardine, B., Harte, R. A., Hinrichs, A. S., Hsu, F., Kober, K. M., Miller, W., Pedersen, J. S., Pohl, A., Raney, B. J., Rhead, B., Rosenbloom, K. R., Smith, K. E., Stanke, M., Thakkapallayil, A., Trumbower, H., Wang, T., Zweig, A. S., Haussler, D., and Kent, W. J. (2008) The UCSC Genome Browser Database: 2008 update. *Nucleic Acids Res* **36**, D773–9.
24. Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2007) REBASE--enzymes and genes for DNA restriction and modification. *Nucleic Acids Res* **35**, D269–70.
25. Chang, H. W., Cheng, Y. H., Chuang, L. Y., and Yang, C. H. (2010) SNP-RFLPing 2: an updated and integrated PCR-RFLP tool for SNP genotyping. *BMC Bioinformatics* **11**, 173.
26. Zweig, A. S., Karolchik, D., Kuhn, R. M., Haussler, D., and Kent, W. J. (2008) UCSC genome browser tutorial. *Genomics* **92**, 75–84.
27. Chang, H. W., Yang, C. H., Chang, P. L., Cheng, Y. H., and Chuang, L. Y. (2006) SNP-RFLPing: restriction enzyme mining for SNPs in genomes. *BMC Genomics* **7**, 30.
28. Grippo, P., Iaccarino, M., Parisi, E., and Scarano, E. (1968) Methylation of DNA in developing sea urchin embryos. *J Mol Biol* **36**, 195–208.
29. Doskocil, J., and Sorm, F. (1962) Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. *Biochim Biophys Acta* **55**, 953–9.
30. Finnegan, E. J., and Kovac, K. A. (2000) Plant DNA methyltransferases. *Plant Mol Biol* **43**, 189–201.
31. Bender, J. (2004) DNA methylation and epigenetics. *Annu Rev Plant Biol* **55**, 41–68.
32. Chan, S. W., Henderson, I. R., and Jacobsen, S. E. (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nat Rev Genet* **6**, 351–60.
33. Hetzl, J., Foerster, A. M., Raidl, G., and Mittelsten Scheid, O. (2007) CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J* **51**, 526–36.
34. Pontecorvo, G., De Felice, B., and Carfagna, M. (2000) Novel methylation at GpC dinucleotide in the fish Sparus aurata genome. *Mol Biol Rep* **27**, 225–30.
35. Kouidou, S., Malousi, A., and Maglaveras, N. (2006) Methylation and repeats in silent and nonsense mutations of p53. *Mutat Res* **599**, 167–77.
36. Rodenhiser, D., Chakraborty, P., Andrews, J., Ainsworth, P., Mancini, D., Lopes, E., and Singh, S. (1996) Heterogenous point mutations in the BRCA1 breast cancer susceptibility gene occur in high frequency at the site of homonucleotide tracts, short repeats and methylatable CpG/CpNpG motifs. *Oncogene* **12**, 2623–9.
37. Vatolin, S., Abdullaev, Z., Pack, S. D., Flanagan, P. T., Custer, M., Loukinov, D. I., Pugacheva, E., Hong, J. A., Morse, H., 3rd, Schrump, D. S., Risinger, J. I., Barrett, J. C., and Lobanenkov, V. V. (2005) Conditional expression of the CTCF-paralogous transcriptional factor BORIS in normal cells results in demethylation and

derepression of MAGE-A1 and reactivation of other cancer-testis genes. *Cancer Res* **65**, 7751–62.

38. Hong, J. A., Kang, Y., Abdullaev, Z., Flanagan, P. T., Pack, S. D., Fischette, M. R., Adnani, M. T., Loukinov, D. I., Vatolin, S., Risinger, J. I., Custer, M., Chen, G. A., Zhao, M., Nguyen, D. M., Barrett, J. C., Lobanenkov, V. V., and Schrump, D. S. (2005) Reciprocal binding of CTCF and BORIS to the NY-ESO-1 promoter coincides with derepression of this cancer-testis gene in lung cancer cells. *Cancer Res* **65**, 7763–74.
39. Klenova, E. M., Morse, H. C., 3rd, Ohlsson, R., and Lobanenkov, V. V. (2002) The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer. *Semin Cancer Biol* **12**, 399–414.

Chapter 7

SIRPH: An HPLC-Based SNuPE for Quantitative Methylation Measurement at Specific CpG Sites

Heike Singer, Nicole Nüsgen, and Osman El-Maarri

Abstract

Genome-wide sequence-specific methylation analysis has become a readily available and affordable procedure in epigenetic laboratories. Most of these procedures are based on immunoprecipitation, microarrays, or next generation deep bisulfite sequencing. However, most of these protocols are far from being quantitative. Moreover, abnormal or specific methylation patterns must always be further validated by quantitative sequence-specific methylation analysis. In this chapter, we describe a detailed and simplified protocol (using one universal HPLC gradient for all separations as well as one SNuPE annealing temperature for all primers) for the previously published SIRPH analysis, which is based on the single nucleotide primer extension combined with high-performance liquid chromatography. This method is highly accurate, reproducible, quantitative, and suitable for analysis of as little as 50 ng of PCR product derived from limited starting materials.

Key words: SIRPH, Bisulfite, CpG methylation, SNuPE, HPLC, Quantitative DNA methylation

1. Introduction

Methylation of cytosine residues is the major DNA modification in humans. It plays an essential and important role in the regulation of gene expression. In somatic tissues, methylated cytosines occur mainly in a CpG context, while in iPS (induced pluripotent stem cells) and embryonic stem cells, it occurs at non-CpG sites. Moreover, it has recently become apparent that 5-hydroxymethyl cytosine exists in different tissues (1–4).

In humans, altered/abnormal levels and patterns of DNA methylation have been observed in specific regions of the genome or genome-wide. These changes or variations in methylation levels are linked to diseases, aging, gender, environmental factors, diet, and DNA polymorphisms (5–10). Thus, accurate quantification of

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_7,

DNA methylation levels is used as surrogate marker to monitor health and disease status.

Available methods to analyze cytosine methylation can be divided into three categories: (1) genome-wide methylated cytosine content, (2) genome-wide sequence specific analysis, and (3) quantitative region-specific analysis. Regarding the first category, several techniques have been developed, such as digestion of genomic DNA to individual nucleosides and their separation by HPLC or capillary electrophoresis (CE) (11, 12). The latter approach was also further modified to greatly increase its sensitivity by fluorescent labeling of the digested product (13). Other genome-wide approaches include the LUMA technique, which is a pyrosequencing-based approach allowing quantification of methylation at only CCGG sequences (14). The second category includes the most advanced methods which have shown tremendous improvements in recent years. The most commonly used methods rely on immunoprecipitation protocols using an antibody reactive against the methylated DNA combined with microarray hybridization of the antibody precipitated DNA (15). A second powerful approach is based on the bisulfite deep sequencing approach, which is basically a "bisulfitome" genome resequencing approach (1, 2). The latter is still expensive and requires major equipment and considerable bioinformatics support.

The third category includes highly reproducible and accurate measurements of methylation (within 1–2% power of detection) at a limited small number of specific CpG sites. Although these provide only limited information, they still have the advantage of being highly quantitative and thus are indispensible in confirming the results of the genome-wide approaches and as molecular epigenetic diagnostic tool. Moreover, using region-specific approaches, a large number of samples may be analyzed and compared quantitatively at a reasonable cost. Examples of such methods are bisulfite-based pyrosequencing and the SIRPH protocol described below (16–18). It is important here to mention that the currently used bisulfite conversion of DNA cannot distinguish 5-methyl cytosine from 5-hydroxymethyl cytosine, as both cytosine modifications will equally prevent the corresponding cytosine from conversion to uracil. As a result, any downstream use of such bisulfite-converted product will fail to distinguish between these two modifications.

SIRPH stands for *S*NuPE combined with *I*on paired *R*everse *P*hase *H*PLC. It is based on the single nucleotide primer extension (SNuPE) (19) technique, where an oligo, just flanking the 5′ end of a CpG site, is extended by either ddCTP or ddTTP for methylated and unmethylated templates, respectively. It is used as a template PCR product of a bisulfite-treated DNA. The SNuPE product can be detected and measured by various methods. Gonzalgo and Jones were the first to apply such an approach for DNA methylation analysis by developing the methylation-sensitive SNuPE (Ms-SNuPE) that uses incorporation of radioactive nucleotides; the extended products are separated on acrylamide gels and quantified by autoradiography (20). Another method is the

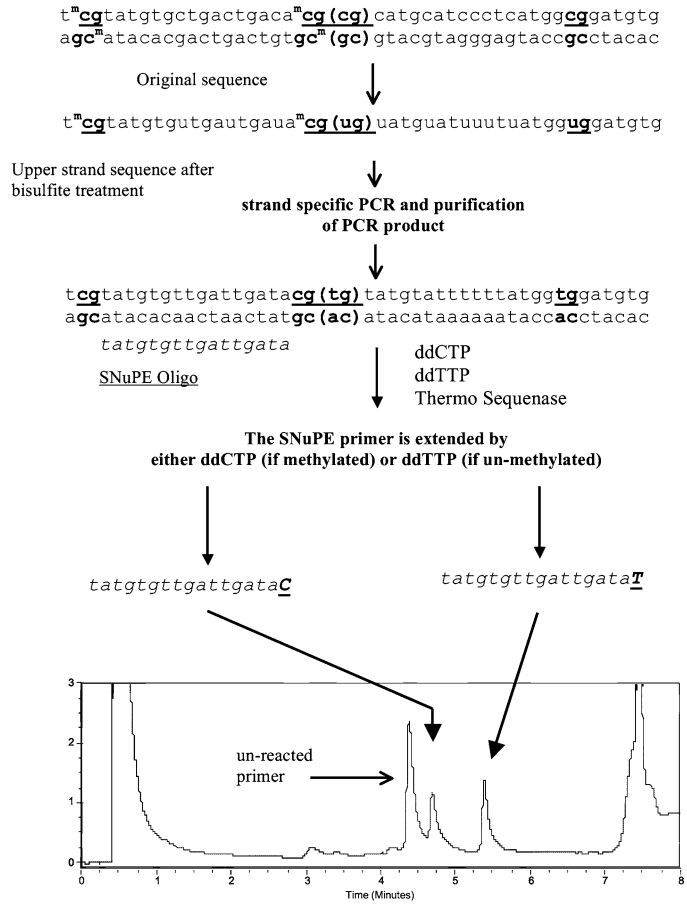
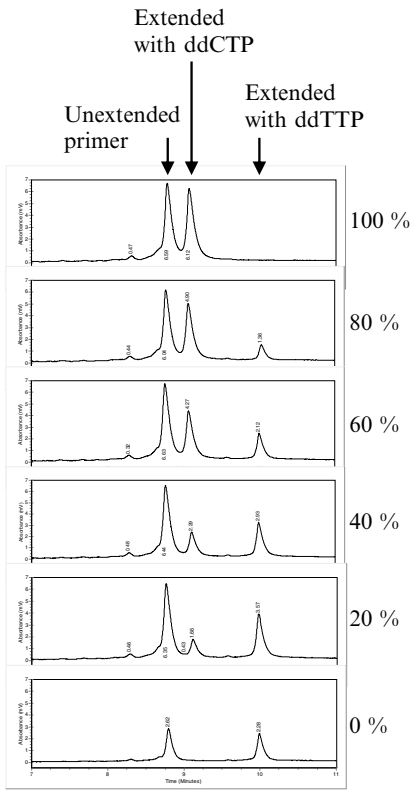
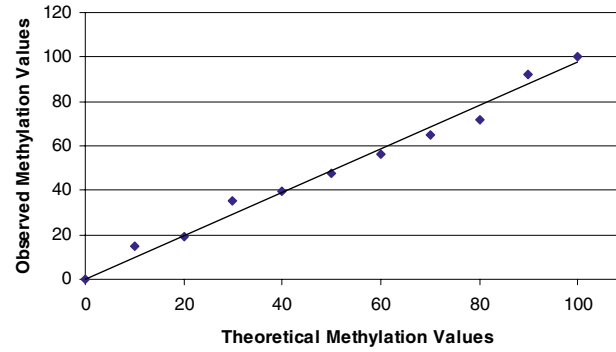


Fig. 1. General scheme of the SIRPH analysis (^{m}CG: methylated cytosine at a CpG site).

MALDI mass spectrometry-based separation method of the SNuPE reaction which is accurate but requires specially modified primers (21). Another application along the same principle is the SNaPshot from Applied Biosystems (ABI) which separates the product on a standard ABI sequencer (20). We (18) developed a cheap, nonradioactive variation of such a SNuPE protocol using ion pair reverse phase high-performance liquid chromatography. As a separation and detection method, this protocol provides an accurate quantification of methylation at specific selected CpG sites.

In this analysis, the unmodified PCR product to be analyzed is purified to remove residual PCR oligos and dNTPs. Subsequently, unmodified primers immediately 5′ to a CpG site are hybridized to the denatured single-stranded PCR product. The used primers are identical in sequence to the bisulfite-treated DNA strand that contains CpGs and/or TpGs (not GCs and/or ACs on the upper strand) and will thus hybridize to the lower strand that contains GCs and/or ACs. This will allow addition of ddCTP and/or ddTTP nucleotides at the 3′ position of the primer (Fig. 1). Alternatively, DNA methylation status can also be determined by





Oligo used: 5' taa ggt tag ttg tgt 3'
Regression analysis y = 0.9745x; R2 = 0.98

Fig. 2. Linearity of the reaction. A PCR product and a primer in the *SNRPN* gene was used with a serial mixes of methylated and umnethylated alleles in 10% increments (units in the *right graph* are % methylation).

incorporation of ddGTP and ddATP, whereby the upper strand of the PCR product serves as a template. However, the C and T extended primers are better separated on the DHPLC than the G and A extended primers. Temperature cycling using a thermo sequenase in the presence of both ddCTP and ddTTP extends the annealed primers. The ddTTP (for unmethylated CpG) or ddCTP (for methylated CpG, see Fig. 1) extended products are then directly loaded onto a DHPLC column (WAVE DNA Fragment Analysis System, Transgenomics). Due to the incorporation of the more hydrophobic ddTTP, the retention time of such an extended product is longer compared to products containing ddCTP (Fig. 1). The amount of the ddTTP and ddCTP extended products can then be quantified by measuring the height of the peaks and calculation of their percentage ratios. The reaction produces highly reproducible results, while maintaining linearity (Fig. 2).

It is also possible to analyze up to three different CpG sites simultaneously by multiplexing three different SNuPE primers. For multiplex analysis of more than one primer, individual primers

should be tested for their retention time and should be well separated so that the extension product of one primer does not overlap with the next. To shift retention time a shorter primer could be chosen (to reduce retention time) or alternatively, a variable number of Ts could be added at the 5′ end (to increase retention time). Combination and length of the primers has to be determined empirically. It is important to mention here that the multiplex approach provides more detailed information on the methylation patterns in a given region. This is particularly useful in regions that have low CpG density and heterogeneous methylation patterns, where one single CpG site may not represent the complete locus.

2. Materials

2.1. Removal of Excess Primers and dNTPs

1. For gel extraction or direct PCR purification (e.g., QIAquick Gel Extraction Kit from Qiagen or QIAquick PCR purification kit).
2. A mixture of exonuclease I and shrimp alkaline phosphatase which will degrade the unreacted primers and inactivate dNTPs (ExoFastAP from Fermentas: Exonuclease I; FastAP).

2.2. Single Nucleotide Primer Extension

1. ddTTP and ddCTP (Fermentas).
2. Thermo sequenase enzyme ["TERMIPol" (Solis Biodyne), or Thermosequenase (Amersham); we tried both enzymes and found them to show equal performance].
3. Standard unmodified oligos are used in the SNuPE reaction, free from n–x secondary products. Usually, HPLC purification is not needed (we routinely order High Purity Salt Free oligos from Eurofin, Germany).

2.3. Reverse Phase Ion Pair High-Performance Liquid Chromatography

For all dHPLC analyses we recommend use of the Wave system from Transgenomic together with the IP RP HPLC column, the DNASep (Transgenomic). The stationary phase in the column is made of alkylated nonporous polystyrene–divinylbenzene 2 μm bead particles (PS/DVB-C18).

1. Triethyl ammonium acetate (TEAA) buffer (20× ready to use buffer, Transgenomic; we found this source to be superior to other providers).
2. Acetonitrile (ROTH).
3. HPLC grade water (Merck).

3. Method

The SIRPH protocol can be divided into three parts: (1) generating the PCR product, (2) performing SNuPE reaction, and (3) separating the products on HPLC and quantification of the peaks. In this chapter, we describe the latter two, while bisulfite treatment is addressed in another chapter in this book.

3.1. Purification of PCR Product

PCR product can be purified by one of two methods (see Note 1):

1. Run the product on 1% agarose gel until optimum separation, excise the specific band, and recover the product by using a standard PCR-gel extraction kit (QIAquick Gel Extraction Kit from Qiagen yields highly pure products with high rates of recovery). Alternatively, in the absence of unspecific PCR product(s), PCR purification kit can be directly used without separation on agarose (QIAquick PCR purification kit).
2. Add 4 µl (for mixture composition see below) of exonuclease I and shrimp alkaline phosphatase (ExoFastAP, Fermentas) to 10 µl of PCR product, heat at 37°C for 15 min followed by 15 min at 85°C to deactivate the enzyme mixture.

   ExoFastAP-Mix (for 1 sample):

Exonuclease I	0.25 µl
FastAP	1 µl
FastAP buffer (10×)	0.4 µl
Distilled water	2.35 µl

3.2. SNuPE Reaction

1. Setup the SNuPE reaction in a total volume of 20 µl with the following components (see Note 2):

SNuPE oligo(s)	1 µl each oligo (12.5 pmol solution/oligo)
PCR product	1–5 µl (50–100 ng of 200–400 bp PCR product)
ddCTP	1 µl (1 mM solution)
ddTTP	1 µl (1 mM solution)
TERMIPol	0.2 µl (of 1 unit/µl stock)
Reaction buffer C	2 µl (10× buffer)
H_2O	up to 20 µl

For SNuPE primers used in the reaction see Notes 3–8.

2. The above mix is subject to the following thermocycles (see Note 9):

   Step 1: 94°C for 2 min

   Step 2: 92°C for 30 s

   Step 3: 40°C for 30 s

   Step 4: 52°C for 1 min

   Repeat steps 2–4 seventy times (since SNuPE reaction is a linear amplification, a relatively large number of cycles is needed to generate sufficient products of the extended primers to be detected on DHPLC).

3.3. Run the Products on HPLC

1. Load 20 μl of the above product directly on the HPLC machine (Wave, Transgenomics). Set oven temperature to 50°C (see Note 10) and elution gradient (mixture of buffers A and B) at 0.9 ml/min over 10 min:

Step	Time (min)	%A (0.1 M TEAA)	%B (0.1 M TEAA, 25% acetonitrile)
Loading	0.0	90	10
Start gradient	0.1	90	10
Stop gradient	10.0	50	50
Start clean	10.01	0	100
Stop clean	11.1	0	100
Start equilibrate	11.2	100	0
Stop equilibrate	12.2	100	0

   At the start gradient setting a mixture of buffer A and B (9:1 ratio) flows through the column. The amount of buffer B will steadily increase over a period of 10 min to reach 50% (buffer A and B ratio 1:1) and elute the DNA molecules from the stationary phase. In our experience, this universal gradient is adequate for separation of all oligos we have worked with.

2. Calculation of the percentage methylation:

   The percentage of methylated portion of the DNA can be calculated according to the formula: $M = (C/(C+T)) \times 100$, whereby C and T are the peak heights of the ddCTP and ddTTP extended oligos, respectively (see Notes 11–12). The WaveMaker software (Transgenomics) automatically calculates the C and T peak heights.

3.4. Multiplex Option

The above protocol lists all basic procedures to perform the SIRPH reaction at a single CpG site. However, the SIRPH reaction may also

be performed in multiplex. Several SNuPE primers can be included simultaneously in one reaction with each primer targeting different CpG sites even from multiplexed PCR product. In our experience, up to three primers can be easily multiplexed provided their retention time is adjusted empirically so that they do not overlap (see Note 7 for primer design). In this protocol, we included two simplified modifications that will make the multiplex reaction easier to perform: First, the SNuPE annealing temperature is set to 40°C regardless of the primer's melting temperature. Second, we recommend a wide HPLC gradient of 10–50% buffer B, by which all possible primers of up to 30 bp will be able to elute, without losing the power of separation of individual primers. Therefore, no optimization of gradient is needed for different sets of primer.

For general information about advantages and disadvantages of the SIRPH method see Note 13.

4. Notes

1. While enzymatic treatment may be more expensive (depending on reagents used), it has the advantage that it is rapid and easier to perform, especially when analyzing large numbers of samples (we are routinely using enzymatic-based cleaning in our laboratory). Gel extraction, on the other hand, is more labor-intensive but has the advantage of concentrating faint PCR product in smaller volume. It also offers the possibility of isolating the specific product in cases where there is more than one unspecific product.
2. The amount of template to be used is flexible with up to 1 μg to be used without affecting the quantification results. However, with less than 50 ng, the yield of SNuPE reaction may not be high enough to provide reproducible quantitative results. The primers used in the SNuPE reaction should always be in excess; their corresponding band (on HPLC separation) can be used as reference for the extended product(s) that should follow shortly after.
3. The 3′ end of the SNuPE oligo has to be just 5′ (flanking) of the specific CpG site to be studied.
4. Avoid placing the oligo on a T-rich region as this could increase mispriming and may lead to inaccuracy in the methylation measurements. However, the 3′ end (region) of the oligo should preferably be placed on a C to T (but not on an initial CpG) converted region so that the specificity to the bisulfite converted product is higher.
5. Oligos should not include a CpG site, as this will bias the linearity of measurements. If this cannot be avoided, due to short

sequence between CpG sites, a C/T (Y) base could be included in the primer. However, several controls with known methylation levels should be tested for each of these positions to assure linearity of measurements.

6. Oligos that are too short have a higher chance of mispriming. Oligos as short as ten bases can still produce accurate data. However, for routine uses we prefer oligos of 15–18 bases whenever possible.
7. For multiplex SNuPE reactions that are run simultaneously on the HPLC, the retention time of the individual oligos and their elongation products should be different. If, for practical reasons, two oligos have to be designed that give similar retention times on the HPLC, we recommend to extend one of the oligos by adding thymidins to its 5′ end. In our experience, this addition has no effect on the annealing to the template or on the SNuPE reaction. The number of Ts to be added has to be determined empirically. Each additional T has a stepwise additional retardation effect in a linear fashion.
8. All oligos have to be tested for self-annealing and self-extension in the absence of a template.
9. The annealing temperature used is 40°C. There is no detectable change either in the yield of the reaction or on the quantification results when using a range of 20°C (from 30 to 50°C) for annealing. Therefore, as a standard procedure for all oligos used in SNuPE reactions, 40°C annealing is used. At this temperature, the extension product of different oligos will vary slightly, depending on the melting temperature of the primer. However, in our experience, this still generates sufficient product for all oligos tested. Extension time of 1 min should allow a good reaction yield for most oligos. However, increasing the extension as well as the annealing times could give higher yields for some oligos. This must be tested individually.
10. An oven temperature of 50°C (compared to 60°C, 70°C, and 80°C) was found to give the highest difference in retention time between the ddCTP and the ddTTP extended oligos.
11. If the unreacted primer – after HPLC separation – is close to the ddCTP extended primer, the integration of the area under the ddCTP extended primer curve may not be accurate. Therefore, it is more accurate to use the peak height for measurements. The gradient used has great influence on both, the separation efficiency and the accuracy of quantification. A slope of at least 2% increase of buffer B over 1 min gives a good separation.
12. Reproducibility of the HPLC measurements: It is recommended to test the accuracy of reproducibility of the measurements on the HPLC machine. This can be done by injecting

the same SNuPE product several times and calculating the standard deviation for each sample. For most oligos and at different ratios of ddCTP to ddTTP, the standard deviation ranges from 1 to 3%.

13. General limitations and advantages of this assay:

    Limitations:

    (a) SIRPH is able to analyze only few CpG sites at a time.

    (b) Not all CpGs can be analyzed, as a SNuPE primer due to the short sequence between the CpG of interest and the one preceding it cannot easily target CpGs in CpG dense regions.

    (c) Low throughput of analysis: an HPLC system is able to analyze a single sample at time, and each sample run requires about 15 min, only a single 96 plate can be analyzed per day.

    (d) The power of separation (resolution) of the extended SNuPE primers depends on the quality of the HPLC column. Thus, the relative methylation value could vary between a new HPLC column and a relatively older one. Therefore, quality of the HPLC column must be monitored regularly. This can be done simply by dividing one control SNuPE (exactly the same) product into two parts to run at position 1 and 96 of a given plate. The methylation values should be essentially the same. If a considerable deviation between the two samples exists (more than 1%) correction for this bias by applying a linear regression correction model should be considered.

    Advantages:

    (a) The assay is highly reproducible and can accurately discriminate methylation differences of about 1%.

    (b) It can analyze an amount of PCR product as small as 50 ng. This is mainly possible because no loss of product accompanies the enzymatic degradation of unused PCR primers and dNTPs. This is in contrast to the pyrosequencing approach which requires a relatively strong PCR product due to the fact that purification and generation of a single PCR product will result in considerable loss of the initial input product.

    (c) Methylation at different CpG sites is measured independently of each other. This has the advantage that a given polymorphism between two sites will not affect the following CpG measurement and no sequence input is needed for the analysis (in contrast to the pyrosequencing approach which is based on the principle of reading by synthesis, requiring input of a specific sequence pre-entered by the

user; thus, the reading is bound to the specific input sequence). Therefore, SIRPH is particularly accurate in the analysis of repeat sequences, such as Alu and LINE-1 that are very polymorphic in their sequences.

References

1. Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A.H., Thomson, J.A., Ren, B., Ecker, J.R. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature.* **462**: 315–22.
2. Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Kin Sung, K.W., Rigoutsos, I., Loring, J., Wei, C.L. (2010) Dynamic changes in the human methylome during differentiation. *Genome Res.* **20**:320–31.
3. Kriaucionis, S., Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science.* 324:929–930.
4. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F., Leonhardt, H. (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**:e181.
5. El-Maarri, O., Seoud, M., Coullin, P., Herbiniaux, U., Oldenburg, J., Rouleau, G., Slim, R. (2003) Maternal alleles acquiring paternal methylation patterns in biparental complete hydatidiform moles. *Hum Mol Genet.* **12**:1405–13.
6. El-Maarri, O., Becker, T., Junen, J., Manzoor, S.S., Diaz-Lacava, A., Schwaab, R., Wienker, T., Oldenburg, J. (2007) Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males. *Hum Genet.* **122**:505–14.
7. El-Maarri, O., Kareta, M.S., Mikeska, T., Becker, T., Diaz-Lacava, A., Junen, J., Nüsgen, N., Behne, F., Wienker, T., Waha, A., Oldenburg, J., Chédin, F. (2009) A systematic search for DNA methyltransferase polymorphisms reveals a rare DNMT3L variant associated with subtelomeric hypomethylation. *Hum Mol Genet.* **18**:1755–68.
8. Fuke, C., Shimabukuro, M., Petronis, A., Sugimoto, J., Oda, T., Miura, K., Miyazaki, T., Ogura, C., Okazaki, Y., Jinno, Y. (2004) Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Ann Hum Genet.* **68**:196–204.
9. Bollati, V., Baccarelli, A., Hou, L., Bonzini, M., Fustinoni, S., Cavallo, D., Byun, H.M., Jiang, J., Marinelli, B., Pesatori, A.C., Bertazzi, P.A., Yang, A.S. (2007) Changes in DNA methylation patterns in subjects exposed to low-dose benzene. *Cancer Res.* **67**:876–80.
10. Wolff, G.L., Kodell, R.L., Moore, S.R., Cooney, C.A. (1998) Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. *FASEB J.* **12**:949–57.
11. Ramsahoye, B.H. (2002) Measurement of genome wide DNA methylation by reversed-phase high-performance liquid chromatography. *Methods.* **27**:156–61.
12. Fraga, M.F., Uriol, E., Borja, Diego, L., Berdasco, M., Esteller, M., Cañal, M.J., Rodríguez, R. (2002) High-performance capillary electrophoretic method for the quantification of 5-methyl 2′-deoxycytidine in genomic DNA: application to plant, animal and human cancer tissues. *Electrophoresis.* **23**:1677–81.
13. Stach, D., Schmitz, O.J., Stilgenbauer, S., Benner, A., Döhner, H., Wiessler, M., Lyko, F. (2003) Capillary electrophoretic analysis of genomic DNA methylation levels. *Nucleic Acids Res.* **31**:e2.
14. Karimi, M., Johansson, S., Stach, D., Corcoran, M., Grandér, D., Schalling, M., Bakalkin, G., Lyko, F., Larsson, C., Ekström. (2006) LUMA (LUminometric Methylation Assay)–a high throughput method to the analysis of genomic DNA methylation. *TJ. Exp Cell Res.* **312**: 1989–95.
15. Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., Schübeler, D. (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet.* **37**:853–62.
16. Tost, J., Dunker, J., Gut, I.G. (2003) Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques.* **35**:152–6.
17. El-Maarri, O., Herbiniaux, U., Walter, J., Oldenburg, J. (2002) A rapid, quantitative, non-radioactive bisulfite-SNuPE- IP RP HPLC assay for methylation analysis at specific CpG sites. *Nucleic Acids Res.* **30**:e25.

18. El-Maarri, O. (2004) SIRPH analysis: SNuPE with IP-RP-HPLC for quantitative measurements of DNA methylation at specific CpG sites. *Methods Mol Biol.* **287**: 195–205.
19. Syvanen, A.C. (1999) From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum Mutat.* **13**:1–10.
20. Gonzalgo, M.L., Jones, P.A. (1997) Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.* **25**:2529–2531.
21. Tost, J., Schatz, P., Schuster, M., Berlin, K. Gut, I.G. (2003) Analysis and accurate quntification by MALDI mass spectrometry. *Nucleic Acids Res.* **31**:e50.

Chapter 8

Restriction Landmark Genome Scanning

Hisato Okuizumi, Tomoko Takamiya, Yasushi Okazaki, and Yoshihide Hayashizaki

Abstract

Restriction landmark genome scanning (RLGS) method is a high-resolution two-dimensional electrophoresis system for analyses of the whole genome DNA which is including methylation status. It has been used for cloning genes of model animals and human genomes, detection of imprinted genes, and genome-wide methylation research in cancer. The conventional RLGS detected both polymorphism and methylated *Not*I sites between samples. Here, we have developed improved RLGS method with isoschizomer restriction enzymes such as *Msp*I and *Hpa*II to specifically detect methylated sites, using differential sensitivity of the restriction enzymes to methylated sequences. Recently, by using the genome database information, the RLGS spot sites were efficiently identified by this improved method. Then, genome methylation sites of *Arabidopsis* were mapped, and a unique inheritance was detected in methylated gene in rice. Now, epigenetic research becomes easy with the improved RLGS and it also can be applied for animal genome. Therefore, RLGS method is useful to explore for novel epigenetic phenomenon.

Key words: Restriction landmark genome scanning, First dimensional electrophoresis, Second dimensional electrophoresis, Isoschizomer, Genome methylation, *in silico* RLGS

1. Introduction

Genomic DNA is not only restricted with genetic information but also contains epigenetic information such as DNA methylation (cytosine methylation). Methylated cytosine which is very common in plant and mammalian genomes plays a major role in the regulation of gene expression. In mammals, methylation patterns change dramatically during gametogenesis and early development (1, 2). In contrast in plants, generational changes in methylation status and inheritance in plants have been unclear. Though, in some genes the methylation statuses were found to be stably inherited (3–5), recent studies show that methylation patterns were

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_8,

altered in first filial generation (F_1) hybrids, and in interspecific hybrids (6–14). Imprinted genes that have sex-dependent methylated cytosine patterns have been identified in endosperm. And they play an important role in the control of flowering or seed formation (15–24). The precise analysis of DNA methylation inheritance may help to identify new imprinted genes and to clarify the biological significance of DNA methylation.

Several methods such as Southern blotting (e.g., ref. 25), PCR (e.g., ref. 26), tiling microarrays (e.g., ref. 27), restriction landmark genome scanning (RLGS) (e.g., refs. 28, 29), and methylation-sensitive amplification polymorphism (e.g., refs. 30, 31) have been commissioned for methylation analysis. Southern blotting and PCR are gene-by-gene type methods analyzing a specific target region. The most comprehensive tools are tiling microarrays, but they are not suitable for exploratory studies, because of their long set-up time and high cost. Among those three genome-wide methods, the methylation-sensitive amplification polymorphism is the most easily applied one. But, the methylation level is not reflected directly in its results, because it uses some PCR amplification steps.

The RLGS method employs high molecular weight DNA that is carefully isolated and their occasional breaks are filled in enzymatically. Then DNA is digested with restriction enzymes, including a methylation-sensitive enzyme and they are radioactively labeled. The so obtained fragments are then separated on subsequent agarose and acrylamide gels and autoradiographed, generating a very reproducible pattern where, missing spots correspond to specific sites methylated in the genome of the analyzed sample. This RLGS is cost effective and yields results in only 3 days, although RLGS analyzes a relatively limited region of the genome than tiling microarrays. In addition, the RLGS spot intensity provides direct quantitation of methylation level and can also detect partial methylation such as imprinting. Consequently, RLGS is appropriate for DNA methylation surveys that broadly sample a genome. The utility of RLGS has been demonstrated in the construction of genetic linkage maps (32, 33), methylation analysis in tumor tissue (34, 35), and identification of imprinted genes in mammals (36–38).

On the other hand, RLGS mainly employs *Not*I (28, 39) that is methylation-sensitive restriction enzyme. The RLGS is also suitable to analyze thousands of methylated sites near genes, because *Not*I sites are frequent in CpG islands (40). However, because it needs to compare the RLGS patterns just between an objective sample and a control, it has been difficult to distinguish DNA polymorphism caused by methylation from sequence polymorphism in conventional RLGS. Then, we proposed an improved RLGS for effective epigenomic methodology. It includes (1) easy identification of methylated sites and their locations in a genome,

(2) efficient methylation survey of coding regions, and (3) applicability to scanning methylation status or stability of transposable elements and centromeres. In order to cover these points, we utilized isoschizomers (*Msp*I and *Hpa*II) to detect methylated sites directly (41). Isoschizomers are restriction enzymes that have the same recognition sequence such as *Msp*I and *Hpa*II. We employed these enzymes because they have different methylation sensitivities. When the second C of CCGG site is methylated ($C^{5m}CGG$), *Msp*I digests the sequence but *Hpa*II does not. Consequently, it is possible to detect many methylated sites in an individual without comparison of objective sample and control, in taking an advantage of isoschizomer in RLGS. This method can also distinguish methylation changes from mutation in an unsequenced species and even in cloned animals or ramets of plants (42).

Here, we describe the protocol for the results in *Arabidopsis* (41), and methylation map with evidence of non-Mendelian inheritance in rice (43, 44) for instance.

2. Materials

2.1. Plant Materials (Arabidopsis and Rice)

Arabidopsis thaliana L. contains rich variations in hundreds of ecotypes that were adapted to particular environmental conditions. Columbia (Col), one of the ecotypes was grown on the pots (peat moss:vermiculite = 1:1) at room temperature for 4 weeks, with 8 h a day of photoperiod in the growth chamber, and above ground parts were collected. Rice F_1 hybrid was developed from the cross between *Oryza sativa* L. subsp. *japonica* "Nipponbare" (pollen parent) and *Oryza sativa* L. subsp. *indica* "Kasalath" (seed parent). F_1 and the parents were grown for 8 weeks from which the leaf blades and sheaths were collected for genomic DNA extraction.

2.2. Genomic DNA

Total genomic DNA was isolated by modified CTAB (43, 45, 46). The plant tissues were placed into a mortar along with required amount of liquid nitrogen and ground them into fine powder. The powdered tissue (0.2 mg) was suspended in 0.8 mL CTAB extraction buffer (1% CTAB, 0.1 M Tris–HCl, pH 8.0, 50 mM EDTA, 1.4 M NaCl, 0.1% β-mercaptoethanol) along with 20 mg SDS and 20 mg insoluble PVP. Two microliters of proteinase K (Merck, Darmstadt, Germany) solution (40 mg/mL) was added to the lysate. The lysate was incubated at 56°C for 30 min followed by extraction with phenol and chloroform. Then the nucleic acid was precipitated by adding equal volume of 2-propanol.

2.3. Specific Reagent for RLGS

10× 1D buffer (for 1DE; 1 M Tris, 0.4 M sodium acetate trihydrate, 0.35 M NaCl, 35 mM $EDTANa_2$ in H_2O, pH to 8.15 with acetic acid with autoclave. The pH of the 1× buffer is 8).

2.4. Electrophoretic Apparatuses

All electrophoretic apparatuses are manufactured by Bio Craft (see Note 1).

1. 1DE tank (including tank anodal top and cathodal bottom).
2. 1D gel holder (including Teflon tubing).
3. 2DE tank.
4. 2DE plate.

2.5. Autoradiography

1. Hydro Tech Gel Drying System™ (Model 583 gel dryer and Hydro Tech vacuum pump) (Bio-Rad, Hercules, USA).
2. BAS 2000™ (FUJIFILM, Tokyo, Japan) (see Note 2); http://www.fujifilm.jp/index.html.

3. Methods

After extraction of genome DNA, following procedures are described such as (1) operation of DNA, (2) first and second dimensional electrophoresis (1DE and 2DE), (3) gel drying and autoradiography, (4) detection of methylated polymorphism, (5) identification of RLGS spots by *in silico* RLGS and spot cloning, (6) confirmation of restriction enzyme sites, and (7) confirmation of methylation status.

3.1. Manipulation of DNA

1. Blocking nicks and gaps: Genomic DNA (3.5–10.5 μg) is treated with 1.75 U of DNA polymerase I (Nippon Gene, Tokyo, Japan) in 10 μL of blocking buffer (50 mM Tris–Cl, pH 7.4, 10 mM $MgCl_2$, 100 mM NaCl, 10 mM DTT, 0.4 mM dGTP[α]S, 0.2 mM dCTP[α]S, 0.4 mM ddATP, and 0.4 mM ddTTP) at 37°C for 20 min and 65°C for 30 min (see Notes 3 and 4).
2. *Not*I Digestion: 20 U of *Not*I (NEB, Beverly, MA, USA) is used to digest the genomic DNA in 20 μL of SH buffer (150 mM NaCl, 0.01% Triton-X-100, 0.01% BSA) (see Note 4).
3. Labeling of *Not*I sites: The cleavage ends are filled in and labeled with ^{32}P in the presence of 1.3 U of Sequenase ver. 2.0™ (USB, Cleveland, OH, USA), 0.33 mM [α-^{32}P] dGTP (3,000 Ci/mmol), 0.33 mM [α-^{32}P] dCTP (6,000 Ci/mmol), and 13 mM DTT at 37°C for 30 min in 20.4 mL. The inactivation of enzyme is done by heating the reaction mixture at 65°C.
4. Second isoschizomer (*Msp*I or *Hpa*II) digestion: The *Not*I digested and labeled DNA is made into two samples in which one tube is digested with 25 U of *Msp*I at 37°C for 1 h and the other one is digested with *Hpa*II separately.

3.2. First and Second Dimensional Electrophoresis

1. On-loading DNA: The genomic DNA that is digested by isoschizomer restriction enzymes samples are separated in agarose disc gel (0.8% Seakem GTG™ agarose, FMC Bioproducts, Rockland, MA, USA) with a dimension of 2.4 mm diameter × 63 cm long tube (see Note 5).
2. 1DE:1D electrophoresis is run at 100 V for 2 h followed by 230 V for 22 h in the 1D buffer (0.1 M Tris–acetate, pH 8.0, 40 mM sodium acetate, 3 mM EDTA, pH 8.0, 36 mM NaCl) (see Note 6).
3. Extruding the gel: The genomic DNA which are separated (in noodle like 1D gel) based on their size was extruded from the tube and soaked for 30 min in the High-buffer (50 mM Tris–Cl, pH 7.4, 10 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT). Thereafter, DNA is digested in the gel with 1,500 U of *Bam*HI for 2 h (see Note 7).
4. 2D gel casting: The gel is fused into the top edge of a 5% polyacrylamide vertical gel, 50 cm (W) × 50 cm (H) × 0.1 cm (thickness) by adding melted agarose (0.8%) at 65°C to connect 1D and 2D gels (see Note 8).
5. 2DE: The 2D electrophoresis is run in Tris–Borate–EDTA (TBE) buffer (50 mM Tris, 62 mM boric acid, 1 mM EDTA), at 100 V for 1 h followed by 150 V for 23 h.

3.3. Gel Drying and Autoradiography

Gel is excised and dried for autoradiography (see Note 9). Autoradiography is carried out for 1–3 days on an imaging plate (FUJIFILM). Then, the imaging plate is analyzed by the BAS-2000™ (FUJIFILM) for printing the spot pattern. An example of the results produced is shown in Fig. 1a, b (41).

3.4. Detection of Methylated Polymorphism

To detect the methylated polymorphic spots, the transparent copies (on OHP sheets) of original printed RLGS patterns are placed over and the spot pattern is compared with each other (see Note 10).

3.5. Identification of RLGS Spots by in silico RLGS and Spot Cloning

We developed the method to map the RLGS spot loci on the chromosome, using information of genome sequence (*in silico* RLGS) (41). The software, "*in silico* RLGS," is originally named by us during the development process and now spread to other researchers (47). "*In silico* RLGS" simulates RLGS spots to identify each spot in actual RLGS. The software searches for restriction enzyme sites that are used in an actual RLGS experiment such as *Not*I, *Msp*I, and *Bam*HI through the whole genome sequence. Then, it calculates fragment length between sites and mobility for simulating a 2D spot pattern. In these days, RLGS in *Arabidopsis* and rice benefits from the availability of the whole genome sequence (48), enabling *in silico* RLGS analysis effectively.

1. Identification of RLGS spots by *in silico* RLGS: At first, the whole genome sequence data such as *Arabidopsis* and rice at NCBI (http://www.ncbi.nlm.nih.gov/) should be acquired.

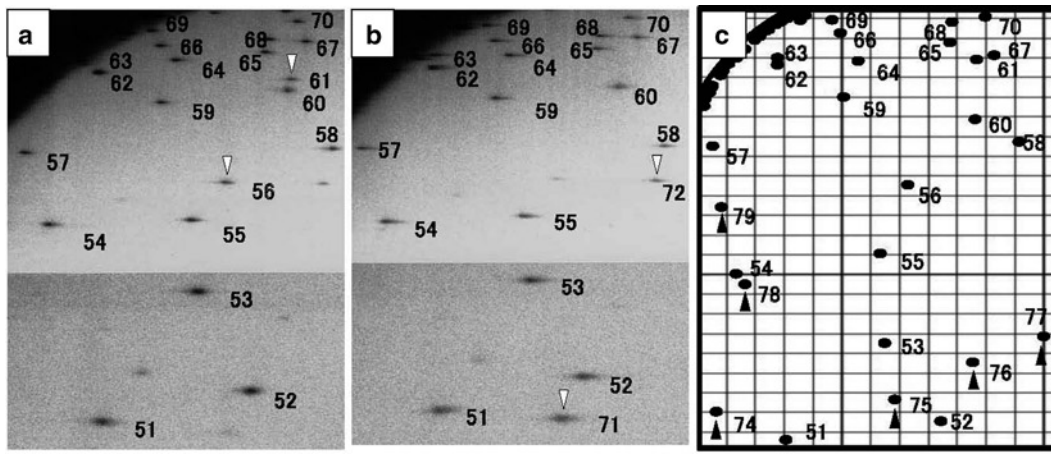


Fig. 1. RLGS spot pattern with *Msp*I or *Hpa*II, detecting methylated sites. (**a**) RLGS pattern of *Arabidopsis thaliana* (Col) with the combination of restriction enzymes, *Not*I–*Msp*I–*Hin*dIII. Twenty distinct spots were detected. (**b**) *Not*I–*Hpa*II–*Hin*dIII combination pattern. There were also 20 spots. In comparison of (**a**) and (**b**), four spots were different in both patterns. Spots 56 and 61 (*open arrowheads*) were detected specifically in (**a**). On the other hand, the spots 71 and 72 (*open arrowheads*) were specific in (**b**). These differences indicate the methylated CpG ($C^{5m}CGG$) in *Msp*I/*Hpa*II site. They are called methylated spot. (**c**) *In silico* RLGS pattern with the combination of *Not*I–*Msp*I–*Hin*dIII. In the pattern, 26 spots were detected in the same range of (**a**) and (**b**). Twenty identical spots were obtained, and six spots were specific in the *in silico* RLGS pattern (reproduced from ref. 41 with permission from WILEY-VCH Verlag GmbH & Co. KGaA).

*Not*I sites are searched through the whole sequence data and given ID numbers. *Msp*I sites that are the nearest to *Not*I sites are also obtained. Length and mobility of each DNA fragment (from *Not*I end to *Msp*I end or from *Not*I end to next *Not*I end) in the 1DE are calculated by the software. The mobility of each DNA fragment in the gel is calculated based on Southern (49). The exact mobility of RLGS spots have been already reconfirmed from electrophoresis of λ DNA fragments as known sequences. The *Bam*HI sites are also obtained and the DNA fragments (from *Not*I end to *Bam*HI end) length of 2D and their mobility on 2DE are calculated, and a 2D graph as "*in silico* RLGS pattern" is drawn. The *in silico* RLGS pattern with its corresponding autoradiographic RLGS pattern with relative spot positions are compared and over half of RLGS spots are identified. An example of the results produced is shown in Fig. 1 (41).

2. Spot cloning: Spots unidentified by *in silico* RLGS analysis can be cloned and sequenced as described previously (41) with specific cloning linkers: *Not*I linker (5′-GGCCGCATGAATG GCGCGCCAAAGA-3′, 3′-CGTACTTACCGCGCGGTTT CT-Biotin-5′) and *Bam*HI linker (5′-GATCCTGTACTGCAC CAGCAAATCC-3′,3′-GACATGACGTGGTCGTTTAGG-5′).

3.6. Confirmation of Restriction Enzyme Sites

First of all, the presence of restriction enzyme sites in the parents should be confirmed in order to compare methylation status among the parents and F_1s. The flanking primers for the *Not*I and *Msp*I/*Hpa*II sites of each RLGS spot are designed. By using 1 ng

of each genomic DNA as a template, PCR is performed with 0.4 U of KOD plus polymerase™ (Toyobo, Tokyo, Japan), 1.5 pmol of flanking primers, 1 mM $MgSO_4$, 0.2 mM dNTPs, and KOD buffer (total volume: 20 μL). PCR conditions were 94°C for 5 min followed by 30 cycles of 94°C for 15 s, 60°C for 30 s, and 68°C for 1 min. Each PCR product is digested by *Not*I or *Msp*I. Then undigested and digested products are electrophoresed in an agarose gel (0.8–3.0%), and the band sizes are compared to reconfirm that the sites are present and does not differ by DNA size polymorphism. Then the methylation status of these sites is analyzed.

3.7. Confirmation of Methylation Status

1. PCR-based method: To confirm the methylation status of *Not*I and *Msp*I/*Hpa*II sites of RLGS spots, template DNA is prepared by digesting the genomic DNA with 30 U of *Not*I, *Msp*I, or *Hpa*II. Undigested genomic DNA is used as a positive control. PCR is performed as described in Subheading 3.6. An example of the results produced is shown in Fig. 2 (43).

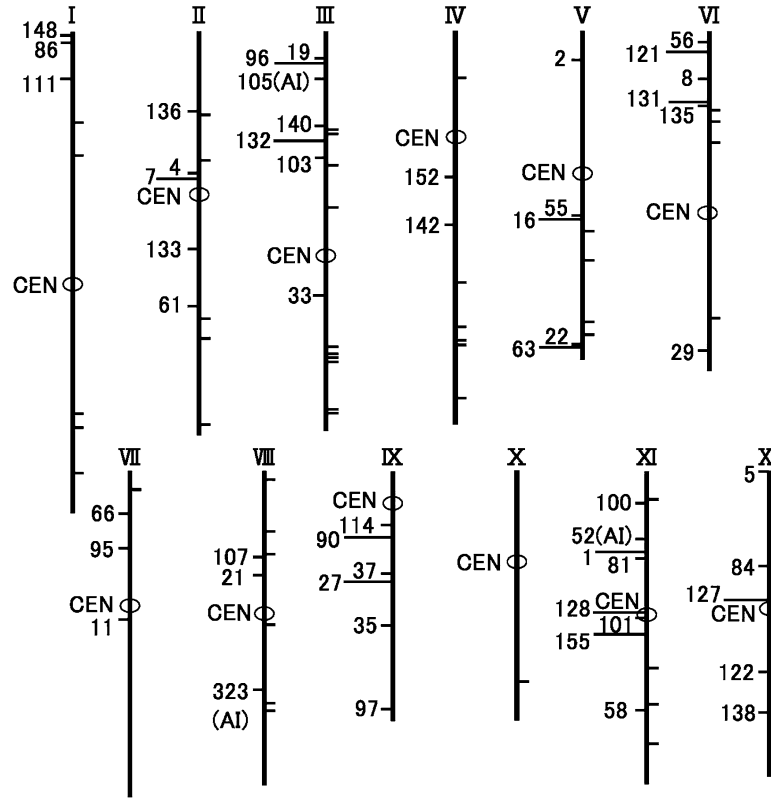


Fig. 2. Map of methylated sites. Numbered loci on the *left horizontal line* had at least one methylation in *Not*I or *Msp*I/*Hpa*II digests of Nipponbare, Kasalath, or F_1s. "AI" indicates three loci with altered inheritance. Inheritance at the other loci appeared to be consistent. The *right horizontal lines* and *ellipses* (CEN) indicate the identified unmethylated RLGS spot loci and centromeres, respectively (reproduced from ref. 43 with permission from WILEY-VCH Verlag GmbH & Co. KGaA).

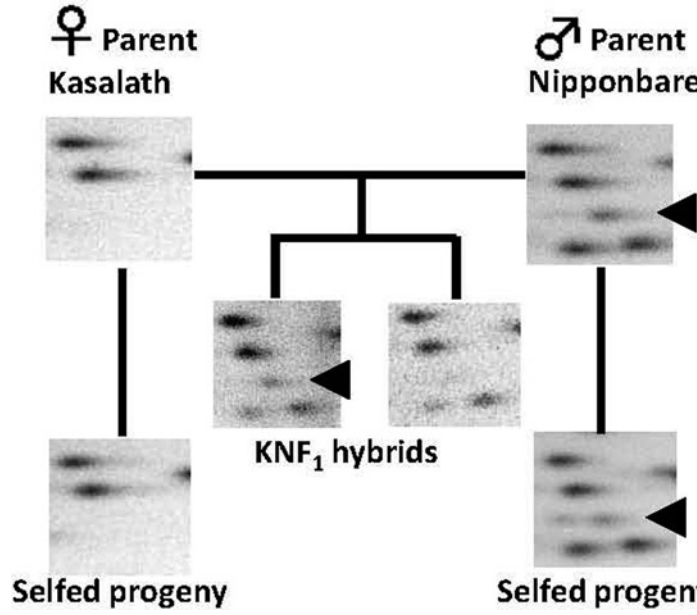


Fig. 3. RLGS [*Msp*I] (*Not*I–*Msp*I–*Bam*HI) combination patterns of the parents, their selfed progeny, and their F_1 hybrid. Spot 200 (*arrowhead*) was detected in the [*Msp*I] patterns and [*Hpa*II] (*Not*I–*Hpa*II–*Bam*HI) patterns of Nipponbare and its selfed progeny. The presence or absence of the spot segregated in F1 population (KNF_1). The spot intensity of this spot was half that of the others (reproduced from ref. 44 with permission from Hindawi Publishing Corporation).

2. Sequence-based method: A sequence-based method is used for analysis of some methylated *Not*I or *Hpa*II sites in the F_1 in rice for instance, to determine which parent the methylation is inherited from. An example of the results produced is shown in Fig. 3 (44). With flanking primers, PCR products are obtained by amplification of genomic DNA of the parents. The products are sequenced directly with the Applied Biosystems 3130 Genetic Analyzer™ (Applied Biosystems, Foster City, CA, USA) and a BigDye Terminator v3.1 Cycle Sequencing Kit™ (Applied Biosystems) to detect single nucleotide polymorphisms (SNPs) between the parents. Then, genomic DNA of F_1s is digested with *Not*I or *Hpa*II to create templates. Using the SNPs to determine the parental origin, the fragments with methylated *Not*I or *Hpa*II sites which were not digested could be sequenced.

4. Notes

1. BIO CRAFT Co. Ltd. (Fuji building 3F, 2-14-9 Itabashi, Itabashi-ku, Tokyo, 173-0004, Japan, Tel: +81-3-3964-6561, Fax: +81-3-3694-6443, HP: http://www.bio-craft.co.jp/).

2. FUJIFILM Co. Ltd. (Midtown West, 7-3, Akasaka 9-chome, Minato-ku, Tokyo 107-0052, Japan, Tel: +81-3-6271-3111 HP: http://fujifilm.jp/) and GE Healthcare UK Ltd. (Amersham Place, Little Chalfont, Buckinghamshir HP7 9NA, England, Tel: 0870-606-1921, Fax: 01494-544350, HP: http://www.gelifesciences.com/APTRIX/upp01077.nsf/Content/default_homepage).
3. If you cannot have nucleotide analogue with [α]S, blocking step can be performed without [α]S, but with normal nucleotide analogue.
4. Mix gently but completely by pipetting, because of viscous high molecular sized genome DNA in the blocking and *Not*I digestion steps.
5. When melting 1D gel using microwave, check out the agarose powder is dissolved completely and avoid lumps. Well-melted 0.8% SeaKem GTG™ agarose gel solution in 1D buffer is kept at 55–60°C. The gel solution is sacked up into the 1D gel folder to reach 1 cm below the top. Keep it stable for 15 min to be solidified the gel, and set the gel folder into the 1DE tank.
6. The loading amount of DNA should be 1–1.5 μg in mice genome and up to 1 μg in *Arabidopsis* and rice. In case you handle smaller genome size organism, lesser amount of DNA should be applied into 1DE.
7. Third restriction enzyme of in-gel digestion requires 1,500 U. The reasonable cost is expected in use of *Bam*HI, *Hin*dIII, and *Eco*RI.
8. Note that the no air bubble should be included in the connecting gel at 65°C.
9. To detach 2D gel from glass completely with no trouble, it is necessary to coat Sigmacote™ (Sigma-Aldrich Inc., St. Louis, MO, USA) on the glass surface. In the summer, Sigmacote is often degraded, and then 2D electrophoresis pattern is often distracted. To overcome this problem, Sigmacote on the glass surface should be removed by soaking the whole glass in 10N NaOH for overnight, washing out completely, and spreading Sigmacote again.
10. To detect polymorphisms, use nonpolymorphic spots as anchor points for comparison of RLGS patterns on the OHP sheets.

5. Conclusion

The efficient scanning method of methylated sites is now applicable to Epigenetics by establishing improved RLGS using isoschizomers (*Msp*I and *Hpa*II). The primary technology for the entire

method was developed by us and we also established the total analysis procedures for practical epigenetic studies in *Arabidopsis* and rice. The bottle neck of obtaining a clear pattern of RLGS in plant genome was because of its difficulty of extracting intact high molecule genomic DNA from plant tissue. Consequently, the techniques of RLGS method applicable in plant genome research can be easily employed in the animal genome research.

Until now, *Not*I was commonly used for landmark and only *Msp*I and *Hpa*II have been used as an isoschizomer in RLGS, however, any other landmark restriction enzymes and isoschizomers can be used. Nevertheless, it has been successful to obtain an evidence of unique epigenetic inheritance in plants. On this point, it enables wider genome scanning for the studies on the gene expression and genome stability (41). This method is also expected to detect novel imprinted genes, even though a few imprinted genes were cloned using conventional RLGS. Therefore, we emphasize that the improved RLGS method has the potential to reveal more various epigenetic phenomena.

Acknowledgments

We thank Paterson, A. H., Nishiguchi, M., Kakutani, T., Mizuno, Y., Yamashita, H., Nakamura, E., Asai, K., Higo, K., Hirochika H., Kawase, M., Tomioka, K., Habu, Y., Ohtake, Y., Ueda, T., Murakami, Y., Iijima, H., Takahashi, S., Nonaka, E., Nakamura, M., Hosobuchi, S., Noguchi, T., Saguchi, T., Fujita, T., and Seetharam, K., for their support.

References

1. Monk, M., Boubelik, M., and Lehnert, S. (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371–382.
2. Tada, S., Tada, T., Lefebvre, L., Barton, S. C., and Surani, M. A. (1997) Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *The EMBO Journal* **16**, 6510–6520.
3. Bender, J. and Fink, G. R. (1995) Epigenetic control of an endogenous gene family is revealed by a novel blue fluorescent mutant of Arabidopsis. *Cell* **83**, 725–734.
4. Jacobsen, S. E. and Meyerowitz, E. M. (1997) Hypermethylated SUPERMAN epigenetic alleles in Arabidopsis. *Science* **277**, 1100–1103.
5. Kakutani, T., Munakata, K., Richards, E. J., and Hirochika, H. (1999) Meiotically and mitotically stable inheritance of DNA hypomethylation induced by ddm1 mutation of *Arabidopsis thaliana*. *Genetics* **151**, 831–838.
6. Matzke, M. A., Mittelsten Scheid, O., and Matzke, A. J. (1999) Rapid structural and epigenetic changes in polyploid and aneuploid genomes. *Bioessays* 21, 761–767.
7. Wendel, J. F. (2000) Genome evolution in polyploids. *Plant Mol. Biol.* **42**, 225–249.
8. Shaked, H., Kashkush, K., Ozkan, H., Feldman, M., and Levy, A. A. (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* **13**, 1749–1759.
9. Madlung, A., Masuelli, R. W., Watson, B., Reynolds, S. H., Davison, J., and Comai, L.

(2002) Remodeling of DNA methylation and phenotypic and transcriptional changes in synthetic Arabidopsis allotetraploids. *Plant Physiol.* **129**, 733–746.

10. Pikaard, C. S. (2001) Genomic change and gene silencing in polyploids. *Trends Genet.* **17**, 675–677.
11. Comai, L., Madlung, A., Josefsson, C., and Tyagi, A. (2003) Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 1149–1155.
12. Akimoto, K., Katakami, H., kim, H. J., Ogawa, E., Sano, C. M., Wada, Y., et al. (2007) Epigenetic inheritance in rice plants. *Ann. Bot.* **100**, 205–217.
13. Liu, Z., Wang, Y., Shen, Y., Guo, W., Hao, S., and Liu, B. (2004) Extensive alterations in DNA methylation and transcription in rice caused by introgression from *Zizania latifolia*. *Plant Molecular Biology* **54**, 571–582.
14. Dong, Z. Y., Wang, Y. M., Zhang, Z. J., Shen, Y., Lin, X. Y., Ou, X. F., et al. (2006) Extent and pattern of DNA methylation alteration in rice lines derived from introgressive hybridization of rice and Zizania latifolia Griseb. *Ther. Appl. Genet.* **113**, 196–205.
15. Grossniklaus, U., Vielle–Calzada, J. P., Hoeppner, M. A., and Gagliano, W. B. (1998) Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. *Science* **280**, 446–450.
16. Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., et al. (1999) Control of fertilization–independent endosperm development by the MEDEA polycomb gene in Arabidopsis. *Proc. Natl. Acad. Sci. USA* **96**, 4186–4191.
17. Kinoshita, T., Yadegari, R., Harada, J. J., Goldberg, R. B., and Fischer, R. L. (1999) Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. *Plant Cell* **11**, 1945–1952.
18. Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J. J., Goldberg, R. B., et al. (2002) DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in Arabidopsis. *Cell* **110**, 33–42.
19. Xiao, W., Gehring, M., Choi, Y., Margossian, L., Pu, H., Harada, J. J., et al. (2003) Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev. Cell* **5**, 891–901.
20. Kohler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W., and Grossniklaus, U. (2003) The Polycomb–group protein MEDEA regulates seed development by controlling expression of the MADS–box gene PHERES1. *Genes Dev.* **17**, 1540–1553.
21. Scott, R. J. and Spielman, M., (2004) Epigenetics: imprinting in plants and mammals–the same but different? *Curr. Biol.* **14**, R201–R203.
22. Kohler, C., Page, D. R., Gagliardini, V., and Grassniklaus, U. (2005) The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nat. Genet.* **37**, 28–30.
23. Baroux, C., Gagliadini, V., Page, D. R., and Grassniklaus, U. (2006) Dynamic regulatory interactions of Polycomb group genes: MEDEA autoregulation is required for imprinted gene expression in Arabidopsis. *Genes Dev.* **20**, 1081–1086.
24. Gehring, M., Huh, J. H., Hsieh, T. F., Penterman, J., Choi, Y., Harada, J. J., et al. (2006) DEMETER DNA glycosylase establishes MEDEA polycomb gene self–imprinting by allele–specific demethylation. *Cell* **124**, 495–506.
25. Marcaud, L., Reynaud, C. A., Therwath, A., and Scherrer, K. (1981) Modification of the methylation pattern in the vicinity of the chicken globin genes in avian erythroblastosis virus transformed cells. *Nucleic Acids Res.* **9**, 1841–1851.
26. Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D., and Baylin, S. B. (1996) Methylation–specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci. USA* **93**, 9821–9826.
27. Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W., Henderson, I. R., et al. (2006) Genome–wide high–resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**, 1189–1201.
28. Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H., and Mukai, T. (1991) A genomic scanning method for higher organisms using restriction sites as landmarks. *Proc. Natl. Acad. Sci. U S A* **88**, 9523–9527.
29. Hayashizaki, Y. and Watanabe, S. (Eds.), (1997) Restriction landmark genomic scanning (RLGS), Springer Verlag, Tokyo, pp. 1–179.
30. Reyna–Lopez, G. E., Simpson, J., and Ruiz–Herrera, J. (1997) Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. *Mol. Gen. Genet.* **253**, 703–710.
31. Xiong, L. Z., Xu, C. G., Saghai Maroof, M. A., and Zhang, Q. (1999) Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation–sensitive

amplification polymorphism technique. *Mol. Gen. Genet.* **261**, 439–446.

32. Okuizumi, H., Okazaki, Y., Ohsumi, T., Hayashizaki, Y., Plass, C., and Chapman, V. M. (1995) Genetic mapping of restriction landmark genomic scanning loci in the mouse. *Electrophoresis* **16**, 233–240.
33. Okuizumi, H., Okazaki, Y., Ohsumi, T., Hanami, T., Mizuno, Y., Muramatsu, M., et al. (1995) A single gel analysis of 575 dominant and codominant restriction landmark genomic scanning loci in mice interspecific backcross progeny. *Electrophoresis* **16**, 253–260.
34. Ohsumi, T., Okazaki, Y., Okuizumi, H., Shibata, K., Hanami, T., Mizuno, Y., et al. (1995) Loss of heterozygosity in chromosomes 1, 5, 7 and 13 in mouse hepatoma detected by systematic genome–wide scanning using RLGS genetic map. *Biochem. Biophys. Res. Commun.* **212**, 632–639.
35. Miwa, W., Yashima, K., Sekine, T., and Sekiya, T. (1995) Demethylation of a repetitive DNA sequence in human cancers. *Electrophoresis* **16**, 227–232.
36. Hayashizaki, Y., Shibata, H., Hirotsune, S., Sugino, H., Okazaki, Y., Sasaki, N., et al. (1994) Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method. *Nature Genetics* **6**, 33–40.
37. Shibata, H., Yoshino, K., Muramatsu, M., Plass, C., Chapman, V. M., and Hayashizaki, Y. (1995) The use of restriction landmark genomic scanning to scan the mouse genome for endogenous loci with imprinted patterns of methylation. *Electrophoresis* **16**, 210–217.
38. Plass, C., Shibata, H., Kalcheva, I., Mullins, L., Kotelevtseva, N., Mullins, J., et al. (1996) Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS–M. *Nat. Genet.* **14**, 106–109.
39. Watanabe, S., Kawai, J., Hirotsune, S., Suzuki, H., Hirose, K., Taga, C., et al. (1995) Accessibility to tissue–specific genes from methylation profiles of mouse brain genomic DNA. *Electrophoresis* **16**, 218–226.
40. Bird, A. (1992) The essentials of DNA methylation. *Cell* **70**, 5–8.
41. Takamiya, T., Hosobuchi, S., Asai, K., Nakamura, E., Tomioka, K., Kawase, M., et al. (2006) Restriction landmark genome scanning method using isoschizomers (*Msp*I/*Hpa*II) for DNA methylation analysis. *Electrophoresis* **27**, 2846–2856.
42. Takamiya, T., Ohtake, Y., Hosobuchi, S., Noguchi, T., Kawase, M., Murakami, Y., et al. (2008) Application of RLGS method for detection of alteration in tissue cultured plants. *JARQ.* **42**, 151–155.
43. Takamiya, T., Hosobuchi, S., Noguchi, T., Asai, K., Nakamura, E., Habu, Y., et al. (2008) Inheritance and alteration of genome methylation in F_1 hybrid rice. *Electrophoresis* **29**, 4088–4095.
44. Takamiya, T., Hosobuchi, S., Noguchi, T., Paterson, A. H., Iijima, H., Murakami, Y., et al. (2009) The application of restriction landmark genome scanning method for surveillance of non–Mendelian inheritance in F_1 hybrids. *Comparative and Functional Genomics* 2009.
45. Kawase, M. (1994) Application of the restriction landmark genomic scanning (RLGS) method to rice cultivars as a new fingerprinting technique. *Theor. Appl. Genet.* **89**, 861–864.
46. Okamoto, H., Takamiya, T., Saito, A., Domon, E., Iimura, K., Tomioka, K., et al. (2006) Development of a new cultivar–discrimination method based on DNA polymorphism in a vegetatively propagated crop. *JARQ.* **40**, 65–69.
47. Matsuyama, T. (2008) Epigenetics : use of *in silico* genome scanning by the virtual image restriction landmark method. *FEBS journal* **275**, 1607.
48. International Rice Genome Sequencing Project (2005) The map–based sequence of the rice genome. *Nature* **436**, 793–800.
49. Southern, E. M. (1979) Measurement of DNA length by gel electrophoresis. *Anal. Biochem.* **100**, 319–323.

Chapter 9

Methylated DNA Immunoprecipitation Genome-Wide Analysis

Mattia Pelizzola and Annette Molinaro

Abstract

DNA methylation is an important and potentially heritable component of the epigenetic machinery that has a major role in the control of gene expression and can be deregulated in many diseases. This modification of genomic DNA can be assessed using the methylated DNA immunoprecipitation (MeDIP) method, based on the quantification of methylated DNA fragments enriched using an antibody specific for methyl-cytosines.

The relationship between the enrichment level and the real DNA methylation status is complex, and only a few methods have been developed to evaluate MeDIP enrichment measures to estimate the absolute or relative number of methyl-cytosines in a given sample. Two such methods are MEDME and BATMAN. This chapter focuses on the description and use of the former with a brief discussion of the latter.

Key words: DNA methylation, Epigenetics, MeDIP, MEDME

1. Introduction

1.1. DNA Methylation

Epigenetics refers to heritable modification of genomic DNA that can have profound effects on the regulation of gene activity. In particular, methylation of cytosines in genomic DNA (1) and post-translational modification of histones (2) are central components of epigenetic regulatory mechanisms. Epigenetic control is intimately associated with chromatin structure and DNA accessibility, and therefore, tight control of the epigenetic patterning is critical for the correct onset of cellular differentiation processes. On the other hand, its deregulation can result in the development of disease (3). For these reasons, it is important to acquire genome-wide maps of epigenetics marks, and to profile their modification over differentiation processes within both healthy and diseased tissues.

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_9,

The first genome-wide base-pair resolution maps of DNA methylation were recently acquired for plants and humans (4, 5) confirming the complexity of epigenetics patterning and clarifying the relevance of DNA methylation in the onset of differentiation processes. Those studies also revealed the importance of DNA methylation in sequence contexts other than the expected CpG di-nucleotide (here referred to as mCpG). In particular, methyl-cytosines in the CHG and CHH context (H being one of A, C, T), usually defined as non-CG methylation, are common in pluripotent cells and in the plant *Arabidopsis thaliana*.

1.2. Methods for Detection of Genome-Wide DNA Methylation

Several methods are available to detect DNA methylation either genome-wide or in specific regions (see ref. 6 for a comprehensive review). Briefly, PCR amplification and subsequent Sanger sequencing of bisulfite converted DNA is the gold standard to detect the location of methyl-cytosines in specific loci. This method is based on the bisulfite selective de-amination of un-methylated cytosines to uracil, while methyl-cytosines remain unconverted.

However, this approach is not practical for analysis of extensive regions. For this reason, several alternative methods were developed which can be distinguished from each other based on their cost, resolution, and genome coverage. These methods can be divided into three main groups based on the high-throughput sequencing of bisulfite-converted genomic DNA; the activity of methylation sensitive and insensitive restriction enzymes; and the enrichment of highly methylated genome fragments. Moreover, variants and combinations also exist. These methods are then coupled with either microarray hybridization or high-throughput sequencing for quantification.

High-throughput bisulfite sequencing methods are unique in that they allow one to obtain base-pair resolution data; nonetheless, the cost is high due to the size of many eukaryotic genomes. A direct result of the cost is a limit on the number of samples that can be analyzed. Methods based on restriction enzymes have a distinct disadvantage due to the inherent dependency on the localization of cutting sites and a consequential bias, as these loci are not uniformly distributed on the genome. Finally, enrichment methods are based on the use of antibodies or binding proteins specific for methyl-cytosines. Genomic DNA is fragmented and enriched for highly methylated DNA fragments, where the amount of enrichment is related to the degree of methylation. Thus, enrichment methods provide the best trade-off between high cost, data resolution, and genome coverage. In respect to bisulfite sequencing methods, the data resolution is certainly lower but so is the cost of the experiment, allowing a greater number of throughput samples. Additionally, enrichment methods provide an unbiased analysis in comparison to methods based on detection of methyl-cytosines in targeted regions of the genome.

1.3. Methylated DNA Immunopreci pitation

Methylated DNA immunoprecipitation (MeDIP) is based on the enrichment of fragmented genomic DNA using an antibody specific for methyl-cytosines (7, 8). Hence, by definition, this methodology can only be employed to identify highly methylated genomic regions and only indirectly can be used to map hypo-methylated ones. Genome coverage and data quality are dependent on the subsequent quantification method, usually either array hybridization (MeDIP-chip) or high-throughput sequencing (MeDIP-seq).

Promoter hyper-methylation is the most studied and clear example of degenerated DNA methylation. In fact, aberrant methylation patterns of promoter regions, rich in CpG islands, usually reveal the strongest effects on the transcription of downstream genes. For this reason, MeDIP is often coupled with hybridization on promoter tiling arrays. This experimental design is relatively inexpensive and extracts the maximum amount of useful information.

On the other hand, when researchers desire to obtain genome-wide unbiased data, they can couple MeDIP with high-throughput sequencing technologies. Nevertheless, there are several challenges to extract information from nonpromoter regions. Included in these challenges are limited data resolution, which is strongly dependent on the fragmented DNA size (typically in the order of few hundred bases); and, the uneven distribution of CpGs on the genome, which are mostly concentrated in gene upstream regions.

1.4. Data Processing

An understanding of the relationship between MeDIP enrichment and the real methylation level is critical for the correct interpretation of the data. This is especially crucial when comparing different samples and quantifying differential methylation. The developers of the MeDIP methodology compared the enrichment level with the actual methylation level for a small number of loci, concluding that the enrichment is linear in respect to the number of methyl-cytosines (7). Nevertheless, the limited number of data points necessitated a more thorough investigation of this relationship.

Two such studies explored the relationship between the MeDIP enrichment and the genome CpG density (9, 10). In both cases, the authors found this relationship to be highly complex and concluded that deriving measures of the absolute and relative methylation levels was not trivial. As a result, both studies proposed alternative strategies on how to process the MeDIP enrichment measures; namely, MEDME (Modeling Enrichment Derived from MeDIP Experiments) and BATMAN (A Bayesian Tool for Methylation Analysis). Here, we focus on MEDME and subsequently offer a brief discussion of BATMAN.

2. Materials

Modeling Enrichment Derived from MeDIP Experiments (MEDME) (10) is a novel methodology implemented in the R programming language (http://www.r-project.org). It is available on the Bioconductor web site, an open-source repository of libraries for the analysis of high-throughput biological data (11) (http://www.bioconductor.org). First, we describe how to install MEDME and then how to use it.

Note in the following: ">" is the R prompt and everything following on the same line is R code that can be typed and executed in the R GUI (see Note 1). The code lines are frequently followed by text output. Functions, objects, and arguments in the chapter text are indicated in italics.

To install the R programming language simply connect to the R web site (http://www.r-project.org), select download, choose a mirror web site, select the appropriate Operating System, and download the *base* installation file. Once R is installed and started, installing MEDME is straightforward and can be done with the following two commands:

```
> source("http://www.bioconductor.org/biocLite.R")
> biocLite("MEDME")
```

The genome of the organism of interest also needs to be installed (see Note 2). For example in the case of human, one would install:

```
> biocLite("BSgenome.Hsapiens.UCSC.hg18")whereas, for mouse:
> biocLite("BSgenome.Mmusculus.UCSC.mm9")
```

Once MEDME and the genome library are installed, the library can be loaded:

```
> library(MEDME)
```

3. Methods

3.1. The MEDME Idea

MEDME is based on the observation that MeDIP-chip enrichment measures are not linearly related to the real methylation level. To illustrate the authors used the *SssI* enzyme to artificially methylate genomic DNA to create a benchmark dataset, where all cytosines in CpG sites were methylated. Using this particular sample, they applied the MeDIP protocol and hybridized both the input and the enriched DNA fraction on two channel arrays with 370 k probes tiled each 100 bp for the entire X chromosome. The log2 ratio between the enriched and the input channels was then determined as a measure of enrichment. In this setting, the enrichment can be directly related to the CpG content of the region around a probe. In fact, in a fully methylated sample all the CpG are methylated and

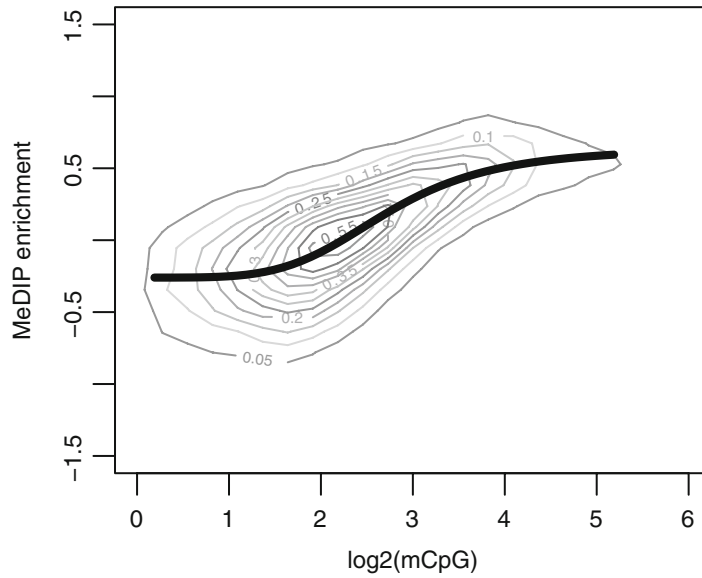


Fig. 1. Modeling the MeDIP enrichment as a function of the DNA methylation. Probe-level MeDIP enrichment is determined for a fully methylated sample and displayed versus the CpG content for each probe region. As the sample is fully methylated, all the CpGs are methylated and CpG and mCpG levels are equivalent. The MeDIP enrichment can be modeled as a sigmoidal function of the log2(mCpG). The scatter plot density is shown and the sigmoidal fit is overlaid.

determining the CpG content is equivalent to determining the mCpG level. Hence, the antibody enrichment and its array-based quantification can be directly related to the absolute methylation level, probe-by-probe. A sigmoidal function is fit to the measure of enrichment versus the log2 CpG content of the region surrounding the probes as shown in Fig. 1. The sigmoidal model based on the fully methylated sample can be deconvoluted determining how the methylation level can be computed from the MeDIP enrichment.

3.2. MEDME Overview

For each experiment, or at least any time the experimental setting (cells, protocols, array type) is changed, a calibration experiment with fully methylated genomic DNA should be performed, in parallel with the MeDIP-chip assay for the experimental samples. As such, the resulting enrichment for each probe is determined and the signal is averaged with the enrichment of the adjacent probes. This is a typical step in the analysis of tiling array data, performed to smooth the data by borrowing information from the surrounding probes. A sigmoidal model is fitted comparing the enrichment of the fully methylated sample with the genomic CpG content of the region around each probe. In particular, the CpG content is determined as the weighted count of the CpG within and around each probe. This is based on the assumption that, when evaluating the antibody binding in a given genomic locus, the relevance for a mCpG is inversely related to its distance from the midposition of the probe

for that locus. The model fit on the fully methylated sample is deconvoluted and is applied to the MeDIP enrichment of the experimental samples in order to estimate the Absolute Methylation Score (AMS). Finally, the probe-level AMS is compared to the probe weighted CpG-content in order to estimate the Relative Methylation Score (RMS) (see Note 3).

3.3. Step by Step Through a MEDME Session

This example MEDME analysis will be carried out on a test dataset available as part of the MEDME library. The test dataset contains both fully methylated calibration data and experimental data for samples with an unknown methylation level, and is a subset of the data used in (10) (see Note 4). First the library and the test dataset have to be loaded in the R workspace:

```
> library(MEDME)
> data(testMEDMEset)
```

The *MEDME.readFiles* function can be used to load real data (see Note 5). For further help see (see Note 6):

```
> ?MEDME.readFiles
```

The MEDIP enrichment data, the log2(cy5/cy3), contained in the testMEDMEset object can be shown using the logR function:

```
> logR(testMEDMEset[1:4, ])
```

	fullyMet	NBMEL	YUSAC2
ldX_86705	0.42	0.17	0.54
ldX_86706	0.76	−0.61	−0.14
ldX_86707	0.67	0.17	0.73
ldX_86708	0.29	0.25	0.81

This dataset contains data normalized within- and between-arrays (see Note 7). Nonetheless, before fitting the MEDME model on the fully methylated data it is convenient to smooth the data, so that each probe can borrow information from the signal of adjacent probes:

```
> testMEDMEset = smooth(data = testMEDMEset, wsize = 1000,
    wFunction = "linear")
```

The *wsize* argument defines the size of the window around a given probe that is used in the averaging procedure. The *wFunction* argument determines if less weight is given to adjacent probes based on their distance from the probe of interest (see the help file for the *smooth* function for additional details). One could optionally try different window sizes (*wsize*) and/or weighting functions (*wFunction*). As anticipated above, in order to fit the model it is necessary to determine the expected methylation level for each probe. Since all CpG are methylated in the fully methylated sample, this is performed determining the probe-level CpG content. The metadata library containing the genomic

sequence has to be installed as explained in the MATERIALS section. In particular, only human and mouse genomes are currently supported (see Note 8). In this example, the human genome is loaded and the CpG content of the region around each probe is determined with the following commands:

```
> library(BSgenome.Hsapiens.UCSC.hg18)
> testMEDMEset = CGcount(data = testMEDMEset)
```

Now it is possible to fit the logistic model to the MeDIP enrichment of the fully methylated sample, to determine the relationship between MeDIP enrichment and the expected methylation level. This can be done as follows:

```
> MEDMEmodel = MEDME(data = testMEDMEset, sample = 1)
```

It is important to compare the plot obtained through the *MEDME* function to the one available in Fig. 1 and in the MEDME vignette. Failure to obtain a correct fit of the model can affect the usefulness of AMS and RMS as estimates of the absolute and relative methylation level.

Finally, the model can be applied to the experimental data in order to predict probe-level AMS and RMS:

```
> testMEDMEset = predict(data = testMEDMEset, MEDMEfit = MEDMEmodel)
```

Log ratios enrichment, Smoothed data, AMS, and RMS are automatically saved in the *logR*, *smoothed, AMS,* and *RMS* components of the *MEDMEset* object, and can be retrieved using methods with the same name of these. Eventually, SGR or GFF files can be exported using the *MEDME.writeFiles* function to visualize these data on genome browsers. Alternatively, the *MEDMEset* object can be further processed with any of the many libraries available in R and Bioconductor for higher-level analysis. For example, the DNA methylation profiles could be integrated with gene expression data.

3.4. Applications

One of the advantages of MEDME is to provide estimates of both the absolute and the relative methylation levels. The importance of using both these measures is exemplified in the study where MEDME was used in combination with MeDIP and tiling promoter arrays to identify promoters differentially methylated in melanoma compared to normal human melanocytes (12). In this study, the authors showed that the promoter CpG content, and consequently the RMS, is an important feature when correlating its methylation status to the transcriptional repression of the down-stream genes. Indeed, an increase in the number of mCpG (estimated based on the MEDME AMS) in low CpG content promoters (LCPs) is not associated with a significant decrease of the expression of the down-stream gene. Rather, a variation in the AMS for intermediate and high CpG content promoters (ICPs and HCPs, respectively) is significantly correlated with variation in the gene expression level. Based on these results and on the observation that transcriptional

repression is greater the shorter the distance between the mCpG and the transcriptional start site, the authors implemented a series of filters to select promoters where differential methylation could be mechanistically related to the gene transcriptional repression. Eventually, differentially methylated markers were identified comparing melanoma strains to normal melanocytes using Linear Model Mixed Effects (LME) based on the MEDME promoter AMS. LME was chosen to take advantage of the experimental design and explicitly model the difference between melanoma and normal cells (a fixed effect) while considering the replication level (eight melanoma strains compared to two normal melanocytes) as the random effect, avoiding pooling or averaging the data for the cell strains within each group.

3.5. Bayesian Tool for Methylation Analysis

BATMAN is an alternative tool that can be used for the processing of MeDIP enrichment data (9). BATMAN relies on different assumptions than MEDME, discussed briefly here and, as expected, has some advantages and disadvantages in respect to MEDME. The main advantage is that it does not require calibration data, since it avoids modeling the enrichment in respect to real measures of DNA methylation. An additional advantage is that it is based on a sophisticated statistical framework that provides relative methylation levels restricted to between 0 and 100%, while MEDME can sometimes generate relative methylation estimates greater than 100%. Finally, BATMAN can also analyze MeDIP-seq data; although it requires a different version of the software that has not yet been released. BATMAN is a stand-alone analysis tool, written in Java. In comparison to MEDME, a disadvantage is that the installation and computer power needed for BATMAN are significantly greater. Moreover, BATMAN does not directly provide measures of absolute methylation levels. While these could in principle be determined afterward, their determination is not trivial.

The starting point in BATMAN is determining the so-called CpG coupling factor, a function of the distance between the probe genomic location and the CpG. This is determined empirically based on the random sampling of a DNA fragment of a given size, based on the expected size distribution, and in silico overlapping it to the array probes.

The relationship between the MeDIP enrichment and this coupling factor is rather complex, but as the authors point out some linearity is evident for probes in low CpG areas. This makes sense from the biological point of view, given that low CpG content regions are usually highly methylated in human. The idea is therefore that a linear model can be fit to model the relationship with the MeDIP enrichment in the whole CpG content range. Finally, a Bayesian inference approach is used to determine the probe-specific methylation score, ranging from 0 to 1.

3.6. Final Remarks

It should be noted that both MEDME and BATMAN rely on the assumption that the methyl-cytosine are only in the CpG sequence context. This can be assumed for the majority of samples, but, has to be carefully evaluated (see Note 3). In fact, it has recently been shown that pluripotent cells can have up to 25% of the methyl-cytosine in sequence context different than CpGs (5). The prevalence of non-CpG methylation needs to be evaluated in other tissues and differentiation stages. Moreover non-CpG methylation is rather common and important in other organisms as seen in plants (4).

Regarding the comparison between MeDIPseq and other methodologies for the genome-wide assessment of DNA methylation, Harris et al. recently reported interesting findings (13). First, when MeDIPseq is compared to the high-throughput sequencing of bisulfite converted DNA it is possible to obtain similar coverage in terms of the number of analyzed CpG and highly correlated results on the corresponding methylation status. The price to pay for this is that the resolution of MeDIPseq data is significantly lower and that studies requiring the evaluation of the methylation status of single nucleotides, as in the methylation of transcription factors binding sites, are not feasible. Finally, MeDIPseq, that by definition allows the identification of highly methylated genomic regions, can be successfully coupled with other methodologies that select for hypo-methylated regions, like those based on restriction enzymes. On one hand, this allows a complete overview of the genome methylation status. On the other hand, this combination allows the identification of allelic DNA methylation when data about the location of SNPs are included.

4. Notes

1. R can be run in a terminal (shell) in Linux and MacOS systems, while it is usually run in a GUI (Graphical User Interface) in Windows and MacOS. The major difference is that several menus are available in the GUI. The same functionalities are available in the terminal versions but are not as accessible.
2. The size of the annotation library can be close to 800 Mb. For this reason, one can expect a significant download time. While downloading the library, the R prompt is not accessible. Alternatively, it is possible to manually download the libraries from the Bioconductor website and install them directly from the downloaded file.
3. The antibody used in the MeDIP methodology is expected to recognize 5mC in the non-CG context too (14). This might complicate the data analysis of stem cell DNA methylation patterns, as this epigenetic mark is rather common in these cell types and MEDME and BATMAN do not directly support it.

4. In principle, the fully methylated data available in the MEDME library should not be used for fitting the model while analyzing the user's own data. This is particularly important, in case, the experimental setting is different from the one used in generating the MEDME example data. See the help page of the test dataset (see *?testMEDMEset*) for complete documentation on the available samples (e.g., NBMEL).
5. A *MEDMEset* object can be built using the *MEDME.readFiles* function based on a set of GFF or SGR files containing the MeDIP enrichment of both the experimental and the fully methylated samples (see *?MEDME.readFiles* for details on GFF and SGR formats). In this example, data derived from Nimblegen two-channel arrays are considered. Nonetheless, any commercial or custom one- or two-channel platform can be used, as long as it is possible to determine enrichment log-ratios. Variations to the published MeDIP methodology, including various antibodies or methyl-binding proteins can be considered. Eventually, the fit of the MEDME model, and in particular the critical evaluation of the plot generated by the *MEDME* function, remain central in determining if MEDME can be correctly applied.
6. The notation of a question mark before an R command will provide access to the corresponding documentation. Please note that the request for help for a function in R will only work if the library to which that function belongs has been loaded.
7. The limma Bioconductor library can be used for these normalization procedures.
8. Please note that only the last genome release should be used and that the LiftOver UCSC tool could be used for batch conversion of old genomic position to the last genome release (http://www.genome.ucsc.edu/cgi-bin/hgLiftOver).

References

1. Suzuki, M. M., and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics, *Nat Rev Genet 9*, 465–476.
2. Barski, A., Cuddapah, S., Cui, K., Roh, T., Schones, D., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome, *Cell 129*, 823–837.
3. Esteller, M. (2008) Epigenetics in cancer, *The New England journal of medicine 358*, 1148–1159.
4. Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., and Ecker, J. R. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis, *Cell 133*, 523–536.
5. Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., and Ecker, J. R. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences, *Nature 462*, 315–322.
6. Laird, P. (2010) Principles and challenges of genome-wide DNA methylation analysis, *Nat Rev Genet 11*, 191.
7. Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W., and Schübeler, D. (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA

methylation in normal and transformed human cells, *Nat Genet 37*, 853–862.

8. Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M., and Schübeler, D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome, *Nat Genet 39*, 457–466.
9. Down, T., Rakyan, V., Turner, D., Flicek, P., Li, H., Kulesha, E., Gräf, S., Johnson, N., Herrero, J., Tomazou, E., Thorne, N., Bäckdahl, L., Herberth, M., Howe, K., Jackson, D., Miretti, M., Marioni, J., Birney, E., Hubbard, T., Durbin, R., Tavaré, S., and Beck, S. (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis, *Nat Biotechnol 26*, 779–785.
10. Pelizzola, M., Koga, Y., Urban, A., Krauthammer, M., Weissman, S., Halaban, R., and Molinaro, A. (2008) MEDME: An experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment, *Genome Research 18*, 1652–1659.
11. Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y., and Zhang, J. (2004) Bioconductor: open software development for computational biology and bioinformatics, *Genome Biol 5*, R80.
12. Koga, Y., Pelizzola, M., Cheng, E., Krauthammer, M., Sznol, M., Ariyan, S., Narayan, D., Molinaro, A., Halaban, R., and Weissman, S. (2009) Genome-wide screen of promoter methylation identifies novel markers in melanoma, *Genome Research 19*, 1462–1470.
13. Harris, R. A., Wang, T., Coarfa, C., Nagarajan, R. P., Hong, C., Downey, S. L., Johnson, B. E., Fouse, S. D., Delaney, A., Zhao, Y., Olshen, A., Ballinger, T., Zhou, X., Forsberg, K. J., Gu, J., Echipare, L., O'Geen, H., Lister, R., Pelizzola, M., Xi, Y., Epstein, C. B., Bernstein, B. E., Hawkins, R. D., Ren, B., Chung, W. Y., Gu, H., Bock, C., Gnirke, A., Zhang, M. Q., Haussler, D., Ecker, J. R., Li, W., Farnham, P. J., Waterland, R. A., Meissner, A., Marra, M. A., Hirst, M., Milosavljevic, A., and Costello, J. F. (2010) Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications, *Nat Biotechnol 26*, 1097–1105.
14. Barres, R., Osler, M. E., Yan, J., Rune, A., Fritz, T., Caidahl, K., Krook, A., and Zierath, J. R. (2009) Non-CpG Methylation of the PGC-1α Promoter through DNMT3B Controls Mitochondrial Density, *Cell Metabolism 10*, 189–198.

Chapter 10

Methylated-CpG Island Recovery Assay

Natalie Mitchell, J. Tyson DeAngelis, and Trygve O. Tollefsbol

Abstract

Alterations in DNA methylation patterns are implicated in playing a major role in the development of cancer, thus highlighting the need to continually develop new technologies to analyze epigenetic marks. Methylated-CpG Island Recovery Assay (MIRA), based on the high affinity of the MBD2b/MDB3L1 complex for double-stranded methylated DNA, allows for the recovery of methylated DNA without the use of bisulfite conversion or antibody recognition. MIRA is capable of detecting low-density methylation of a single methylated CpG nucleotide. This technique can be used in conjunction with microarrays or next-generation sequencing to analyze recovered methylated DNA on a genome-wide scale.

Key words: DNA methylation, Methylated-CpG island recovery assay, Methyl-CpG binding protein, Microarray, Next-generation sequencing

1. Introduction

DNA methylation at CpG dinucleotides is an important epigenetic modification in mammalian cells. Epigenetic modifications describe the heritable changes in gene expression that occur without a change in the primary DNA sequence (1). Regions of DNA where a large density of CpG dinucleotides exist are known as CpG islands which are found in the proximal promoter region in nearly half of the genes in the mammalian genome (2). In normal cells, hypermethylated promoters serve to maintain genes in an inactive state that are not required for that specific cell type. During oncogenesis, genome-wide hypomethylation combined with site-specific hypermethylation leads to alterations in the expression of genes associated with tumor development (3, 4). Because changes in DNA methylation have been implicated in playing a central role in tumorigenesis, it is important that techniques be available to readily analyze changes in DNA methylation at CpG islands (5, 6).

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_10,

While numerous methods exist for analyzing DNA methylation, the Methylated-CpG Island Recovery Assay (MIRA) is of high interest. It is important to note that MIRA is sensitive enough to detect low-density methylation of a single methylated CpG nucleotide; however, it is the spacing between individual methylated CpG nucleotides that possibly limits the sensitivity of MIRA (7). This technique can be used in combination with a variety of microarrays to analyze changes in the methylome (8).

In 1975, the first assay was developed to analyze DNA methylation at specific nucleotide sequences using methylation-sensitive restriction enzymes called isoschizomers. Isoschizomers are capable of cutting the same unmethylated DNA sequence; however, only one of the two restriction endonucleases will cut the nucleotide sequence if it is methylated. Thus, the presence of DNA methylation at specific cytosine residues can affect the results obtained using restriction endonucleases. Methylation of a specific genomic sequence can be observed through Southern blot hybridization (9).

Since the initial attempt of locating DNA methylation in the mid-1970s, a wide variety of other methods have been created. Although many techniques for detecting methylated CpG islands exist, sodium bisulfite treatment of genomic DNA has become the preferred method for studying DNA methylation. Bisulfite treatment of genomic DNA converts unmethylated cytosines to uracils by deamination and leaves methylated cytosines unaffected. After the area of interest is amplified through PCR, it is usually cloned and sequenced. Using PCR, this technique analyzes only a small region of the genome, obviously presenting a large limitation of the procedure (10, 11). In addition, the loss of a base through bisulfite sequencing compromises the complexity of data generated and requires comparison to the original sample sequence. Although bisulfite treatment of DNA has been incredibly valuable over the last few decades, the inability to compare DNA methylation density between large discrete genomic regions at the single nucleotide level has limited its usefulness it the postgenomic era.

The first next-generation assay for studying genome-wide DNA methylation is the methylated DNA immunoprecipitation microarray (MeDIP) technique. MeDIP is capable of detecting DNA methylation at CpG islands across the entire genome. MeDIP relies on antibodies that recognize 5-methylcytosine (5mC) to immunoprecipitate single-stranded DNA that contains heavily methylated CpG sites. The methylated DNA can then be further analyzed with DNA microarrays or parallel sequencing techniques by simply quantifying the regions that were recovered. Because MeDIP relies on genomic regions that are heavily methylated, regions of DNA with a low CpG dinucleotide density may not be accurately analyzed by MeDIP despite the fact that DNA

methylation changes may exist. In addition, antibodies can be expensive and rather variable in their recognition response (12).

The latest and most efficient means of analyzing genomic DNA methylation patterns is MIRA. MIRA does not rely on bisulfite conversion of DNA or antibody recognition of single-stranded DNA. MIRA can analyze DNA methylation patterns in double-stranded DNA and has a higher sensitivity and specificity than the aforementioned methods. Mammalian genomes encode several proteins that are capable of binding to methylated CpG dinucleotides and serve to repress transcription of the downstream gene (13, 14). While the first methyl-CpG binding (MBD) protein discovered was MeCP2, MIRA is based on the recognition of methylated CpG dinucleotides by the methyl-CpG-binding domain protein-2b (MBD2b). The binding of MBD2b to methylated DNA is sequence-independent, making it a more attractive candidate for isolating methylated DNA than MeCP2 (15, 16). MBD2b has one of the highest affinities for binding methylated DNA; however, the binding affinity is further increased by the addition of the methyl-CpG-binding domain protein 3-like-1 (MBD3L1) (7, 15, 17). MBD3L1 binds with MBD2b in vivo and in vitro to promote the formation of methylated-DNA-binding complexes (17). This complex is the central core that allows MIRA to successfully function.

The first step in MIRA is to fragment double-stranded DNA to an average length of 0.35 kb by either sonication or restriction endonuclease digestion. Methylated DNA fragments are then bound by the glutathione-S-transferase (GST)-tagged recombinant MBD2b protein/His-tagged MBD3L1 complex. The DNA–protein complex is then captured using glutathione beads in an affinity column. Following capture, DNA is eluted from DNA–protein complexes and amplified with PCR for site-specific analysis (7). See Fig. 1 for an overview of the MIRA procedure.

MIRA has several advantages over other current DNA methylation analysis techniques and may be applicable in a variety of clinical and diagnostic situations, including characterization of DNA methylation in tissues or body fluids. Unlike MeDIP, MIRA works on double-stranded DNA and does not require the use of antibodies against 5-methylcytosine that is specific to single-stranded DNA. In addition, MIRA does not require the occurrence of methylation-sensitive restriction sites within the targeted sequence. Because it is not extremely laborious, MIRA is useful in genome-wide methylation analysis in cancers. MIRA may also be used to detect cell-type dependent differences in DNA methylation for a large number of genes through subsequent microarray analysis. In addition, the number of false-positives generated by the MIRA method is extremely low, making it an even more attractive candidate for analyzing DNA methylation at CpG dinucleotides (7, 8).

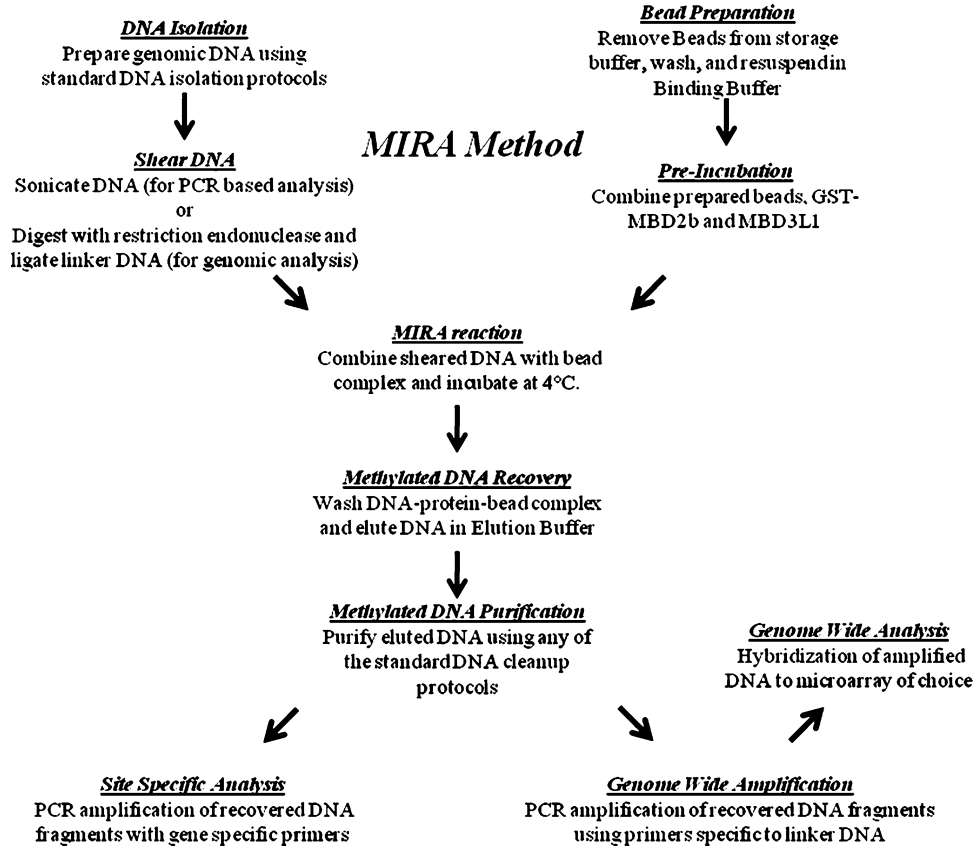


Fig. 1. Schematic of the steps necessary to complete MIRA. The products of DNA shearing and the bead/protein complex are combined to perform the MIRA reaction. After the methylated DNA is recovered, purified, and amplified with PCR, it can be analyzed via a variety of assays.

Recently, Active Motif has developed a commercially available kit (*MethylCollector Ultra*) based on the MIRA assay. DNA is sheared and incubated with the His-tagged MBD2b/MBD3L1 protein complexes. The protein–DNA complex is captured with nickel-coated magnetic beads. Unmethylated DNA and protein are washed away and methylated DNA is eluted and ready for analysis. Active Motif has also developed an *UnMethylCollector* kit that recovers only unmethylated DNA. The combination of these two kits provides a very powerful tool able to analyze genome-wide DNA methylation patterns. The use of *UnMethylCollector* further validates experimentation results obtained by *Methyl-Collector Ultra*. Instead of assuming the status of DNA is unmethylated by negative identification with the *MethylCollector Ultra* kit, *UnMethylCollector* provides a positive method of ensuring collected data is unmethylated.

Multiple commercially available arrays exist for the analysis of recovered DNA fragments; this chapter focuses on describing

methods to analyze the products of MIRA and current trends in epigenetic research.

2. Materials

1. 1 μg purified glutathione-S-transferase-tagged recombinant Methyl-CpG-binding domain (MBD) protein-2b (GST-MBD2b) (see Note 1).
2. 1 μg purified His-tagged MBD3L1 (see Note 1).
3. Glutathione sepharose CL-4B matrix (Amersham Biosciences).
4. Binding buffer (10 mM Tris–HCl (pH 7.5), 50 mM NaCl, 1 mmol/L EDTA, 1 mM DTT, 3 mM $MgCl_2$, 0.1% Triton-X 100, 5% glycerol, 25 μg/mL, bovine serum albumin, and 1.25 μg sonicated JM110 (*dcm* minus) bacterial DNA).
5. Wash buffer (binding buffer containing 700 mM NaCl).
6. Guanidinium hydrochloride-containing Elution buffer.
7. Qiaquick PCR purification kit.
8. Genomic DNA.
9. Ultrasonic homogenizer with a microtip.
10. Benchtop microcentrifuge.

3. Methods

3.1. DNA Preparation

1. Genomic DNA from cell cultures or tissue samples should be prepared using any of the standard protocols.
2. Shear genomic DNA using Ultrasonic homogenizer to an average length of 0.35 kb by sonicating each sample five times for 10 s at an output setting of 30%. Sonicate samples in an ice bath to prevent overheating. See Note 2 for additional information regarding shearing and linker ligation for microarray analysis.

3.2. Bead Preparation

1. Centrifuge 100 μL of GST-sepharose slurry for 5 min at 500 × *g*. See Note 3 for additional information bead preparation.
2. Carefully remove the supernatant.
3. Wash the beads three times with 500 μL of binding buffer.
4. Resuspend the beads in 400 μL of binding buffer.

3.3. Preincubation

1. Combine 1 µg of GST-MBD2b and 1 µg of MBD3L1 with GST sepharose beads prepared in the previous step and incubate for 20 min at 4°C on a rotating platform (see Note 4).

3.4. MIRA reaction

1. Combine 500 ng of sonicated DNA with binding reaction from the previous step and incubate for 4 h at 4°C on a rotating platform (see Note 5).
2. Pellet beads by centrifugation at 500 × *g* for 5 min.
3. Wash beads three times with 500 µL of wash buffer. For each wash, incubate at room temperature (RT) for 5 min on a rotating platform followed by centrifugation at 500 × *g* for 5 min.

3.5. Elution and DNA purification

1. Following the third and final wash, resuspend the beads in 100 µL of Elution buffer and incubate at 50°C for 1 h.
2. Pellet beads by centrifugation at 500 × *g* for 5 min.
3. Carefully remove the supernatant and transfer to a fresh tube.
4. Purify methylated genomic DNA using the *Qiaquick PCR Purification Kit* following the manufacturer's protocol.

3.6. Amplification of Recovered DNA

1. To analyze a specific region of genomic, amplify with PCR using gene primers.
2. For microarray analysis, amplify all recovered DNA using primers specific to linker DNA (see Note 6).

3.7. Analyzing the Genomic Products of MIRA

The MIRA-chip method can be used to determine the DNA methylation status of large numbers of genes and chromosomal regions of both normal and cancerous tissues. Rauch and colleagues characterized the DNA methylation status of the human HOX gene clusters by combining the MIRA method with genome-wide CpG island tiling arrays (18). Site-specific PCR amplification can also be used in conjunction with MIRA to analyze DNA methylation changes at a particular gene of interest.

It was originally assumed that functionally important DNA methylation changes occurred in promoter regions and CpG islands. However, as recently noted by Feinberg and colleagues, many methylation alterations occur in sequences up to 2 kb from CpG islands; such areas are termed CpG island shores. It is methylation changes at CpG shores that have been shown to discriminate normal tissue from cancerous tissue more so than CpG island methylation status (19). Thus, a distinct limitation exists in using only CpG island tiling arrays in conjunction with MIRA to analyze DNA methylation changes as all possible sites where methylation may occur are not included in the array.

This problem is partially solved through the use of a human promoter array, which allows the user to analyze promoter DNA

from over 25,000 genes. The Affymetrix product covers 7.5 kb upstream and 2.45 kb downstream of the 5 transcriptional start site. For over 1,300 cancer-associated genes, the promoter coverage was expanded to include 10 kb upstream through 2.45 kb downstream. Thus, the ability to analyze DNA methylation at CpG shores is possible; however, it does not test for all genomic regions, including exonic and intronic DNA.

Of further interest, however, is whole-genome tiling arrays (WGAs) which interrogate an entire genome in an unbiased fashion through the use of nonoverlapping or partially overlapping probes spaced at regular intervals. The technique is unbiased due to the ability to include as yet unidentified genes. WGAs have never been used in conjunction with MIRA but have been used with MeDIP to map the *Arabidopsis thaliana* methylome (20). After immunoprecipitation with an anti-5-methyl cytosine antibody, the complex is hybridized to a tiling array to map the methylome. However, the extremely high costs of WGAs make them an impractical choice for use in mammalian genomes. In addition, because WGAs have such a high level of sensitivity and assume no underlying gene models or annotations, results may contain an overlap of noise and signal (21). It is important to note that Rauch and colleagues recently used MIRA in combination with WGAs to characterize the entire B cell methylome of an individual human at 100-bp resolution (22). Although progress is being made on current microarray technologies, most study of DNA methylation patterns in mammalian genomes is currently limited to CpG islands and promoters.

Several techniques have been developed to analyze DNA methylation patterns on a genome-wide scale. Current methods involve either cleavage by methylation-sensitive restriction endonucleases (3), bisulfite sequencing (23), or precipitation of methylated DNA with an antibody (12). However, these microarray methods are not capable of providing a high-resolution DNA methylation map of the mammalian genome.

Next-generation sequencing (NGS) allows for the sequencing of short DNA fragments at a cost much lower than the traditional Sanger method of sequencing. It is capable of single-nucleotide resolution and consequently is challenging microarrays as the tool of choice for analyzing the genome. Known as ChIP sequencing (ChIP-seq), NGS has been combined with chromatin-immunoprecipitation (ChIP) to sequence the genomic DNA fragments bound by transcription factors (24). In addition, MIRA-seq provides extremely quantitative data about genome-wide methylation. MIRA-enriched fragments undergo bisulfite conversion and analyzation by NGS (25). Thus, NGS has future implications of further analyzing epigenetic marks of the genome.

4. Notes

1. Plasmids expressing both of the recombinant MBD proteins used in the MIRA assay are available upon request by contacting the original authors (see ref. 7).
2. If microarrays are to be used to identify methylated genomic regions, genomic DNA should be digested with Mse1 and linker DNA ligated. Ligation of linker DNA allows for amplification of recovered DNA prior to microarray hybridization.
3. GST-beads are shipped in an ethanol slurry and bead preparation is performed to remove all traces of ethanol that may interfere in the MIRA reaction.
4. The high binding affinity of GST-MBD2b and MBD3L1 allows formation of the MBD2b-MBD3L1 complex in vitro. This step is critical to the recovery of methylated DNA and steps should be taken to ensure this part of the procedure is performed properly.
5. To ensure even mixing, samples should be lightly mixed by hand at each hour interval during the course of the 4-h incubation.
6. For genome wide analysis of recovered DNA, a much greater amount of DNA is required. Linker DNA that is ligated following enzymatic digestion can be used to amplify all recovered DNA.

Acknowledgments

This work was supported in part by a grant from the National Cancer Institute (RO1 CA129415). J.T.D. was supported by an NCI Cancer Prevention and Control Training Program (R25CA047888).

References

1. Holliday, R. (1994) Epigenetics: an overview., *Dev Genet* ***15***, 453–457.
2. Bird, A. (1986) CpG-rich islands and the function of DNA methylation., *Nature* ***321***, 209–213.
3. Costello, J. F., Fruhwald, M. C., Smiraglia, D. J., Rush, L. J., Robertson, G. P., Gao, X., Wright, F. A., Feramisco, J. D., Peltomaki, P., Lang, J. C., Schuller, D. E., Yu, L., Bloomfield, C. D., Caligiuri, M. A., Yates, A., Nishikawa, R., Su Huang, H., Petrelli, N. J., Zhang, X., O'Dorisio, M. S., Held, W. A., Cavenee, W. K., and Plass, C. (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns, *Nat Genet,* ***24***, 132.
4. Jones, P., and Baylin, S. (2002) The fundamental role of epigenetic events in cancer., *Nat Rev Genet* ***3***, 415–428.
5. Fraga, M., and Esteller, M. (2002) DNA methylation: a profile of methods and applications., *Biotechniques* ***33***, 632, 634, 636–649.

6. Shiraishi, M., Oates, A., and Sekiya, T. (2002) An overview of the analysis of DNA methylation in mammalian genomes., *Biol Chem* ***383***, 893–906.
7. Rauch, T., and Pfeifer, G. (2005) Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer., *Lab Invest* ***85***, 1172–1180.
8. Rauch, T., Li, H., Wu, X., and Pfeifer, G. (2006) MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells., *Cancer Res* ***66***, 7939–7947.
9. Singer, J., Roberts-Ems, J., and Riggs, A. (1979) Methylation of mouse liver DNA studied by means of the restriction enzymes msp I and hpa II., *Science* ***203***, 1019–1021.
10. Clark, S., Harrison, J., Paul, C., and Frommer, M. (1994) High sensitivity mapping of methylated cytosines., *Nucleic Acids Res* ***22***, 2990–2997.
11. Frommer, M., McDonald, L., Millar, D., Collis, C., Watt, F., Grigg, G., Molloy, P., and Paul, C. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands., *Proc Natl Acad Sci USA* ***89***, 1827–1831.
12. Weber, M., Davies, J., Wittig, D., Oakeley, E., Haase, M., Lam, W., and Schübeler, D. (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells., *Nat Genet* ***37***, 853–862.
13. Wade, P. (2001) Methyl CpG binding proteins: coupling chromatin architecture to gene regulation., *Oncogene* ***20***, 3166–3173.
14. Hendrich, B., and Bird, A. (2000) Mammalian methyltransferases and methyl-CpG-binding domains: proteins involved in DNA methylation., *Curr Top Microbiol Immunol* ***249***, 55–74.
15. Fraga, M., Ballestar, E., Montoya, G., Taysavang, P., Wade, P., and Esteller, M. (2003) The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties., *Nucleic Acids Res* ***31***, 1765–1774.
16. Klose, R., Sarraf, S., Schmiedeberg, L., McDermott, S., Stancheva, I., and Bird, A. (2005) DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG., *Mol Cell* ***19***, 667–678.
17. Jiang, C., Jin, S., and Pfeifer, G. (2004) MBD3L1 is a transcriptional repressor that interacts with methyl-CpG-binding protein 2 (MBD2) and components of the NuRD complex., *J Biol Chem* ***279***, 52456–52464.
18. Rauch, T., Wang, Z., Zhang, X., Zhong, X., Wu, X., Lau, S., Kernstine, K., Riggs, A., and Pfeifer, G. (2007) Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay., *Proc Natl Acad Sci USA* ***104***, 5527–5532.
19. Irizarry, R., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., Ji, H., Potash, J., Sabunciyan, S., and Feinberg, A. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores., *Nat Genet* ***41***, 178–186.
20. Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S., Chen, H., Henderson, I., Shinn, P., Pellegrini, M., Jacobsen, S., and Ecker, J. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis., *Cell* ***126***, 1189–1201.
21. Gregory, B., and Belostotsky, D. (2009) Whole-genome microarrays: applications and technical issues., *Methods Mol Biol* ***553***, 39–56.
22. Rauch, T., Wu, X., Zhong, X., Riggs, A., and Pfeifer, G. (2009) A human B cell methylome at 100-base pair resolution., *Proc Natl Acad Sci USA* ***106***, 671–678.
23. Meissner, A., Mikkelsen, T., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B., Nusbaum, C., Jaffe, D., Gnirke, A., Jaenisch, R., and Lander, E. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells., *Nature* ***454***, 766–770.
24. Hurd, P., and Nelson, C. (2009) Advantages of next-generation sequencing versus the microarray in epigenetic research., *Brief Funct Genomic Proteomic* ***8***, 174–183.
25. Pfeifer, G. P., and Rauch, T. A. (2009) DNA methylation patterns in lung carcinomas, *Semin Cancer Biol* ***19***, 181–187.

Chapter 11

Global DNA Methylation Analysis Using the Luminometric Methylation Assay

Mohsen Karimi, Karin Luttropp, and Tomas J. Ekström

Abstract

Epigenetic alterations regulate the utilization of the genome by permitting or inhibiting access of transcription factors and associated complexes. Although there are several different types of epigenetic alterations, such as acetylation and methylation of histone tails, the one which has been the most extensively studied is DNA-methylation, wherein the cytosine residue in a CpG dinucleotide is methylated.

Luminometric Methylation Assay (LUMA) enables researchers to study global methylation by using methylation-sensitive restriction enzymes followed by Pyrosequencing® which quantitates the number of cuts in the genome relative to an internal standard. The relative measurement of global methylation levels is simple and enables up to 96 samples to be analyzed at the same time.

Key words: LUMA, Global DNA methylation, Epigenetics, Pyrosequencing

1. Introduction

1.1. Background

Several assays for measuring global 5′-CpG methylation have been reported previously (1). However, these are either labor intensive, involve the use of radioactive isotopes or require large amounts of DNA. An early DNA methylation assay utilized methylation-sensitive and -insensitive restriction endonucleases, where the amount of genome-wide CpG methylation was obtained by comparing the amount of restriction performed by these enzymes (2). For this type of method, isoschizomer endonucleases (restriction enzymes with the same recognition sequence) are used; most commonly *Hpa*II and *Msp*I. Both enzymes have the recognition sequence CCGG. However, *Hpa*II digestion is inhibited if the internal cytosine

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_11, © Springer Science+Business Media, LLC 2011

in the DNA is methylated (CmCGG), while *Msp*I always cuts the recognition sequence, regardless of methylation status (2). The human genome is generally methylated to between 50 and 70% in CCGG sequences (see Note 1). DNA methylation analysis methods using *Hpa*II and *Msp*I are numerous, including self-primed in situ labeling (SPRINS) (3), methylation-sensitive arbitrarily primed polymerase chain reaction (AP-PCR) (4), nonmethylated genomic sites coincidence cloning (NGSCC) (5), differential methylation hybridization (DMH) (6), and methylation target array (MTA) (7). A more comprehensive overview of techniques for DNA methylation analysis can be found in reference 8.

Based on the properties of these endonucleases, Pogribny et al. developed a "cytosine extension assay," which combines the restriction of *Hpa*II and *Msp*I endonucleases with single nucleotide extension. Radioactively labeled [3H]dCTP is incorporated following restriction and is thus inversely correlated with DNA methylation (9). Following this report, a modified version with biotinylated dCTP was developed (10), wherein the amount of DNA methylation was defined as the *Hpa*II/*Msp*I ratio. The *Hpa*II/*Msp*I ratio would be 1.0 (i.e., the amount of restriction is equal) when the DNA is completely unmethylated, and would approach 0.0 when the DNA is fully methylated. LUMA has been successfully applied in several studies in addition to our own (11–21). This chapter describes the Luminometric Methylation Assay, or LUMA, in which the radioactive single nucleotide extension assay has been replaced by a Pyrosequencing® reaction (22).

1.2. Principle of LUMA

The LUMA method utilizes *Hpa*II (or some other methylation-sensitive isoschizomer enzyme leaving a 5′-overhang) and *Msp*I (or some other methylation-insensitive isoschizomer enzyme leaving a 5′-overhang) to perform enzymatic restriction of DNA. The enzymatically treated DNA is then analyzed by a luminometric polymerase extension assay to quantify the amount of restriction cleavage by each of the enzymes. The relative amount of DNA methylation is then expressed as an *Hpa*II/*Msp*I ratio. Parallel reactions, one with *Hpa*II and one with *Msp*I, are run. To enable normalization between runs and for DNA input, *Eco*RI is included in all reactions. *Eco*RI has the recognition sequence GAATTC, and is thus unaffected by CpG methylation. After *Hpa*II or *Msp*I restriction at their recognition sequence, there is a resulting 5′-CG overhang, whereas *Eco*RI restriction yields a 5′-AATT overhang. Using the Pyrosequencing® platform, nucleotides are added stepwise in a predetermined order. As the nucleotides (dNTP) are incorporated with DNA polymerase, inorganic pyrophosphate (PPi) is released and converted to ATP by ATP-sulfurylase and adenosine-5′-phosphosulfate. The resulting ATP is then utilized by luciferase to convert luciferin to oxyluciferen by luciferase, which produces visible light. The light is proportional to the amount of

dNTP incorporation in the original restriction enzyme produced 5′-overhang. The light is then detected by a charge-coupled device (CCD) camera and visualized as peaks in the software (23). For LUMA, dNTPs are added in four sequential steps (Step 1: dATPαS, Step 2: dGTP + dCTP, Step 3: dTTP, and Step 4: dGTP + dCTP). Peaks following dATPαS (Step 1) and dTTP (Step 3) dispensations both correspond to the amount of *Eco*RI restriction and are expected to be equal to one another since the amount of dATP and dTTP incorporation should be the same. dCTP and dGTP are added simultaneously in dispensation 2, and the corresponding peak represents the amount of *Hpa*II or *Msp*I restriction. Finally, dCTP and dGTP are added again in Step 4 to ensure that all *Hpa*II or *Msp*I overhangs were filled in during Step 2. The peak height following Step 4 should be zero or close to zero. Figure 1 illustrates the schematic principle of LUMA.

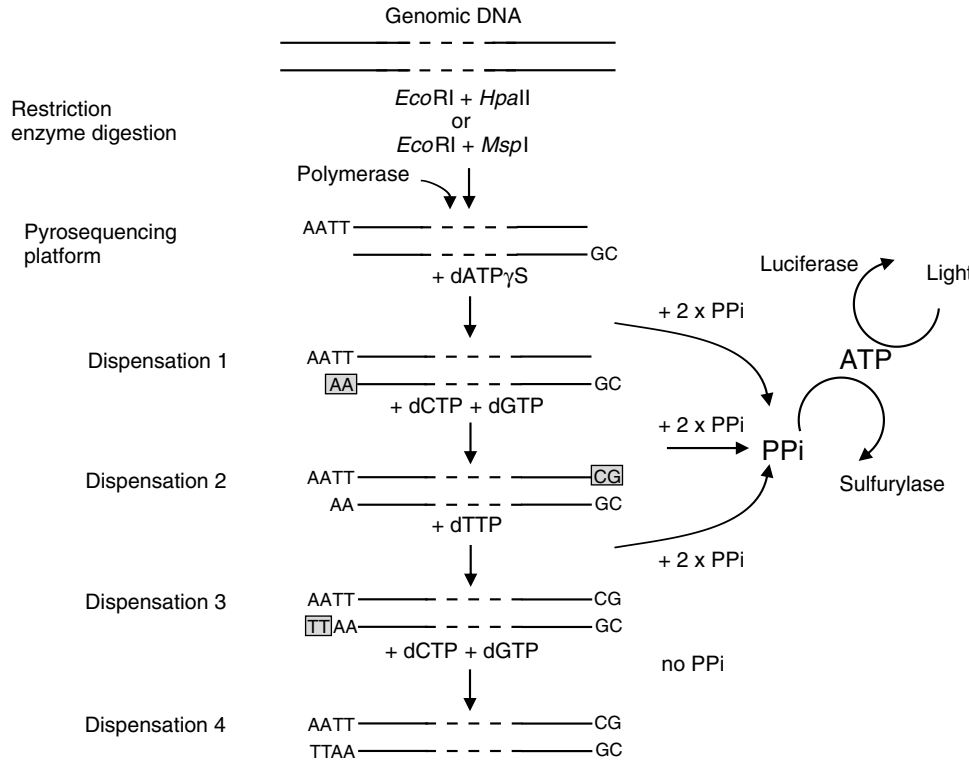


Fig. 1. Illustrative schematic picture of the LUMA assay. Restriction enzymes (either *Hpa*II + *Eco*RI or *Msp*I + *Eco*RI) digest the genomic DNA, whereupon the degree of enzymatic restriction is quantified by a polymerase extension assay based on a four-step Pyrosequencing® reaction. Inorganic pyrophosphate (PPi) is released after each nucleotide incorporation, which is consequently converted to ATP by ATP-sulfurylase. The resulting ATP is then used by luciferase to activate luciferin. The amount of luciferin activation produces a proportional amount of visible light which is detected by a CCD camera. Thus, the number of overhangs produced by the respective restriction enzymes correspond to the amount of light produced.

2. Materials

2.1. Assessment of DNA Quality and Integrity

1. 10× TBE buffer, 0.89 M Tris–borate pH 8.3, 20 mM Na_2EDTA.
2. Standard agarose.
3. Suitable DNA ladder (e.g., 1 kb) to be visualized on the agarose gel.
4. GelRed dissolved in water (Biotium). Ethidium bromide (EtBr) can also be used (see Note 2).
5. UV-table for visualizing agarose gels.
6. NanoDrop for determining DNA concentration and purity.

2.2. Enzyme Restriction of Genomic DNA

1. Restriction enzyme *Hpa*II, 10 U/μl (New England Biolabs).
2. Restriction enzyme *Msp*I, 20 U/μl (New England Biolabs).
3. Restriction enzyme *Eco*RI, 20 U/μl (New England Biolabs).
4. Tango™ buffer 10× (33 mM Tris–acetate, pH 7.9, 10 mM Mg-acetate, 66 mM K-acetate, 0.1 mg/ml BSA) (Fermentas).
5. DNase-free water.
6. Pyrosequencing plates (Qiagen).

2.3. Pyrosequencing® Analysis

1. Pyrosequencing instrument. Other instruments can optionally be used (see Notes 3 and 4).
2. PyroMark Annealing buffer for Pyrosequencing® (Qiagen).
3. Dispensation cartridge for Pyrosequencing® machine (Qiagen).
4. Pyrosequencing Gold Kit (Qiagen).

3. Methods

The LUMA workflow is illustrated in Fig. 2.

LUMA is normally performed using *Hpa*II/*Msp*I. However, other methylation-sensitive restriction enzymes can also be used (see Note 5). The method described below can be performed on all different Pyrosequencing® platforms (see Notes 3 and 4).

3.1. Assessment of DNA Quality and Integrity

1. Since LUMA relies on enzymatically created cuts in the genome, it is vital that the DNA is of high quality. To check the integrity of DNA, it is highly recommended that the samples are run on a 1% agarose gel with either EtBr or GelRed staining prior to LUMA analysis (see Note 2). DNA of sufficient quality will appear as a strong band with high molecular weight, without any visible smear below the band.

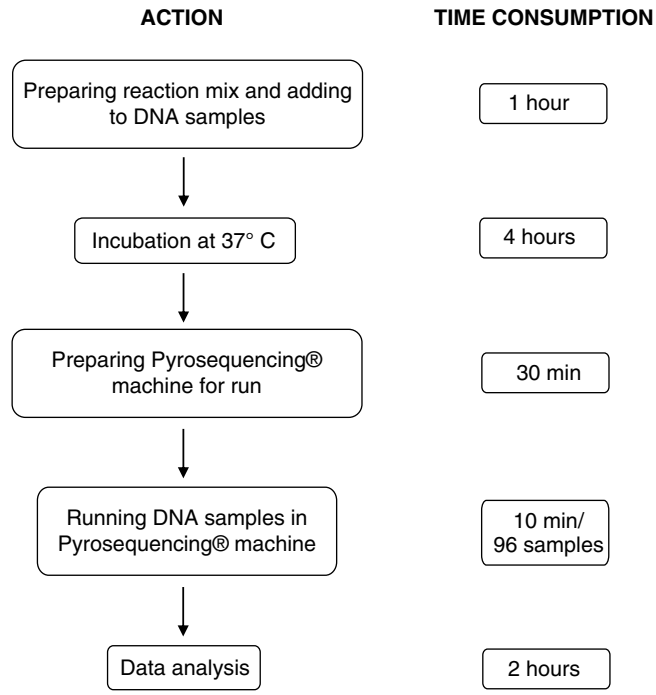


Fig. 2. Flowchart of the LUMA process. Time consumption of each step is indicated.

2. To ensure that the DNA is pure and that the concentration is sufficient, a spectrophotometric measurement is used. DNA samples with protein and/or RNA contamination should not be included in the LUMA analysis, and any samples with a concentration below 50 ng/μl should be avoided.

   Stop point 1: The samples can be frozen at –20°C and used later.

3.2. Preparation of Genomic DNA for LUMA

1. DNA samples that have been tested according to Subheading 3.1 are prepared for LUMA analysis. The preferred concentration is between 50 and 150 ng/μl, but it is not necessary that all samples have equal concentrations since the *Eco*RI internal normalization control will compensate for differences (see step 3 of Subheading 3.5). Samples should be diluted in DNAse-free water.
2. 5 μl of each diluted sample (at a concentration of 50–150 ng/μl) is dispensed onto Pyrosequencing® plates for enzyme restriction. It is highly recommended that all samples are run in duplicates throughout the procedure. Also, keep in mind that all samples are present in two reactions for the enzyme restriction – one for the *Hpa*II restriction and one for the *Msp*I restriction. Including the recommended duplicates, this means that each sample is dispensed in four wells, amounting to a total volume of 20 μl for each sample.

Stop point 2: The samples can be used for LUMA immediately, or the plates can be sealed and stored at –20°C.

3.3. Enzyme Restriction of Genomic DNA

1. Master mixes containing enzymes are prepared for restriction digestion of genomic DNA. One mix (Mix A) contains *Hpa*II, *Eco*RI, Tango buffer, and water, whereas the other (Mix B) has the same ingredients except that *Hpa*II is substituted for *Msp*I. Both mixes are calculated based on a total volume of 20 µl including 5 µl DNA. Master mixes specified below can be scaled up to an appropriate number of samples.
   (a) Mix A is prepared in the following way (volumes per sample):
   13 µl DNAse-free water.
   2 µl 10× Tango™ buffer.
   5 U *Eco*RI (0.25 µl).
   5 U *Hpa*II (0.5 µl).
   (b) Mix B is prepared in the following way (volumes per sample):
   13 µl DNAse-free water.
   2 µl 10× Tango™ buffer.
   5 U *Eco*RI (0.25 µl).
   5 U *Msp*I (0.25 µl).
2. Add 15 µl of either Mix A or Mix B to the 5 µl DNA sample on a Pyrosequencing® plate. Mix by pipetting up and down, seal the plates, and incubate at 37°C for 4 h. No heat inactivation is required after the incubation.

   Stop point 3: The reactions can be frozen at this step, preferably at –20°C.

3.4. Pyrosequencing® Assay

1. Program the instrument by setting up a run in SNP mode. The sequence to analyze should be set as AC/TCGA, which results in an ACTCGA dispensation order for the nucleotides (see Note 6).
2. Add 20 µl of Pyrosequencing® Annealing Buffer to each reaction. This should be done as soon as possible following the enzyme incubation as described in Subheading 3.3.
3. Prepare the nucleotides used in the Pyrosequencing® reaction. The volumes specified below are for 96 samples (1 plate):
   Dilute 50 µl dATPαS with 50 µl ddH_2O.
   Dilute 50 µl dTTP with 50 µl ddH_2O.
   Mix 50 µl dCTP and 50 µl dGTP.
   Add the nucleotides to the Pyrosequencing® cartridge as shown in Fig. 3.

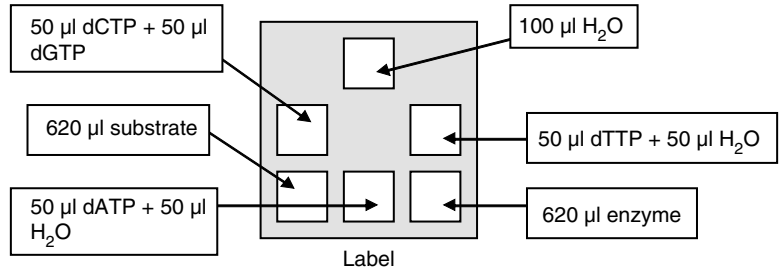


Fig. 3. Addition of reagents to the Pyrosequencing® cartridge, designed especially for LUMA. Note that the label on the cartridge should face front. The volumes given will be sufficient for one 96-well plate.

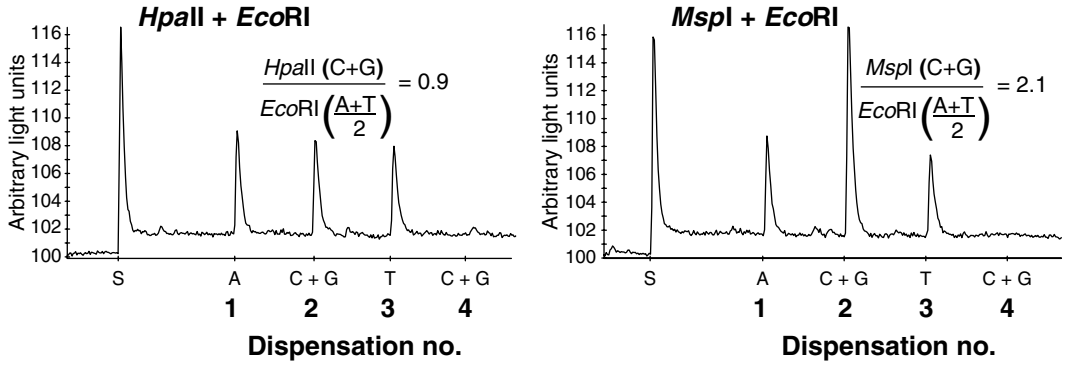


Fig. 4. Typical LUMA results for a sample of human genomic DNA analyzed by a PSQ96 MA Pyrosequencer. The *graphs* show the luminometric output from two representative LUMA runs of human lymphocyte DNA, using *Hpa*II + *Eco*RI, and *Msp*I + *Eco*RI. The A and T peaks denote the additions of dATPαS and dTTP, respectively, which extend all overhangs generated by *Eco*RI. The second peak denotes the second nucleotide dispensation and is an indication of the degree of *Hpa*II or *Msp*I restriction. The first peak, designated "S," is the peak resulting from addition of substrate mix to the enzyme mix, and is always present at the starting point of Pyrosequencing® reactions.

4. Reconstitute the enzyme and substrate mix from the Pyro Gold kit by adding 620 µl ddH_2O to each of the vials. One vial is enough for 96 samples. Add the enzyme mix to the "E" compartment of the Pyrosequencing® cartridge, and the substrate mix to the "S" compartment, as shown in Fig. 3.
5. Start the run after inserting the cartridge and the plate in the instrument. A typical run is displayed in Fig. 4.

3.5. Data Analysis

1. When the Pyrosequencing® run is finished, open the resulting data file and select "AQ mode" for the data analysis. After the analysis is finished, export the peak height data from the "Peak height" menu.
2. Open the exported data in Excel. The peak heights for all six nucleotide dispensations (numbered as 1–6 in the order of dispensation) are specified. First, perform a quality check of your data. The first three peaks, following the substrate peak, should be substantially higher than the last three (CGA),

which should be very small and preferably nonexistent. If these last peaks are high, especially the second C and A dispensations, this usually indicates fragmented DNA (see Notes 7 and 8).

3. To obtain the *Hpa*II/*Msp*I ratio as an indication of degree of methylation, do as follows:
   (a) Calculate the average of peak 1 and peak 3 to get an average peak height for *Eco*RI restriction (see Note 9).
   (b) Calculate the ratio of peak 2 to the averaged *Eco*RI peak for *Hpa*II and *Msp*I digestions separately, to get normalized values for both enzymes (see **Note 10**).
   (c) Calculate the ratio of normalized *Hpa*II to normalized *Msp*I to get the *Hpa*II/*Msp*I-ratio. This value indicates the degree of methylation; the higher the ratio, the lower the amount of methylation (since the *Hpa*II digestion increases with decreasing methylation levels; see Note 11).

4. Notes

1. *Hpa*II recognition sites are distributed throughout the genome. Although *HpaII* sites are enriched 15-fold in CpG islands (24) these sites represent only 12% of the total *Hpa*II sites in the whole human genome (25). Therefore, *Hpa*II/*Msp*I ratios are representative for whole genome methylation.
2. GelRed is generally preferred to EtBr as the latter is carcinogenic, whereas the former is less toxic.
3. Pyrosequencing® instruments may not be available to all research laboratories, but it is possible to apply LUMA using other platforms. Preliminary data in our hands indicate that a 96-well luminometer equipped with an automatic dispenser can substitute the Pyrosequencing® platform. A Luminoskan Ascent microplate Luminometer from Thermo Scientific may replace a Pyrosequencer instrument for running LUMA.
4. Using a high sensitivity Pyrosequencing instrument, less DNA (200 ng) may be used for the analysis.
5. While LUMA as described in this chapter is based on the isoschizomers *Hpa*II and *Msp*I, this will not provide full coverage of CpG sites in the genome since these enzymes only recognize the CCGG sequence. However, LUMA allows the use of other methylation-sensitive restriction enzymes, given that they yield suitable 5′-overhangs. Thus, using enzymes with another recognition sequence in addition to *Hpa*II and *Msp*I, could provide an improved coverage of CpG or CpNpG sites. For example, cytosine methylation in C*C*WGG sequences has been analyzed using the restriction enzymes Psp6I and AjnI (11).

6. The Pyrosequencing® software requires the position of a SNP (single nucleotide polymorphism) to be specified in the dispensation order when running the machine in SNP mode. Therefore, the dispensation sequence includes a "fake" SNP which does not have any effect on the dispensation of the nucleotides.
7. In Pyrosequencing, the dATP nucleotide has been chemically modified to dATPαS. This enables it to become incorporated in the growing DNA strand and yields a light signal, but prevents it from interacting with the ATP-dependent enzyme luciferase. However, the dATPαS is still a weak substrate for luciferase which results in slightly higher A peaks than T peaks, but this is normal and not a source of error.
8. It is not uncommon to see small peaks among the last three (CGA) nucleotide dispensation; this is especially true for peak 6 due to the chemical modification of dATPαS. Minor signals can be tolerated, as long as they are substantially smaller than peaks 1–3; they should be no more than approximately 10% of the peak height seen in peaks 1–3.
9. While it is normal for peak 1 to be somewhat higher than peak 3 due to the chemically modified dATPαS, these two peaks should not differ too much.
10. The theoretical ratio of *Msp*I/*Eco*RI is 2.8–3.0 in mammalian genomes. Therefore, the *Msp*I/*Eco*RI ratio should not exceed 3, as there are no more than three times as many *Msp*I restriction sites as there are *Eco*RI sites. A ratio >3 indicates that the restriction has failed for one or both of the enzymes, or that the DNA was degraded. In genomes from other classes, the ratios may be different.
11. It is recommended that the percentage of methylation is calculated based on the LUMA results with the following formula: Methylation % = 100(1 – HpaII/MspI). This will provide the percentage of methylation and a more clear and logical data presentation.

References

1. Fraga, M. F., and Esteller, M. (2002) DNA methylation: a profile of methods and applications, *Biotechniques* **33**: 632, 4, 6–49.
2. Cedar, H., Solage, A., Glaser, G., and Razin, A. (1979) Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI, *Nucleic Acids Res* **6**: 2125–32.
3. Andersen, C. L., Koch, J., and Kjeldsen, E. (1998) CpG islands detected by self-primed in situ labeling (SPRINS), *Chromosoma* **107**: 260–6.
4. Liang, G., Gonzalgo, M. L., Salem, C., and Jones, P. A. (2002) Identification of DNA methylation differences during tumorigenesis by methylation-sensitive arbitrarily primed polymerase chain reaction, *Methods* **27**: 150–5.
5. Azhikina, T., Gainetdinov, I., Skvortsova, Y., Batrak, A., Dmitrieva, N., and Sverdlov, E. (2004) Non-methylated Genomic Sites Coincidence Cloning (NGSCC): an approach to large scale analysis of hypomethylated CpG patterns at predetermined genomic loci, *Mol Genet Genomics* **271**: 22–32.
6. Yan, P. S., Wei, S. H., and Huang, T. H. (2002) Differential methylation hybridization using CpG island arrays, *Methods Mol Biol* **200**: 87–100.

7. Chen, C. M., Chen, H. L., Hsiau, T. H., Hsiau, A. H., Shi, H., Brock, G. J., Wei, S. H., Caldwell, C. W., Yan, P. S., and Huang, T. H. (2003) Methylation target array for rapid analysis of CpG island hypermethylation in multiple tissue genomes, *Am J Pathol* **163**: 37–45.
8. Ammerpohl, O., Martin-Subero, J. I., Richter, J., Vater, I., and Siebert, R. (2009) Hunting for the 5th base: Techniques for analyzing DNA methylation, *Biochim Biophys Acta* **1790**: 847–62.
9. Pogribny, I., Yi, P., and James, S. J. (1999) A sensitive new method for rapid detection of abnormal methylation patterns in global DNA and within CpG islands, *Biochem Biophys Res Commun* **262**: 624–8.
10. Fujiwara, H., and Ito, M. (2002) Nonisotopic cytosine extension assay: a highly sensitive method to evaluate CpG island methylation in the whole genome, *Anal Biochem* **307**: 386–9.
11. Barres, R., Osler, M. E., Yan, J., Rune, A., Fritz, T., Caidahl, K., Krook, A., and Zierath, J. R. (2009) Non-CpG Methylation of the PGC-1 alpha Promoter through DNMT3B Controls Mitochondrial Density, *Cell Metabolism* **10**: 189–98.
12. El-Maarri, O., Becker, T., Junen, J., Manzoor, S. S., Diaz-Lacava, A., Schwaab, R., Wienker, T., and Oldenburg, J. (2007) Gender specific differences in levels of DNA methylation at selected loci from human total blood: a tendency toward higher methylation levels in males, *Human Genetics* **122**: 505–14.
13. El-Maarri, O., Kareta, M. S., Mikeska, T., Becker, T., Diaz-Lacava, A., Junen, J., Nusgen, N., Behne, F., Wienker, T., Waha, A., Oldenburg, J., and Chedin, F. (2009) A systematic search for DNA methyltransferase polymorphisms reveals a rare DNMT3L variant associated with subtelomeric hypomethylation, *Human Molecular Genetics* **18**: 1755–68.
14. Fish, J. E., Yan, M. S., Matouk, C. C., Bernard, R. S., Ho, J. J. D., Gavryushova, A., Srivastava, D., and Marsden, P. A. (2010) Hypoxic Repression of Endothelial Nitric-oxide Synthase Transcription Is Coupled with Eviction of Promoter Histones, *Journal of Biological Chemistry* **285**: 810–26.
15. Perisic, T., Zimmermann, N., Kirmeier, T., Asmus, M., Tuorto, F., Uhr, M., Holsboer, F., Rein, T., and Zschocke, J. (2010) Valproate and Amitriptyline Exert Common and Divergent Influences on Global and Gene Promoter-Specific Chromatin Modifications in Rat Primary Astrocytes, *Neuropsychopha-rmacology* **35**: 792–805.
16. Pilsner, J. R., Lazarus, A. L., Nam, D. H., Letcher, R. J., Sonne, C., Dietz, R., and Basu, N. (2010) Mercury-associated DNA hypomethylation in polar bear brains via the LUminometric Methylation Assay: a sensitive method to study epigenetics in wildlife, *Molecular Ecology* **19**: 307–14.
17. Romermann, D., Hasemeier, B., Metzig, K., Gohring, G., Schlegelberger, B., Langer, F., Kreipe, H., and Lehmann, U. (2008) Global increase in DNA methylation in patients with myelodysplastic syndrome, *Leukemia* **22**: 1954–6.
18. Sadeghi, R. N., Zojaji, H., Mohebbi, S. R., Chiani, M., Vahedi, M., Mirsattari, D., Molaei, M., Mashayekhi, R., and Zali, M. R. (2009) Evaluation of Global Genome Methylation in Gastritis Lesion and its Correlation With Clinicopatological Findings, *Oncology Research* **17**: 549–58.
19. Thompson, R. F., Atzmon, G., Gheorghe, C., Liang, H. Q., Lowes, C., Greally, J. M., and Barzilai, N. (2010) Tissue-specific dysregulation of DNA methylation in aging, *Aging Cell* **9**: 506–18.
20. Thompson, R. F., Fazzari, M. J., and Greally, J. M. (2010) Experimental approaches to the study of epigenomic dysregulation in ageing, *Experimental Gerontology* **45**: 255–68.
21. Zechner, U., Nolte, J., Wolf, M., Shirneshan, K., El Hajj, N., Weise, D., Kaltwasser, B., Zovoilis, A., Haaf, T., and Engel, W. (2009) Comparative methylation profiles and telomerase biology of mouse multipotent adult germline stem cells and embryonic stem cells, *Molecular Human Reproduction* **15**: 345–53.
22. Karimi, M., Johansson, S., and Ekstrom, T. J. (2006) Using LUMA: a Luminometric-based assay for global DNA-methylation, *Epigenetics* **1**: 45–8.
23. Ronaghi, M. (2001) Pyrosequencing sheds light on DNA sequencing, *Genome Res* **11**: 3–11.
24. Bjornsson, H. T., Sigurdsson, M. I., Fallin, M. D., Irizarry, R. A., Aspelund, T., Cui, H., Yu, W., Rongione, M. A., Ekström, T. J., Harris, T. B., Launer, L. J., Eiriksdottir, G., Leppert, M. F., Sapienza, C., Gudnason, V., and Feinberg, A. P. (2008) Intra-individual change over time in DNA methylation with familial clustering, *JAMA* **299**: 2877–83.
25. Fazzari, M. J., and Greally, J. M. (2004) Epigenomics: beyond CpG islands, *Nat Rev Genet* **5**: 446–55.

Chapter 12

Inhibition of DNA Methylation in Somatic Cells

Angelica M. Giraldo and Kenneth R. Bondioli

Abstract

DNA methylation plays a significant role in the expression of the genetic code and affects early growth and development through its influence on gene expression. DNA methyltransferase 1 (Dnmt1) is the enzyme responsible for maintaining the methylation marks through cell division. However, the de novo methyltransferases, Dnmt3a and Dnmt3b, can also contribute to the maintenance of the methylation pattern. Manipulation of these enzymes, especially Dnmt1, provides a means to alter DNA methylation levels. Manipulation of the DNA methylation pattern of somatic cells will allow a better understanding of the different molecular process associated with chromatin structure and gene expression. Different approaches to artificially manipulate the expression of Dnmt1 in somatic cells include the addition of 5-azacytidine, culture of cells for an extended period of time, and the use of small interfering RNA technologies.

Key words: DNA methylation, DNA methyltransferase 1, siRNA, 5-Azacytidine, Cell culture

1. Introduction

Methylation of DNA plays an important role in the regulation of gene expression, genomic imprinting, chromatin structure, and is essential for mammalian development (1). Manipulation of the DNA methylation marks of somatic cells will allow a better understanding of the different molecular processes associated with chromatin structure, gene expression, and nuclear reprogramming. Additionally, rearrangement of the methylation pattern should improve our understanding of the differentiation-associated cellular changes, and facilitate the manipulation of stem cell differentiation into a desired cell type, or conversely, the de-differentiation of specific cell types into pluripotent stem cells (2, 3).

Methylation of the DNA consists of the covalent addition of a methyl group to the number 5 carbon of the cytosine pyrimidine ring.

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_12, © Springer Science+Business Media, LLC 2011

This reaction is catalyzed by DNA methyltransferase (Dnmt) enzymes and is generally associated with transcriptional silencing due to the inability of transcription factors to bind to methylated DNA sequences. Dnmt1 enzyme is most likely responsible for maintaining the methylation states of sites during cell division. However, recent studies suggest that the de novo DNA methyltransferases, Dnmt3a and Dnmt3b, are also involved in the maintenance of DNA methylation (4).

Global genomic hypomethylation has been achieved by down-regulation of Dnmt1 (1, 5–9). Different approaches to artificially manipulate the gene expression of Dnmt1 in somatic cells have been employed. Gene knock-out models generated by interrupting the Dnmt1 gene have been successfully reported in mice (6, 10, 11). However, for large animals, knockout of Dnmt1 would require the application of low-efficiency homologous recombination technology combined with somatic cell nuclear transfer, making this endeavor very labor intensive, and costly to produce and maintain. In livestock species, expression of Dnmt1 can be down-regulated by the addition of the chemical inhibitor 5-azacytidine (5), culture of cells for extended periods of time (7, 9, 12, 13), and small interfering RNA (siRNA) (1, 7–9, 14). The methodologies described in this chapter have been successfully used to down-regulate DNA methylation levels in bovine and porcine fibroblast cells. However, inhibition of Dnmt3a and Dnmt3b may be necessary to achieve global hypomethylation levels in other species and cell lines.

2. Materials

2.1. Establishment and Maintenance of a Cell Line

1. Rubbing alcohol.
2. Collection and dissection tools: Razors, biopsy punches, scissors, scalpels, blades.
3. Culture supplies: 15- or 50-mL conical tubes, culture dishes, Pasteur pipettes, pipettors, tips.
4. Holding medium: Dulbecco's phosphate-buffered saline (PBS, calcium and magnesium free) containing 100 U of penicillin and 100 μg of streptomycin.
5. Cell culture medium: Dulbecco's Modified Eagle's Medium (DMEM) high glucose containing 10% of fetal bovine serum, 100 U of penicillin, and 100 μg of streptomycin.
6. Trypsin, 0.05% with EDTA.
7. Equipment: Inverted microscope, aspiration pump, CO_2 incubator, centrifuge, and hemocytometer.

2.2. Downregulation of Dnmt1 After Extended Cell Culture

1. Cell culture supplies, media, trypsin, and equipment (see Subheading 2.1).

2.3. Pharmacologic Inhibition of DNA Methylation

1. Cell culture supplies, media, trypsin, and equipment (see Subheading 2.1).
2. 5-Azacytidine stock solution: 100 μM in cell culture medium (see Subheading 2.1). Filter, aliquot, and freeze at −20°C.

2.4. Induction of DNA Hypomethylation Using Dnmt1-Specific siRNA

2.4.1. Optimization of the Transfection Conditions

1. Cell culture supplies, media, trypsin, and equipment (see Subheading 2.1).
2. RNAiFect transfection reagent and EC-R buffer (Qiagen, Cat No. 301605).
3. 3′Fluorescein-labeled nonsilencing siRNA (20 μm): Add 250 μL of the siRNA suspension buffer to the tube containing 5 nM of lyophilized siRNA (Qiagen). Heat the tube to 90°C for 1 min, aliquot, and store at −20°C.
4. Additional equipment: Epifluorescence microscope or flow cytometer.

2.4.2. Dnmt1-siRNA Design and Transfection

1. Cell culture supplies, media, trypsin, and equipment (see Subheading 2.1).
2. RNAiFect transfection reagent and EC-R buffer (see Subheading 2.4.1).
3. Nonsilencing siRNA and Dnmt1-specific siRNA (Qiagen): For preparation see Subheading 2.4.1.

2.5. Analysis of Global DNA Methylation

2.5.1. Immunolabeling of Cells with Anti-5-Methylcytidine

1. Cell culture supplies, media, trypsin, and equipment (see Subheading 2.1).
2. Permeabilization solution: PBS (without calcium and magnesium) containing 1% of BSA (fraction V) and 0.1% of Tween-20.
3. Fixatives: 0.25% of paraformaldehyde (EM grade) in PBS (without calcium and magnesium) prepared on the day of use, and ice-cold absolute methanol (acetone free).
4. Depurination solution: 2 N HCl in PBS (without calcium and magnesium).
5. 0.1 M borate in PBS (without calcium and magnesium).
6. Blocking buffer: PBS (without calcium and magnesium) containing 1% of BSA.
7. Mouse anti-5-methylcytidine (1 ng/mL, Serotec) diluted in blocking buffer.

8. Anti-mouse IgG conjugated with Alexa Fluor® 488 (40 μg/mL) diluted in blocking buffer.
9. Propidium iodide (50 μg/μL) diluted in PBS (without calcium and magnesium) with 1% of BSA.

2.5.2. Quantification of DNA Methylation by Flow Cytometry

1. Washing medium: PBS (calcium and magnesium free) with 0.1% of BSA.
2. Cell strainer of 40 μm.
3. Flow cytometer.

2.5.3. Quantification of DNA Methylation by Epifluorescence Microscopy

1. See Subheading 3.5.1.
2. Glass bottom cell culture dishes.
3. Mounting medium.
4. Epifluorescence microscope.

3. Methods

3.1. Establishment of the Fibroblast Cell Culture

1. Fetuses should be decapitated and eviscerated before processing. If adult animals are used to establish the cell culture, the skin should be shaved and disinfected prior to the collection of the sample. The skin sample, of approximately 10-mm in diameter, can be collected using a biopsy punch or a scalpel. All the instruments used during the procedure should be sterile.
2. Place the skin sample or fetus in a sterile tube containing holding medium. The sample can be processed immediately or refrigerated overnight (see Note 1).
3. Transfer the sample to a culture dish. Finely chop the tissue using scissors or scalpels to about 1-mm cubes. Rinse the pieces with holding medium.
4. Place the pieces onto a tissue culture dish and add just enough culture medium to cover the culture surface of the dish and hold the tissue pieces in place.
5. Culture the explants under 5% CO_2 in air and 90% humidity at 39°C.
6. Check the culture dish 3–5 days later. If the pieces have adhered, add enough culture medium to cover the explants. Cells will start to migrate from the explants 5–10 days after seeding.
7. After 5–7 days in culture, the medium should be changed every 3 days. With a sterile Pasteur pipette attached to a vacuum pump, remove the medium from the dish and add fresh medium.

8. After the primary culture is established and the cells reach 80–100% of confluence, the cell line can be subcultured. Discard the culture medium of the culture dish.
9. Add enough PBS to cover the surface of the dish and dilute the remaining culture medium. Gently, agitate the dish in an effort to rinse the cell culture.
10. Remove the PBS and the explants that come loose during the process. Repeat this process two more times.
11. Add enough prewarmed trypsin to cover the cell culture. Incubate the dish at 39°C for 5 min.
12. Check the dish under an inverted microscope to verify that the cells have de-attached from the bottom of the dish (see Note 2).
13. Transfer the cell suspension into a 15-mL tube containing prewarmed culture medium.
14. Centrifuge the cells at 350 × *g* for 5 min and discard the supernatant.
15. Add enough culture medium to resuspend the cell pellet. Pipette up and down several times to obtain a cell suspension free of cell aggregates.
16. Count cells using a hemocytometer. Seed the desired number of cells into a new culture dish containing fresh medium (see Note 3).
17. Steps 8–16 can be repeated every time the cells reach 80–100% of confluence.

3.2. Downregulation of Dnmt1 in Somatic Cells After Extended Cell Culture

Epigenetic marks can be modified by culturing somatic cells for an extended period of time. Wilson and Jones (12) reported decreased levels of methylated DNA in cells cultured for extended periods. Additionally, a significant decline in mRNA coding for Dnmt1 has been noted in cells at late population doublings (PDs) when compared with cultured fibroblasts at early PDs (9, 13). However, down-regulation of Dnmt1 occurs at different PD for every cell line. Consequently, repeated passages, a standard culture technique, can be used to generate cells containing low concentrations of Dnmt1.

1. After the primary culture is established and the cells reach 80–100% of confluence, a portion of the cells can be used to determine the level of DNA methylation at early passage (see Subheading 3.5), while the remaining cells can be subcultured (see Subheading 3.1).
2. Count cells using a hemocytometer. Seed the desired number of cells (see Note 3) into a new culture dish containing fresh

medium and use the remaining cells for DNA methylation analysis (*see* Subheading 3.5).

3. Passage the cells every time they reach 80–100% of confluence until they reach senescence. A portion of cells can be collected at every passage to determine the level of DNA methylation. However, changes in the level of global methylation are generally more noticeable after PD 30 (see Note 4).

3.3. Pharmacologic Inhibition of DNA Methylation

Several pharmacologic products have been used to inhibit DNA methylation in cells. Cytidine analogs such as 5-azacytidine, 5-aza-2′deoxycytidine, and 5-fluoro-2′-deoxycytidine (15), as well the non-nucleoside inhibitors procaine and procainamide (16) have been successfully used to downregulate levels of DNA methylation.

Several studies have demonstrated that the addition of 5-azacytidine inhibit the activity of Dnmt1 and consequently reduce the methylated DNA content of cells in culture in a dose-depended manner (17, 18). The concentration of 5-azacytidine used in a cell culture is critical and varies between cell types and cell lines. Inappropriate concentration of 5-azacytidine could also target other proteins or enzymes, affect other pathways in addition to the targeted system, and have cytotoxic effects (18). Additionally, 5-azacytidine is degraded by a nucleoside deaminase. Thus, cells that express high levels of this enzyme are less sensitive to this pharmacologic reagent (19).

1. Seed an adequate concentration of cells in at least 12 culture dishes and culture until the cells reach 80% of confluence.
2. At least three different concentrations of 5-azacytidine and three different incubation times should be tested in order to determine the optimal dose and incubation time. We recommend incubating three cell culture dishes with 0.1, 1.0, and 10 μM of 5-azacytidine. At 24, 48, and 72 h postaddition of 5-azacytidine, one dish of every concentration should be used to determine the level of DNA methylation (see Subheading 3.5 and Note 5).

3.4. Induction of DNA Hypomethylation Using Dnmt1-Specific siRNA

Interference RNA (RNAi) technology has become a powerful and widely accepted tool for the analysis of gene function in many organisms including plant, invertebrate, and mammalian cells. The specificity of small siRNA is comparable to gene knock-out experiments but much less time consuming and expensive. RNAi is based on double-stranded RNAs (dsRNAs) that trigger sequence-specific degradation of mRNA. A siRNA, consisting of an in vitro-synthesized 21-base pair oligonucleotide duplex, can mediate RNA interference, and gene knockdown, in a sequence-specific manner in cultured mammalian cells. RNAi technology has been

previously used to reduce the expression of Dnmt1 in human cancer cells, as well as, murine and bovine fibroblasts (7–9, 14). However, the effectiveness of siRNA treatments depends on several factors including RNA-target region, confluence of the cells, and siRNA:liposome ratio during transfection. Then, the optimization of a protocol to transfect siRNA into somatic cells is essential to achieve significant downregulation of Dnmt1 and consequently decrease the levels of DNA methylation. This chapter describes methodologies for the transient transfection of Dnmt1-specific siRNAs. Results from these experiments can be utilized to identify specific target sequences resulting in efficient down regulation. These sequences could then be utilized in the design of short hairpin encoding expression plasmids which can be transfected or delivered by viral infection. These plasmids can also contain selection cassettes and be used to establish stable inhibited cell lines.

3.4.1. Optimization of the Transfection Conditions

1. Seed 3.0×10^4 cells in nine wells of a 24-well plate with standard culture medium with antibiotics added.
2. Incubate the cells for 2 days or until they reach 80% of confluence (see Note 6).
3. Transfect the cells using different ratios of siRNA and transfection reagent (μg:μL; 1:3, 1:6 and 1:9) as well as siRNA doses (0.5, 1.0, and 1.5 μg for 1, 2, and 3 nM, respectively) using a 3′fluorescein-labeled nonsilencing siRNA (see Note 7). First, add the adequate concentration of siRNA in the appropriate volume of buffer EC-R to give a final volume of 100 μL, and mix by vortexing (*see* Table 1).

Table 1
Pipetting scheme for optimizing siRNA transfection

	Ratio of siRNA to lipofection reagent (μg:μL)		
siRNA amounts (μg)	**1:3**	**1:6**	**1:9**
0.5	98.1 μL EC-R 1.9 μL siRNA 1.5 μL liposome	98.1 μL EC-R 1.9 μL siRNA 3 μL liposome	98.1 μL EC-R 1.9 μL siRNA 4.5 μL liposome
1.0	96.1 μL EC-R 3.9 μL siRNA 3 μL liposome	96.1 μL EC-R 3.9 μL siRNA 6 μL liposome	96.1 μL EC-R 3.9 μL siRNA 9 μL liposome
1.5	94.2 μL EC-R 5.76 μL siRNA 4.5 μL liposome	94.2 μL EC-R 5.76 μL siRNA 9 μL liposome	94.2 μL EC-R 5.76 μL siRNA 13.5 μL liposome

4. For complex formation, add the appropriate volume of liposome reagent to the diluted siRNA and vortex (*see* Table 1).
5. Incubate the samples for 15 min at room temperature to allow formation of transfection complexes.
6. Aspirate the cell culture medium from the wells and add 300 μL of fresh medium.
7. Add the complexes to the cells and swirl the plate to ensure uniform distribution of the transfection complexes. Incubate the cells for 6–12 h.
8. Remove the transfection complexes and incubate the cells with cell culture medium.
9. Monitor fluorescence 24–72 h after transfection. Fluorescence can be analyzed using epifluorescence microscopy or flow cytometry.

3.4.2. Dnmt1-siRNA Design and Transfection

1. Use the Dnmt1 transcript sequence to select the target site and design the siRNA. Any available siRNA design software can be used to design the oligonucleotide (see Note 8). Alternatively, Dnmt1 siRNAs have been designed and are commercially available for various species. Additionally, some siRNA sequences have been published and proven to be effective to downregulate Dnmt1 levels in several species and cell lines (1, 7, 9, 14).
2. A nonsilencing siRNA should be included as a negative control.
3. Seed 3.0×10^4 cells in several wells of a 24-well plate with standard medium and antibiotics.
4. Incubate the cells for 2 days or until they reach 80% of confluence.
5. Transfect at least four wells with each of the Dnmt1-specific siRNAs using optimized conditions (*see* Subheading 3.4.1).
6. Analyze the cells at 12, 24, 48, and 72 h post-transfection (*see* Subheading 3.5). Levels of DNA methylation should decrease approximately 24 h post-transfection and remain low for at least one or two PDs. However, the dynamics of the transient downregulation of DNA methylation is dependent on the siRNA and cell line.

3.5. Analysis of Global DNA Methylation

Immunofluorescence using antibodies against 5-methylcytidine has been widely used to characterize the global methylation levels of cells in culture and embryos. Immunolabeled cells can be analyzed using epifluorescence microscopy or flow cytometry. This method can be used for the large-scale screening of genome-wide methylation and has several advantages over other global methylation analysis techniques using extracted genomic DNA such as

chromatography and ELISA-like reactions. Immunolabeling of 5-methylcytidine can be not only used for the analysis of very small and heterogeneous samples, but also the methylation profiles can be evaluated in individual cells or embryos and even in parental sets of chromosomes uniquely visible at syngamy (20).

3.5.1. Immunolabeling of Cells with Anti-5-Methylcytidine

1. Trypsinize the cells as described in Subheading 3.1. Resuspend the cells in culture medium and centrifuge to obtain a pellet.
2. Resuspend the cells with 2 mL of permeabilization solution. Centrifuge and discard the supernatant.
3. Fix the cells with 2 mL of 0.25% paraformaldehyde for 10 min at room temperature. Centrifuge and discard the supernatant.
4. Add 200 µL of PBS. Cool and maintain at 4°C for 10 min.
5. Add 1.8 mL of cold methanol to the cell suspension. Place the tube at –20°C for at least 30 min (see Note 9). Centrifuge and discard the supernatant.
6. Wash the cells with 2 mL of permeabilization solution. Centrifuge and discard the supernatant.
7. Add 2 mL of 2N HCl to the pellet. Incubate the cell suspension at room temperature for 30 min (see Note 10). Centrifuge and discard the supernatant.
8. Neutralize the pH of the solution by adding 2 mL of 0.1 M borate buffer for 5 min at room temperature. Centrifuge and discard the supernatant.
9. Block nonspecific binding by incubating the cells with PBS containing 1% of BSA for a minimum of 30 min. Centrifuge and discard the supernatant.
10. Incubate the cells with 1 ng/mL of mouse anti-5-methylcytidine antibody for at least 1 h at room temperature or overnight at 4°C. Add 2 mL of blocking buffer, centrifuge, and discard the supernatant.
11. Incubate the cells with 40 µg/mL of anti-mouse IgG conjugated with Alexa Fluor® 488 for at least 1 h at room temperature. Add 2 mL of blocking buffer, centrifuge, and discard the supernatant.
12. Incubate the cells with 50 µg/µL of propidium iodide for 10 min at room temperature. Add 2 mL of blocking buffer, centrifuge, and discard the supernatant.

3.5.2. Quantification of DNA Methylation by Flow Cytometry

1. Rinse the immunolabeled cells with 2 mL of washing medium. Centrifuge and discard the supernatant.
2. Add 1–2 mL of washing medium to the cell pellet. Resuspend the pellet by pipetting up and down.

3. If necessary, pass the cells through a 40-μm cell strainer to remove visible cell aggregates.
4. Quantify the level of fluorescence by flow cytometry (see Note 11). WinMDI, a free on line software (http://facs.scripps.edu/software.html), can be used to quantify the fluorescence level of a cell population.

3.5.3. Detection of DNA Methylation Patterns by Epifluorescence Microscopy

1. Seed cells in a glass bottom cell culture dish until the cells reach the desired confluence.
2. Permeabilize, fix, and immunolabel the cells as described in Subheading 3.5.1 with some minor modifications. The cells should remain attached to the dish (no trypsinization required) and all the solutions should be added directly to the culture dish and removed using a Pasteur pipette connected to a vacuum pump. Additionally, all the solutions must be prepared with PBS containing calcium and magnesium.
3. Rinse the immunolabeled cells with 2 mL of washing medium. Aspirate the medium.
4. Add a drop of mounting medium on the center of the dish. Carefully place a coverslip on top.
5. Seal the coverslip with nail polish and analyze the cells by epifluorescence microscopy.
6. Save the pictures as RGB images. The fluorescence level of each cell can be analyzed using ImageJ (http://rsbweb.nih.gov/ij/docs/index.html). This software allows transforming RGB pixels to brightness values for every cell analyzed. Then, brightness values can be compared between cells and treatments.

4. Notes

1. If kept at 4°C, biopsy samples will survive for at least 24 h and even up to 3–4 days, although the longer the time from collection to culture, the lower the likelihood of establishing a healthy cell line.
2. Cells in primary cultures sometimes require longer incubation time in trypsin. If cells are not completely de-attached after 10 min, tap the sides of the dish vigorously.
3. Generally, skin fibroblasts can be subcultured with a split ratio of 1:5 or 1:10. However, some cell lines may require higher or lower initial seeding densities to obtain a satisfactory growth curve.
4. Passage number can be used to calculate the age of the culture; however, this system can be inexact since the number of times

that a cell divides per passage depends of the initial seeding concentration, which can vary greatly between passages. However, measuring the PDs of the culture allows an accurate quantification of the number of times that a cell has divided. PD can be calculated using the following equation: log (final concentration/initial concentration) × 3.33.

5. Although we recommend 0.1, 1.0, and 10 μM of 5-AZA as a starting point for optimization, some cells may be more susceptible or resistant to the effects of this hypomethylation reagent. In this case, broader range of concentrations should be tested.
6. Consistency in the level of confluence at the time of transfection is crucial for the repeatability and effectiveness of the siRNA treatment. siRNA transfection of low cell density cultures can lead to cytotoxic effects, while transfection of over-confluent cells can produce a reduced response to the treatment.
7. These are only general guidelines for liposome-mediated transfection, specific transfection protocols are provided by the liposome supplier.
8. A minimum of four siRNAs, each targeting a different segment of the transcript sequence, should be designed using the Dnmt1 transcript sequence. Studies indicate that regardless of optimized transfection conditions, only a limited fraction of siRNAs appear capable of producing an effective reduction of specific-gene expression in mammalian cells. The biological activity of the siRNAs may be influenced by local characteristics of the target RNA and cellular resistance to siRNA.
9. The cell suspension can be stored at this point for several days without compromising the fluorescence results.
10. Cells become very fragile after the addition of HCl and a great portion of the cells may be lost during this step. To minimize the cells lost, add only 200 μL of HCl to the side of the tube, pipette the cells up and down very gently using a tip of wide orifice. Once the cells are resuspended in the HCl solution, add the remaining volume of HCl.
11. Additionally, cells can be incubated with 50 μg/μL of propidium iodide for 10 min at room temperature to discriminate between the different phases of the cell cycle.

References

1. Giraldo, A. M., Lynn, J. W., Purpera, M. N., Vaught, T. D., Ayares, D. L., Godke, R. A., and Bondioli, K. R. (2009). Inhibition of DNA methyltransferase 1 expression in bovine fibroblast cells used for nuclear transfer. *Reprod. Fertil. Dev.* **21**, 785–795.
2. Koukalova, B., Fojtova, M., Lim, K. Y., Fulnecek, J., Leitch, A. R., and Kovarik, A. (2005). Dedifferentiation of tobacco cells is associated with ribosomal RNA gene hypomethylation, increased transcription, and chromatin alterations. *Plant Physiol* **139**, 275–286.

3. Yeo, S., Jeong, S., Kim, J., Han, J. S., Han, Y. M., and Kang, Y. K. (2007). Characterization of DNA methylation change in stem cell marker genes during differentiation of human embryonic stem cells. *Biochem. Biophys. Res Commun.* **359**, 536–542.
4. Chen, T., Ueda, Y., Dodge, J. E., Wang Z., Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol.* **23**, 5594–5605.
5. Enright, B. P., Kubota, C., Yang, X., and Tian, X. C. (2003). Epigenetic characteristics and development of embryos cloned from donor cells treated by trichostatin A or 5-aza-2'-deoxycytidine. *Biol Reprod* **69**, 896–901.
6. Gaudet, F., Hodgson, J. G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J. W., Leonhardt, H., and Jaenisch, R. (2003). Induction of tumors in mice by genomic hypomethylation. *Science* **300**, 489–492.
7. Leu, Y. W., Rahmatpanah, F., Shi, H., Wei, S. H., Liu, J. C., Yan, P. S., and Huang, T. H. (2003). Double RNA interference of DNMT3b and DNMT1 enhances DNA demethylation and gene reactivation. *Cancer Res.* **63**, 6110–6115.
8. Suzuki, M., Sunaga, N., Shames, D. S., Toyooka, S., Gazdar, A. F., and Minna, J. D. (2004). RNA interference-mediated knockdown of DNA methyltransferase 1 leads to promoter demethylation and gene re-expression in human lung and breast cancer cells. *Cancer Res* **64**, 3137–3143.
9. Giraldo, A. M., Hylan, D. A., Ballard, C. B., Purpera, M. N., Vaught, T. D., Lynn, J. W., Godke, R. A., and Bondioli, K. R. (2008). Effect of epigenetic modifications of donor somatic cells on the subsequent chromatin remodeling of cloned bovine embryos. *Biol. Reprod.* **78**, 832–840.
10. Li, E., Bestor, T. H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926.
11. Lei, H., Oh, S. P., Okano, M., Juttermann, R., Goss, K. A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**, 3195–3205.
12. Wilson, V. L. and Jones, P. A. (1983). DNA methylation decreases in aging but not in immortal cells. *Science* **220**, 1055–1057.
13. Moore, K., Wroclawska, E., Kramer, J. M., and Goica, S. L. (2006). Cell type and culture over time effect DNA methyltransferase 1 expression in bovine donor fibroblast cells. *Reprod. Fertil. Dev.* **19**, 151.
14. Adams, A. M., Pratt, S. L., and Stice, S. L. (2005). Knockdown of the Dnmt1s transcript using small interfering RNA in primary murine and bovine fibroblast cells. *Mol. Reprod Dev.* **72**, 311–319.
15. Jones, P. A., Taylor, S. M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. *Cell.* **20**, 85–93.
16. Villar-Garea, A., Fraga, M. F., Espada, J., Esteller, M. (2003). Procaine is a DNA-demethylating agent with growth-inhibitory effects in human cancer cells. *Cancer Res.* **63**, 4984–4989.
17. Christman, J. K. (2002). 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21**, 5483–5495.
18. Juttermann, R., Li, E., and Jaenisch, R. (1994). Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proc. Natl. Acad. Sci. U. S. A* **91**, 11797–11801.
19. Li, L. H., Olin, E. J., Buskirk, H. H., Reineke, L. M. (1970). Cytotoxicity and mode of action of 5-azacytidine on L1210 leukemia. *Cancer Res.* **30**, 2760–2769.
20. Santos, F., Dean, W. (2006). Using immunofluorescence to observe methylation changes in mammalian preimplantation embryos. *Methods Mol Biol.* **325**, 129–37.

Chapter 13

DNA Methyltransferase Assays

Renata Z. Jurkowska, Alexandre Ceccaldi, Yingying Zhang, Paola B. Arimondo, and Albert Jeltsch

Abstract

DNA methyltransferases are important enzymes and their inhibition has many potential applications. The investigation of DNA methyltransferases as well as screening for potential inhibitors requires specialized enzyme assays. In this chapter, we describe three DNA methyltransferase assays, each of them based on a different method: (1) An assay using radioactively labeled AdoMet and biotinylated DNA substrates that is ideal for enzymatic characterization of these enzymes. (2) An assay using bisulfite conversion of in vitro methylated DNA that is ideal to determine details of the methylation pattern introduced by DNA-(cytosine C5)-methyltransferases. (3) A novel fluorescence-coupled, restriction-based assay suitable for high-throughput screening of DNA methyltransferase inhibitors.

Key words: DNA methylation, DNA methyltransferase, Enzyme assay, Bisulfite sequencing, High-throughput screen, Methyltransferase inhibitors

1. Introduction

DNA methylation is an important and essential modification of DNA present in most Prokaryotes and Eukaryotes that has many biological functions (see refs. 1–6 for review). In Prokaryotes, DNA methylation is used to coordinate DNA replication and cell cycle, to direct postreplicative mismatch repair, and to distinguish between self and foreign DNA. In Eukaryotes, it is involved in gene regulation, maintenance of genome integrity, X-chromosome inactivation, and regulation of development. Erroneous DNA methylation contributes to the development of various human diseases including cancer (7–10). DNA methyltransferases (MTases) catalyze the transfer of a methyl group to the N^6 position of adenine, N^4 position of cytosine, and C^5 position of cytosine by using *S*-adenosyl-L-methionine (AdoMet) as a donor for an

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_13,

activated methyl group (see refs. 2, 11 for review). In mammals, only the cytosine C^5 methyltransferases are present. Because of its diverse biological functions, methods for analysis of DNA methylation are very important and deserve continuous refinement. Several assay systems have been developed to study the activity of DNA MTases (see refs. 12, 13 for review), including digestion of DNA by methylation-sensitive restriction enzymes (14–16), bisulfite conversion for detection of 5-methylcytosine (17, 18), or the separation and quantitative determination of modified nucleosides by HPLC (19–21). Another class of methylation assays relies on the use of AdoMet that carries a radioactive label on its methyl group that is transferred to the DNA by the MTase. These assays require a separation of methylated DNA and unused cofactor which can be achieved by (1) spotting the reaction mix onto a DE-cellulose filter sheet (22); (2) coupling of DNA to cellulose (23); (3) thin layer chromatography (24); or (4) using biotinylated DNA bound to avidin-coated plates or beads (25). In this chapter, we describe one assay for each of these types:

- An assay using radioactively labeled AdoMet that is ideal for enzymatic characterization of different DNA MTases.
- An assay using bisulfite conversion of in vitro methylated DNA that is ideal to determine details of the methylation pattern introduced by DNA-(cytosine-C5)-methyltransferases.
- A novel fluorescence-coupled, restriction-based assay suitable for high-throughput screening of DNA MTase inhibitors.

1.1. Radioactive Biotin Microplate DNA Methylation Assay

The biotin–avidin microplate assay is a sensitive method to measure methylation of biotinylated substrates by DNA MTases (25, 26). The outline of the assay is shown in Fig. 1. The methylation reaction is carried out in solution using radioactively labeled [methyl-^{3}H]-AdoMet and a biotinylated DNA substrate (a double stranded oligonucleotide or a PCR product). During the reaction, the radioactively labeled methyl group is transferred from the cofactor to the DNA. Afterward, the biotinylated DNA is immobilized on the surface of an avidin-coated microplate. During this step, the incorporation of new radioactivity into the DNA is quenched by the addition of an excess of unlabeled AdoMet in the binding buffer. In the next step, the unreacted and the enzyme-bound cofactor are removed by intensive washing with a high salt buffer. Finally, the bound DNA is digested with a nonspecific nuclease to release the incorporated radioactivity. After digestion the radioactivity in the solution is measured by liquid scintillation counting to quantify the amount of methyl groups enzymatically transferred to the DNA. Since the capacity of the plate and the sensitivity of the scintillation reaction can vary with conditions, each plate should contain a sample of completely methylated DNA that can be used for calibration of the results.

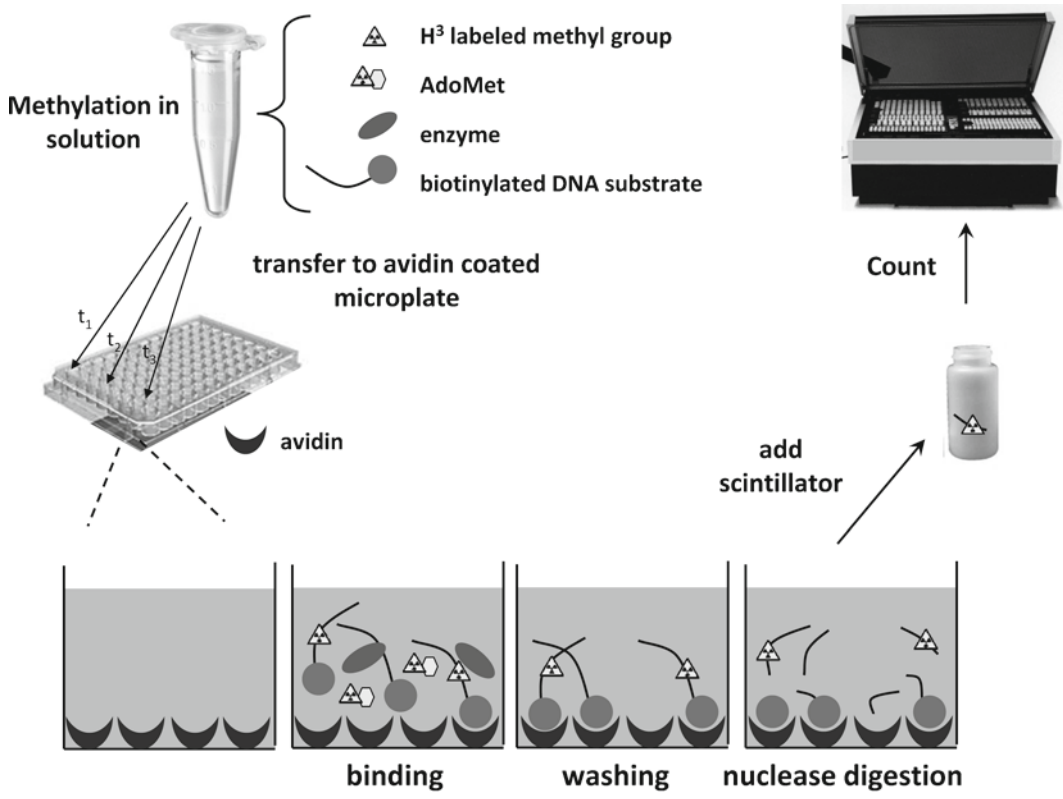


Fig. 1. Schematic drawing of the biotin/avidin microplate DNA methylation assay described here.

This assay is a very sensitive in vitro method that provides kinetic resolution and allows measuring low amounts of DNA methylation in a fast, inexpensive, robust, and accurate way. Additionally, it can be used with both adenine and cytosine DNA methyltransferases. This method has many advantages in comparison to other published protocols: first, the detection of [^{3}H] by liquid scintillation counting is highly efficient, which makes the assay very sensitive and allows the detection of low amounts of DNA methylation, at a level of smaller than 0.1% of total methylation of the DNA. Second, the background of radioactivity is low, because of the efficient removal of unreacted AdoMet. Third, the results are reliable, as they can be accurately reproduced with small deviations of (±10%). Thus, the assay provides quantitative data of high accuracy and reproducibility. It also gives a good kinetic resolution, because methylation can be measured at many time points for each reaction. Furthermore, it exploits the microplate format to process many samples in parallel for washing and readout in a fast and inexpensive way.

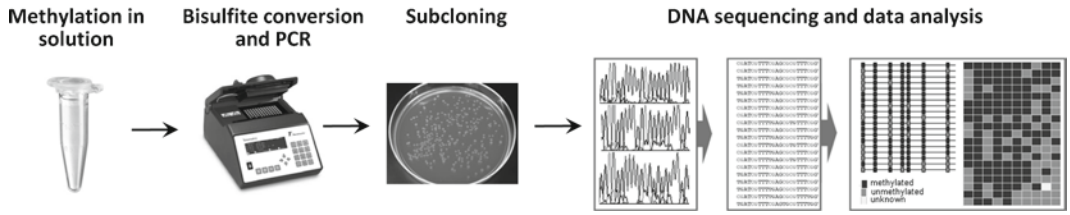


Fig. 2. Outline of the bisulfite analysis of the in vitro methylated DNA method.

1.2. Bisulfite Analysis of In Vitro Methylated DNA

The bisulfite sequencing method is based on the selective deamination of cytosine residues, but not 5-methylcytosines, by treatment of DNA with sodium bisulfite, and the sequencing of subcloned samples generated by PCR with primers specific for bisulfite-treated DNA. (17, 27, 28). This method is commonly used to study the methylation patterns after isolation of genomic DNA from cells. However, it can also be used for studying the methylation activity of DNA-(cytosine C5)-MTases in vitro. Figure 2 shows the outline of the method. First the substrate DNA is in vitro methylated by the methyltransferase in the presence of the cofactor AdoMet. The reaction is stopped by flash freezing the sample in liquid nitrogen and digestion with proteinase K. The DNA is then purified and subjected to bisulfite conversion. In the presence of NaOH and sodium bisulfite, the substrate DNA molecules are denatured to single-stranded DNA and the unmethylated cytosines are chemically converted to uracil. In the next step, the DNA sequence under investigation is amplified by PCR with primer pairs specific for the upper or lower strand of the bisulfite converted DNA (after bisulfite conversion the upper and lower strands of the DNA are no longer complementary). During the PCR reaction, uracil will be amplified as thymine but 5-methylcytosine will yield cytosine. The amplified PCR product is a pool of DNA molecules and each of them could have a unique methylation pattern. Therefore, the PCR product is subsequently subcloned and several or many randomly selected clones are sequenced. From the sequencing results, it can be distinguished if the cytosines in the substrate DNA were methylated or not. A variation of this method, called hairpin bisulfite (29, 30), allows determination of the methylation pattern in both strands simultaneously, as it contains an additional ligation step, in which a hairpin loop is ligated to the substrate DNA to couple both its strands.

The most important advantage of the bisulfite analysis is that it provides precise information about the methylation state of each individual site. This is of relevance if the substrate DNA contains more than one potential sites of methylation. However, it is relatively time consuming and expensive and, therefore, it cannot give good kinetic resolution. Therefore, it is not commonly used for

checking the activity of the MTases, but rather to determine their specificity, flanking sequence preferences or processivity. Additionally, it cannot be used with adenine methyltransferases.

1.3. Restriction-Based Fluorescence Assay

The restriction-based fluorescence assay is a novel method developed recently for screening of MTase inhibitors. In this method, double-modified DNA substrates containing a fluorophore at one end and biotin at the other end on the other strand are used. The substrate contains a target site (CpG in the case of human DNA MTases) overlapping with a site for a methylation-sensitive restriction enzyme, like HhaI cleaving GCGC or HpaII cleaving CCGG which both are inhibited by the methylation of the inner cytosine (Fig. 3). The substrate DNA is immobilized on an avidin-coated 96-well plate and used for the methylation reaction in the presence of a potential inhibitor compound. Two enzymatic reactions are achieved sequentially: DNA methylation and restriction digestion with a methylation-sensitive enzyme. When methylation occurs, the restriction enzyme cannot cleave the DNA in the following step and the fluorescence signal is preserved after final washing. In contrast, if an inhibitor molecule blocks the methylation reaction,

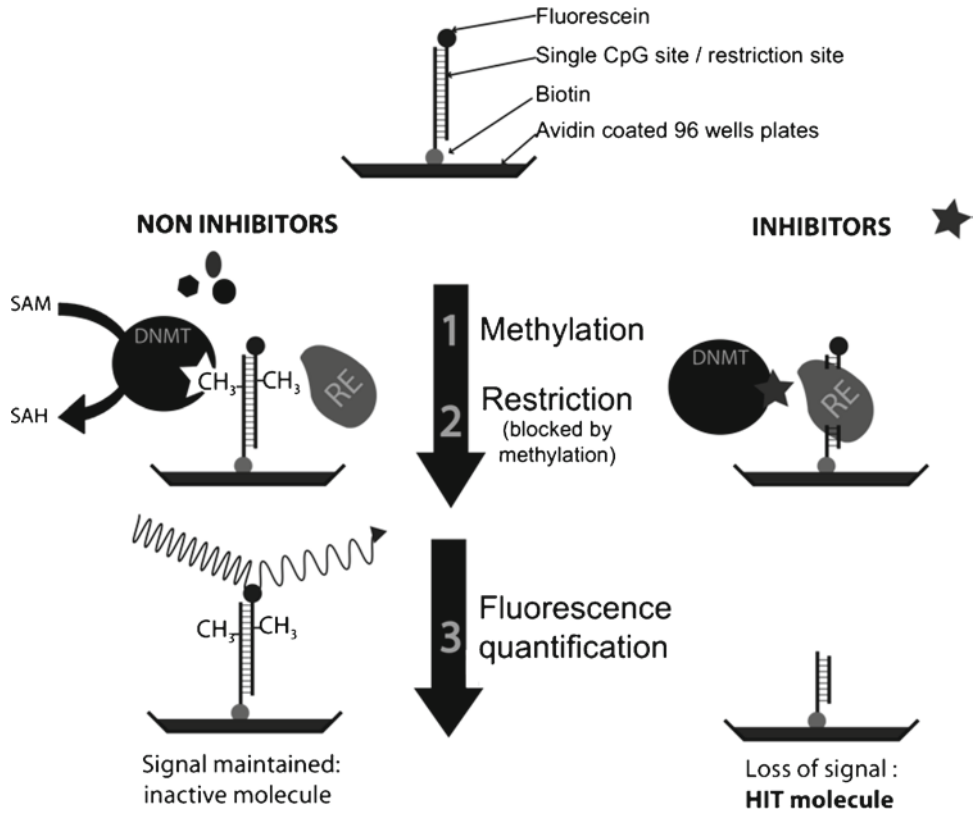


Fig. 3. Principle of the restriction-based fluorescent assay designed for screening libraries of inhibitors.

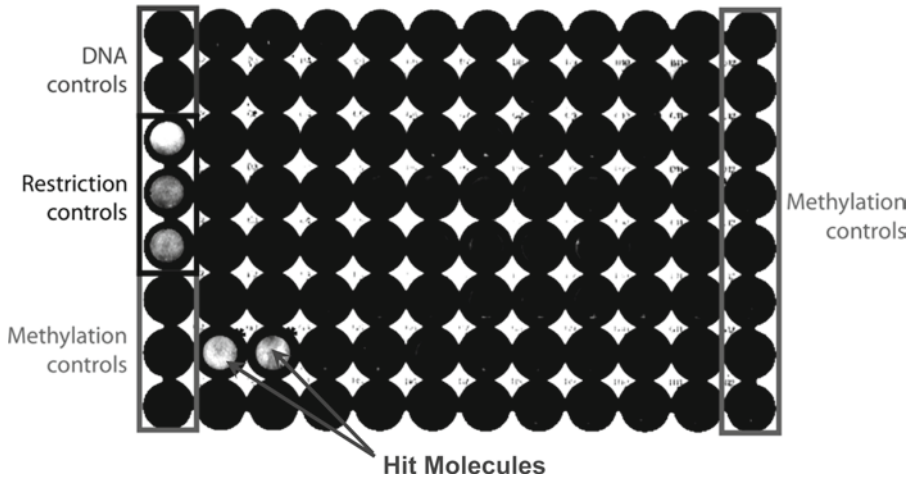


Fig. 4. An example of a screening plate for 80 molecules on Dnmt3a/Dnmt3L complex. For details refer to the text.

the DNA will be cleaved and the fluorescence signal will be lost or depleted after the final washing step. Figure 4 shows an example of a typical primary screening result of 80 compounds, which are candidates to inhibit the Dnmt3a/Dnmt3L methyltransferase complex. In this example, two hit molecules are easily identified. Three types of controls are used to ensure the reliability of each experiment. The maximal fluorescence signal is given by DNA controls that do not undergo any treatment except the washing steps. Restriction of unmethylated DNA indicates the background fluorescence signal. Methylation controls undergo all the enzymatic steps and allow checking the activity of the MTases during the assay; they also correspond to the fluorescence signal of inactive compounds. During the setup phase of the assay, isoschizomeric restriction enzymes not sensitive to methylation may be used to ensure that a loss of cleavage is indeed related to DNA methylation. Additionally, known active inhibitors like sinefungin (an AdoMet analog) can be used on the plate as additional controls.

In our experience, experiments showing DNA protection by methylation of at least 50% of the substrate allow a reliable identification of hit compounds. For data analysis, the Z factor is computed, which is a simple statistical parameter very useful to validate the overall quality of an high-throughput screen, by resuming assay dynamics and signal-to-noise ratio (31). The Z factor is defined as $1 - \frac{3 \times (SD_{drugs} + SD_{restri})}{(\mu_{drugs} + \mu_{restri})}$ in which SD_{drugs} and SD_{restri} are the observed standard deviations of the signals of the tested drugs and restriction controls, respectively, and μ_{drugs} and μ_{restri} are the means of the corresponding fluorescence signals. The Z' factor is based only on

the control wells and serves as an intrinsic quality parameter of the assay, without any assumptions like the number of expected hits. It is defined as $1-\frac{3\times(SD_{meth}+SD_{restri})}{(\mu_{meth}+\mu_{restri})}$, in which μ_{meth} and SD_{meth} are the mean signal corresponding to methylation controls and the associated standard deviation. Experiments showing a Z' factor < 0.20 are ignored. Larger screening campaigns using this assay showed an excellent global Z factor > 0.50, including the signals of the tested compounds (31). The optimization of these parameters is a crucial issue when it comes to high-throughput strategies. This means that experimental noise needs to be reduced to a minimum, by ensuring the optimal work of the automated pipetting system, but also the appropriate treatment of all enzymes, for example, by avoiding long preincubation times under conditions where the enzyme is not fully stable.

The restriction-based fluorescent assay is a robust, versatile, and reliable method that provides a direct read-out of the results for up to 800 compounds (ten plates) in a single run on a fluorescence scanner (Typhoon, GE Healthcare). It can be used as a screen to find new DNA MTases inhibitors and also as a quantitative assay to determine apparent IC_{50} (concentration of drug needed to obtain 50% of inhibition) of the hit molecules by testing a range of concentration for each molecule. A typical assay with Dnmt3a/Dnmt3L only lasts 1 h 30 min. Importantly, since the design includes the immobilization of the DNA substrate on the microplate surface, which allows washing after each step, it is fully compatible with all kinds of tested chemical compounds, including intrinsically fluorescent molecules or compounds that would inhibit the restriction step. The screen can lead to the discovery of new DNA MTase inhibitors with various modes of actions: substrates competition (DNA and/or AdoMet competition), noncompetitive inhibition, conformational changes and allosteric effects, inhibition of protein–protein interactions, etc. It can be applied to screens with purified MTases or cellular extracts. Finally, this assay is cheap (<0.5 €/tested compound), inoffensive (no use of radioactivity or toxic chemicals) and can be easily automated. According to preliminary tests, up-scaling to 384-wells plates is also possible. The manual throughput for a single operator is about 500 tested compounds per day with 96-wells plates. Automation drastically increases this productivity, making the assay suitable for screening of large libraries of inhibitors. One critical parameter of this method is the design of the DNA duplex, because it must meet the requirements of the MTase of interest and the restriction enzyme at the same time. In our study, the choice of the DNA sequence was crucial and was based on previous results on the DNA sequence preferences of the enzymatic complex. Accordingly, the restriction enzyme has to be selected in order to recognize the cleavage site

and respond to methylation in the overlapping MTase site. However, many restriction enzymes are commercially available that are sensitive to methylation state of target sites as short as 4 bp, providing a wide choice of alternatives.

2. Materials

2.1. Radioactive Biotin Microplate Assay

1. Microplates (e.g., transparent plates: E.I.A./R.I.A. plate, flat bottom, high binding, Cat. No. 9018, Costar Corp., Cambridge, MA, USA; white plates: LIA-plates, flat bottom, high binding, Cat. No. 655074, Greiner bio-one; or comparable product from any other manufacturer); store coated plates at 4°C.
2. Avidin (Sigma) 2.5 mg/ml in ddH_2O; store at 4°C.
3. Unlabeled AdoMet (Sigma); 10 mM solution dissolved in 10 mM H_2SO_4; store in small aliquots at –20°C.
4. Coating buffer: 100 mM $NaHCO_3$ pH 9.6.
5. Washing buffer: PBST: 140 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM K_2HPO_4, 0.05% Tween 50, pH 7.2, supplemented with 500 mM NaCl.
6. Multichannel pipette.
7. Biotinylated oligonucleotide substrates; anneal and prepare a stock solution that is stored at –20°C. Longer DNA substrate may also be used, for example, PCR products prepared using biotinylated primers.
8. Reaction buffer for the methylation reaction, composition depends on the enzyme studied.
9. [Methyl-^{3}H]-AdoMet (3.22 TBq/mmol, Perkin Elmer), store at –20°C in small aliquots.
10. Nonspecific endonuclease, e.g., *Serratia marcescens* nuclease in 1× Serratia buffer: 50 mM Tris–HCl pH 8, 5 mM $MgCl_2$; the enzyme is commercially available as Benzonase from Merck, alternatively DNaseI might be used as well.
11. 0.05% HCl.
12. Liquid scintillator solution Rotizint® eco plus (Carl Roth GmbH + co. KG, Karlsruhe, Germany) or Microscint 20 solution (Perkin Elmer) if the microplate counter is used for readout.
13. Liquid Scintillation Counter and counter vials or Microplate counter like Top Count.

2.2. Bisulfite Analysis of In Vitro Methylated DNA

Methylation reaction

1. Reaction buffer: the composition of the reaction buffer depends on the enzyme studied.
2. Unlabeled AdoMet (Sigma); 10 mM solution in 10 mM H_2SO_4; store in small aliquots at −20°C.
3. Substrate DNA: typically a longer DNA substrate amplified by PCR is used (see Note 9).
4. Liquid nitrogen.
5. Proteinase K (NEB), 20 mg/ml, NEB buffer 2.
6. DNA purification kit (e.g., Nucleospin Extract II, Machery and Nagel or ChargeSwitch® PCR Clean-Up Kit, Invitrogen).

Primers design for bisulfite converted DNA

1. BiSearch: http://bisearch.enzim.hu/.
2. MethPrimer: http://www.urogene.org/methprimer/index1.html.

Bisulfite conversion of in vitro methylated DNA

1. Freshly prepared 2 M NaOH and 0.3 M NaOH in sterile distilled H_2O.
2. Solution I: 1.9 g $NaHSO_3$ (Sigma) is dissolved in the mixture of 2.5 ml sterile water and 750 µl 2 M NaOH (see Note 10). It needs to be freshly prepared.
3. Solution II: 98.6 mg of 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Sigma) is dissolved in 2 ml dioxane. It needs to be freshly prepared.
4. 1× TE buffer: 10 mM Tris–HCl, pH 8, 1 mM EDTA.
5. Microcon Ultracel YM-50 columns (Millipore).

PCR and gel electrophoresis (agarose and polyacrylamide)

1. HotStarTaq DNA Polymerase (Qiagen).
2. Gel Extraction Kit (Nucleospin Extract II, Machery and Nagel or QIAquick Gel Extraction Kit, Qiagen).
3. ChargeSwitch® PCR Clean-Up Kit (Invitrogen).
4. 1.2% agarose gels in 1× TPE.
5. PAGE running buffer: 10× TPE buffer: Tris–HCl 0.9 M, EDTA 20 mM, pH 8.2 adjusted with H_3PO_4.
6. For PAGE: Rotiphorese gel 40 (40% solution of acrylamide/bis-acrylamide 29:1, Roth), *N*,*N*,*N*′,*N*′-tetramethylethylenediamine (TEMED, Roth) should be stored at 4°C. 10% ammonium persulfate (APS): prepare 10% solution in water and store the aliquots at −20°C.

Cloning

1. Luria–Bertani (LB) medium: 1.0% NaCl, 1.0% tryptone, and 0.5% yeast extract, pH 7.0.
2. LB agar (1.5%) plates containing ampicillin (100 mg/ml).
3. StrataClone™ PCR Cloning Kit (Agilent-Stratagene).

Sequencing

1. 3730×l ABI 96-capillary sequencer systems (ABI Weiterstadt).
2. ABI BigDye Terminator kit vers. 3–1 (ABI Weiterstadt).

Result analysis and presentation

1. Chromas: http://www.technelysium.com.au/.
2. FinchTV (Geospiza): http://www.geospiza.com/finchtv/.
3. BISMA http://biochem.jacobs-university.de/BDPC/BISMA (32).

Data compilation and presentation: BDPC http://biochem.jacobs-university.de/BDPC/ (33, 34).

2.3. Restriction-Based Fluorescence Assay

1. Microplates (e.g., E.I.A./R.I.A. plate, flat bottom, high binding, Cat. No 9018, Costar Corp., Cambridge, MA, USA; or comparable product from any other manufacturer); store coated plates at 4°C.
2. Greiner™ 96-well V-form microplate for the preincubation of MTase with inhibitor.
3. Avidin (Sigma) 2.5 mg/ml in ddH_2O; store at 4°C.
4. Unlabeled AdoMet (Sigma); 10 mM solution in 10 mM H_2SO_4; store in small aliquots at –20°C.
5. Coating buffer: 100 mM $NaHCO_3$ pH 9.6.
6. Washing buffer PBST: 140 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM K_2HPO_4, 0.05% Tween 50, pH 7.2; supplemented with 500 mM NaCl.
7. Multichannel pipette.
8. The oligonucleotide substrates biotinylated at one end and fluorescently labeled on the other end; anneal and prepare a stock solution that is stored at –20°C.
9. Reaction buffer for the methylation reaction, composition depends on the enzyme studied; for Dnmt3a and Dnmt3L experiments we used 20 mM HEPES pH 7.2, 50 mM KCl, 1 mM EDTA.
10. Inhibitors stocks.
11. Methylation-sensitive restriction enzyme and the appropriate digestion buffers can be identified at Rebase (http://rebase.neb.com).

12. As an example, substrates designed for the assay with DNMT3a and DNMT1 have the following sequences (the CpG site is in bold):

    DNMT3a: GCTATATATA**CG**TACTGTGAACCCTACCAG ACATGCACTG

    DNMT1: GCATATATATGA**CG**ATCCTGTAGGTCACTAC CAGACATGCACTG (used in hemimethylated form with the methylation in the upper strand)

    Methylation was probed with HpyCH4IV and BfuCI, respectively.

3. Methods

3.1. Radioactive Biotin Plate Assay

1. Preparation of plates: coat the microplates with avidin; for each 96-well plate combine 40 µl of the avidin solution and 10 ml of the coating buffer (see Note 1), pipette 100 µl per well and incubate overnight at 4°C. Coated plates can be used for up to 2 weeks if stored at 4°C.
2. Before use, wash the wells five times with 200 µl PBST supplemented with 500 mM NaCl to remove unbound avidin (see Note 2).
3. Prepare the binding buffer consisting of 5 µl of 10 mM unlabeled AdoMet in 35 µl PBST supplemented with 500 mM NaCl per well (see Note 3) and distribute it in each well of the microplate to quench the incorporation of [^{3}H] into the DNA after the methylation reaction.
4. Methylation reactions are carried out in 10–50 µl in a reaction tube. Typically, 0.1–10 µM biotinylated oligonucleotide and 1 nM to –10 µM enzyme are used in a buffer adapted to the needs of the enzyme to be studied, e.g., 100 mM HEPES, pH 8.0, 50 mM NaCl, 1 mM EDTA, 0.5 mM DTT, 200 ng/µl BSA in the presence of 0.75 µM labeled [methyl-^{3}H]-AdoMet (3.22 TBq/mmol, Perkin Elmer) (see Note 4). Methylation reactions can be carried out at different temperatures. The reaction can be started by the addition of enzyme, DNA, or coenzyme, depending on the purpose of the experiment.

   To measure a time course of methylation (see Note 5) remove aliquots of 1–5 µl from the reaction mixture at each time point and pipette them into the wells of the microplate that contain the binding buffer with an excess of unlabeled AdoMet (see Note 3) to quench the incorporation of new [^{3}H] into the DNA. After the last time point, incubate the plate for 30 min at room temperature to allow binding of the biotinylated substrates to the avidin on the microplate.

Prepare a fully methylated sample by using the same amount of substrate DNA and AdoMet, but incubate with a high amount of the MTase for a long time. This sample may be applied on each plate for calibration.

5. Wash the wells five times with 200 µl PBST supplemented with 500 mM NaCl to remove unreacted AdoMet and the enzyme (see Notes 2 and 6).
6. Digest the immobilized DNA by adding 0.7 µg *S. marcescens* nuclease (see Note 7) in 1× Serratia buffer per well and slowly shake the microplate for 30 min at room temperature (see Note 8).
7. After digestion, the solution from each well of the microplate is transferred into a scintillation vial. Each sample is mixed with 2 ml liquid scintillator solution and subjected to liquid scintillation counting to quantify the amount of methyl groups transferred to the DNA.
8. Alternatively, if the Top Count scintillation counter for the microplate format is available, digest the immobilized DNA by adding 0.7 µg *S. marcescens* nuclease in 40 µl of 1× Serratia buffer per well, transfer the solution into fresh microplate after digestion, add 160 µl of MicroScint solution to each well, and count the plate after shaking (see Note 8).

3.2. Bisulfite Analysis of In Vitro Methylated DNA

Primer design

1. Design a pair of primers in the region of interest. The length of the PCR products to be amplified is suggested not to exceed 500 bp (see Note 11). BiSearch (35) and MethPrimer (36) are two online primer-designing programs with unique properties of designing primers for bisulfite converted DNA (see Note 12).

In vitro methylation of substrate DNA

1. Prepare a reaction mix containing the methylation buffer, the substrate DNA, unlabeled AdoMet, and the enzyme of interest and incubate for required amount of time (see Note 13). In parallel, prepare a control sample having the same composition but without the enzyme or containing inactive enzyme (mutant or heat-inactivated enzyme), this sample will serve as conversion control for the bisulfite treatment (see Note 14).
2. Stop the reaction by flash freezing the sample in liquid nitrogen.
3. Treat the sample with proteinase K (3 mg/ml) in NEB buffer 2 for 2–3 h at 37°C to degrade the protein and purify the DNA over a column using a PCR/DNA purification kit, elute the sample in TE buffer.
4. Determine the concentration of the DNA by measuring the absorbance at 260 nm.

Bisulfite conversion of methylated DNA

1. Add 187 µl solution I to the 20 µl (20–100 ng) of your in vitro methylated DNA to denature the DNA and convert unmethylated cytosine residues into uracil. Mix by pipetting up and down.
2. Add 73 µl solution II to the mixture and gently mix by pipetting up and down.
3. Incubate the mixture in a thermocycler using the following conditions: 15 min at 99°C, 30 min at 50°C, 5 min at 99°C, 1.5 h at 50°C, 5 min at 99°C, 1.5 h at 50°C (see Note 15).
4. Add 150 µl sterile distilled H_2O to the mixture. After mixing by pipetting up and down, transfer the reaction mixture to the Microcon Ultracel YM-50 columns. Place the column in a collection tube (provided in the kit) and centrifuge at 14,000 × *g* for 15 min (see Note 16).
5. Carefully separate the column from the collection tube. Discard the filtrate. Place the YM-50 column back into the same collection tube.
6. Add 500 µl 1× TE buffer and centrifuge at 14,000 × *g* for 10 min. Separate the column from the collection tube. Discard the filtrate. Place the YM-50 column back into the same collection tube.
7. Add 500 µl 0.3 M NaOH to the column and incubate at room temperature for 10 min to desulfonate the DNA. Afterward centrifuge at 14,000 × *g* for 10 min. Discard the filtrate.
8. Add 500 µl TE buffer, centrifuge at 14,000 × *g* for 10 min. Discard the filtrate.
9. Place the column upside-down in a new collection tube provided. Add 50 µl 1× TE (50°C) in the middle of the sample reservoir. Incubate at room temperature for 1 min and centrifuge at 1,000 × *g* for 10 min (see Note 17).
10. Collect the DNA in the collection tube and measure the concentration of the single-stranded bisulfite-treated genomic DNA by UV spectrometry. Store it at –20°C (see Note 18).

PCR

1. 1–2 µl of the bisulfite converted DNA is used as template for PCR in a 25-µl reaction mixture (1× PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.4 µM of each primer, and 2.5 U of HotStarTaq polymerase). Perform PCR with the following program: 15 min at 95°C, 5× (30 s at 94°C, 30 s at 65°C, 90 s at 72°C), 5× (30 s at 94°C, 30 s at 60°C, 90 s at 72°C), 35× (30 s at 94°C, 30 s at 55°C, 90 s at 72°C), 5 min at 72°C (37) (see Note 19).

2. After PCR, 5 µl of the PCR product is electrophoresed on a 1.2% agarose gel. For better resolution an 8% PAGE gel can be prepared. Prepare a 1.5-mm thick gel by mixing 1 ml of 10× TPE buffer, 7 ml of deionised water, 2 ml acrylamide (40%), 25 µl TEMED, and 25 µl APS (10%) (see Note 20).
3. Stain the gel and visualize the PCR product under UV-light.
4. Purify the PCR product following the manufacturer's instructions of Nucleospin Extract II or ChargeSwitch® PCR Clean-Up Kit (see Note 21).

Cloning and shipping clones for sequencing

1. Subclone the purified PCR product using the StrataClone™ Kit (see Note 22).
2. The sequencing format depends on the requirements of the sequencing companies. For example, the colonies can be picked and transferred to the wells of 96- or 384-well plates containing LB ampicillin agar (1.5%) in each well. After overnight incubation at 37°C, the plates are sealed and shipped by overnight mail. A second possibility is to perform PCR on selected clones with primers enclosing the insert (colony PCR). Colony PCR products positive for the insert are shipped by overnight mail for sequencing.

Result analysis

1. Extract the sequencing results in fasta format results using a trace file viewer like Chromas or FinchTV (see Note 23 and 24). This is necessary if the subsequent analysis is done with the BiQ Analyzer software.
2. The BISMA software tool can be used to import the sequences, make the alignment, perform basic statistics of the methylation level, and generate the methylation pattern http://biochem.jacobs-university.de/BDPC/BISMA (32).
3. To allow the compilation and presentation of results from several clones, the BDPC Web interface can be used. In addition, the program provides a summary file containing all data for further downstream analysis. http://biochem.jacobs-university.de/BDPC/ (33, 34).

3.3. Restriction-Based Fluorescence Assay

1. Preparation of plates: coat the microplates with 1 µg/well of avidin; for each 96-well plate combine 40 µl of the avidin solution and 10 ml of the coating buffer (see Note 25), pipette 100 µl per well and incubate overnight at 4°C. Coated plates can be used for up to 2 weeks if stored at 4°C.
2. Before use, wash the wells five times with 200 µl PBST supplemented with 500 mM NaCl to remove unbound avidin.
3. To further coat the substrate on the plate, incubate 100 pmol/well of DNA duplex in 100 µl of PBST at RT for at least

30 min. Then wash the plate three times with PBST + 500 mM NaCl and three times with PBST (see Note 26).

4. Preincubate the methyltransferase of interest with the inhibitor in the methylation buffer in a total volume of 55 µl/well of a Greiner™ 96-well V-form microplate. For the Dnmt3a/3L reaction, the C-terminal catalytic domain of the murine Dnmt3a (623–908) and the C-terminal domain of DNMT3L (208–421), obtained as described in (38) were preincubated together at 200 nM during 20 min at RT in reaction buffer in the presence of the inhibitor. Add AdoMet at a final concentration of 20 µM.
5. Transfer instantly 50 µl of the MTase preincubated with inhibitor into the corresponding well of the testing plate coated with the DNA substrate and incubate at required temperature for the appropriate amount of time.
6. Prepare control wells: (1) two wells coated with DNA that will neither undergo methylation nor restriction digestion (DNA controls), (2) three wells, in which the substrate will be methylated and digested (methylation controls), and (3) three wells in which the DNA will not be methylated but will undergo cleavage (restriction controls) (see Note 27).
7. Wash each well three times with PBST + 500 mM NaCl and three times with PBST.
8. Perform the restriction digestion step with the methylation-sensitive enzyme on the testing plate. In the DNMT3a/3L assay, 2 U/well of the methylation-sensitive restriction enzyme HpyCH4 IV (New Englands Biolabs) were incubated in 50 µl of total volume of restriction buffer (10 mM Bis-Tris-propane–HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.0) for 1 h at 37°C (see Note 28).
9. Wash each well three times with PBST + 500 mM NaCl and three times with PBST.
10. Measure the fluorescence signal of the DNA substrate by scanning the plate in a fluorescence reader (e.g., Typhoon scanner (GE Healthcare)). In our system, quantification was done automatically by measuring the sum of pixel intensities in each well.
11. Automatically analyze and quantify the data to find hit molecules. A typical tested compound, fluorescence signal intensity of which is lower or equal to a threshold of $(\mu_{restri} + 6 \times SD_{restri})$ should be considered as an active inhibitor at the tested concentration. In the case of IC_{50} determination tests, each experimental set should be performed in triplicate and the results plotted as relative methylation activity against log (inhibitor concentration). IC_{50} values can be evaluated after fitting the dose–response plots by nonlinear regression, for example, using the GraphPad Prism™ software.

4. Notes

1. Wrong pH leads to a lower binding efficiency of avidin to the surface of the microplate.
2. Avoid scratching the bottom of the plate with the pipette tips. Tilting of the plate facilitates complete the removal of avidin or washing buffer.
3. Prepare the binding buffer which contains an excess of unlabeled AdoMet to quench the incorporation of [^{3}H] into the DNA directly before using. Unlabeled AdoMet should be stored at −20°C in small aliquots in 10 mM H_2SO_4 and thawed only once to avoid degradation. Use high salt buffer to prevent binding of the methyltransferase to the DNA after stopping the reaction.
4. Labeled AdoMet should be aliquoted, stored at −20°C, and thawed only once.
5. Carry out each measurement at least in duplicates to get more accurate data.
6. Complete removal of the MTase and unreacted labeled AdoMet is very important for a low background of radioactivity. After each washing step no buffer should be left in the wells.
7. The purification of *S. marcescens* nuclease was performed similarly as described (39). The His_6-Nuclease fusion protein was expressed in TGE900 *E. coli* cells. Protein overexpression was induced at a cell density of OD_{600nm} 0.5–0.6 by changing the temperature from 28°C to 42°C and the cells were grown for an additional 2 h. All following steps were carried out at 4°C. The cells were harvested by centrifugation (15 min and 3,000 × *g*) and washed with STE buffer (10 mM Tris–HCl (pH 8.0), 0.1 mM EDTA, 0.1 mM NaCl). The cell pellet was resuspended in buffer A (10 mM Tris–HCl, pH 8.2) and the cells were disrupted by sonication. Cell debris were removed by centrifugation (1 h at 15,000 × *g*) and the pellet was resuspended in buffer B (6 M urea, 10 mM Tris–HCl, pH 8.2, 10 mM imidazole) and kept overnight at 4°C on a shaker. After a new centrifugation (1 h at 15,000 × *g*), the supernatant was applied onto a Ni-NTA column (Qiagen) equilibrated with buffer B. The column was washed with 150 ml buffer B and the protein eluted with buffer B containing 200 mM imidazole. Elution fractions containing the nuclease were pooled and dialyzed overnight against 10 mM Tris–HCl, pH 8.2. The concentration of the nuclease was determined using an extinction coefficient of $\varepsilon_{280nm} = 44{,}620/M/cm$. The enzyme was stored in small aliquots at −80°C. After thawing, keep the nuclease at 4°C.

For checking the activity of the nuclease and estimating how much enzyme is required for full digestion, perform a methylation assay as described above and digest the biotinylated oligonucleotides using different amounts of nuclease.

Alternatively, the enzyme is commercially available as Benzonase from Merck.

8. Proper mixing of the sample with the scintillation solution is crucial for the accurate counting of the sample. Mix the solution obtained after digestion and liquid scintillator solution by vortexing or inverting the scintillation vials before counting. If using the microplate, after addition of the Microscint, cover the plate with a plastic seal and place on a shaker for at least 30 min to mix the scintillator solution and the sample.
9. When choosing substrate DNA it is important to check if it contains parts free of CG sites that can be used for designing primers for bisulfite converted DNA (see also Note 4). Alternatively, adaptor sequences can be ligated to the substrate DNA after the methylation step providing priming sites for the primers. This method can also be used with in vitro reconstituted nucleosomal array (40) as substrate for methylation.
10. Sodium bisulfite dissolves completely after the addition of NaOH.
11. The bisulfite conversion rate is critical to determine the accuracy of the method to define the methylation status. To ensure complete bisulfite conversion, it is not recommended to start with more than 500 ng of DNA in a single reaction.
12. It is also possible to design primers on in silico converted DNA manually. In principle, the primer should target a region, which contains several cytosines ideally located in the 3′ part. These cytosines are substituted by thymines, such that converted DNA is amplified specifically. The primers should not contain CpG sites within their sequence to avoid discrimination against methylated or unmethylated DNA. The BiSearch software allows including CpG sites in primer design. In this case, Y is used to represent C or T in the sense chain and R is used to represent A or G in the antisense chain. (35). When designing primers, remember that after bisulfite treatment the two DNA strands are no longer complementary. To analyze the methylation status of both strands, a pair of primers specific for each strand needs to be designed.
13. The exact parameters for the methylation reaction need to be determined experimentally, as they vary greatly between the enzymes tested and the purpose of experiment.
14. The control reaction will serve to determine the bisulfite conversion rate. Ideally, catalytically inactive variant should be

used; alternatively the tested enzyme can be used after heat inactivation. No unconverted cytosines should be left in the control.

15. In this step, the reaction mixture should be split, such that the reaction volume fits the size of the wells of the thermocycler.
16. All centrifugation steps are carried out at room temperature.
17. It is important to store the bisulfite-treated genomic DNA in TE buffer, rather than water. In TE buffer, the DNA is more stable and can be stored at −20°C for up to 6 months.
18. To avoid DNA degradation by thawing and freezing, it is suggested to make aliquots of the bisulfite-treated genomic DNA.
19. It is necessary to use a DNA polymerase, which can use templates containing uracil like Taq polymerase. If there is a difficulty to get the PCR products with the described parameters, alternative parameters could be used: 15 min at 95°C, 5× (30 s at 94°C, 30 s at 60°C, 90 s at 72°C), 5× (30 s at 94°C, 30 s at 55°C, 90 s at 72°C), 35× (30 s at 94°C, 30 s at 50°C, 90 s at 72°C), 5 min at 72°C. Alternatively, 15 min at 95°C, 5× (30 s at 94°C, 30 s at 55°C, 90 s at 72°C), 5× (30 s at 94°C, 30 s at 50°C, 90 s at 72°C), 35× (30 s at 94°C, 30 s at 45°C, 90 s at 72°C), 5 min at 72°C. Also nested-PCR or seminested PCR may help to get specific PCR products. Using a gradient PCR cycler can further optimize the annealing temperature.
20. In acrylamide gels the separation of DNA usually is better than in agarose gels. But one needs to consider that converted DNA in acrylamide gels migrates differently as compared to normal DNA (41).
21. In case that unspecific PCR by-products appear that cannot be removed by optimization of the PCR protocol, the PCR product may be purified from agarose gels using commercially available kits such as QIAquick Gel Extraction Kit or Nucleospin Extract II Kit.
22. The subcloning kits based on topoisomerase ligation technique are highly efficient even using small amount of DNA substrate. One can split the material provided in the kit for one reaction to perform three reactions.
23. We suggest using Chromas as a trace file viewer and for DNA sequence extraction. With this software one can edit the annotation if necessary. Typical vector sequence can be defined in Chromas and the DNA insert sequence can be extracted by cutting the vector automatically on export. Furthermore, many sequencing results can be exported in batch.
24. During the result analysis, it happens that there are misread bases in the sequencing result. For example, there could be

NG-sites, CN-sites, or TN sites at CG dinucleotide positions. It is recommended to check the chromatogram of the corresponding clones and correct the errors manually wherever possible.

25. Wrong pH leads to a lower binding efficiency of avidin to the surface of the microplate.
26. Avoid scratching the bottom of the plate with the pipette tips. Tilting of the plate facilitates the complete removal of avidin or washing buffer.
27. A control with DNA methylated but not restricted may be used to avoid artifacts, like nuclease contamination in the MTase preparation. Other practical controls include the usage of known inhibitors or application of restriction enzymes not inhibited by methylation.
28. The restriction digestion conditions need to be optimized individually for each restriction enzyme used.

References

1. Miranda, T. B., and Jones, P. A. (2007) DNA methylation: the nuts and bolts of repression, *J Cell Physiol* **213**, 384–390.
2. Jeltsch, A. (2002) Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases, *Chembiochem* **3**, 274–293.
3. Hermann, A., Gowher, H., and Jeltsch, A. (2004) Biochemistry and biology of mammalian DNA methyltransferases, *Cell Mol Life Sci* **61**, 2571–2587.
4. Jeltsch, A., Jurkowska, R. Z., Jurkowski, T. P., Liebert, K., Rathert, P., and Schlickenrieder, M. (2007) Application of DNA methyltransferases in targeted DNA methylation, *Appl Microbiol Biotechnol* **75**, 1233–1240.
5. Lobner-Olesen, A., Skovgaard, O., and Marinus, M. G. (2005) Dam methylation: coordinating cellular processes, *Curr Opin Microbiol* **8**, 154–160.
6. Jeltsch, A., and Jurkowska, R. Z. (2010) DNA Methylation, in *The Chemical Biology of Nucleic Acids* (Mayer, G., Ed.), pp 307–322, Wiley-VCH.
7. Robertson, K. D. (2005) DNA methylation and human disease, *Nat Rev Genet* **6**, 597–610.
8. Jones, P. A., and Baylin, S. B. (2007) The epigenomics of cancer, *Cell* **128**, 683–692.
9. Feinberg, A. P., and Tycko, B. (2004) The history of cancer epigenetics, *Nat Rev Cancer* **4**, 143–153.
10. Sharma, S., Kelly, T. K., and Jones, P. A. (2010) Epigenetics in cancer, *Carcinogenesis* **31**, 27–36.
11. Cheng, X. (1995) Structure and function of DNA methyltransferases, *Annu Rev Biophys Biomol Struct* **24**, 293–318.
12. Oakeley, E. J. (1999) DNA methylation analysis: a review of current methodologies, *Pharmacol Ther* **84**, 389–400.
13. Fraga, M. F., and Esteller, M. (2002) DNA methylation: a profile of methods and applications, *Biotechniques* **33**, 632, 634, 636–649.
14. Bird, A. P., and Southern, E. M. (1978) Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis, *J Mol Biol* **118**, 27–47.
15. Bird, A. P. (1978) Use of restriction enzymes to study eukaryotic DNA methylation: II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern, *J Mol Biol* **118**, 49–60.
16. Singer, J., Roberts-Ems, J., and Riggs, A. D. (1979) Methylation of mouse liver DNA studied by means of the restriction enzymes msp I and hpa II, *Science* **203**, 1019–1021.
17. Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual

DNA strands, *Proc Natl Acad Sci U S A* **89**, 1827–1831.

18. Clark, S. J., Harrison, J., Paul, C. L., and Frommer, M. (1994) High sensitivity mapping of methylated cytosines, *Nucleic Acids Res* **22**, 2990–2997.
19. Kuo, K. C., McCune, R. A., Gehrke, C. W., Midgett, R., and Ehrlich, M. (1980) Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA, *Nucleic Acids Res* **8**, 4763–4776.
20. Gehrke, C. W., McCune, R. A., Gama-Sosa, M. A., Ehrlich, M., and Kuo, K. C. (1984) Quantitative reversed-phase high-performance liquid chromatography of major and modified nucleosides in DNA, *J Chromatogr* **301**, 199–219.
21. Eick, D., Fritz, H. J., and Doerfler, W. (1983) Quantitative determination of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography, *Anal Biochem* **135**, 165–171.
22. Rubin, R. A., and Modrich, P. (1977) EcoRI methylase. Physical and catalytic properties of the homogeneous enzyme, *J Biol Chem* **252**, 7265–7272.
23. Hubscher, U., Pedrali-Noy, G., Knust-Kron, B., Doerfler, W., and Spadari, S. (1985) DNA methyltransferases: activity minigel analysis and determination with DNA covalently bound to a solid matrix, *Anal Biochem* **150**, 442–448.
24. Jeltsch, A., Friedrich, T., and Roth, M. (1998) Kinetics of methylation and binding of DNA by the EcoRV adenine-N6 methyltransferase, *J Mol Biol* **275**, 747–758.
25. Roth, M., and Jeltsch, A. (2000) Biotin-avidin microplate assay for the quantitative analysis of enzymatic methylation of DNA by DNA methyltransferases, *Biol Chem* **381**, 269–272.
26. Liebert, K., and Jeltsch, A. (2008) Detection and quantitation of the activity of DNA methyltransferases using a biotin/avidin microplate assay, *Methods Mol Biol* **418**, 149–156.
27. Grunau, C., Clark, S. J., and Rosenthal, A. (2001) Bisulfite genomic sequencing: systematic investigation of critical experimental parameters, *Nucleic Acids Res* **29**, E65-65.
28. Zhang, Y., Rohde, C., Tierling, S., Stamerjohanns, H., Reinhardt, R., Walter, J., and Jeltsch, A. (2009) DNA methylation analysis by bisulfite conversion, cloning, and sequencing of individual clones, *Methods Mol Biol* **507**, 177–187.
29. Laird, C. D., Pleasant, N. D., Clark, A. D., Sneeden, J. L., Hassan, K. M., Manley, N. C., Vary, J. C., Jr., Morgan, T., Hansen, R. S., and Stoger, R. (2004) Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules, *Proc Natl Acad Sci U S A* **101**, 204–209.
30. Jurkowska, R. Z., Anspach, N., Urbanke, C., Jia, D., Reinhardt, R., Nellen, W., Cheng, X., and Jeltsch, A. (2008) Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L, *Nucleic Acids Res* **36**, 6656–6663.
31. Zhang, J. H., Chung, T. D., and Oldenburg, K. R. (1999) A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays, *J Biomol Screen* **4**, 67–73.
32. Rohde, C., Zhang, Y., Reinhardt, R., and Jeltsch, A. (2010) BISMA--fast and accurate bisulfite sequencing data analysis of individual clones from unique and repetitive sequences, *BMC Bioinformatics* **11**, 230.
33. Rohde, C., Zhang, Y., Jurkowski, T. P., Stamerjohanns, H., Reinhardt, R., and Jeltsch, A. (2008) Bisulfite sequencing Data Presentation and Compilation (BDPC) web server--a useful tool for DNA methylation analysis, *Nucleic Acids Res* **36**, e34.
34. Rohde, C., Zhang, Y., Stamerjohanns, H., Hecher, K., Reinhardt, R., and Jeltsch, A. (2009) New clustering module in BDPC bisulfite sequencing data presentation and compilation web application for DNA methylation analyses, *Biotechniques* **47**, 781–783.
35. Tusnady, G. E., Simon, I., Varadi, A., and Aranyi, T. (2005) BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes, *Nucleic Acids Res* **33**, e9.
36. Li, L. C., and Dahiya, R. (2002) MethPrimer: designing primers for methylation PCRs, *Bioinformatics* **18**, 1427–1431.
37. Yamada, Y., Watanabe, H., Miura, F., Soejima, H., Uchiyama, M., Iwasaka, T., Mukai, T., Sakaki, Y., and Ito, T. (2004) A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q, *Genome Res* **14**, 247–266.
38. Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., and Cheng, X. (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation, *Nature* **449**, 248–251.
39. Friedhoff, P., Gimadutdinow, O., Ruter, T., Wende, W., Urbanke, C., Thole, H., and Pingoud, A. (1994) A procedure for renaturation and purification of the extracellular

Serratia marcescens nuclease from genetically engineered Escherichia coli, *Protein Expr Purif* **5**, 37–43.

40. Zhang, Y., Jurkowska, R., Soeroes, S., Rajavelu, A., Dhayalan, A., Bock, I., Rathert, P., Brandt, O., Reinhardt, R., Fischle, W., and Jeltsch, A. (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail, *Nucleic Acids Res.*
41. Handa, V., and Jeltsch, A. (2004) Anomalous mobility of polymerase chain reaction products after bisulfite treatment of DNA, *Anal Biochem* **333**, 196–198.

Chapter 14

A Chromatin Immunoprecipitation Protocol for Small Cell Numbers

Philippe Collas

Abstract

Chromatin immunoprecipitation (ChIP) is a widely used technique to get a snap-shot of protein–DNA interactions in cells. ChIP has notably been used for mapping the location of modified histones, transcription factors, or chromatin remodeling enzymes in the genome, often in relation to transcription or differentiation. Conventional ChIP protocols however, have for a long time required large numbers of cells, which has limited the applicability of ChIP to rare or small cell samples. In recent years, ChIP assays for small cell numbers (in the 10,000–100,000 range) have been recently reported by us and others. This chapter describes a micro (μ)ChIP procedure for multiple parallel ChIPs from a single chromatin batch from 1,000 cells.

Key words: Chromatin immunoprecipitation, ChIP, Histone, Acetylation, Methylation, Epigenetics

1. Introduction

Interactions between proteins and DNA are essential for many cellular functions such as genomic stability, DNA replication and repair, chromosome segregation, transcription, and epigenetic silencing of gene expression. ChIP has become a technique of choice in the study of protein–DNA interactions and for unraveling transcriptional regulatory circuits within the cell (1). ChIP has been used for mapping the location of post-translationally modified histones, transcription factors, chromatin modifiers, and other nonhistone proteins which (in)directly interact with DNA. This mapping can be restricted to specific genomic sites (2–8) or expanded to a genome-scale (9–16).

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_14,

In a classical ChIP assay, DNA and proteins are reversibly cross-linked to maintain the association of proteins with target DNA sequences. However, when analyzing histone modifications cross-linking can be omitted (native ChIP) (3, 17). Chromatin is subsequently sheared, usually by sonication, to ~200–500 bp fragments and large complexes are removed by sedimentation. The supernatant, i.e., the chromatin preparation, is used for immunoprecipitation of specific protein–DNA complexes using antibodies coupled to beads. Immunoprecipitated complexes are washed under stringent conditions, the cross-link is reversed, proteins are digested and the precipitated DNA is eluted. Genomic sequences associated with the precipitated protein can be identified by polymerase chain reaction (ChIP-PCR), cloning and sequencing, high-throughput sequencing (ChIP-seq), or hybridization to microarrays (ChIP-chip). Parameters and variations of the ChIP assay, and tools implemented to investigate the profiles of DNA–protein interactions have been addressed in other recent publications (1, 18–25).

In spite of the versatility in the nature of DNA-bound proteins and cell types that can be examined by ChIP, the assay has been hampered by a requirement for large cell numbers (in the 10^6–10^7 range), which has prevented the application of ChIP to rare cell samples. Another drawback has been the length of the procedure which can take up to 4 days. These limitations have prompted the development of variations on the ChIP assay. (1) A carrier ChIP (CChIP) assay (4) relies on a single immunoprecipitation from 100 cells and involves the inclusion of carrier chromatin from *Drosophila* cells to reduce loss and facilitate precipitation. However, the assay is cumbersome and entails radioactive labeling of PCR products for detection. It is also unclear whether it is suitable for precipitation of transcription factors. Furthermore, the use of foreign carrier chromatin implies that primers used for detection of immunoprecipitated sequences must be highly species-specific. (2) Still with the aim of reducing cell numbers for ChIP, a microChIP protocol for 10,000 cells without carrier chromatin was reported (15). This assay allows the analysis of histone or RNA polymerase II (RNAPII) binding throughout the genome by ChIP-on-chip – but it assay takes 4 days. (3) A fast ChIP assay (6, 26) has shortened two steps of conventional ChIP and reduced the assay to 1 day. An ultrasonic bath has been applied to speed up antibody binding to target proteins, and DNA isolation has been sped up by the use of a resin-based (Chelex-100) DNA isolation (26). Nonetheless, the fast protocol requires large numbers of cells (10^6–10^7 range). (4) We have set up a quick and quantitative (Q^2) ChIP assay suitable for up to 1,000 histone ChIPs or 100 transcription factor ChIPs from 100,000 cells (7). Q^2ChIP can be done in 1 day. (5) Recently, a microplate-based ChIP assay (matrix-ChIP) was reported, which increases throughput and simplifies the assay

(27). All steps are carried out in microplate wells without sample transfers. Matrix ChIP enables 96 ChIPs for histones and DNA bound proteins in 1 day (27). (6) Moreover, a procedure for whole-genome mapping of histone modifications from as few as ~25,000 cells or 50 ng of ChIP DNA was recently reported (28).

(7) The lower limit on cell numbers has been further stretched in our micro- (μ)ChIP assay suitable for up to eight parallel ChIPs of histones and/or RNAPII from a single batch of 1,000 cells, or for a single ChIP from 100 cells without carrier chromatin (29) (Fig. 1). The assay has been validated by assessing several

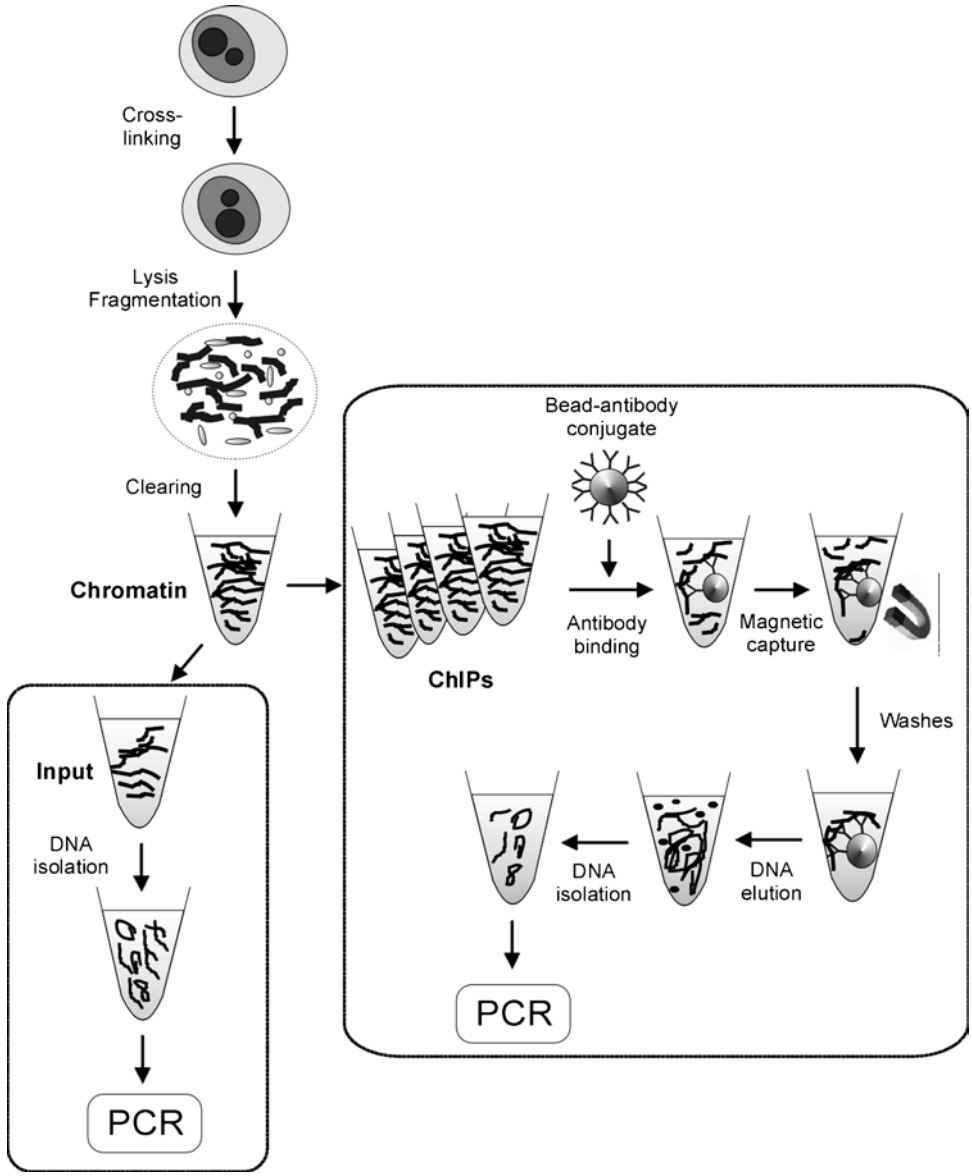


Fig. 1. Overview of the μChIP assay.

post-translational modifications of histone H3 and RNAPII binding to developmentally regulated promoters in embryonal carcinoma cells and biopsies (29). The profiles of histone modifications identified from chromatin prepared from 1,000 cells and from starting batches of 100 cells are similar and reflect the expression status of the genes (29). This chapter describes the μChIP assay as it is used in our laboratory. Applications of the assay to small tissue biopsies have been reported elsewhere (29).

2. Materials

2.1. Laboratory Equipment

1. Siliconized pipette tips.
2. Filtered pipette tips (10, 200, and 1,000 μL).
3. Magnetic rack for 200-μL tube strips (Diagenode).
4. 200-μL PCR tubes in eight-tube strip format (Axygen).
5. 0.6- and 1.5-mL centrifuge tubes.
6. Magnetic holder for 1.5-mL tubes.
7. Probe sonicator (Sartorius Labsonic M sonicator with 3-mm diameter probe, or similar).
8. Rotator placed at 4°C.
9. Table top centrifuge.
10. Minicentrifuge.
11. Vortex.
12. Thermomixer (Eppendorf).
13. Heating block.
14. Thermal cycler with real-time capacity.

2.2. Reagents

1. 36.5% formaldehyde.
2. Dynabeads® Protein A (Invitrogen). The beads should be well suspended before pipetting. Use Dynabeads® Protein A beads with rabbit IgGs and Dynabeads® Protein G (Invitrogen) with mouse IgGs.
3. 5 M NaCl.
4. 400 mM EGTA.
5. 500 mM EDTA.
6. 1 M Tris–HCl, pH 7.5.
7. 1 M Tris–HCl, pH 8.0.
8. Glycine: 1.25 M stock solution in PBS.
9. Chelex-100 (BioRad, cat. no. 142–1253): 10% (wt/vol) Chelex in MilliQ water.

10. Acrylamide carrier (Sigma-Aldrich).
11. Proteinase K: 20 mg/mL solution in MilliQ water.
12. Protease inhibitor mix (Sigma-Aldrich).
13. PMSF: 100 mM stock solution in 100% ethanol.
14. Sodium butyrate: 1 M stock solution in MilliQ water. Na-butyrate is a histone deacetylase inhibitor and should be used for anti-acetylated epitope ChIPs.
15. Phosphate-buffered saline (PBS).
16. PBS/Na-butyrate solution 20 mM butyrate in 1× PBS. Make immediately before use.
17. PBS/Na-butyrate/formaldehyde fixative: 20 mM butyrate, 1% (vol/vol) formaldehyde, 1 mM PMSF, and protease inhibitor mix in 1× PBS. Make immediately before use.
18. Phenol–chloroform isoamylalcohol (25:24:1).
19. Chloroform isoamylalcohol (24:1).
20. 3 M NaAc.
21. IQ SYBR® Green (BioRad).
22. Antibodies of choice. Use ChIP-grade antibodies when available (see Note 1).

2.3. Buffers

1. Lysis buffer: 50 mM Tris–HCl, pH 8.0, 10 mM EDTA, 1% (wt/vol) SDS, protease inhibitor mix (1:100 dilution from stock), 1 mM PMSF, 20 mM Na-butyrate. Protease inhibitor mix, PMSF, and Na-butyrate should be added immediately before use.
2. RIPA buffer: 10 mM Tris–HCl, pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% (vol/vol) Triton X-100, 0.1% (wt/vol) SDS, 0.1% (wt/vol) Na-deoxycholate.
3. RIPA ChIP buffer: 10 mM Tris–HCl, pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% (vol/vol) Triton X-100, 0.1% (wt/vol) SDS, 0.1% (wt/vol) Na-deoxycholate, protease inhibitor mix (1:100 dilution from stock), 1 mM PMSF, 20 mM Na-butyrate. Protease inhibitor mix, PMSF, and Na-butyrate should be added immediately before use.
4. TE buffer: 10 mM Tris–HCl, pH 8.0, 10 mM EDTA.
5. Elution buffer**:** 20 mM Tris–HCl, pH 7.5, 5 mM EDTA, 50 mM NaCl.
6. Complete elution buffer: 20 mM Tris–HCl, pH 7.5, 5 mM EDTA, 50 mM NaCl, 20 mM Na-butyrate, 1% (wt/vol) SDS, 50 μg/mL proteinase K. Na-butyrate, SDS, and proteinase K should be added just before use.

3. Methods

3.1. Preparation of Antibody–Bead Complexes

1. Prepare a slurry of Dynabeads® Protein A (if using rabbit IgGs). For 16 ChIPs, including two negative controls, place 180 µL of well-suspended Dynabeads® Protein A stock solution into a 1.5-mL tube, place the tube in the magnetic holder, allow beads to be captured, remove the buffer, remove from the magnet, and add 500 µL RIPA buffer. Ensure the stock bead suspension is homogenous before pipetting.
2. Vortex, capture the beads, remove the buffer, add another 500 µL RIPA buffer.
3. Vortex, capture the beads, remove the buffer, add 170 µL RIPA buffer.
4. Vortex the beads and place the tube on ice.
5. Aliquot 90 µL RIPA buffer into 200-µL PCR tubes (one tube per ChIP) place on ice and add 10 µL washed Dynabeads® Protein A-bead slurry from step 4 and 2.4 µg antibody to each tube. To the negative control samples, do not add the antibody, but add a pre-immune antibody preferably of the same isotype as the ChIP antibodies. Place at 40 rpm on a rotator for 2 h at 4°C (see Note 2).

3.2. Cross-linking of DNA and Proteins

1. Add 20 mM of the histone deacetylase inhibitor Na-butyrate from the 1 M stock to the cell culture and mix gently. Na-butyrate only needs to be included when acetylated epitopes are assessed.
2. Discard the medium to remove dead cells (if cells are growing adherent) and add room temperature (20–25°C) PBS/Na-butyrate (10 mL per 175 cm^2 culture flask).
3. Harvest cells by trypsinization or as per your standard protocol according to cell type. Trypsin or other harvesting solution should contain 20 mM Na-butyrate.
4. Count cells and resuspend 1,000 (or 100) cells in 500 µL PBS/Na-butyrate in a 0.6-mL tube at room temperature (see Note 3).
5. Add 13.5 µL formaldehyde (1% vol/vol final concentration), mix by gentle vortexing, and incubate for 8 min at room temperature (see Note 4).
6. Add 57 µL of the 1.25 M glycine stock (125 mM final concentration) and incubate for 5 min at room temperature. Pellets of cross-linked cells can be stored at –80°C for at least 1 month.

3.3. Preparation of Chromatin from 1,000 Cells

The procedure described here is for preparing chromatin from 1,000 cells (starting material). It is however also suited for up to 50,000 cells with adjustments in sonication conditions. A procedure

for assessing chromatin fragmentation by sonication of small cell numbers has recently been published (30).

1. Centrifuge formaldehyde-cross-linked cells at 470×*g* for 10 min at 4°C in a swing-out rotor with soft deceleration settings. Slowly aspirate and discard the supernatant, leaving ~30 µL of the solution with the cell pellet to ensure that none of the loosely packed cells are aspirated.
2. Resuspend the cells in 500 µL ice-cold PBS/Na-butyrate by gentle vortexing and centrifuge at 470×*g* for 10 min at 4°C as in step 1.
3. Repeat the washing procedure (step 2) once. Upon aspiration of the last wash, leave 20 µL PBS/Na-butyrate with the cell pellet.
4. Add 120 µL room temperature lysis buffer, vortex for 2 x 5 s, leave on ice for 5 min and resuspend cells by vortexing. Ensure that no liquid is trapped in the lid.
5. Sonicate on ice for 3 x 30 s, with 30 s pauses on ice between each 30 s session, using the probe sonicator. With the Labsonic M sonicator, use the following pulse settings: cycle 0.5, 30% power (see Note 5).
6. Add 400 µL RIPA ChIP buffer to the tube (which contains ~140 µL lysate) and mix by vortexing.
7. Centrifuge at 12,000×*g* for 10 min at 4°C, aspirate the supernatant (chromatin), and transfer it into a clean 1.5-mL tube chilled on ice (see Note 6). To avoid aspirating the sedimented material, leave ~50 µL supernatant in the tube after aspiration.
8. Add 410 µL RIPA ChIP buffer to the remaining volume, mix by vortexing, and centrifuge at 12,000×*g* for 10 min at 4°C.
9. Aspirate the supernatant, leaving ~20 µL with the (invisible) pellet and pool it with the first supernatant. This yields ~930 µL of chromatin suitable for eight parallel ChIPs and one input reference. Discard the pellets. Diluting the chromatin reduces SDS concentration to 0.1%, which is suitable for immunoprecipitation with most antibodies.
10. Aliquot 100 µL chromatin each into, e.g., eight chilled 0.2-mL tubes (in strip format) containing antibody–bead complexes held to the wall in the magnetic rack (on ice), and from which the RIPA buffer has been pipetted out.
11. Add 100 µL chromatin to a tube chilled on ice. This is used as input chromatin. A 1.5-mL tube is used in this step if DNA is to be purified with phenol–chloroform isoamylalcohol. For DNA isolation using Chelex-100, a 0.6-mL tube is preferred.

3.4. Preparation of Chromatin from 100 Cells

This procedure is for preparing chromatin when starting with 100 cells, but can also be applied to up to 1,000 cells. When starting with 100 cells, only one immunoprecipitation can be performed per sample. Prepare an additional sample for reference input chromatin.

1. Centrifuge formaldehyde-cross-linked cells at 470×*g* for 10 min at 4°C in a swing-out rotor with soft deceleration settings. Aspirate the supernatant; leave ~30 μL of the solution with the pellet.
2. Add 500 μL ice-cold PBS/Na-butyrate, resuspend the cells by gentle vortexing and centrifuge at 470×*g* for 10 min at 4°C using a swing-out rotor with soft deceleration settings.
3. Repeat the washing procedure (step 2) once. Leave ~20 μL of PBS/Na-butyrate with the pellet (invisible) after removing the last wash.
4. Add 120 μL lysis buffer, vortex for twice 5 s, and incubate for 3 min on ice (see Note 7).
5. Centrifuge the nuclei at 860×*g* for 10 min at 4°C using a swing-out rotor with soft deceleration settings and discard the supernatant; leave 20–30 μL of lysis buffer in the tube.
6. Add 120 μL RIPA ChIP buffer and vortex for 10 s.
7. Sonicate each tube on ice for twice 30 s, with 30 s pauses on ice between each 30 s session, using the probe sonicator (cycle 0.5 and 30% power with the Labsonic M). Repeat for each tube while leaving the sonicated samples on ice. Note that when starting with 100 cells, it is impossible to visualize chromatin fragmentation by agarose gel electrophoresis. Instead, we use a PCR-based assay (30).
8. Pipette the lysate several times using a siliconized pipette tip and transfer into a 0.2-mL PCR tube containing antibody-coated beads and from which the RIPA buffer has been removed.

3.5. Immunoprecipitation and Washes

1. Remove the tube strip from the magnetic rack to release the antibody–bead complexes into the chromatin suspension and place the tubes on a rotator at 40 rpm for 2 h at 4°C. This step can be carried out overnight at 4°C if necessary, but prolonged incubation may enhance background.
2. Centrifuge the tubes in a minicentrifuge for 1 s to bring down any solution trapped in the lid during the incubation on the rotator, and capture the immune complexes by placing the tubes in the chilled magnetic rack.
3. Discard the supernatant, add 100 μL ice-cold RIPA buffer, and remove the tubes from the magnetic rack to release the immune complexes into the buffer. Resuspend the complexes

by gentle manual agitation and place the tubes on a rotator at 40 rpm for 4 min at 4°C.

4. Repeat step 2 and step 3 twice. Briefly spin the tubes in a minicentrifuge for 1 s to bring down any liquid trapped in the lid prior to placing the tubes in the magnetic rack.
5. Centrifuge the tubes in a minicentrifuge for 1 s.
6. Remove the supernatant, add 100 µL TE buffer, incubate on a rotator at 4°C for 4 min at 40 rpm.
7. Centrifuge the tubes in a minicentrifuge for 1 s.
8. Place the tubes on ice (not in the magnetic rack), transfer the content of each tube into separate clean 0.2-mL tubes on ice, capture the complexes in the magnetic rack, and remove the TE buffer.

3.6. DNA Recovery by Phenol–Chloroform–Isoamylalcohol Extraction

We have used two procedures for recovering DNA from the ChIP material and from input chromatin. Involving (1) a phenol–chloroform–isoamylalcohol extraction and (2) a resin-mediated DNA isolation (Chelex-100).

3.6.1. DNA Recovery from ChIP Material: Combined DNA Elution, Cross-link Reversal, Proteinase K Digestion, Followed by DNA Purification by Phenol–Chloroform–Isoamylalcohol Extraction

1. Place the tubes from step 8 of Subheading 3.5 in a rack and add 150 µL complete elution buffer to each tube.
2. Incubate for 2 h on the Thermomixer at 68°C, 1,300 rpm. Meanwhile, prepare the input sample as described in Subheading 3.6.2. DNA elution from immune complexes, cross-link reversal, and protein digestion is combined into one step.
3. Remove tubes from the Thermomixer and centrifuge for 3 s with a minicentrifuge.
4. Capture the beads using the magnetic rack, collect the supernatant, and place it into a clean 1.5-mL tube.
5. Add 150 µL complete elution buffer to the remaining ChIP material and incubate on the Thermomixer for 5 min at 68°C, 1,300 rpm.
6. Remove the tubes from the Thermomixer, capture the beads using the magnetic rack, collect the supernatant, and combine it with the first supernatant.
7. Add 200 µL elution buffer to the eluted ChIP material.
8. Extract DNA once with an equal volume of phenol–chloroform–isoamylalcohol, centrifuge at 15,000 × *g* for 5 min to separate the phases and transfer 460 µL of the aqueous (top) phase to a clean tube.
9. Extract once with an equal volume of chloroform isoamylalcohol, centrifuge at 15,000 × *g* for 5 min, and transfer 400 µL of the aqueous phase to a clean tube. Use filtered tips when adding phenol–chloroform–isoamylalcohol and chloroform isoamylalcohol to prevent dripping during transfer.

10. Add 44 µL of 3 M NaAc (pH 7.0), 10 µL of 0.25% (wt/vol) acrylamide carrier, and 1 mL 96% ethanol at –20°C. Mix thoroughly and incubate for at least 1 h at –80°C. DNA can be left at –80°C for several hours or days if more convenient.
11. Thaw the tubes and centrifuge at 20,000×*g* for 15 min at 4°C.
12. Remove the supernatant, add 1 ml of 70% ethanol at –20°C, and vortex briefly to wash the DNA pellet. Centrifuge at 20,000×*g* for 10 min at 4°C. Repeat this step once more.
13. Remove the supernatant and dissolve the DNA in 30 µL TE for ChIPs from chromatin from 100 cells or 60 µL for a ChIP from chromatin from 1,000 cells. DNA can be immediately used for PCR or stored at –20°C for up to 1 week (see Note 8).

3.6.2. DNA Recovery from Input Chromatin

1. To input chromatin samples, add 200 µL of elution buffer and 7.5 µL of a 10× dilution (2 mg/mL) of the proteinase K solution, vortex and incubate for 2 h on a heating block at 68°C.
2. Remove samples from the heating block and add 200 µL elution buffer.
3. Continue from step 8 in Subheading 3.6.1, processing the input samples and the ChIP samples in parallel.

3.7. DNA Recovery Using Chelex-100

This DNA recovery procedure describes a Chelex-100-mediated DNA purification reported previously (26), with modifications for small cell number ChIP and to speed up handling.

3.7.1. DNA Recovery from ChIP Samples

1. To the washed ChIP samples, add 40 µL of 10% Chelex-100, release immune complexes, and vortex for 10 s. Make sure the Chelex-100 beads are in suspension while pipetting and that the opening of the pipette tip is large enough not to hinder the beads.
2. Boil ChIP samples and input samples (prepared as described in step 4, Subheading 3.7.2) for 10 min in a PCR machine and cool to room temperature.
3. Add 1 µL proteinase K solution, vortex, and incubate at 55°C, 30 min, at 1,300 rpm in the Thermomixer.
4. Boil for 10 min, centrifuge for 10 s in a minicentrifuge and keep tubes upright for ~1 min on the bench, with no magnet, to allow beads to settle.
5. Using a siliconized tip, transfer 30 µL of the supernatant into a clean 0.6-mL tube chilled on ice. Take great care to avoid transfer of beads.
6. Add 10 µL MilliQ H_2O to the remaining beads, vortex, and centrifuge for 10 s in a minicentrifuge.
7. After the beads settle, collect 12 µL of the supernatant, pool with the first supernatant and vortex (see Note 9).

3.7.2. DNA Recovery from Input Chromatin

1. To input chromatin samples, add 10 µL acrylamide carrier and 250 µL 96% ethanol at −20°C. Vortex thoroughly and place at −80°C for 30 min.
2. Thaw, immediately centrifuge at 20,000 × *g* for 15 min at 4°C, and wash the pellet in 500 µL of 70% ethanol. Dry the pellet.
3. Add 40 µL of 10% (wt/vol) Chelex-100 to the dried pellet and vortex for 10 s.
4. Continue from step 2, Subheading 3.7.1, processing input and ChIP samples in parallel.

3.8. Set Up of Real-Time PCR and Analysis of Data

1. Prepare a master mix and aliquot for individual 25 µL qPCR reactions (MilliQ water 6.5 µL; SYBR Green Master Mix (2×) 12.5 µL; forward primer (20 µM stock) 0.5 µL; reverse primer (20 µM stock) 0.5 µL; DNA template, 5 µL) for all ChIP and input samples with each primer pair.
2. Prepare a standard curve with genomic DNA. Make sure to include a wide range of DNA concentrations (e.g., 0.005–20 ng/µL) to cover the range of your ChIP DNA samples. Use 5 µL DNA in each PCR. Establish one standard curve for each primer pair and for each PCR plate.
3. Set up a real-time PCR program, using your real-time PCR system, with a 40-cycle program.
4. Acquire the data using your real-time PCR data acquisition program.
5. Calculate the amount of DNA in each sample using the standard curve.
6. Export the data into Excel spreadsheets.
7. Determine the amount of precipitated DNA relative to input as ([amount of ChIP DNA]/[amount of input DNA]) × 100. We analyze at least three independent ChIPs, each in duplicate qPCRs and express the data as percent (±SD) precipitated DNA relative to input DNA (Fig. 1) (see Notes 10 and 11).

4. Notes

1. We have used with this protocol the following anti-histone antibodies: anti-H3K9ac (Upstate, cat. no. 06–942), anti-H3K9me2 (Upstate, cat. no. 07–441), anti-H3K9me3 (Upstate, cat. no. 07–442), anti-H3K27me3 (Upstate, cat. no. 05–851), anti-H3K9me3 (Diagenode, cat. no. pAb-056-050), anti-H3K4me2 (Abcam, cat. no. Ab7766), anti-H3K4me3 (Abcam, cat. no. Ab8580). We have also used an anti-RNAPII antibody (Santa Cruz

Biotechnology, cat. no. sc-899); the procedure should be tested for other antibodies.

2. This incubation step should be carried out during cross-linking, cell lysis, and chromatin preparation and if necessary can be prolonged until all chromatin samples are ready for immunoprecipitation. We recommend using 0.2-mL PCR tubes in an eight-tube strip format, which fits in the magnetic rack.
3. Up to 50,000 cells can be used using the same protocol. More cells allow the analysis of more genomic loci by PCR. To prevent cell lysis during pipetting, use a 1,000-µL pipette tip or a 200-µL pipette tip with a cut end.
4. Formaldehyde cross-links DNA to proteins located within 2 Å of DNA (31). To simplify the cross-linking step and enhance cell recovery, we consistently cross-link cells in suspension. Time of cross-linking may vary with the protein to be immunoprecipitated, but for most applications, 8–10 min cross-linking is sufficient. Please note that the user might crosslink DNA and proteins prior to trypsinization of the cells if they are concerned for changes in histone modifications that might originate from altered signaling from the cell surface. However, how cross-linking prior to cell collection by trypsinization might affect the yield of harvest has not been determined.
5. Sonication should produce chromatin fragments of ~200–500 bp. The sonication regime indicated is suitable for a variety of cultured cell lines but must be optimized for each cell type, particularly for primary cells. Do not allow samples to foam as foaming reduces sonication efficiency. If foaming occurs, ensure that the sonicator probe is placed deep enough, a few millimeters from the bottom of the tube, or reduce sonication intensity.
6. To avoid aspirating the sedimented material, leave ~50 µL of supernatant in the tube after aspiration.
7. Keeping cells in lysis buffer for over 3 min prior to centrifugation increases the chance of SDS precipitating. If the SDS precipitates during centrifugation, remove the lysis buffer, add 200 µL RIPA ChIP buffer, dissolve the SDS by vortexing, and centrifuge the nuclei as in step 5, Subheading 3.4.
8. TE volume depends on the number of cells in the ChIP. Note that low DNA concentrations lead to degradation of the DNA more rapidly than at high concentrations. Thus, we recommend to immediately use DNA for PCR for ChIPs from 1,000 cells or less.
9. The volumes collected must be identical between samples if ChIP results are to be compared. Chelex-100 enhances DNA

recovery but yields larger volumes than phenol–chloroform–isoamylalcohol extraction. Final ChIP results are similar with either isolation method (26, 29). The DNA can be immediately used for PCR or stored at −20°C for up to 1 week.

10. If no PCR signal is detected, several factors may be implicated. (1) There is not enough chromatin in the ChIP assay: increase the amount of cells or chromatin (note that it may be difficult to extract all chromatin from certain primary cell types); (2) the ChIP did not work: use ChIP-grade antibodies if possible; do an antibody titration; (3) the PCR did not work: setup a control qPCR with the same primers on genomic DNA and optimize PCR conditions; ensure there is no carry-over Chelex-100 with the template. If PCR signals are weaker than expected, there might not be enough DNA template. If variations in PCR signal intensity are detected between ChIP replicates, this may be due to (1) inconsistent chromatin preparations between samples: ensure that insoluble debris are removed by sedimentation after fragmentation; do not to carry over debris when aspirating the chromatin supernatant; (2) inconsistent sonication: practice sonication on larger cell numbers (e.g., 100,000) until fragmentation is reproducible; (3) variable amounts of Dynabeads between samples: ensure magnetic beads are well suspended while pipetting; (4) too little and variable amounts of input DNA template (high Ct values): increase the amount of input DNA template in the PCR and ensure consistency between replicates; ensure that ethanol-precipitated DNA is fully dissolved before PCR.
11. It should be mentioned that the ChIP protocol described here has been used for various cell types (see also Note 5). Its application to 1,000 cells or fewer may however depend on cell type, as chromatin configuration, the abundance of a given histone post-translational modification, and accessibility of epitopes to antibodies may vary from cell type to cell type. Therefore, the application of this protocol to histone modifications or chromatin-bound proteins other than those listed here should carefully be determined; this can be done using the present protocol as a starting point.

Acknowledgments

Dr. John Arne Dahl is thanked for setting up the protocol in the laboratory. Our work is supported by the Research Council of Norway, the Norwegian Cancer Society, and the University of Oslo.

References

1. Collas, P. and Dahl, J. A. (2008) Chop it, ChIP it, check it: the current status of chromatin immunoprecipitation. *Front Biosci.* **13**, 929–943.
2. O'Neill, L. P. and Turner, B. M. (1995) Histone H4 acetylation distinguishes coding regions of the human genome from heterochromatin in a differentiation-dependent but transcription-independent manner. *EMBO J.* **14**, 3946–3957.
3. O'Neill, L. P. and Turner, B. M. (1996) Immunoprecipitation of chromatin. *Methods Enzymol.* **274**, 189–197.
4. O'Neill, L. P., Vermilyea, M. D. and Turner, B. M. (2006) Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. *Nat. Genet.* **38**, 835–841.
5. Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M. and Fisher, A. G. (2006) Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8**, 532–538.
6. Nelson, J. D., Denisenko, O., Sova, P. and Bomsztyk, K. (2006) Fast chromatin immunoprecipitation assay. *Nucleic Acids Res.* **34**, e2.
7. Dahl, J. A. and Collas, P. (2007) Q^{2}ChIP, a quick and quantitative chromatin immunoprecipitation assay unravels epigenetic dynamics of developmentally regulated genes in human carcinoma cells. *Stem Cells* **25**, 1037–1046.
8. Attema, J. L., Papathanasiou, P., Forsberg, E. C., Xu, J., Smale, S. T. and Weissman, I. L. (2007) Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis. *Proc. Natl. Acad. Sci U. S. A* **104**, 12371–12376.
9. Bernstein, B. E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D. K., Huebert, D. J., McMahon, S., Karlsson, E. K., Kulbokas, E. J., III, Gingeras, T. R., Schreiber, S. L. and Lander, E. S. (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181.
10. Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R. and Young, R. A. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956.
11. Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L. and Lander, E. S. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326.
12. Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., Wong, K. Y., Sung, K. W., Lee, C. W., Zhao, X. D., Chiu, K. P., Lipovich, L., Kuznetsov, V. A., Robson, P., Stanton, L. W., Wei, C. L., Ruan, Y., Lim, B. and Ng, H. H. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431–440.
13. Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K., Koseki, H., Fuchikami, T., Abe, K., Murray, H. L., Zucker, J. P., Yuan, B., Bell, G. W., Herbolsheimer, E., Hannett, N. M., Sun, K., Odom, D. T., Otte, A. P., Volkert, T. L., Bartel, D. P., Melton, D. A., Gifford, D. K., Jaenisch, R. and Young, R. A. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313.
14. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. and Young, R. A. (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88.
15. Acevedo, L. G., Iniguez, A. L., Holster, H. L., Zhang, X., Green, R. and Farnham, P. J. (2007) Genome-scale ChIP-chip analysis using 10,000 human cells. *Biotechniques* **43**, 791–797.
16. Zhao, X. D., Han, X., Chew, J. L., Liu, J., Chiu, K. P., Choo, A., Orlov, Y. L., Sung, W. K., Shahab, A., Kuznetsov, V. A., Bourque, G., Oh, S., Ruan, Y., Ng, H. H. and Wei, C. L. (2007) Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**, 286–298.
17. O'Neill, L. P. and Turner, B. M. (2003) Immunoprecipitation of native chromatin: NChIP. *Methods* **31**, 76–82.
18. Hudson, M. E. and Snyder, M. (2006) High-throughput methods of regulatory element discovery. *Biotechniques* **41**, 673, 675, 677.
19. Dunn, J. J., McCorkle, S. R., Everett, L. and Anderson, C. W. (2007) Paired-end genomic signature tags: a method for the functional analysis of genomes and epigenomes. *Genet. Eng.* **28**, 159–173.
20. Aiba, K., Carter, M. G., Matoba, R. and Ko, M. S. (2006) Genomic approaches to early embryogenesis and stem cell biology. *Semin. Reprod. Med.* **24**, 330–339.

21. Clark, D. J. and Shen, C. H. (2006) Mapping histone modifications by nucleosome immunoprecipitation. *Methods Enzymol.* **410**, 416–430.
22. Negre, N., Lavrov, S., Hennetin, J., Bellis, M. and Cavalli, G. (2006) Mapping the distribution of chromatin proteins by ChIP on chip. *Methods Enzymol.* **410**, 316–341.
23. Wu, J., Smith, L. T., Plass, C. and Huang, T. H. (2006) ChIP-chip comes of age for genome-wide functional analysis. *Cancer Res.* **66**, 6899–6902.
24. Bulyk, M. L. (2006) DNA microarray technologies for measuring protein-DNA interactions. *Curr. Opin. Biotechnol.* **17**, 422–430.
25. O'Geen, H., Nicolet, C. M., Blahnik, K., Green, R. and Farnham, P. J. (2006) Comparison of sample preparation methods for ChIP-chip assays. *Biotechniques* **41**, 577–580.
26. Nelson, J. D., Denisenko, O. and Bomsztyk, K. (2006) Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat. Protoc.* **1**, 179–185.
27. Flanagin, S., Nelson, J. D., Castner, D. G., Denisenko, O. and Bomsztyk, K. (2008) Microplate-based chromatin immunoprecipitation method, Matrix ChIP: a platform to study signaling of complex genomic events. *Nucleic Acids Res.* **36**, e17.
28. Goren, A., Ozsolak, F., Shoresh, N., Ku, M., Adli, M., Hart, C., Gymrek, M., Zuk, O., Regev, A., Milos, P. M. and Bernstein, B. E. (2010) Chromatin profiling by directly sequencing small quantities of immunoprecipitated DNA. *Nat. Methods* **7**, 47–49.
29. Dahl, J. A. and Collas, P. (2008) MicroChIP – A rapid micro chromatin immunoprecipitation assay for small cell samples and biopsies. *Nucleic Acids Res.* **36**, e15.
30. Dahl, J. A. and Collas, P. (2008) A rapid micro chromatin immunoprecipitation assay (μChIP). *Nat. Protoc.* **3**, 1032–1045.
31. Orlando, V. (2000) Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* **25**, 99–104.

Chapter 15

Native Chromatin Immunoprecipitation

Céline Cosseau and Christoph Grunau

Abstract

Native chromatin immunoprecipitation refers to a method that allows for identification and quantification of DNA that is associated with specific chromatin proteins without altering the structure of these proteins. The method has been used with great success in the past and has some advantages over the more widely used cross-linking chromatin immunoprecipitation. We describe here a protocol that was specifically optimized for low cell numbers.

Key words: Native chromatin, Histones, Immunoprecipitation

1. Introduction

In contrast to cross-linking chromatin immunoprecipitation (xChIP), native ChIP relies on relatively strong interaction between the target proteins and DNA under physiological salt conditions (1, 2). Histones are a typical example of such proteins, but nChIP has also successfully been used with transcription factors (3). In this protocol, we focus on histone isoforms that are typical examples for a chromatin marking system. When histone-style proteins are targeted, we see three advantages of nChIP compared to xChIP: (1) the proteins remain in their native form and there is no danger that crosslinking fixes interactions that do not occur systematically in the cell; (2) nChIP is 10–100 times more sensitive than xChIP and less starting material is required; and (3) since enzymatic fragmentation of chromatin is used, no expensive equipment such as a sonicator is necessary. Nevertheless, xChIP and nChIP are complementary, and ideally both approaches should be used in parallel.

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_15,

2. Materials

2.1. Enzymes and Enzyme Inhibitors

1. 100 mM Dithiothreitol (DTT) in distilled water, aliquot to 1 ml (store at –20°C). DTT is a reducing agent that prevents the formation of disulfide bonds in and between proteins.
2. 25 mM Phenylmethanesulfonylfluoride (PMSF) in isopropanol, 10 ml (store at –20°C). PMSF is a serine protease inhibitor.
3. Roche Complete Protease Inhibitor tablets (ref: 11 697 498 001) (store at 4°C).
4. 2.5 M Sodium butyrate in distilled water (Sigma B5887 1 g) (store at 4°C). Sodium butyrate is an histone deacetylase (HDAC) inhibitor. It is not very stable and should not be stored for more than 4 weeks in solution. The product is irritant and the solution has a nauseating odor, wear gloves!
5. 15 U/µl Micrococcal nuclease (MNase) (EC 3.1.31.1) (USB 70196Y) in sterile 50% glycerol, aliquot to ~10 µl and store at –20°C. Do not refreeze. MNase digests DNA between nucleosomes (4, 5).
6. 2% NaN_3 in water (store at 4°C) as preservative. Sodium azide is very toxic.

2.2. Antibodies and Sepharose-Protein A

1. Aliquot antibodies on arrival to 2–4 µl and store at –20°C.
2. Preparation of sepharose-protein A:
   (a) Weight 250 mg sepharose-protein A in 15 ml falcon tube.
   (b) Wash with 10 ml sterile water.
   (c) Centrifuge for 10 min at 1,700 × *g*.
   (d) Remove supernatant.
   (e) Repeat washing step four times.
   (f) Add sterile water to 5 ml.
   (g) Add NaN_3 to 0.02% and store at 4°C.

250 mg Protein A-sepharose swell to approximately 1 ml gel and bind about 20 mg human IgG. You will need 50 µl of the homogeneously mixed sepharose-protein A per ChIP. Protein A has strong affinity to human, mouse, and rabbit IgG, for antibodies raised in goat and sheep, sepharose-protein G or protein A/G mixtures should be used. Paramagnetic sepharose particles are available but in our hands, background was on average 20 times higher than in a centrifugation-based separation.

2.3. Other Stock Solutions

1. 1 M KCl, autoclave.
2. 5 M NaCl, filter to prevent formation of crystals later and autoclave.

3. 1 M MgCl, autoclave.
4. 1 M Tris/Cl pH 7.4–7.6, autoclave.
5. 0.5 M EDTA, autoclave.
6. 1 M $CaCl_2$, 10 ml, sterile filter or autoclave.
7. 20% SDS, sterile filter.
8. 20 g/l glycogen solution, aliquot to 100 µl (store at –20°C).

2.4. Other Equipment

1. Microdialysis units (Slide-a-Lyzer 3500 D cut-off, Pierce 69550).
2. If microdialysis units are not available, prepare dialysis tubes:
   (a) Cut tube (e.g., VWR international dialysis tube 0.5 mm) into pieces of 10–20 cm length.
   (b) Boil for 10 min in a large volume of 2% (w/v) sodium bicarbonate and 1 mM EDTA.
   (c) Rinse the tube thoroughly with distilled water.
   (d) Boil for 10 min in 1 mM EDTA.
   (e) Cool and store in this solution at 4°C.
   (f) Before use, wash tube inside and outside with distilled water.
3. Centrifuges, stirrers, waterbath or similar, PCR machine, gel electrophoresis.

3. Methods

3.1. Step 1: Testing the Antibodies

By their very nature, immunoprecipitation methods rely on antibody–antigen interactions. Specificity and the strength of this interaction will determine the quality of the ChIP experiment in terms of specificity and sensitivity. Since histones, the principle targets of nChIP, are highly conserved proteins, most laboratories will prefer to buy antibodies instead of producing their own. Several monoclonal and polyclonal antibodies against histone isoforms are now commercially available. Some of them are "ChIP-certified," that is to say the manufacturer guaranties success of the immunoprecipitation if the "certified" animal or plant species are used. Nevertheless, antibody specificity and efficiency must always be tested. A simple Western blot gives a good indication of specificity. Already a straightforward protein extraction and separation on a standard SDS-PAGE will provide sufficient information (see Note 1). If bands can be detected outside the expected size range of the target proteins, and if the secondary antibody is not the origin of the problem, then antibodies from another source should be tested. Some companies provide antibody samples. While it cannot be excluded that antibodies that show unspecific binding in the

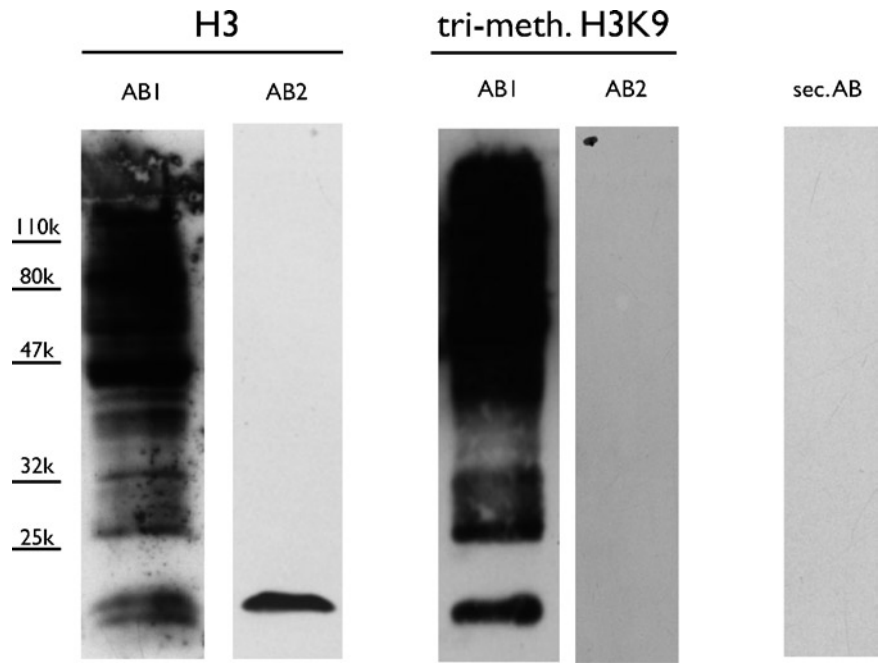


Fig. 1. Luminescence-revealed Western blots of *Schistosoma mansoni* protein extracts, separated on 15% SDS-PAGE and incubated with antibodies against histone H3 (*left*) and histone H3 trimethylated at lysine 9 (*middle*). In each lane, the same amount of protein was applied. Antibodies from two different suppliers (AB1 and AB2) were used and revealed with a peroxidase coupled secondary antibody. On the *right*, only the secondary antibody was used as control.

Western blot are specific in the IP conditions, such antibodies should be avoided. The nonspecific activity is probably due to immunogenic protein carriers used during immunization such as keyhole limpet hemocyanin KLH (6). An example is shown in Fig. 1. Ideally, a single band of the expected size should be visible. In the example in Fig. 1, antibodies from supplier AB1 should not be used for ChIP since additional bands are detected. Antibodies from supplier AB2 are suitable. However, also absence of any detection does not indicate that the antibody does not work in immunoprecipitation. If the preimmune serum is available, IgG from this serum must be used as control in all steps of the here described procedure.

3.2. Step 2: Chromatin Preparation

The principle of the procedure is shown in Fig. 2. The procedure follows tightly earlier protocols (7, 8) but has improved sensitivity and lower background (we use it routinely with less than 1,500 *Schistosoma mansoni* larvae, i.e., roughly 150,000 cells (9)). The following procedure describes nChIP for three samples and one control. One should use 10^6–10^7 (at least 10^5) cells for each sample. Cells can be aliquoted and stored at −80°C for up to 6 month or in liquid nitrogen for up to 12 month. However, it is preferable to use fresh biological material.

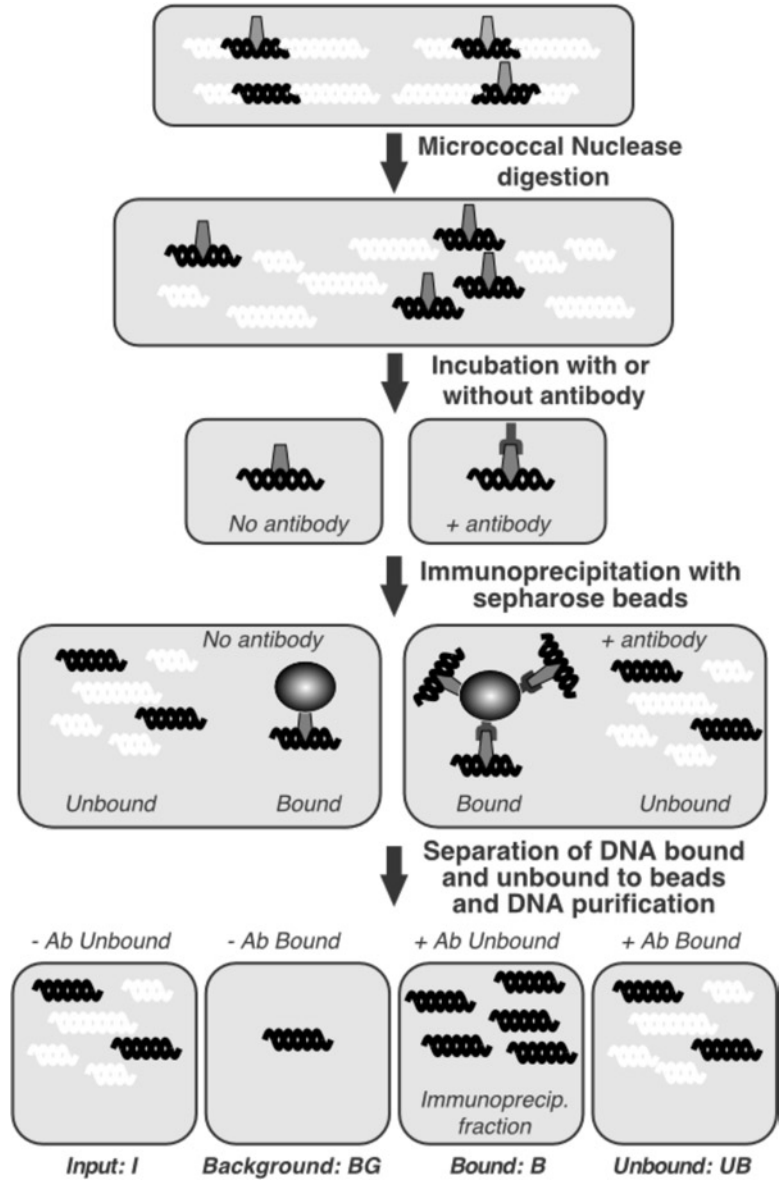


Fig. 2. Schematic representation of the nChIP procedure.

1. Prepare a centrifuge for 10 or 50-ml tubes, up to 7,700–7,800 ×*g* and cool it down to 4°C.
2. Preheat a water bath to exactly 37°C.
3. Prepare the solutions in Tables 1–6 (always freshly).

3.2.1. Cell Lysis and Purification of Nuclei

1. For culture cells or cell suspensions:
   (a) Multiply 10^6–10^7 cells with the number of antibodies to be used in immunoprecipitation (IP), for the control use the same amount of cells as for one antibody.

Table 1
2× base buffer

1x: 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15 mM Tris/Cl, pH 7.5	
6 ml	1 M KCl
1.5 ml	1 M Tris/Cl
0.3 ml	5 M NaCl
0.5 ml	1 M $MgCl_2$
20 µl	500 mM EDTA
to 50 ml with autoclaved distilled water, put on ice	
2 Roche protease inhibitor tablets	

Table 2
Buffer 1

0.3 M sucrose, 5 mM $CH_3CH_2CH_2COONa$, 0.1 mM PMSF, 0.5 mM DTT, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15 mM Tris/Cl, pH 7.5	
2.58 g	sucrose
12.5 ml	2× base buffer
50 µl	2.5 M sodium butyrate
100 µl	25 mM PMSF
125 µl	100 mM DTT
to 25 ml with autoclaved distilled water, put on ice	

Table 3
Buffer 2

0.3 M sucrose, 5 mM $CH_3CH_2CH_2COONa$, 0.1 mM PMSF, 0.5 mM DTT, 0.8% (v/v) NP40, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15 mM Tris/Cl, pH 7.5
10 ml buffer 1
Put on 37°C to allow Nonidet P-40 (NP40) to be pipetted into the buffer
80 µl NP40 or replacement product (Sigma-Aldrich has replaced Nonidet P-40 with a nonionic, nondenaturing substitute detergent (Fluka 74385). The principle is that the detergent will destroy the cytoplasmic membrane but not the nuclear membrane. Centrifugation separates the intact nuclei by precipitation.)
Put on 37°C to fully dissolve NP40, mix well but do not vortex and put on ice

Table 4
Buffer 3 for three samples and one control

1.2 M sucrose, 5 mM $CH_3CH_2CH_2COONa$, 0.1 mM PMSF, 0.5 mM DTT, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15 mM Tris/Cl, pH 7.5	
20.55 g	Sucrose
25 ml	2× base buffer
100 µl	2.5 M sodium butyrate
200 µl	25 mM PMSF
250 µl	100 mM DTT
to 50 ml with autoclaved distilled water, put on ice	

Table 5
MNase digestion buffers

0.3 M sucrose, 5 mM $CH_3CH_2CH_2COONa$, 0.2 mM PMSF, 4 mM $MgCl_2$, 1 mM $CaCl_2$, 50 mM Tris/Cl, pH 7.5	
1.1 g	Sucrose
0.5 ml	1 M Tris/Cl
80 µl	25 mM PMSF
40 µl	1 M $MgCl_2$
20 µl	2.5 M sodium butyrate
10 µl	1 M $CaCl_2$ (essential for the enzyme)
to 10 ml with autoclaved distilled water, put at 37°C	

Table 6
Dialysis buffers

5 mM $CH_3CH_2CH_2COONa$, 0.2 mM PMSF, 0.2 mM EDTA, 1 mM Tris/Cl, pH 7.5	
10–20 ml	Autoclaved distilled water
400 µl	25 mM PMSF
100 µl	2.5 M sodium butyrate
50 µl	1 M Tris/Cl
20 µl	500 mM EDTA
to 50 ml with autoclaved distilled water, put on ice	

(b) For each IP and for the control, transfer the corresponding amount of cells into a 2-ml Eppendorf tube and wash with 1 ml 150 mM NaCl (centrifuge for 670×*g* for 10 min at 4°C, remove supernatant).

2. For tissue samples: cut into small pieces and wash as above, use 10^6–10^7 cells per antibody.
3. Resuspend pellets completely in 1 ml buffer 1.
4. Add 1 ml buffer 2 and homogenize for 3 min with Dounce (pestle A) on ice.
5. Put on ice 7 min.
6. Fill 8 ml buffer 3 into a 14- or 50-ml corex centrifugation tube, label the tubes appropriately.
7. Use two corex tubes for each sample.
8. Overlay the 8-ml buffer 3 with 1 ml cell suspension so that the tubes are ready for centrifugation 10 min after buffer 2 has been added to the cells, disturb the interface between buffer 3 and the sample a little bit with the pipette tip.
9. If you do not use a swing-out rotor, mark the tubes on the exterior side (to know where to look for the nuclei).
10. Centrifuge 7,700–7,800×*g*, 20 min, 4°C (ideally in a swing-out rotor).
11. Remove the supernatant completely and carefully.

3.2.2. Chromatin Fragmentation with MNase

1. Resuspend the two nuclei-containing pellets for each sample in 1 ml MNase digestion buffer.
2. Aliquot 500 µl of this suspension into two 1.5 ml Eppendorf tubes.
3. Add 1 µl MNase (15 U) and incubate for 2–6 min at 37°C (see Note 2).
4. To stop the reaction add 20 µl 0.5 M EDTA to each 500 µl MNase digest and put the tube on ice.
5. Centrifuge 13,000×*g*, 10 min, 4°C.
6. Transfer the supernatant to a new tube (S1) and keep the pellet (P1) on ice.
7. Quantify chromatin in S1 by measuring OD at 260 nm against MNase buffer. (In general, we find about 50 µg/ml DNA in the undiluted S1, OD260/280 values can be bad because there is a lot of protein in the solution. DNA quantification is therefore not precise but sufficient for reproducibility).
8. Store S1 at –20°C.
9. Humidify Slide-a-Lyzer with 50 µl dialysis buffer.
10. Resuspend the pellet P1 in 100 µl dialysis buffer and dialyze overnight at 4°C against 50 ml dialysis buffer with gentle stirring. This dialysis step is important for the liberation of chromatin

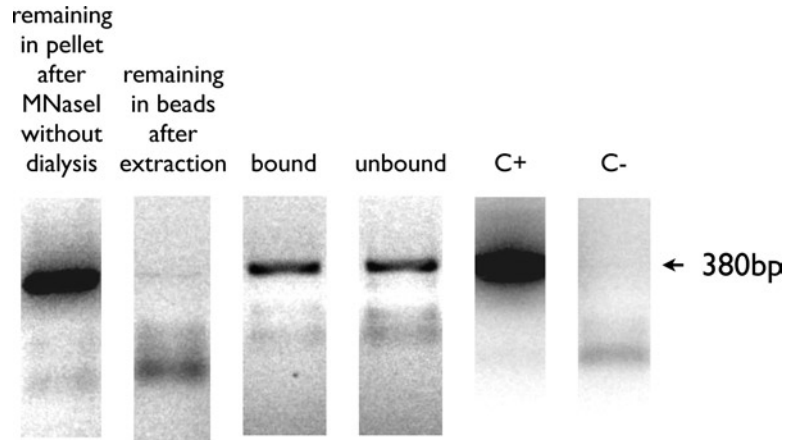


Fig. 3. PCR products separated on agarose-gels used for the detection of 28S rDNA genes in different fractions during the nChIP process. Stained with ethidiumbromide. Image inverted. On the left (*first lane*) PCR on DNA extracted from the S1 pellet before dialysis. Clearly, much chromatin is still present. In contrast, no DNA remains on the beads after DNA extraction that follows immunoprecipitation (*second lane*). C+ is genomic DNA, c- PCR without template (negative control).

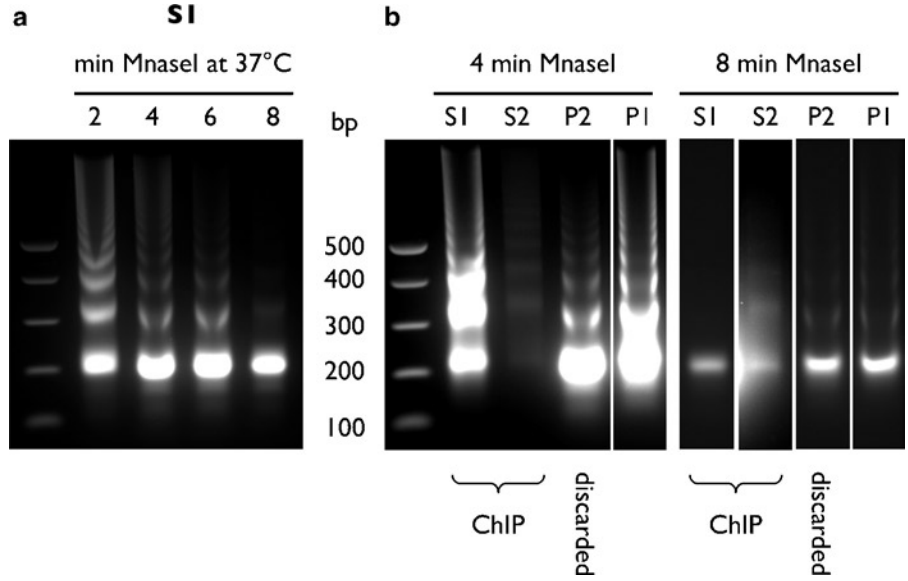


Fig. 4. (**a**) Optimization of the Mnasel digestion process. DNA was extracted from the S1 fraction and fragments were separated on a 2% agarose gel (50 V, 90 min). (**b**) Comparison of DNA extracted from fractions S1, S2, P1, and P2. Dialysis of P1 gives soluble fraction S2 and pellet P2. P2 is discarded because it contains cell debris that would interfere with the subsequent centrifugation steps. S2 fraction of 8 min digest was barely visible and the photo was software improved.

fragments from the nuclear debris. For some cell types this step can be skipped, if no DNA can be detected in the pellet. In general, the dialysis step is however necessary (Figs. 3 and 4).

11. The next day, prepare a 2% 0.5× TBE agarose gel with 20 µl slots.

12. Thaw yesterdays supernatant S1, transfer dialyzed sample to Eppendorf tubes.
13. Centrifuge both at 13,000 × *g*, 10 min, 4°C.
14. Transfer the supernatants to new tubes and repeat the centrifugation and transfer steps two times, discard the pellets (P2). These triple centrifugations are important! They reduce the unspecific background! If you still observe a pellet after the third centrifugation, repeat the centrifugation step.
15. The final supernatants are fractions S1 (nondialyzed) and S2 (dialyzed).
16. Use 50 µl of S1 and S2 for phenol/chloroform extraction, centrifuge, and load 20 µl of supernatant on a 2% 0.5× TBE agarose gel (100 V, 25 min) to verify the presence of mono- to pentanucleosomes.

3.2.3. Incubation with the Antibody

When the antibodies have passed the initial Western blot test, the right antibody-to-chromatin ratio must be determined. For a given amount of chromatin, the antibody must be in excess, and the amount of immunoprecipitated DNA must not depend on the amount of antibody used. This titration procedure can be done with the below outlined procedure, using a constant quantity of chromatin and increasing amounts of antibodies. From a certain quantity of antibody on, the amount of immunoprecipitated DNA should remain constant (Fig. 5). For further experiments, an antibody concentration

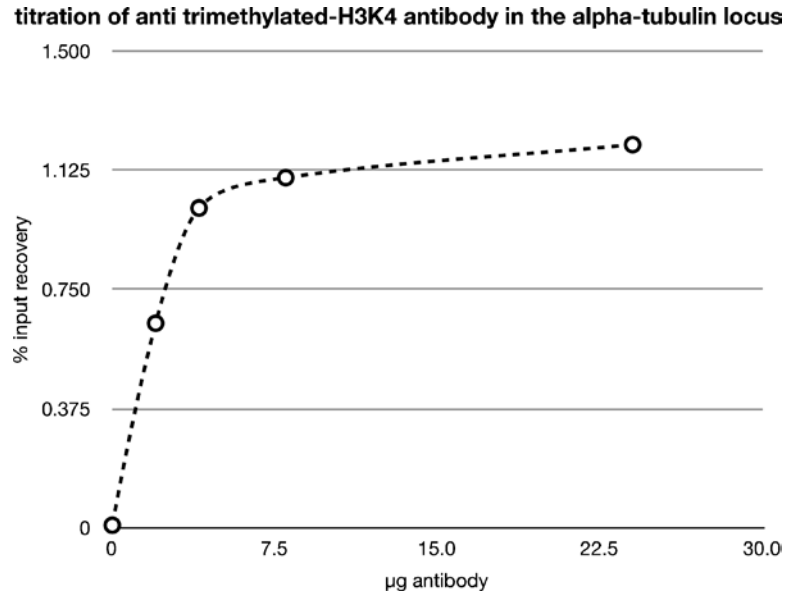


Fig. 5. Example for the results of a titration experiment. In this case, antibody saturation was reached at around 10 µg antibody. The antibody had been sold as "ChIP-grade" and supplier recommended amount was 2 µg.

above this threshold must be used. If no threshold can be reached, the antibody is not suitable for IP since the amount of precipitated DNA will be a function of the used antibody amount. The advantage is that even if antibody concentrations are unknown, the suitable amount (in µl) can easily be determined. The procedure is costly since relatively large amounts of antibodies are consumed and much biological material is required, but it assures that in the following steps reproducible results can be achieved. Failure to determine the correct antibody to chromatin ratio would lead to many difficulties in subsequent experiments.

1. Prepare a dilution series of your chromatin in MNase buffer starting with 20–40 µg chromatin DNA for histone ChIP and use 2–20 µg antibody (if the concentration is not known use 2–20 µl antibody).
2. Add appropriate amounts of stock solutions to generate the antibody incubation buffer (150 mM NaCl, 20 mM sodium butyrate, 5 mM EDTA, 0.2 mM PMSF, 20 mM Tris/Cl pH 7.5) by taking into consideration the amount of S1 and S2 and their respective buffers.
3. Dilute S1 and S2 in 1 ml final volume of antibody incubation buffer.
4. Add the appropriate amount of antibody (typically 4–8 µg). Add an equal volume to the control tube without antibody.
5. Incubate overnight at 4°C on a slowly rotating wheel.

3.3. Immunoprecipitation and DNA Extraction

1. Prepare 50 µl of sepharose-protein A for each tube.
2. Wash the beads to remove NaN_3: short spin, remove supernatant, and replace with equal volume of sterile water.

3.3.1. Immunoprecipitation

3. Add 50 µl of sepharose-protein A to each tube, including the control tube.
4. Incubate at least for 4 h at 4°C on a rotating wheel.
5. Prepare washing buffers (10 ml/tube) and cool down to 4°C (Table 7).
6. Centrifuge chromatin–antibody–beads mixture for 10 min, 4°C, 11,600×*g*.
7. Keep the supernatant in a 2-ml tube. This is the unbound fraction UB.
8. Resuspend the pellet in approximately 1 ml washing buffer A and transfer into a 15-ml Falcon tube containing 9 ml washing buffer A.
9. Mix for 10 min on a rotating wheel at 4°C.
10. Centrifuge for 10 min 1,700×*g* 4°C and pour off supernatant.

Table 7
Washing buffers A, B, and C (example 100 ml)

5 mM $CH_3CH_2CH_2COONa$, 10 mM EDTA, 75 mM (A)/125 mM (B)/175 mM (C) NaCl, 50 mM Tris/Cl, pH 7.5	
1.5 ml (A) or 2.5 ml (B) or 3.5 ml (C)	5 M NaCl
200 µl	2.5 M sodium butyrate
5 ml	1 M Tris/Cl
2 ml	500 mM EDTA
to 100 ml with autoclaved distilled water, put on ice	

Table 8
Elution buffer

1% SDS, 50 mM NaCl, 5 mM EDTA, 20 mM sodium butyrate, 0.1 mM PMSF, 20 mM Tris/Cl pH 7.5	
500 µl	20% SDS
200 µl	1 M Tris/Cl
100 µl	5 M NaCl
80 µl	2.5 M sodium butyrate
40 µl	25 mM PMSF
to 10 ml with autoclaved distilled water	

11. Add 10 ml washing buffer B, mix for 10 min on a rotating wheel at 4°C, and centrifuge for 10 min 1,700 × *g*, 4°C.
12. Pour off the supernatant.
13. Add 10 ml washing buffer C, mix for 10 min on a rotating wheel at 4°C and centrifuge for 10 min, 1,700 × *g*, 4°C.
14. Pour off the supernatant.
15. Centrifuge for 10 min, 1,700 × *g*, 4°C.
16. Remove the remaining supernatant completely (centrifuge if necessary).
17. Resuspend the pellet in 500 µl elution buffer (Table 8) at room temperature.
18. Transfer to a 1.5-ml tube and incubate for 15 min at room temperature on a rotating wheel.

19. Centrifuge for 10 min, 11,600 × *g*, 18–20°C.
20. Transfer the supernatant to a 1.5-ml tube. This is the bound fraction B.

To make sure, that the correct protein was immunoprecipitated, it is advisable to do a Western blot for fraction B. This is often not trivial since Protein A will leach into the SDS gel, and bands corresponding to the heavy and light chains of the antibody will be detected. Using beads coated with recombinant Protein G can in some cases provide a solution.

3.3.2. DNA Extraction

1. Use a standard phenol/chloroform extraction of DNA (10) in fractions B and UB.
2. Add 1 µl of a 20-g/l glycogen stock solution.
3. Add NaCl to 250 mM (26 µl and 52 µl for B and UB, respectively) and add 1 volume isopropanol.
4. Put overnight at –20°C or 30 min at –80°C.
5. Precipitate by centrifugation and wash with 70% ethanol.
6. Dry the pellet and resuspend in 40 µl 10 mM Tris/Cl.
7. For single-copy loci use 2.5 µl of this DNA for PCR in 10 µl reactions (quantitative real-time PCR) or continue with massive sequencing (see Chapter 22). In qPCR, single locus DNA should deliver Ct values of around 25 when starting with 150,000 cells.

3.4. Step 5: Quantification of Precipitated DNA

The amount of genomic region of interest (RoI) must be determined in the unbound and bound fractions of the control without antibody, and in the bound fraction with antibody (B). The unbound control fraction corresponds to the input (I), and the bound control fraction is the background (BG), i.e., DNA that sticks to the beads (Fig. 2). If no DNA can be detected in fraction B, the antibody unbound fraction UB must be analyzed, in order to exclude that DNA was entirely degraded. Essentially, there are two possibilities to determine the quantity of RoI, either by comparing to an internal standard (reference locus method), or by determining the proportion of immunoprecipitated DNA compared to the input (% input recovery method).

3.4.1. Relative Quantification (Reference Locus Method)

In this case, in fraction B, I, and BG a reference RoI and target RoI are quantified by quantitative PCR. The relative enrichment is calculated according to the following simplified formula:

$$\text{Enrichment factor} = 2^{-(\Delta Ct)},$$

where

$$\Delta Ct = (Ct_{\text{target}} - Ct_{\text{reference}})_B - (Ct_{\text{target}} - Ct_{\text{reference}})_I.$$

In this case, it is assumed that the amplification efficiency of the PCR is close to 2 and identical for the two qPCR (for a detailed description of the underlying assumptions of the method and possible pitfalls see ref. 11). It could also be envisaged to use an efficiency-corrected comparative quantification or multiple reference genes.

By using a standard curve, the amount for each locus can be determined based on a dilution series of DNA of known concentration. The PCR efficiencies can be different. In this case,

$$\text{Enrichment factor} = [\text{ng target(B)} / \text{ng reference(B)}] / [\text{ng target(I)} / \text{ng reference(I)}].$$

The standard curve method should be preferred if sufficient qPCR reactions can be performed simultaneously. The advantage of this quantification method is that target to reference DNA ratio in the same tube is measured. Even if DNA is lost during the purification process, the relative enrichment remains the same. Despite the relatively complex procedure, enrichment factors are very reproducible; in our hands standard errors are around 10%. The reference locus can, for instance, be in the body of a housekeeping gene. The enrichment factor described how much more chromatin modifications are associated with the target locus compared to the reference. A principal caveat of this quantification method is that it assumes that chromatin structure in the reference locus does not change. This is probably true for regions in the body of housekeeping genes, but it cannot be excluded, that under particular conditions, these regions change their chromatin status in parallel with the target locus. In this case, the relative quantification method would not be appropriate.

3.4.2. Absolute Quantification (% Input Recovery Method)

A solution could be used to quantify directly the amount of precipitated target RoI (bound) and to compare it to the amount of target RoI in the input I (unbound fraction of control). In general, this ratio is expressed in % input recovery (12):

$$\%\ \text{input recovery} = 100 \times \text{PCR} - \text{efficiency}^{(\text{Ct - input} - \text{Ct - bound})}.$$

The problem with this method is that the DNA in two different tubes is compared (B and UB-Control). If more DNA is lost during the preparation process in one tube, this will induce errors that cannot be detected. The method is sensitive to pipetting errors, requires careful standardization and, naturally, several technical and biological duplicates must be performed. Even then, day-to-day variations can be high (Fig. 6). Input recovery of BG should be ≤0.1%.

In conclusion, for a successful nChIP, it is indispensable to test the specificity of the antibodies and to determine the optimal amount by titration, prior to the experiment. For titration, a suitable target region in the genome must be known. For histone isoforms this will be relatively straightforward, for other chromatin proteins with restricted target regions, this might be more difficult.

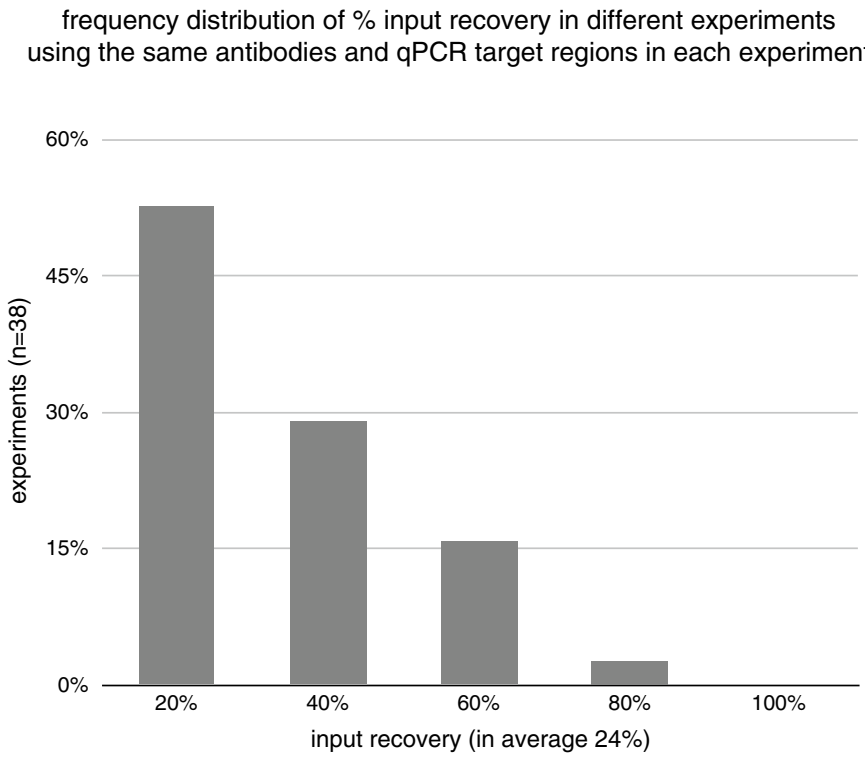


Fig. 6. Frequency distribution of % input recovery for anti-acetyl H3K9 in the *Schistosoma mansoni* alpha-tubulin promotor region in 38 experiments. The graph illustrates the relatively high day-to-day variations with the quantification method. The % input recovery corresponds to the proportion of DNA in a specific locus that is precipitated (fraction B), compared to the unbound fraction in the tube without antibody (input I). Differences in these fractions do not only reflect differences in binding versus nonbinding to the beads but can be introduced through experimental errors (e.g., pipetting errors, loss of precipitated material). In half of the shown experiments, input recovery in this locus was around 20%, but in other experiments it was higher. In one case it exceeded 80%, probably due to a loss of material during the purification of the input DNA that served as reference.

Nevertheless, and for obvious reason, this positive control region is necessary. In the experiment itself, input recovery in the background must be low (≤0.1%). In our hands, quantification via the reference locus method gives better reproducibility. The % input recovery method has relatively high day-to-day variations and, apart from the comparison to the background, it should only be applied if a reference locus cannot be used.

4. Notes

1. Protein extraction and Western Blot: For high-resolution analysis of histones and histone isoforms, classical acid extraction and Triton–acetic acid–urea (TAU) gels should be used (13). For antibody testing we recommend the faster following

method: Roughly 3×10^5 cells are suspended in denaturation buffer (0.2% bromophenol, 10% sucrose, 3% SDS, and 0.2 M DTT, 62.5 mM Tris/Cl, pH 6.8) treated by sonication (e.g., 60% intensity of Vibra cell sonicator 75185, six times 15 s, with cooling intervals on ice) and boiled for 10 min at 95°C. Proteins are separated by conventional 15% SDS-PAGE and transferred for 1 h to nitrocellulose membranes (e.g., Amersham RPN203D) by the semidry method. The membranes are blocked overnight at 4°C in blocking buffer (150 mM NaCl, 0.05% v/v Tween 20, 5% w/v fat-free dry milk, 20 mM Tris/Cl, pH 7.5) and incubated with the desired antibody for 2 h in blocking buffer. The membranes are washed with washing buffer (150 mM NaCl, 0.05% v/v Tween 20, 20 mM Tris/Cl, pH 7.5) and incubated with a peroxidase-coupled secondary antibody for 30 min in blocking buffer. Bands are revealed by chemical luminescence (e.g., ECL Pierce) and direct exposure to x-ray film (Amersham EmNo.27304). For stripping, incubate the membrane at 50°C for 1 h in stripping solution (2% SDS, 0.8% beta-mercaptoethanol, 62 mM Tris/Cl pH 6.8) in a tightly closed 50 ml tube and wash with distilled water several times until the smell of beta-mercaptoethanol has completely disappeared. A membrane can be used three to four times.

2. Optimization of MNase digestion: The digestion time must be determined experimentally and can be different for each cell type and each MNase preparation. To determine the optimal digestion time, transfer every minute 100 µl of the reaction mix into Eppendorf tubes containing 10 µl 0.5 M EDTA on ice. Mix 50 µl of each fraction with 50 µl phenol–chloroform, vortex, centrifuge, and apply 20 µl of the upper phase on the 2% 0.5× TBE agarose gel. Separate the fragments (100 V, 25–35 min) and observe after ethidiumbromide or other DNA staining. Optimal digestion is achieved when DNA fragments corresponding to 1–5 nucleosomes are visible. For ChIP followed by massive sequencing (ChIP-Seq), mononucleosomes are preferred, however, these mononucleosomes should not be produced by prolonged digestion with MnaseI since this will introduce a bias in heterochromatin to euchromatin ratio. Other methods such as sucrose gradient centrifugation should be used (7). In our hands, typical incubation times for 10^6 cells and 37°C are 2–6 min. An example is shown in Fig. 4a. In this case, liver tissue of a Syrian hamster (*Mesocricetus auratus*) was used. After 8 min, virtually all chromatin is digested to mononucleosomes. Fig. 4b illustrates the distribution of DNA fragments in the different fractions of the chromatin preparation. Fraction S2 is enriched in larger fragments. It should be noted that MnaseI incubation time influences also on the relative

Table 9
% DNA in the different fractions after 4 and 8 min Mnase I digest at 37°C

	4 min Mnase I				8 min Mnase I				
Locus	**S1**	**S2**	**S1 + S2**	**P2**	**S1**	**S2**	**S1 + S2**	**P2**	**Quantification method**
Total DNA	49	18	67	33	37	15	52	48	Nanodrop
P53 (Genbank U08134)	74	15	89	11	13	69	82	18	qPCR (176 bp)
5S (Genbank J00063)	66	8	74	25	11	4	15	85	qPCR (101 bp)
Heterochromatic repeat (Genbank AB185090)	79	6	85	15	10	3	13	86	qPCR (151 bp)

amount of DNA in the different fractions. This is shown in Table 9. In this example, after 4 min MnaseI treatment, the distribution of different loci in the hamster genome is relatively homogenous. In contrast, after incubation for 8 min with MnaseI, distribution is uneven. Since the majority of the chromatin in S1 and S2 is digested to mononucleosomes of around 150 bp, this might also reflect problems in PCR amplification for larger primer distances.

References

1. Gottesfeld,J.M. and Luger,K. (2001) Energetics and affinity of the histone octamer for defined DNA sequences. *Biochemistry* 40: 10927–10933.
2. Khrapunov,S.N., Dragan,A.I., Sivolob,A.V. and Zagariya,A.M. (1997) Mechanisms of stabilizing nucleosome structure. Study of dissociation of histone octamer from DNA. *Biochim. Biophys. Acta* 1351: 213–222.
3. Roca,H. and Franceschi,R.T. (2008) Analysis of transcription factor interactions in osteoblasts using competitive chromatin immunoprecipitation. *Nucleic Acids Res.* 36: 1723–1730.
4. Greil,W., Igo-Kemenes,T. and Zachau,H.G. (1976) Nuclease digestion in between and within nucleosomes. *Nucleic Acids Res.* 3: 2633–2644.
5. Hewish,D.R. and Burgoyne,L.A. (1973) Chromatin sub-structure. The digestion of chromatin DNA at regularly spaced sites by a nuclear deoxyribonuclease. *Biochem. Biophys. Res. Commun.* 52: 504–510.
6. Oulehlova,D., Hala,M., Potocky,M., Zarsky,V. and Cvrckova,F. (2009) Plant antigens cross-react with rat polyclonal antibodies against KLH-conjugated peptides. *Cell Biol Int* 33: 113–118.
7. O'Neill,L.P. and Turner,B.M. (2003) Immunoprecipitation of native chromatin: NChIP. *Methods* 31: 76–82.
8. Umlauf,D., Goto,Y. and Feil,R. (2004) Site-specific analysis of histone methylation and acetylation. *Methods Mol. Biol.* 287: 99–120.
9. Cosseau,C., Azzi,A., Smith,K., Freitag,M., Mitta,G. and Grunau,C. (2009) Native chromatin immunoprecipitation (N-ChIP) and ChIP-Seq of Schistosoma mansoni: Critical experimental parameters. *Mol Biochem Parasitol* 166: 70–76.
10. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular cloning : a laboratory manual. Cold Spring Harbor Laboratory.

11. Livak,K.J. and Schmittgen,T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) Method. *Methods* 25: 402–408.
12. Edmunds,J.W., Mahadevan,L.C. and Clayton,A.L. (2008) Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* 27: 406–420.
13. Ryan,C.A. and Annunziato,A.T. (2001) Separation of histone variants and post-translationally modified isoforms by triton/acetic acid/urea polyacrylamide gel electrophoresis. *Curr Protoc Mol Biol* Chapter 21: Unit 21.2.

Chapter 16

Q-PCR in Combination with ChIP Assays to Detect Changes in Chromatin Acetylation

Ryan A. Irvine, Cindy Okitsu, and Chih-Lin Hsieh

Abstract

Quantitative polymerase chain reaction (Q-PCR) allows for the accurate and reproducible determination of the amount of target DNA in a sample through the measurement of PCR product accumulation in "real time." This method determines starting target DNA quantity over a large assay dynamic range and requires no post-PCR sample manipulation. When used in combination with the method of chromatin immunoprecipitation (ChIP), the amount of protein binding to a specific region of DNA can be accurately and rapidly determined. A method for quantifying the presence of acetylated histones H3 and H4 on different regions of a target locus using Q-PCR after ChIP is described.

Key words: Quantitative PCR, Chromatin immunoprecipitation, Histone acetylation, Chromatin structure

1. Introduction

This chapter details a method for using quantitative polymerase chain reaction (Q-PCR) in combination with chromatin immunoprecipitation (ChIP) to determine the extent of histone acetylation at discrete regions of nucleosomal DNA. This method was used in a series of studies to assess the impact of DNA methylation on gene transcription and local chromatin structure and the impact of transcription on histone H3 lysine 4 (H3K4) modifications in cultured human cells (1–3). Although the described method was optimized for the analysis of DNA sequences on a replicating episome in human cells, it is generally applicable to the analysis of genomic loci. The ChIP technique provides a powerful in vivo tool for detecting protein–DNA interactions through formaldehyde

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_16,

crosslinking of proteins to DNA, followed by immunoprecipitation with specific antibodies (see Chapter 14 and ref. 4). Theoretically, any protein and its in vivo binding site can be detected by this technique. In our studies, the presence and absence of specific histone H3 modifications being influenced by DNA methylation can be determined within a distance of less than 200 bp of DNA indicating the potential resolution of this method. After immunoprecipitation of crosslinked chromatin, the formaldehyde crosslinks are reversed and the DNA is purified and analyzed with various methods. ChIP assay in combination with microarray (ChIP-Chip) or deep sequencing (ChIP-seq) allows the global mapping of specific protein binding sites (see Chapter 21). Determination of the amount of immunoprecipitated DNA from a specific genomic or extrachromosomal region by Q-PCR provides a direct assessment of the frequency of protein association with that region. In this way, not only can a detailed "map" of protein binding to a particular gene or region be generated, but the relative amount of protein binding among different samples can also be assessed.

In addition to its application to DNA isolated by ChIP, Q-PCR has been used successfully to quantify gene expression, gene copy number, DNA damages, viral load, and minimal residual disease in cancer patients (5–8). The Q-PCR strategy used in the described method employs a dual-labeled fluorescent TaqMan probe that anneals to the target sequence within the PCR amplicon (9). During the extension phase of PCR, the 5′ exonuclease activity of *Taq* polymerase cleaves the probe, liberating the 5′-fluorescent reporter molecule from the linked 3′-quencher moiety. With successive rounds of PCR amplification, the reporter fluorescence in the reaction increases. This change in fluorescent emission over time (or cycle number) is detected and measured by a charge-coupled device (CCD) camera equipped with filters that absorb light at the appropriate wavelength. Importantly, during the exponential phase of PCR amplification, a linear relationship exists between the log of the amount of starting template DNA and the cycle at which the reporter fluorescence significantly exceeds background levels (i.e., the threshold cycle). Consequently, using a standard curve of known amounts of the target sequence (e.g., a dilution series of a plasmid that contains the sequence), the quantity of target DNA in an unknown sample can be accurately determined from its threshold cycle. Although the cost of custom probes for TaqMan assay has reduced recently, the overall cost may be substantial when a large number of targets are of interest. The use of intercalating dyes such as SYBRGreen or the Universal Probe Library (Roche) is alternatives to the TaqMan assay using custom probes described in this protocol. Special attention to the specificity of the assay is highly recommended if intercalating dyes are used because all amplified DNA is detected indiscriminately.

Q-PCR is a highly reproducible and sensitive technique that offers clear advantages over traditional PCR, semiquantitative PCR, and Southern blotting methods. Traditional PCR is not quantitative and is primarily used for endpoint analysis by assaying the amplified product from the "plateau phase" of the reaction. By amplifying a dilution series of template DNA over a limited number of cycles, some degree of quantification has been achieved with the so-called semiquantitative PCR. Q-PCR, on the other hand, provides a direct measurement of amplified product at every PCR cycle and therefore ensures that template quantification occurs during the exponential phase of the reaction. Q-PCR, furthermore, requires no postamplification handling of the reaction products. Southern blotting is another method that provides semiquantification of target DNA, although it is generally less preferable than the PCR methods owing to the relatively large amounts of starting material that are required.

2. Materials

2.1. Chromatin Immunoprecipitation

1. Phosphate-buffered saline (PBS), pH 7.2.
2. Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum and hygromycin (200 μg/mL).
3. Trypsin/EDTA solution: 0.5 g trypsin, 0.2 g EDTA per liter (Irvine Scientific).
4. TE: 10 mM Tris–HCl, 1 mM EDTA, pH 8.0.
5. 37% Formaldehyde solution (JT Baker).
6. 10% Sodium dodecyl sulfate (SDS).
7. Radioimmunoprecipitation assay (RIPA) buffer: 10 mM sodium phosphate, pH 7.2, 2 mM EDTA, 150 mM NaCl, 50 mM NaF, 0.2 mM Na_3OV_4, 1% sodium deoxycholate, 1% Nonidet P40, 0.1% SDS. RIPA buffer is stable for up to 1 year at 4°C. The sodium orthovanadate should be added immediately before use from a 0.2-M stock solution.
8. Protease inhibitor cocktail tablets (Roche). One tablet is dissolved in 10 mL of RIPA buffer or PBS immediately before use.
9. Antiacetyl-histone H3/H4 antibodies (Millipore).
10. 10 mg/mL salmon sperm DNA (Sigma). The DNA is sheared by sonication to a range of 200–500 bp, phenol/chloroform extracted, ethanol precipitated, and resuspended in water.
11. Protein G Sepharose 4 Fast Flow (GE Healthcare). The Sepharose beads are equilibrated in RIPA buffer through repeated washing steps. A final 50% (v/v) slurry in RIPA buffer is prepared for use.

12. Proteinase K (Roche). A 20 mg/mL stock solution is prepared in H_2O.
13. Elution buffer: 0.1 M $NaHCO_3$, 1% SDS.
14. 3 M Sodium acetate, pH 5.2.
15. 20 mg/mL glycogen (Roche).
16. Phenol (TE, pH 8.0 buffered, saturated).
17. Chloroform.
18. Ethanol.

2.2. Quantitative PCR

1. 10× PCR buffer: 150 mM Tris–HCl, pH 8.0, 500 mM KCl.
2. 10× PCR stabilizer: 0.5% gelatin, 0.1% Tween-20.
3. 40% Glycerol solution.
4. 350 mM $MgCl_2$ stock solution.
5. 2.5 mM Deoxyribonucleotide triphosphate (dNTP) mix.
6. AmpliTaq Gold DNA polymerase (PE Applied Biosystems).
7. Fluorophore-labeled TaqMan probe (5′-FAM [6-carboxyfluorescein]) labeled, 3′-Black Hole Quencher [BHQ]-1 labeled (Biosearch Technologies). Fluorogenic probes are dissolved in H_2O or TE, stored at −20°C, and protected from direct light exposure.
8. Oligonucleotide primers (Operon Technologies).
9. iQ Supermix (Bio-Rad) and be used to replace nos. 1, 2, 3, 4, 5, and 6 above.

3. Methods

3.1. Chromatin Immunoprecipitation

This ChIP protocol is derived from previously published methods (10, 11).

1. Grow human 293/EBNA1 cells (a human embryonic kidney carcinoma cell line that expresses the viral replication factor EBNA1 (12)) transfected with replicating episomes, in DMEM that contains 10% FBS and 200 μg/mL hygromycin.
2. Wash exponentially growing cultures that are 70–80% confluent (approx 3×10^6 cells per 10-cm dish; see Note 1) once with PBS and then trypsinize with 1 mL trypsin/EDTA solution for 2–5 min at room temperature (RT).
3. To neutralize the trypsin, add 3 mL of media to the dish and transfer the resulting 4 mL cell suspension to a 14-mL culture tube (Sarstedt).

4. To fix the cells, add 108 μL of 37% formaldehyde solution to the cell suspension (i.e., 1% final concentration) followed by gentle rotary mixing for exactly 10 min at RT (see Note 2).
5. Collect fixed cells by centrifugation at 2,000 × *g* in a Beckman GS-6KR swinging (bucket centrifuge) for 5 min at 4°C.
6. Discard the supernatant and resuspend the cell pellet in 4 mL of ice-cold PBS and collect by centrifugation as in step 5.
7. Repeat step 6.
8. Following complete removal of the second PBS wash, snap freeze the cell pellet in liquid nitrogen, and store at −80°C until use (see Note 3).
9. Thaw the cell pellet on ice, thoroughly resuspend in 1 mL of RIPA buffer that contains protease inhibitors, and transfer to a 1.9-mL Eppendorf tube.
10. Perform sonication with a Branson Sonifier 450, equipped with a microtip, using an output power setting of 5 and a constant duty cycle (i.e., 100%). Subject the cell suspension to 20 consecutive 10-s bursts with cooling on a dry ice/ethanol bath between each sonication event. The cell lysate must be kept cool throughout the sonication procedure, as high temperatures can denature proteins. It cannot be overstated that the sonication regimen is critical to the success of the ChIP technique and must be empirically determined for each study (see Note 4). In our hands, this procedure produces average chromatin fragment sizes of about 1 kb or less for 293/EBNA1 cells fixed in 1% formaldehyde for 10 min at RT.
11. Following sonication, centrifuge the lysate at 16,000 × *g* in a microcentrifuge for 15 min at 4°C to pellet insoluble material. The supernatant is the soluble chromatin sample.
12. For preclear, add 100 μL of the 50% slurry of Protein G Sepharose (prepared in RIPA buffer) to the soluble chromatin sample and incubate at 4°C with gentle rotary mixing for at least 1 h (see Note 5).
13. Remove a 50-μL aliquot of the soluble chromatin, mix with 50 μL TE, and store as the total chromatin fraction (TCF) at −20°C until use (see step 22).
14. Divide the remainder of the soluble chromatin sample into two 475 μL fractions (see Note 6) in 1.9-mL Eppendorf tubes on ice: fraction 1, antiacetyl H3/H4 and fraction 2, no antibody (Ab).
15. To the antiacetyl H3/H4 tube only, add 5 μg each of antiacetylated histone H3 and antiacetylated histone H4 IgG. Add nothing to the no-Ab tube (alternatively, a completely unrelated antibody of the same isotype may be used) (see Note 7).

16. Incubate both tubes for 2 h or longer at 4°C with gentle rotary mixing.
17. To each tube, add 40 μg of sheared salmon sperm DNA and 100 μL of the 50% slurry of Protein G Sepharose (prepared in RIPA buffer).
18. Incubate the samples overnight at 4°C with rotary mixing as before.
19. Collect the Sepharose beads by centrifugation at 2,000 × *g* in a microcentrifuge for 2 min at 4°C, and then transfer the unbound soluble chromatin to a new tube and store at –20°C.
20. Wash the Sepharose beads sequentially for 10 min each at 4°C, once with 0.5 mL of ice-cold RIPA buffer and twice with 0.5 mL of ice-cold PBS.
21. Elute immunocomplexes from the Protein G Sepharose in 0.5 mL of elution buffer by incubation for 15 min at RT with rotary mixing. Spin down the Sepharose, and transfer the supernatant to a clean tube. Repeat the elution step and combine the eluates (see Note 8).
22. To each eluate, add 25 μL of 5 M NaCl and then incubate at 65°C for 4 h to reverse the formaldehyde crosslinks. To the 100 μL TCF sample, add 5 μL 5 M NaCl and incubate at 65°C for 4 h as well.
23. Add 10 μL of 0.5 M EDTA, 20 μL 1 M Tris–HCl (pH 6.5) and 2 μL proteinase K (20 mg/mL) to the 500 μL eluate. Add 2 μL of 0.5 M EDTA, 4 μL 1 M Tris–HCl (pH 6.5) and 0.4 μL of 20 mg/mL proteinase K to the 100 μL TCF (see step 13) samples. Incubate at 55°C for at least 1 h.
24. Extract the samples with 1 volume of phenol/chloroform (1:1), ethanol-precipitate the DNA in the presence of 20 μg of glycogen and 0.3 M Sodium acetate (pH 5.2), wash with 70% ethanol, and dissolve in 50 μL TE.

3.2. Quantitative PCR

1. Design TaqMan probes and primers using the Primer Express (v3.0) software package from PE Applied Biosystems. This program facilitates the selection of probe/primer sets that result in high-efficiency target amplification using a simple two-step cycling program. In general, a TaqMan probe should be selected that is less than 40 nucleotides in length, has a G/C content in the 20–80% range, has a melting temperature (T_m) near 70°C that is 8–10°C higher than the primers, has no runs of four or more of an identical nucleotide (especially G), and does not have a G at the 5′ end. Primers should be 10–20 nucleotides in length and designed as close as possible to the probe without overlapping it so that the amplicon is between 50 and 150 bp in size. Small amplicon size serves to maximize PCR efficiency. In addition, to avoid nonspecific amplification,

primers should not be rich in G and/or C bases at their 3′ ends. A general guideline is that primers should not have more than two G and/or C bases in the last five nucleotides at the 3′ end. All TaqMan probes used in our studies were synthesized with a 5′-FAM (or other) reporter fluorophore and a 3′-BHQ-1 quencher (Biosearch Technologies). All primers were synthesized by Operon Technologies.

2. Amplify ChIP DNA samples (see Subheading 3.1 above) in 20 μL reactions that contain 1 μL of DNA template (see Note 9), 15 mM Tris–HCl, pH 8.0, 50 mM KCl, 3.5 mM $MgCl_2$, 200 μM dNTPs, 0.05% gelatin, 0.01% Tween-20, 8% glycerol, 50 nM TaqMan probe, 300 nM of each primer, and 1 U of AmpliTaq Gold DNA polymerase (see Note 10). Alternatively, the 20 μL reactions can be set up to contain 1 μL of DNA template, 50 nM TaqMan probe, 300 nM of each primer, and 10 μL of iQ Supermix. Prepare reactions at RT in iCycler iQ PCR 96-well plates (Bio-Rad) and then seal them with Microseal "B" film (Bio-Rad). Perform PCR cycling and fluorescence detection with a Bio-Rad iCycler using the following two-step program after an initial 10-min incubation at 95°C to activate the polymerase: 95°C for 15 s and 60°C for 1 min for 40 cycles. In general, the PCR conditions and cycling parameters listed here produced high-efficiency amplification of the target DNA sequences used in our studies. It should be recognized, however, that probe and primer concentrations might need to be optimized for the efficient amplification of other target sequences (see Note 11).

3. To quantify the amount of target DNA in a given ChIP sample, simultaneously amplify a standard curve consisting of titrations of a known amount of the target sequence on the same Q-PCR plate. A fivefold dilution series of the 12 kb pCLH22 episome, from 100 to 0.00128 pg, served as the standard curve in our experiments (1). This range of input target DNA is compatible with the known dynamic range of linear response for the iCycler. The efficiency of a particular Q-PCR reaction is inferred from the slope of the standard curve graph (log of input DNA [pg] versus threshold cycle) generated by the iCycler iQ software package (v3.1). Q-PCR reactions with standard curve slopes of −3.0 to −3.7 were arbitrarily accepted for our experiments (a slope of −3.32 indicates 100% efficiency). Standard curve slopes outside of this range are indicative of either inefficient Q-PCR reactions or of reactions generating substantial nonspecific products, both of which must be optimized before continuing with ChIP sample analysis. It should be stated that some Q-PCR applications may require more stringent efficiency parameters (e.g., absolute mRNA quantification via reverse-transcriptase Q-PCR). The standard curve is a critical

feature of Q-PCR as the amount of DNA in all unknown samples is extrapolated from it. Therefore, it must be empirically validated before reliable quantitative results can be attained.

4. Perform all Q-PCR reactions (unknown ChIP samples and standards) in duplicate or triplicate and use mean quantities for all subsequent data analyses. The fraction of immunoprecipitated DNA was calculated by subtracting the amount of DNA in the no-Ab control from that in the antiacetyl H3/H4 sample and dividing by the amount of DNA in the corresponding TCF sample. Using this calculation, we determined the amount of acetylated histones H3 and H4 that were associated with different regions of the pCLH22 episome in 293/EBNA1 cells.

4. Notes

1. For our experiments, the target DNA sequences for immunoprecipitation (IP) were on episomes that were maintained at 10–100 copies per cell. Therefore, the sensitivity of target DNA detection per cell in our assays was greater than that for standard ChIP experiments, in which targets are endogenous genomic sequences with two copies per cell. Consequently, only about 3×10^6 cells were required for each IP described here, compared with approx 10^8 cells recommended for other ChIP strategies (e.g., ref. 13).
2. It is our experience that the duration of formaldehyde fixation is a critical parameter of this technique. Modest variations in fixation times can have a significant, negative impact on assay reproducibility. Thus, once a fixation time has been validated for a particular cell type and ChIP experiment, it should be used consistently. In general, crosslinking times of 10 min or less are suitable for ChIP analyses of nucleosomal proteins; longer times (i.e., 30 min to 1 h) may be required for other DNA binding factors. The formaldehyde fixation may be stopped by the addition of glycine to a final concentration of 125 mM. We found that washing the cells with PBS after centrifugation without the addition of glycine works well; therefore, the addition of glycine was not included in this protocol. It is important to note that the additional steps after the crosslinking step are needed if tissue or plants are used (14, 15). It should be mentioned that in the study of histone modifications, some protocols omit the formaldehyde crosslinking step, using native chromatin directly for IP (see Chapter 16). This is a reasonable variation of the ChIP protocol in this context, as core histones are tightly bound to DNA. The potential advantages of skipping the crosslinking step include shorter

processing times and the avoidance of epitope masking by formaldehyde crosslinks, which can result in suboptimal IP efficiency. The potential disadvantages of skipping the crosslinking step, however, include lowered IP efficiency owing to potentially significant core histone dissociation from target chromatin. In the study of DNA binding proteins other than core histones, formaldehyde fixation is almost always required.

3. Usually, several dishes of cells are collected simultaneously for a single ChIP experiment. Thus, freezing and storing the fixed cell pellets at this step in the protocol may be convenient. We have noticed no negative impact on results by freezing the fixed pellets, although the protocol can be continued without this step.
4. The reproducible fragmentation of formaldehyde crosslinked chromatin into approximately 500–1,000 bp fragments by sonication depends on several factors including the type of sonicator, the cell type used in the experiment, and the duration of formaldehyde crosslinking. The fixation and sonication parameters detailed above should be used only as a starting point for new ChIP experiments. It is highly recommended that a formaldehyde fixation time course be run and that the intensity and duration of sonication be fully investigated with each new ChIP assay. A small amount of soluble chromatin should be electrophoresed on a 1% agarose gel, after reverse crosslinking, proteinase K treatment, phenol/chloroform extraction, EtOH precipitation, and RNase treatment, to assess chromatin fragment size after sonication. Some laboratories use bioruptor instead of sonicator to generate mononucleosomal chromatin fragments.
5. The preclear step reduces nonspecific binding of the chromatin to the Sepharose beads. We have not detected much difference in our assays with and without the preclear step, potentially because the episomal targets of our interest do not bind to the Sepharose beads nonspecifically and/or the specificity of the TaqMan assays used is sufficient to eliminate noise contributed by nonspecific binding. However, in applications such as ChIP-seq, it is important to include the preclear step to reduce nonspecific binding that can lead to artifacts. Also, the nonspecific binding of chromatin to the Sepharose beads may potentially lead to higher noise if intercalating dyes were used for Q-PCR.
6. The amount of soluble chromatin used in each IP experiment depends on the efficiency of the antibody binding to the target protein and the potential abundance of the target protein at the DNA sites of interest. Therefore, the amount of soluble chromatin as well as the amount of antibody used in each IP should be determined empirically based on the antibody used and the Q-PCR results.

7. Ideally, the specificity of different antibodies used should be tested with cells transfected with an expression vector of the protein of interest or tested with cells known to express the protein of interest and to compare them with cells known not to express the protein of interest. At least, the antibodies should be tested by using increasing amounts of the antibody in a series of pull-down assays and recover the protein for both silver stain and Western blot to assess the specificity of the antibody. Specific PCR assays may also be used to determine the specificity of the antibody if DNA targets and nontargets are clearly known.
8. If the quantitation of the target protein is not of interest or necessary for the experimental design, the elution step may be skipped. NaCl in step 21 can be added immediately after adding the elution buffer and incubate the tube at 65°C for at least 4 h.
9. Theoretically, 1 μL of a given TCF sample contains about 36 ng of genomic DNA (i.e., [3×10^6 cells] [12 pg genomic DNA per tetraploid human 293/EBNA1 cell][1/20 total sample volume] = 18×10^5 pg DNA per 50 μL TCF sample). For an autosomal target gene, 36 ng of genomic DNA contains about 12,000 gene copies (i.e., four copies or 12 pg per tetraploid genome). For our experiments, target genes were on replicating episomes, which were maintained at high, although variable copy numbers in 293/EBNA1 cells. On average, 35 pg (i.e., 2.6×10^6 copies; nearly 90 copies/cell) of target DNA were detected in 1 μL of a given TCF sample. Immunoprecipitations are carried out on 475 μL (47.5%) of the total DNA sample (i.e., about 17 μg of genomic DNA). Again, for an autosomal target gene, 17 μg of genomic DNA contains about 5.7×10^6 copies of a given gene target. Assuming 100% IP efficiency, we would detect 2.56×10^6 copies of the episomal targets in 1 μL of a given antiacetyl H3/H4 sample in our experiments. In general, we detect 1.4×10^5 copies on average (nearly five copies/cell) were detected in 1 μL of each antiacetyl H3/H4 sample, depending on the histone acetylation status of the episomal target gene. Background levels of target DNA, detected in the no-Ab control samples, were routinely 10–100-fold lower than experimental values. Obviously, the histone acetylation status of the target chromatin largely determines the background level. Thus, if the target gene contains mostly deacetylated histones, background and experimental values will be similar. Finally, since we were detecting different target genes on episomes maintained at different copy numbers within cells, it was important to verify that IP efficiencies were similar across ChIP samples (1). This was achieved by quantifying the amount of an endogenous target locus (i.e., XRCC1) in all samples.

10. Our own PCR reaction buffer as described and the 10× PCR Gold Buffer (PE Applied Biosystems) for use with AmpliTaq Gold DNA polymerase do not change the PCR results. The iQ Supermix from Bio-Rad gives more robust PCR results in side-by-side comparison, and it works well in reactions with dual fluorophores.
11. PE Applied Biosystems recommends varying probe and primer concentrations within the 50–250 nM and 50–900 nM ranges, respectively, to optimize Q-PCR efficiencies (resources available at http://www/appliedbiosystems.com). As with all other quantitative assays, reactions need to be set up in a consistent manner in order to maximize accuracy and reproducibility. Although the detection system allows for the precise measurement of the amplified product, it cannot overcome manual errors introduced into the reaction during setup.

References

1. Irvine, R. A., and Hsieh, C.-L. (2002) DNA methylation has a local effect on transcription and histone acetylation. *Mol. Cell. Biol.* **22**, 6689–6696.
2. Okitsu, C. Y., and Hsieh, C.-L. (2007) DNA methylation dictates histone h3k4 methylation. *Mol. Cell. Biol.* **27**, 2746–2757.
3. Okitsu, C. Y., Hsieh, J. C., and Hsieh, C.-L. (2010) Transcriptional activity affects the H3k4me3 level and distribution in the coding region. *Mol. Cell. Biol.* **30**, 2933–2946.
4. Orlando, V. (2000) Mapping chromosomal proteins in vivo by formaldehyde crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* **25**, 99–104.
5. Ginzinger, D. G. (2002) Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Exp. Hematology* **30**, 503–512.
6. Stevens, S. J., Verschuuren, E. A., Pronk, I., van Der Bij, W., Harmsen, M. C., The, T. H., Meijer, C. J., van Den Brule, A. J., and Middeldorp, J. M. (2001) Frequent monitoring of Epstein-Barr virus DNA load in unfractionated whole blood is essential for early detection of posttransplant lymphoproliferative disease in high-risk patients. *Blood* **97**, 1165–1171.
7. Santos, J. H., Meyer, J. N., Mandavilli, B.S., and Van Houten, B. (2006) Quantitative PCR-based measurement of nuclear and mitochondrial DNA damage and repair in mammalian cells. *Methods Mol. Biol.* **314**, 183–199.
8. Smith, C. J., and Osborn, A. M. (2009) Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol. Ecol.* **67**, 6–20.
9. Heid, C. A., Stevens, J., Livak, K. J., and Williams, P. M. (1996) Real Time Quantitative PCR. *Genome Res.* **6**, 986–994.
10. Solomon, M. J., Larsen, P. L., and Varshavsky, A. (1988) Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell.* **53**, 937–947.
11. Braunstein, M., Rose, A. B., Holmes, S. G., Allis, C. D., and Broach, J. R. (1993) Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes Dev.* **7**, 592–604.
12. Hsieh, C.-L. (1994) Dependence of transcriptional repression on CpG methylation density. *Mol. Cell. Biol.* **14**, 5487–5494.
13. Chen, H., Lin, R. J., Xie, W., Wilpitz, D., and Evans, R. M. (1999) Regulation of hormone-induced histone hyperacetylation and gene activation via acetylation of an acetylase. *Cell.* **98**, 675–686.
14. Acevedo, L. G., Bieda, M., Green, R., and Farnham, P. J. (2008) Analysis of the mechanisms mediating tumor-specific changes in gene expression in human liver tumors. *Cancer Res.* **68**, 2641–2651.
15. Haring, M., Offermann, S., Danker, T., Horst, I., Peterhansel, C., and Stam, M. (2007) Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant Methods* **3**, 11.

Chapter 17

Sequential Chromatin Immunoprecipitation Assay and Analysis

Ricardo B. de Medeiros

Abstract

Sequential chromatin immunoprecipitation (SeqChIP) assays have been developed for the study of interactions of two or more proteins (or simultaneous histone modifications) at genomic sites. It is based on the principle that chromatin and associated proteins can be first immunoprecipitated with a first antibody and the obtained immunoprecipitate can be subjected to a second antibody. At the end of the assay the immunoprecipitated material contains only chromatin that concomitantly carries both DNA-associated proteins (or both histone modifications). The SeqChIP protocol described here combines speed (minimum of 3–4 h to perform the complete assay), sensitivity (known targets can be detected with only about 20,000 cell equivalents), and avoidance of antibody–antigen disruption after the first ChIP step. In addition, specific SeqChIP controls and potential shortcomings are discussed, the main characteristics of different SeqChIP protocols are described and several examples of protein complexes and protein–protein interactions at genomic sites that have been solved by SeqChIP in the recent years are presented.

Key words: Chromatin immunoprecipitation, Sequential ChIP, Re-ChIP, Two-step ChIP

1. Introduction

At the current stage of the epigenetic code deciphering it has been already realized that one of its hallmarks is the presence of intricate and regulated multicomponent complexes of DNA-associated proteins (histones, histone modifiers, transcription factors, adapters, etc.) in a given genetic locus (1, 2). In parallel, several histone post-translational modifications (phosphorylation, methylation, acetylation) can be simultaneously identified in a given histone molecule (3–5). Sequential chromatin immunoprecipitation (SeqChIP) is one of the methodologies that have been developed for the detection and analysis of these protein interactions or simultaneous histone modifications at genomic sites (6–8).

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_17, © Springer Science+Business Media, LLC 2011

SeqChIP (also referred to as two-step ChIP, Re-ChIP, ChDIP, or double ChIP) is a relatively simple assay that permits the investigator to determine if two or more proteins are close enough in a chromatin context or whether they are interacting in a certain genomic region. It addresses the cohabitation, or co-occupancy, of two or more proteins in the genome (7). It is based on the principle that chromatin and associated proteins can be first immunoprecipitated with a first antibody and the eluted material can be subjected to a second antibody. This second antibody recognizes a protein that is suspected or known to be in direct or indirect association with the first immunoprecipitated antigen, as part of a complex or simply associating to genomic sites separated by a short distance. At the end of the assay the immunoprecipitated material contains only chromatin that concomitantly carries both DNA-associated proteins, or both histone modifications, in opposite to standard ChIP or ChIP-chip assays (9). Standard ChIP assays do not determine whether two proteins simultaneously occupy a given DNA locus, but merely determine the relative levels of different proteins at genomic regions. Occupation of the same genetic locus in different subsamples, as detected in genome-wide ChIP-chip assays, is often considered a strong indication of co-occupancy, but requires validation via complementary approaches (10). SeqChIP, however, cannot address if the interaction is direct, for which case in vitro pull-down assays are a more appropriated experimental approach.

In a typical SeqChIP assay several controls are recommended: (a) the order of the antibodies used for ChIP can be inverted and generally the same results should be expected; (b) negative genomic controls should be used in PCR reactions, such as a set of primers to a genomic region for which no interaction is expected with the protein under study, and this genomic control can also be used in a duplex PCR format; (c) normal IgG should be included as the first or the second antibody, like in a standard ChIP assay, and no immunoprecipitation over background should be expected; (d) members of the RNA Pol II complex (such as TBP, TFIIA, and TFIIB) are well characterized, known to completely co-occupy active promoters, and can be used as positive controls (7, 11).

In addition, crosslinking and chromatin shearing deserve special attention. Sonication of crosslinked chromatin often results in a randomized population of fragments that average 0.4–0.5 kb in size. If the proteins under study associate to the DNA at a distance of 0.4 kb, the sonication will result in a substantial proportion of DNA fragments that contain only one biding site, but not the other. Therefore, for two proteins which co-occupy a given DNA region, independent of tissue type or physiological condition, but whose binding sites are separated about 0.4 kb, the outcome from a SeqChIP assay will appear to be no co-occupancy, when compared to controls (7). Such intrinsic limitation of the assay can be addressed by increasing the fragments size and changing the antibodies order

in the SeqChIP assay. In addition, SeqChIP can be performed in native conditions, by omitting formaldehyde crosslinking, an option that is preferred to proteins with high-DNA affinity, such as histones (12).

The SeqChIP protocol and analysis method described here have been recently reported (13) and combines speed (minimum of 3–4 h to perform the complete assay), sensitivity (known targets can be detected with only about 20,000 cell equivalents), and avoidance of antibody–antigen disruption after the first ChIP step (schematic diagram shown in Fig. 1). For an overall comparison several SeqChIP protocols that have been reported in the last few years (3, 7, 8, 13–18) and their main characteristics are described in Table 1. Most SeqChIP protocols utilize, for immunocapture after the first ChIP step, disruption buffers containing dithiothreitol (DTT) or similar reagents, or epitope-specific peptides, to elute the immunocomplexes from the beads used in the first round of immunoprecipitation. In contrast, the protocol described here uses two methods of physical separation of the targeted DNA fragments: microcentrifugation of protein G or protein A-sepharose/agarose beads and magnetic separation of streptavidin–biotin complexes. Consequently, the subsequent rounds of immunoprecipitations can be carried out in the same, intact samples. Figure 1 illustrates the principle of the current SeqChIP protocol, which is described next. If only DNA fragments bound simultaneously to factors X and Y are the target, the first ChIP is performed with a nonbiotinylated anti-X antibody, which is captured with protein A/G-sepharose beads by microcentrifugation. In this step, all DNA fragments

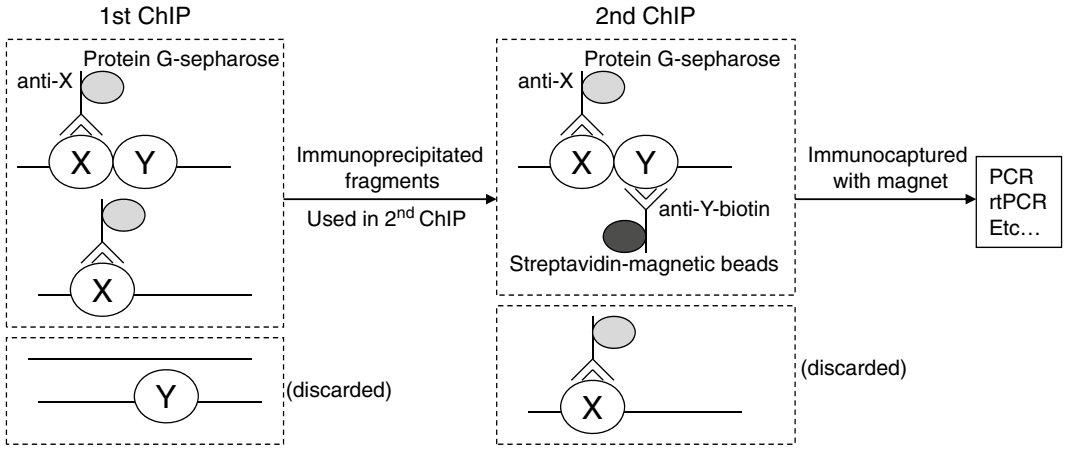


Fig. 1. Schematic diagram of the SeqChIP assay. In the first ChIP, the DNA fragments bound both to factor X only and to both X and Y factors are immunoprecipitated with a nonbiotinylated anti-X antibody, which are captured by protein G/A-sepharose beads (fragments associated with factor Y only are not immunocaptured and, therefore, discarded). This material is then used in a second ChIP, in which a biotinylated anti-Y antibody is used. Streptavidin-magnetic beads are added, and DNA fragments are captured with a magnet. At this time fragments associated with protein X only are not attracted to the magnet. These fragments end up precipitating to the bottom of the tube by gravity and, therefore, are discarded. In the end, only DNA fragments associated to factors X and Y simultaneously are isolated and used in PCR assays.

Table 1
Main characteristics of several reported SeqChIP protocols

	Characteristics						
Protocols	**Preclearing**	**Shearing**	**Immunocapture**	**Elution**	**X-link reversal**	**Prot. K/RNase**	**Reference**
1	No	Enzymatic	Protein G-Magnetic	10 mM DTT	Yes	Yes	(14)
2	Yes	Sonication	Protein A/G-Sepharose	10 mM DTT	Yes	Yes	(8)
3	Yes	Sonication	Protein A/G-Sepharose	50 mM $NaHCO_3$	No	No	(15)
4	Yes	Sonication	Protein G-Sepharose	10 mM DTT	No	No	(16)
5	No	Sonication	Protein A-Sepharose	10 mM DTT/SDS	Yes	Yes	(3)
6	Yes	Sonication	Protein A-Sepharose	Epitope-spec. peptide	Yes	No	(7)
7	Yes	Enzymatic	Protein G-Magnetic	1% Triton	No	No	(17)
8	No	Sonication	Protein G-Sepharose	1% SDS	Yes	No	(18)
9	No	Sonication	Ptn G-Seph./Streptav.-Mag.	No disruption	No	No	(13)

bound to the X factor are captured, including fragments bound to X only and fragments bound to X and Y. Fragments bound to Y factor only are discarded, since no anti-Y antibody was added to the first ChIP. The sample resulting from the first ChIP is used then in a second ChIP. At this time a biotinylated anti-Y antibody is utilized, and captured with streptavidin-magnetic beads. Only the fragments that contain the Y factor will be captured by the magnet, and because of the previous ChIP they will contain the X factor as well. The fragments that contain the X factor only (bound to sepharose beads from previous step) will be discarded since they will not move toward the magnet and will precipitate by gravity to the bottom of the tube. The benefits obtained from this approach are anticipated to be lower experimental noise, higher specificity, and the possibility of using smaller sample size more efficiently. This assumption stands from the fact that this SeqChIP protocol is based on a standard ChIP protocol that, when compared to a commonly used ChIP protocol side-by-side, using the exact same biotinylated antibodies, showed all those advantages (13). Table 1 also shows that a wide variety of experimental approaches can be followed in a SeqChIP assay and indicates that several of the main steps can be successfully modified to accommodate the specific characteristics of the proteins under study and/or specific laboratory settings.

Finally, numerous important protein complexes and specific protein–protein interactions at genomic sites and simultaneous histone modifications have been solved by SeqChIP in recent years. Several examples are shown in Table 2. SeqChIP, given its simplicity and these various successful outcomes, has become an increasingly important tool in a broad range of research areas, currently including stem cell biology, tumorigenesis, immunology, neurobiology, and cell biology.

2. Materials

2.1. Tissue Culture

1. Tissue culture media: according to type of cell line or primary cells used.
2. Trypsin solution (0.25%, Gibco).
3. Cell scrapers (Fisher).

2.2. Sequential Chromatin Immunoprecipitation

1. Antibodies: nonbiotinylated and biotinylated primary antibodies, to be used in the first and second ChIP steps, respectively (see Note 1).
2. Streptavidin-magnetic beads and magnet (R&D Systems).
3. Chemicals: PBS buffer, SDS, EDTA, Tris buffer, Triton X-100, NaCl, Chelex-100, LiCl, Na deoxycholate, protease inhibitors (Leupeptin, phenyl-methanesulfonyl-floride (PMSF),

Table 2
Examples of protein–protein interactions and histone modifications reported by SeqChIP

Association/modification	Genetic locus	Cell type	Reference
PolII/Pit1 and Pit1/N-CoP/T_3Rβ	Mouse *growth hormone*	Pituitary gland	(6)
H2A/FoxA	Mouse *albumin* enhancer	Liver	(19)
Tup1/SWI/SNF or SAGA	*GRE2*	Yeast	(20)
TBP/H3	*α1 antitrypsin*	$Caco_2$	(21)
TBP/HNF3β/c/EBPα/TFIIB	*HNF4α*	$Caco_2$	(16)
p300/CBP/TIP60/PRMT1/ CARM1/SRC1/p/CIP/hERα/ TBP/TFIIA	*pS2*	MCF-7	(22)
Mot1/TBP or TFIIB	Several	Yeast	(11)
MEL18/YY1	*Hoxd4*	P19	(23)
H3K4me3/H3K27me3	Mouse genome	ES cells	(3)
CREB/cFOS	*OPN*	Smooth muscle	(24)
c/EBP/PPARδ	*14-3-3*	Endothelial	(25)
P53/SMAD2/4/mSIN3A	*AFP*	Hepatomas	(26)
EFI16/NM23	*p53 and c-myc*	Melanoma	(27)
PU.1/BCL6	Several	B cells	(28)
H3.3/H2A.Z	Several	HeLa	(5)
ERα/ERβ	Several	MCF-7	(29)
CBX7/CBX8	*INK4a*	Fibroblasts	(30)
SWI/SNF/HIC1/ARID1A	*E2F1*	Fibroblasts	(31)
SOX2/NANOG/OCT3/4/cMYC and RUNX2/BMI-1/ SMAD2/3	Several	ES cells	(13)
TFIID/SAGA	Several	Yeast	(18)
Brg1/14-3-3	Mouse *Jun, Cox2, Fos*	Fibroblasts	(17)
p300/Ets1	Mouse *Npr1*	Mesangial	(14)

and aprotinin), 37% formaldehyde, NP-40, glycine, dimethylsulfoxide (DMSO) (Sigma).

4. QIAQuick® DNA purification kit (Qiagen).
5. Protein A, protein G-agarose/sepharose beads (Thermo Scientific, Pierce) (see Note 2).
6. Sonicator (Heat Systems-Ultrasonic).

7. Ultrasonic bath (Branson).
8. Eppendorf tubes rotator (Labnet).
9. Refrigerated microcentrifuge and heat block.

2.3. PCR and Real-Time PCR

1. Standard PCR: AmpliTaq® Gold Polymerase PCR kit and dNTP mix (Applied Biosystems).
2. Real-time PCR: Power SYBR® Green PCR Master Mix (Invitrogen).
3. PCR plates (96-well) and 96-well plate foil (Eppendorf).
4. Primers (IDT) at 200 nM for real-time PCR or at 1 μM for standard PCR.
5. Eppendorf Mastercycler® for standard PCR and ABI 7900 HT instrument for real-time PCR.
6. ImageJ software (distributed through National Institutes of Health, http://www.rsbweb.nih.gov/ij/).

3. Methods

1. Cell samples are crosslinked (fixed) by adding formaldehyde to the tissue culture media to a 1–2% final concentration and incubating (gentle rocking) for 15 min at room temperature (RT) (see Note 3).
2. Samples (1×10^6 cells) are quenched by adding glycine to 125 mM final concentration for 5 min (RT).
3. Samples are centrifuged down to eliminate media and resuspended in 0.5 mL of Lysis Buffer (1% SDS, 10 mM EDTA, 50 mM Tris–HCl, pH 8) containing protease inhibitors (10 μg/mL Leupeptin, 10 μg/mL aprotinin, and 1 mM PMSF) and incubated on ice (10 min) (see Note 4).
4. Samples are sonicated at RT to shear chromatin to an average length of about 0.5–1 kb, transferred to 1.5-mL tubes, microcentrifuged for 10 min at max speed (see Note 5).
5. Supernatants are collected in 1.5-mL tubes containing 1 mL of the Dilution Buffer (0.01% SDS, 1.1% Triton, 1.2 mM EDTA, 167 mM NaCl, 17 mM Tris–HCl, pH 8).
6. 5 μg of the first antibody (nonbiotinylated) are added, samples are incubated 15 min in an ultrasonic bath or overnight at 4°C on a rotator, depending on the antibody immunoprecipitation efficiency, determined previously (see Note 6).
7. 50 μL of protein G or protein A-agarose/sepharose beads (depending on the primary antibody type of IgG) are added, for 2 h at 4°C on a rotator.

8. Beads are collected by microcentrifugation (1 min, max speed) and washed four times as follows.
9. First wash: beads are resuspended with 1 mL of Wash Buffer 1 (0.1% SDS, 1% Triton, 2 mM EDTA, 150 mM NaCl, 20 mM Tris–HCl, pH 8) and collected by microcentrifugation (1 min, max speed).
10. Second wash: beads are resuspended with 1 mL of Wash Buffer 2 (0.1% SDS, 1% Triton, 2 mM EDTA, 500 mM NaCl, 20 mM Tris–HCl, pH 8) and collected by microcentrifugation (1 min, max speed).
11. Third wash: beads are resuspended with 1 mL of Wash Buffer 3 (0.25 M LiCl, 1% NP-40, 1% Na deoxycholate, 1 mM EDTA, 10 mM Tris–HCl, pH 8) and collected by microcentrifugation (1 min, max speed).
12. Final wash: beads are resuspended with 1 mL of Wash Buffer 4 (10 mM Tris–HCl, pH 8, 1 mM EDTA) and collected by microcentrifugation (1 min, max speed).
13. Protein A or protein G beads are resuspended in 1.5-mL tubes containing 1 mL of the Dilution Buffer (0.01% SDS, 1.1% Triton, 1.2 mM EDTA, 167 mM NaCl, 17 mM Tris–HCl, pH 8).
14. 5 μg of the second antibody (biotinylated) are added, samples are incubated 15 min in an ultrasonic bath or overnight at 4°C on a rotator, depending on the antibody immunoprecipitation efficiency, determined previously (see Note 6).
15. 50 μL of streptavidin-magnetic beads are added and incubated for 30 min (4°C) on a rotator.
16. Beads are collected with a magnet, washed four times as described in steps 8–11 above (see Note 7).
17. After the last wash, 100 μL of a 10% Chelex-100/PBS solution are added to the beads.
18. Samples are boiled (10 min) in a heat block, microcentrifuged (1 min, max speed), supernatants transferred to a new 1.5-mL tube.
19. MilliQ water (120 μL) is added back to the beads, samples are microcentrifuged again (1 min, max speed), and the new supernatant is pooled with the previous one.
20. DNA samples are then cleaned up with QIAQuick® kit, resuspended in 50 μL of Wash Buffer 4 and 1–2 μL subsamples (containing about 20,000 cell equivalents) can be used per PCR reaction (see Note 8).

SeqChIP Analysis

In a set of preliminary experiments, the investigator can identify the most appropriated method for analysis of the SeqChIP assays. Probably the most popular and straightforward analysis method is

real-time PCR. Recently (13), we have developed an alternative analysis method based on the use of standard (semiquantitative) PCR coupled to densitometric analysis of PCR bands. This analysis method matched real-time PCR results in terms of known targets detected above background for several transcription factors in standard ChIP and SeqChIP assays (13). Densitometric analysis using the ImageJ software is simpler, generally more practical and requires less time and resources in preliminary calibration experiments. It offers a viable alternative to real-time PCR when absolute quantification is not required, since it presents a quantification limitation and lower fold-enrichment values might be obtained with densitometry analysis, probably resulting from pixel saturation in DNA bands from PCR reactions and consequent lower measurement capacity. For qualitative analysis, however, such as identification of potential target genes, this alternative analysis method does not show any apparent disadvantage, at least in small scale settings, and for several antibodies, such quantification limitation was minimum (13). A brief description of each analysis method is outlined below only for reference. Numerous real-time PCR methods are described in the literature and the choice of a specific one depends on general experimental conditions.

3.1. Standard (Semiquantitative) PCR and Densitometry with ImageJ Software

1. DNA samples, from SeqChIP assays or from aliquots of whole cell lysates (Input DNA, positive control), primers (1 μM), and DMSO (5%) are added to the additional standard PCR reagents (according to the manufacturer's instructions, AmpliTaq® Gold Polymerase PCR kit and dNTP mix) in a 50-μL reaction volume. When available primers designed to known target genes should be used as additional positive controls. All primers should be validated previously.
2. Samples are subjected to the appropriate cycle, according to the primers used in the study and additional experimental conditions. One general example is 95°C/9 min, 43× (95°C/1 min, 60°C/1 min), and 60°C/10 min, using 96-well plates and an Eppendorf Mastercycler® (see Note 9).
3. PCR products were subjected to a 1.5% agarose gel electrophoresis, visualized with ethidium bromide staining and photographed. Images were saved as TIFF files for analysis with ImageJ software.
4. Densitometric Image Analysis: signal intensities from semiquantitative (standard) PCR data obtained from SeqChIP assays or from whole cell lysates (Input DNA) are quantified from the TIFF images with ImageJ software, as previously described (13). Enrichment values are equal to (antibody SeqChIP signal minus IgG SeqChIP signal)/Input DNA ratio. As usual, statistical analysis should be applied according to the experimental design, number of samples, etc.

3.2. Real-Time PCR

1. DNA samples, from SeqChIP assays or from aliquots of whole cell lysates (Input DNA), and primers (200 nM, IDT) are added to the additional real-time PCR reagents (according to the manufacturer's instructions, Power SYBR® Green PCR Master Mix) in a 20-μL reaction volume. All primers should be validated previously.
2. Samples are subjected to the appropriate cycle, according to the primers used in the study and additional experimental conditions. One general example is 50°C/2 min, 95°C/2 min, 40× (95°C/15 s, 60°C/1 min), using 96-well plates and an ABI 7900 HT instrument (see Note 9).
3. When available primers designed to known target genes should be used as positive controls. As usual, statistical analysis should be applied according to the experimental design, number of samples, etc.

4. Notes

1. Antibody quality and specificity should be determined previously with appropriate controls by Western blotting and immunocytochemistry, for example. The antibody is the most critical reagent in a ChIP assay and, therefore, one must spend all required efforts to be certified of its quality and specificity. The following biotinylated antibodies have been successfully tested with this SeqChIP protocol: SOX2, NANOG, OCT3/4, cMYC, BMI-1, RUNX2, SMAD2/3, PKC*mu*, and HDAC8 (all from R&D Systems and previously ChIP-validated), using human embryonic stem (ES) cells, PBMCs, isolated human dendritic cells, and Jurkat T cells. Protocols have been described in the literature for the biotinylation of even small amounts of antibody (32).
2. Protein A is considered a better choice to capture rabbit polyclonal antibodies, while protein G should be used to capture sheep or goat polyclonals as well as mouse or rat monoclonal antibodies.
3. If using adherent cells formaldehyde can be added after the addition of the trypsin solution, or cells can be fixed and removed with cell scraper. The time of incubation with formaldehyde may be critical in some experimental conditions. If the sheared chromatin is insufficient or degraded, when visualized by gel electrophoresis, it is recommended to reduce crosslinking time, 5 min may be sufficient. Likewise, over-crosslinking (characterized by average DNA fragments smaller than the distance between the primers used in the study) might be the problem,

if the incubation with formaldehyde is too long. Several ChIP protocols addressing cross-linking conditions and optimization can be found in the literature (9, 10, 33, 34).

Caution: formaldehyde is flammable, highly toxic by inhalation, contact with skin or if swallowed, causes burns; and is potentially carcinogenic. It should be used with appropriate safety measures, such as protective gloves, glasses and clothing, and adequate ventilation. Formaldehyde waste should be disposed of according to regulations for hazardous waste.

4. Lysis will not happen in this step and the SDS may even precipitate; lysis will occur during sonication, which is carried out at RT.
5. Sonication settings will vary with manufacturer. DNA shearing should be checked by gel electrophoresis and ethidium bromide staining, optimization may be required. For example, DNA of 1 kb average size can be obtained by setting a Heat Systems-Ultrasonics sonicator to 4% output power, 70% duty, output control 3, performing four rounds of 15 pulses (2 s pulses), resting the samples on icy water for 2 min between rounds (to avoid DNA denaturation). In addition, for standard sonicators, to avoid solution foaming during sonication, the sonicator tip should be localized in the tube as deep as possible without touching the bottom or walls of the tube. Immerse tube in an appropriate size ice-water bath, which is most easily done by keeping the tip and tube fixed, placing the bath on a height-adjustable platform and raising it into position. If there is significant foaming, remove bubbles by high-speed centrifugation and resuspend all material.
6. Antibody quantity and incubation period should be the subject of preliminary experiments for optimization: antibodies of good quality targeting abundant antigens normally work well at 5 μg in ultrasonic bath 15 min incubations (ultrasound accelerates molecules in solution, resulting in more efficient antibody–antigen recognition). Antibodies of lower quality and/or of scarce antigens may require longer incubation periods (overnight at 4°C in a rotating device) and/or higher quantities. Normal IgG replaces the antibody in negative controls.
7. The protein G or protein A-agarose/sepharose beads actually end up collected with the streptavidin-magnetic beads at this point. The evidence for this assumption is that the appearance of the beads carried to the magnet is whitish at this point of the protocol, instead of a dark appearance if there were only magnetic beads in the sample captured by the magnet.
8. Therefore, one million cells (sample size) are sufficient for up to 50 PCR or rtPCR reactions.

9. Troubleshooting guidelines for standard and real-time PCR cycles and conditions are numerous, variable and can be easily found in the literature and in manufacturer's websites, therefore they are not shown here.

Acknowledgments

The author would like to thank colleagues at R&D Systems who contributed for the development of this protocol, including Ernesto Resnik, Brian Hoium, and Jessie Ni.

References

1. Narlikar, G.J., Fan, H.Y., and Kingston, R.E. (2002) Cooperation between complexes that regulate chromatin structure and transcription. *Cell.* **108**: 475–487.
2. Kim, T.H. and Ren, B. (2006) Genome-wide analysis of protein-DNA interactions. *Annu. Rev. Genomics Hum. Genet.* 7: 81–102.
3. Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., and Cuff, J. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell.* **125**: 315–326.
4. Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., and John, R.M. (2006) Chromatin signatures of pluripotent cell lines. *Nat. Cell. Biol.* **8**: 532–538.
5. Jin, C., Zang, C., Wei, G., Cui, K., Peng, W., Zhao, K. and Rosenfeld, G. (2009) H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat. Genet.* **41**: 941–945.
6. Scully, K.M., Jacobson, E.M., Jepsen, K., Lunyak, V., Viadiu, H., and Carriere, C. (2000) Allosteric effects of Pit1 DNA sites on long-term repression in cell type specification. *Science.* **290**: 1127–1131.
7. Geisberg, J.V., and Struhl, K. (2004) Quantitative sequential chromatin immunoprecipitation, a method for analyzing co-occupancy of proteins at genomic regions in vivo. *Nuc. Acids. Res.* **32**: e151.
8. Furlan-Magaril, M., Rincon-Arano, H. and Recillas-Targa, F. (2009) Sequential chromatin immunoprecipitation protocol: ChIP-reChIP. In: Moss, T. and Leblanc, B. (Eds.) *Methods in Molecular Biology, DNA-Protein Interactions*, vol. 543, Humana Press, NJ.
9. Lee, T.I, Johnstone, S.E. and Young, R.A. (2006) Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protocols.* **1**: 729–748.
10. Collas, P. (2010) The current state of chromatin immunoprecipitation. *Mol. Biotech.* **45**: 87–100.
11. Geisberg, J.V., and Struhl, K. (2004) Cellular stress alters the transcriptional properties of promoter-bound Mot1-TBP complexes. *Mol. Cell.* **14**: 479–489.
12. Thorne, A.W., Myers, F.A., and Hebbes, T.R. (2004) Native chromatin immunoprecipitation. *Methods. Mol. Biol.* **287**: 21–44.
13. Medeiros, R.B., Papenfuss, K.J., Hoium, B., Coley, K., Jadrich, J., Goh, S.K. et al. (2009) Novel sequential ChIP and simplified basic ChIP protocols for promoter co-occupancy and target gene identification in human embryonic stem cells. *BMC Biotech.* **9**: 59.
14. Kumar, P. and Pandey, K.N. (2010) Cooperative activation of Npr1 gene transciption and expression by interaction of Ets1 and p300. *Hypertension.* **54**: 172–178.
15. Cui, R., Nguyen, T.T., Taube, J.H., Stratton, S.A., Feuerman, M.H. and Barton, M.C. (2005) Family members p53 and p73 act together in chromatin modification and direct repression of α-fetoprotein transcription. *Jour. Biol. Chem.* **280**: 39152–39160.
16. Hatziz, P., and Talianidis, I. (2002) Dynamics of enhancer-promoter communication during differentiation-induced gene activation. *Mol. Cell.* **10**: 1467–1477.
17. Drobic, B., Perez-Cadahia, B., Yu, J., Kung, S.K.-P. and Davie, J.R. (2010) Promoter chromatin remodeling of immediate-early genes is mediated through H3 phosphorylation at either serine 28 or 10 by the MSK1 multiprotein complex. *Nuc. Acids. Res.* **1–13** (doi:10.1093/nar/gkq030).

18. Ohtsuki, K., Kasahara, K., Shirahige, K. and Kokubo, T. (2010) Genome-wide localization analysis of a complete set of Tafs reveals a specific effect of the taf1 mutation on Taf2 occupancy and provides indirect evidence for different TFIID conformations at different promoters. *Nuc. Acids Res.* **38**:1805–20.
19. Chaya, D., Hayamizu, T., Bustin, M. and Zaret, K.S. (2001) Transcription factor FoxA (HNF3) on a nucleosome at an enhancer complex in liver chromatin. *Jour. Biol. Chem.* **276**: 44385–44389.
20. Proft, M., and Struhl, K. (2002) Hog1 kinase converts the Sko1-Cyc8-Tup1 repressor complex into an activator that recruits SAGA and SWI/SNF in response to osmotic stress. *Mol. Cell.* **9**: 1307–1317.
21. Soutoglou, E., and Talianidis, I. (2002) Coordination of PIC assembly and chromatin remodeling during differentiation-induced gene activation. *Science* **295**: 1901–1904.
22. Metivier, R., Penot, G., Hubner, M.R., Reid, G., Brand, H., Kos, M. and Gannon, F. (2003) Estrogen receptor-α directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell.* **115**: 751–763.
23. Kobrossy, L., Rastegar, M. and Featherstone, M. (2006) Interplay between chromatin and trans-acting factors regulating the hoxd4 promoter during neural differentiation. *Jour. Biol. Chem.* **281**: 25926–25939.
24. Jalvy, S., Renault, M.A., Lam Shang Leen, L., Belloc, I., Reynaud, A., Gadeau, A.P. and Desgranges, C. (2007) CREB mediates UTP-directed arterial smooth muscle cell migration and expression of the chemotactic protein osteopontin via its interaction with activator protein-1 sites. *Circ. Res.* **100**: 1292–1299.
25. Brunelli, L., Cieslik, K.A., Alcorn, J.L., Vatta, M. and Baldini, A. (2007) Peroxisome proliferator-activated receptor-delta upregulates 14-3-3 epsilon in human endothelial cells via CCAAT/enhancer binding protein-beta. *Circ. Res.* **100**: e59–71.
26. Wilkinson, D.S., Tsai, W.-W., Schumacher, M.A. and Barton, M.C. (2008) Chromatin-bound p53 anchors activated Smads and the mSin3A corepressor to confer transforming growth factor b-mediated transcription repression. *Mol. Cell. Biol.* **28**: 1988–1998.
27. Egistelli, L., Chichiarelli, S., Gaucci, E., Eufemi, M., Schinina, M.E., and Giorgi, A. (2009) IFI16 and NM23 bind to a common DNA fragment both in th p53 and the cMyc gene promoters. *Jour. Cell. Biochem.* **106**: 666–672.
28. Wei, F., Zaprazna, K., Wang, J. and Atchison, M.L. (2009) PU.1 can recruit BCL6 to DNA to repress gene expression in germinal center B cells. *Mol. Cell. Biol.* **29**: 4612–4622.
29. Papoutsi, Z., Zhao, C., Putnik, M., Gustafsson, J.-A. and Dahlman-Wright, K. (2009) Binding of estrogen receptor a/b heterodimers to chromatin in MCF-7 cells. *Jour. Mol. Endocrinol.* **43**: 65–72.
30. Maertens, G.N., El Messaoudi-Aubert, S., Racek, T., Stock, J.K., and Nicholls, J. (2009) Several distinct polycomb complexes regulate and co-localize on the INK4a tumor suppressor locus. *PLoS ONE* **4**: e6380.
31. van Rechem, C., Boulay, G. and Leprince, D. (2009) HIC1 interacts with a specific subunit of SWI/SNF complexes, ARID1A/BAF250A. *Biochem. Biophys. Res. Comm.* **385**: 586–590.
32. Strachan, E., Mallia, A.K., Cox, J.M., Antharavally, B., Desai, S., and Sykaluk, L., (2004) Solid-phase biotinylation of antibodies. *Jour. Mol. Recognit.* **17**: 268–76.
33. Collas, P., and Dahl, J.A. (2008) Chop it, ChIP it, check it: the current status of chromatin immunoprecipitation. *Front. Biosci.* **13**: 929–943.
34. Nelson, J.D., Denisenko, O., Bomsztyk, K. (2006) Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nature Protocols.* **1**: 179–185.

Chapter 18

Combined Chromatin Immunoprecipitation and Bisulfite Methylation Sequencing Analysis

Yuanyuan Li and Trygve O. Tollefsbol

Abstract

Epigenetic mechanisms control gene transcription primarily through regulating chromatin structures and DNA methylation. Transcription factors can also affect gene transcription through binding of the key transcriptional machinery to the gene promoter. These factors normally jointly influence the transcriptional processes, leading to silencing or activation of gene expression. A novel technique has been recently explored in our laboratory, which is a combination of conventional chromatin immunoprecipitation (ChIP) with bisulfite methylation sequencing assays, so-called ChIP and bisulfite methylation sequencing (ChIP-BMS). This technique provides precise information of DNA methylation status at the selected DNA fragments precipitated by the antibodies to histone molecules or transcription factors of interest. This method also helps to investigate the interactions between histone modification and DNA methylation, and how this crosstalking can affect gene expression. More importantly, it is easy to determine potential methylation-sensitive transcription factors that influence transcription mainly depending on methylation status of the binding sites. In this chapter, we discuss the detailed procedures of this novel technique and its broad application in epigenetic and genetic fields.

Key words: DNA methylation, Histone modification, Transcription factor, Bisulfite sequencing, Chromatin immunoprecipitation, ChIP-BMS

1. Introduction

The epigenetic changes involving DNA methylation and chromatin structures profoundly affect both physiological and pathological processes mainly via regulating gene transcription. DNA methylation, for example, most often occurs at a cytosine base located 5′ to a guanosine (CpG dinucleotide), which is found mainly located in the proximal promoter regions of almost half of the genes in the mammalian genome (1). DNA methylation interferes

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_18,

with gene transcription principally through the prevention of the binding of the basal transcriptional machinery and of ubiquitous transcription factors to the gene promoters (2, 3). This effect can be implemented though the direct and indirect mechanisms of DNA methylation. For the direct mechanism, methylated CpG sites can block the binding of certain transcription factors that are sensitive to the methylation status in their recognition sites, resulting in silenced transcription (4–10). The indirect mechanism is more involved in the accessibility of the transcription repressor complex at the hypermethylated promoter modulated by a group of proteins such as methyl-cytosine-binding proteins (MBPs) (11, 12). As a consequence, DNA hypermethylation is mainly associated with gene repression, whereas DNA hypomethylation at the promoter allows gene expression.

Another important epigenetic mechanism is chromatin modification, which also plays a key role in controlling gene transcription (13). Different types of chromatin modification patterns have been identified such as acetylation, methylation, phosphorylation, and ubiquitination on the specific lysine residues of core histone tails (14). Two main mechanisms by which these modifications on chromatin structure influence gene transcription are well documented. One is that the alteration of chromatin packing can directly change the conformation of the DNA polymer, thus facilitating the accessibility of DNA-binding proteins such as transcription factors to the core gene regulatory region. The other mechanism is involved with the recruitment of the transcription factor machinery triggered by the altered chemical moieties on the nucleosome surface during chromatin remodeling (15, 16). Therefore, epigenetic events working on gene transcription primarily rely on regulation of a dynamic equilibrium between the conformation change of DNA or chromatin and binding ability of transcription factors to the core gene regulatory region. This crosslink working model is more likely to generate a complicated interaction between the epigenetic and genetic mechanisms for gene transcription through a "bridge," transcription factors (Fig. 1). For example, many methylation-sensitive transcription factors have been identified in which their binding abilities predominately depend on the methylation status of the consensus binding sites on the gene promoter by which binding gene expression will be turned on or off (Table 1). Thus, a new method that can provide the detailed information of these interaction patterns is needed.

Bisulfite genomic sequencing analysis is classic technology for the detection of DNA methylation by which an unmethylated cytosine residue in single-stranded DNA will be converted to uracil and methylated cytosine will remain as cytosine (17). The precise DNA methylation information will be conveyed by the subsequent PCR amplification and sequencing. The chromatin immunoprecipitation (ChIP) assay is commonly used in the determination

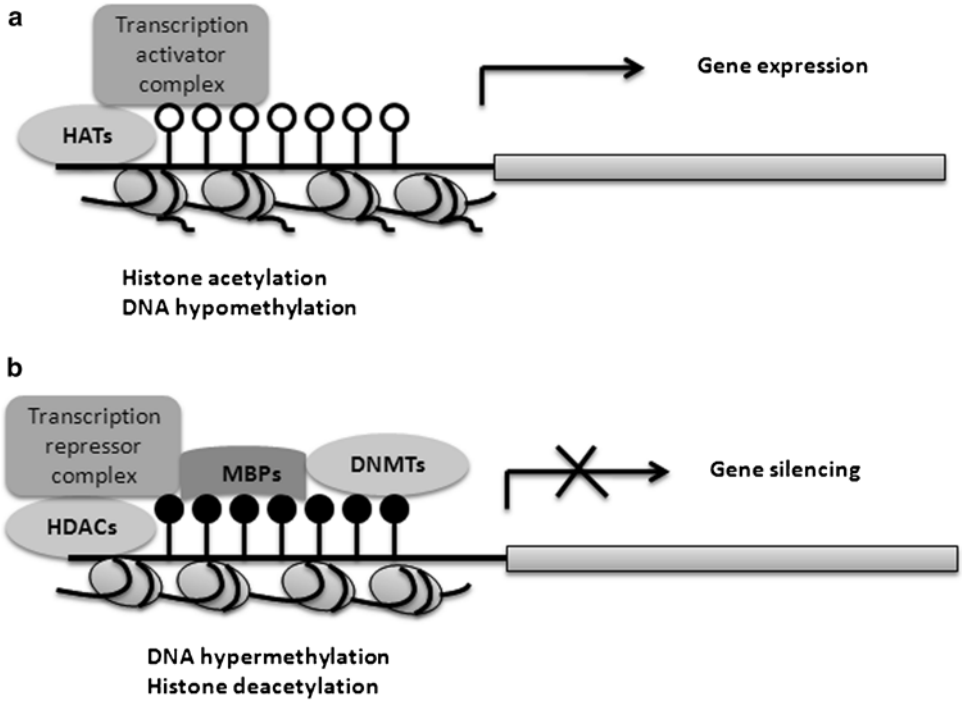


Fig. 1. Crosstalking between chromatin modifications and DNA methylation through transcription factors during gene transcription regulation. (**a**) Histone acetylation by histone acetyltransferases (HATs) and DNA hypomethylation can assemble active transcription factors on gene regulatory regions and induce gene expression. (**b**) Hypermethylated CpG sites by DNA methyltransferases (DNMTs) is bound by methyl-cytosine-binding proteins (MBPs), with the presence of histone deacetylases (HDACs), which recruits the complex of repressive transcription factors to the gene promoter, leading to gene repression.

of chromatin modification patterns or transcription factor binding (18). Several new technologies have recently emerged on the basis of the ChIP assay on a genome-wide scale, such as ChIP on chip (19, 20). However, there is no available technology to further test the DNA methylation patterns on the ChIP DNA. Recently, we developed a novel ChIP-bisulfite methylation sequencing (ChIP-BMS) approach on the basis of combining the conventional ChIP and bisulfite methylation sequencing assays (10). This new technology detects the methylation status of ChIP DNA pulled-down by a specific antibody (histone markers or transcription factors) (Fig. 2). ChIP-BMS is believed to provide an excellent opportunity to investigate the interaction patterns between histone modification and DNA methylation, transcription factor binding, and methylation of recognition sites, as well as multiple interactions between genetic and epigenetic factors. It could be widely used in various research fields such as determination of the crosstalking between histone modification and DNA methylation, candidate methylation-sensitive transcription factors and epigenetic regulation of gene transcription.

Table 1
Selective methylation-sensitive transcription factors

Transcription Factors	E2F-1	c-Myc	CTCF	Sp1
Putative recognition site[a]	TTTCC**CG**C	E-box CA**CG**TG	C**CG**C**G**C**GG**C**G**GCAG	ACTC**CG**CC**CG**
Regulated gene	*Cyclin E*, *ATM*, *p73*, *hTERT*	*NFκB*, *TGFβ*, *hTERT*	*c-MYC*, *Igf2*, *hTERT*	Housekeeping genes
Binding preference	Hypomethylation	Hypomethylation	Hypomethylation	Hypomethylation
Methylation effects on transcription	Repression	Repression	Repression	Activation
Primary function	Cellular cycle regulation	Cell proliferation Oncogenesis	Imprinting regulation Tumor suppression	Universal transcription regulation
Reference	(9, 10)	(7)	(4, 5)	(8)

[a]The consensus sequence of methylation-sensitive transcription factors contains at least one CpG dinucleotide (*bold*), which is recognized by the methyltransferase and methylation regulatory proteins, thus influencing binding ability of these transcription factors

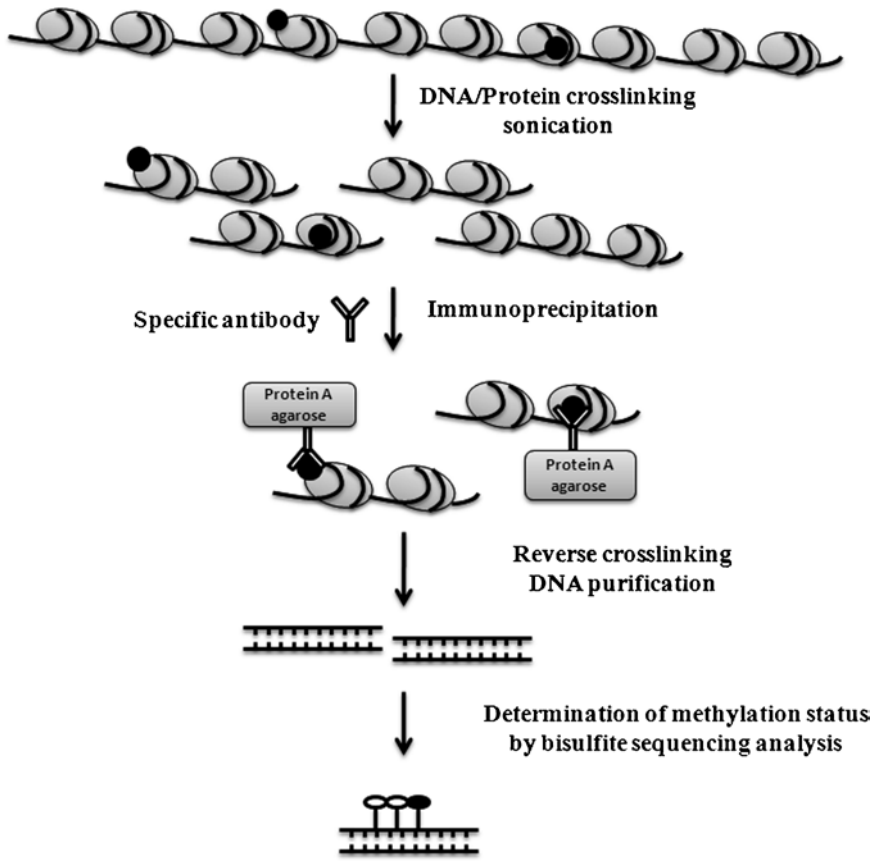


Fig. 2. Schematic representation of the procedure of chromatin immunoprecipitation-bisulfite methylation sequencing (Chip-BMS).

2. Materials

2.1. Chromatin Immunoprecipitation

2.1.1. DNA–Protein Crosslinking and Sonication

1. 37% Formaldehyde.
2. Cold 1× PBS buffer.
3. 2.5 M Glycine.
4. SDS lysis buffer (Millipore).
5. Protease inhibitor cocktail (Sigma).
6. Ultrasonic homogenizer (Biologics Inc.).
7. Agarose gel and electrophoresis apparatus.
8. Ethidium bromide.

2.1.2. Immunoprecipitation

1. ChIP dilution buffer (Millipore).
2. Salmon sperm DNA/Protein A Agarose-50% slurry (Millipore).
3. Specific antibody (histone marker or transcription factor) and control mouse IgG.

2.1.3. Wash

1. Low salt immune complex wash buffer (Millipore).
2. High salt immune complex wash buffer (Millipore).
3. LiCl immune complex wash buffer (Millipore).
4. TE buffer.
5. Shaker at 4°C at room temperature (RT).
6. Microcentrifuge.

2.1.4. DNA Elution and Reverse Crosslinking

1. ChIP elution buffer (freshly prepared): 0.1 M $NaHCO_3$ and 1% SDS.
2. 5 M NaOH.
3. Incubator at 65°C.

2.1.5. Immunoprecipitated DNA Purification

1. Proteinase K buffer (freshly prepared): 10 mM Tris–HCl, pH 6.5; 5 mM EDTA, 10 μg/μl Proteinase K.
2. Wizard DNA clean-up system (Promega).
3. Disposable 5-ml lure-lock syringes.
4. Deionized water or TE buffer.

2.1.6. ChIP PCR

1. Regular 2× PCR Mastermix (Promega) for semiquantitative PCR and SYBR Green qPCR Supermix (Invitrogen) for Realtime-quantitative PCR.
2. Specific ChIP primers.
3. Agarose gel and electrophoresis apparatus.
4. Roche Realtime LC480.

2.2. ChIP-Bisulfite Methylation Sequencing

1. Quantification of Chip-purified DNA: Spectrophotometer (Bio-Rad).
2. Bisulfite reaction kit, EpiTect Bisulfite Kit (Qiagen).

2.2.1. Bisulfite Reaction

3. Wizard DNA clean-up system (Promega) for purification of bisulfite-treated DNA.
4. Disposable 5-ml lure-lock syringes.
5. Deionized water or TE buffer.

2.2.2. PCR Purification, Cloning, and Sequencing Analysis of ChIP DNA Fragment

1. Regular 2× PCR Mastermix (Promega).
2. QIAquick PCR Purification Kit (Qiagen) for purification of PCR product.
3. QIAquick Gel Extraction Kit (Qiagen) for purification of target PCR fragment from multiple unspecific PCR products.
4. pGEM-T Easy vector system II (Promega).
5. For bacterial culturing and positive cloning selection, bacto tryptone (BD), yeast extract, sodium chloride, ampicillin solution, isopropyl-β-D-thiogalactoside (IPTG), X-Gal (Bio-Rad), and bacterial shaker incubator at 37°C are required.

6. QIAprep Spin Miniprep Kit (Qiagen).
7. ABI 3730 DNA Analyzer.

3. Methods

3.1. Chromatin Immunoprecipitation

3.1.1. DNA–Protein Crosslinking and Sonication

The method described below is suitable for the analysis of cultured cells. The cells grown in 100 mm culture plates should reach 90% confluence (around 1×10^6 cells) prior to the harvest point. To obtain adequate ChIP DNA for the subsequent bisulfite sequencing analysis, several extra plates of cells may be required (see Note 1). All procedures should be performed on ice to prevent potential protein loss.

1. Add 270 μl of 37% formaldehyde to 10 ml of growth medium (1% final concentration) for crosslinking for 10 min in the incubator at 37°C (see Note 2).
2. Add 0.5 ml 2.5 M glycine (0.125 M of final concentration) to the medium to quench the crosslinking and incubate at room temperature for 5 min.
3. Remove the medium and wash the attached cells with cold PBS twice.
4. Scrape the cells from the plates to a microtube and centrifuge at $13{,}400 \times g$ to pellet the cells.
5. Discharge the supernatant and resuspend the cell pellet with 500 μl in SDS lysis buffer containing protease inhibitor, then incubate on ice for 10 min.
6. Shear DNA using an Ultrasonic homogenizer at 20–30% output. Each sample is sonicated for four to six cycles (10 s sonication, 30 s pause). The conditions for sonication should be carefully optimized. The optimal crosslinked chromatin DNA should be sheared at 200–1,000 bp in length as determined by routine electrophoresis gel analysis (see Note 3). This sheared DNA sample can be stored in −80°C for half a year or in liquid nitrogen for longer storage.

3.1.2. Immunoprecipitation of Crosslinked Chromatin DNA–Protein

1. Dilute 100 μl of sheared DNA–protein sample from Subheading 3.1.1, step 6 with 900 μl 10× ChIP dilution buffer containing protease inhibitor. A small fraction of chromatin DNA after dilution (~50 μl) will be extracted for the future use of internal control (input). Proceed to Subheading 3.1.3, step 1 for continuation of processing of the input.
2. Optimally, a preclean step should be included prior to immunoprecipitation in order to remove the nonspecific background. In this step, 50 μl of Protein A Agarose slurry is

added to the chromatin followed by incubation for 1 h at 4°C with rotation (see Note 4).

3. Spin down the agarose beads and remove the supernatant to a new micro-tube.
4. Add the specific antibody (5–10 μg) to the supernatant and incubate overnight at 4°C with rotation (see Note 5).
5. Incubate 60 μl of Protein A Agarose with the chromatin DNA–protein complex for 2–3 h at 4°C with rotation. Briefly spin down the agarose and discharge the supernatant (see Note 6).
6. Wash the Protein A Agarose-antibody/chromatin complex by suspending the agarose with the following commercially-available wash buffers in sequence at room temperature with rotation and collect the agarose beads by brief centrifugation (1,560 × *g*).
   (a) Low Salt Immune Complex Wash Buffer, 1 ml, one wash, 10 min.
   (b) High Salt Immune Complex Wash Buffer, 1 ml, one wash, 10 min.
   (c) LiCl Immune Complex Wash Buffer, 1 ml, one wash, 10 min.
   (d) TE Buffer, 1 ml, two washes, 10 min.
7. After the final wash with TE, the agarose is incubated with 250 μl freshly made ChIP Elusion buffer for 20 min at room temperature with rotation for twice. The supernatant containing specifically pulled down-chromatin DNA–protein complex will be collected together for a total volume of 500 μl.

3.1.3. Purification of Immunoprecipitated Chromatin DNA and ChIP-PCR

1. Reverse crosslinking: Add 20 μl 5 M NaCl to the 500 μl eluent. Add 2 μl 5 M NaCl to the input DNA from Subheading 3.1.2, step 1. Incubate the eluent and input at 65°C for 6 h or overnight.
2. Add 30 μl of freshly made Proteinase K buffer to the eluent. Add 3 μl of Proteinase K buffer to the input. Incubate the eluent and input at 45°C for 1 h.
3. We routinely use the Wizard DNA clean-up kit from Promega to purify the immunoprecipitated chromatin DNA. The detailed procedure involving this step is followed by the manufacturer's protocol.
4. The purified immunoprecipitated (IP) DNA is ready for PCR analysis. ChIP-PCR is performed as a regular PCR reaction. To verify the ChIP-PCR results, appropriate controls should be set up along with the whole procedure (see Note 7). The PCR results can be determined by gel-based electrophoresis as a semiquantitative PCR, or if possible, a quantitative PCR can be performed when the specific realtime ChIP primers are available.

The final results will be normalized to input DNA and calibrated to levels in control samples if necessary.

5. The ChIP-PCR results are considered as positive binding at the selected region of the DNA when the abundance of the DNA band is more than tenfold higher than the negative control. Therefore, the DNA methylation status of this selected binding region can be detected by the use of IP DNA (see Note 8).

3.2. Bisulfite Modification, PCR Amplification and Sequencing

1. Quantification of IP DNA: at least 2 μg of IP DNA is required for bisulfite treatment per sample. Several attempts may be needed to obtain an adequate amount of IP DNA (see Note 1).
2. Use the EpiTect Bisulfite Kit (Qiagen) to convert and purify the bisulfite modified-IP DNA. The detailed protocol is provided by the manufacturer's protocol (see Note 9).
3. ChIP-Bisulfite PCR amplification can be performed as a normal PCR reaction. As a general rule for any PCR reaction, the PCR conditions should be carefully optimized (see Note 10). The PCR results will be verified by gel-based electrophoresis and a single, bright, and specific band will be considered as a successful PCR amplification.
4. After a successful PCR amplification, PCR products should be purified by using commercially available kits such as QIAquick PCR Purification Kit (Qiagen) or QIAquick Gel Extraction Kit (Qiagen), which can help purify the target PCR product from multiple nonspecific PCR bands.
5. The purified PCR products can be directly sequenced. Subcloning sequencing is necessary to observe methylation patterns of the single molecules. We prefer to use the pGEM-T Easy vector system II (Promega) for cloning purposes and all detailed procedures are followed by the manufacturer's protocol.

3.3. Data Interpretation

Following successful bisulfite PCR amplification or subcloning procedures, IP DNA methylation status can be interpreted by subsequent sequencing analysis. After bisulfite treatment, all unmethylated cytosines (C) convert to thymine (T) and the presence of C-peaks indicate the presence of 5mC in the genome. If a band appears in both the C- and T-peaks, then this indicates partial methylation or potentially incomplete bisulfite conversion. For example, to observe the methylation status of the binding site of E2F-1, a methylation-sensitive transcription factor, ChIP assay is performed followed by a bisulfite sequencing analysis on IP DNA. Therefore, by comparing the original DNA sequence, the methylation status of E2F-1 binding sites with E2F-1 binding on certain gene promoters can be determined by bisulfite sequencing analysis. Alternatively, the proportion of methylation changes of E2F-1 binding sites with E2F-1 binding can be interpreted by comparing a large segment containing the E2F-1 binding site for DNA methylation changes.

4. Notes

1. One trial of ChIP assay (1×10^6) will yield about 1 μg of IP DNA. However, IP DNA has been sheared to 200–1,000 bp by sonication, which will increase DNA damage during bisulfite modification. Therefore, ChIP-BMS may require more DNA (>2 μg) for the bisulfite reaction to compensate for the potential DNA loss. Several repeat ChIP reactions will help to collect enough IP DNA for the bisulfite reaction if necessary.
2. The conditions of crosslinking should be carefully optimized. The amount of formaldehyde, the fix time, and incubation temperature should be determined for different types of cells, as well as proteins and target DNA regions of interest (21). For high-affinity DNA binding proteins such as histones, the native chromatin can be used for the ChIP assay (22).
3. Optimal conditions for sonication are required that guarantee the sheared and crosslinked DNA will be in the 200–1,000 bp range. Variable parameters such as processing time and the power setting should be adjusted every time if the cell type and cell lysis concentration have been altered. Agarose-gel based electrophoresis can be used to determine the desired length of sheared DNA.
4. If the enrichment of ChIP product is weak, this step can be omitted to reduce subsequent loss of DNA/protein complexes.
5. The primary antibody for immunoprecipitation should specifically recognize the protein of interest. Monoclonal antibodies are desired due to their specificity, whereas polyclonal antibodies can also be applied if no monoclonal antibody is available although nonspecific binding may occur with the use of a polyclonal antibody. The appropriate amount of primary antibody for immunoprecipitation should be determined to yield enough IP DNA. Normally, 2–5 μg of antibody will produce enough IP DNA, but low specific antibody may reduce the amount of IP DNA.
6. Protein A Agarose is suitable for binding most antibodies. However, Protein G Agarose has more affinity to bind mouse IgG than Protein A Agarose. A combined Protein A/G Agarose can be applied since this combination has the additive properties of Protein A and G Agarose.
7. To avoid nonspecific binding, appropriate controls are very important. Optimally, there should be a set of four different controls in ChIP PCR: positive control (RNA polymerase antibody), negative control (mouse IgG), "no antibody" control, and an internal control (input). The control primers (GAPDH) and specific primers will be used to determine the outcome of ChIP-PCR (Table 2).

Table 2
Interpretation of ChIP results

Primers		Positive control	Negative control	"No antibody" control	Sample	Internal control	Indicated consequence	Solution
Control primer	GAPDH	+	–	–	n/a	+	Good working system	Proceed to testing the specific primer
	GAPDH	+	+/slight	–	n/a	+	Weak working system	Need to adjust the crosslinking and sonication conditions
	GAPDH	+	+	+	n/a	+	Potential contamination	Change PCR reagent
Specific primer	Specific primer	n/a	–	–	+	+	Good result	n/a
	Specific primer	n/a	+/slight	–	+	+	Nonspecific antibody	Need to adjust annealing temperature or other PCR conditions; change antibody
	Specific primer	n/a	+	+	+	+	Potential contamination	Change PCR reagent

8. Only IP DNA-produced positive binding at the specific location can be used for subsequent bisulfite analysis. For the negative binding samples, a regular bisulfite genomic sequencing can be used to determine DNA methylation status by which the ChIP assay is omitted.
9. As mentioned previously, IP DNA should be sheared to 200–1,000 bp by sonication and subjected to the subsequent immunoprecipitation and purification procedures. The quality and quantity of IP DNA likely has been dramatically reduced during the ChIP treatment. For a limited amount of IP-DNA sample, several modifications can be applied to reduce further DNA damage and loss such as the use of low-melting point agarose block during bisulfite reaction (23).
10. The basic principles for designing of methylation PCR primers and adjustments of PCR conditions have been well described previously (24, 25). However, ChIP-BMS uses DNA fragments (200–1,000 bp) from IP DNA rather than genomic DNA to initiate bisulfite conversion. Different lengths of bisulfite primers covering the area of the IP DNA are necessary to determine the desired outcome of bisulfite PCR.

Acknowledgments

This work was supported in part by grants from the National Cancer Institute (R01 CA 129415), the Susan G. Komen for the Cure, and a Postdoctoral Award (PDA) sponsored by the American Institute for Cancer Research (AICR).

References

1. Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21.
2. Kass, S. U., Pruss, D., and Wolffe, A. P. (1997) How does DNA methylation repress transcription? *Trends Genet.* **13**, 444–449.
3. Bird, A. P., and Wolffe, A. P. (1999) Methylation-induced repression belts, braces, and chromatin. *Cell* **99**, 451–454.
4. Ohlsson, R., Renkawitz, R., and Lobanenkov, V. (2001) CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet.* **17**, 520–527.
5. Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., and Tilghman, S. M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **405**, 486–489.
6. Mikovits, J. A., Young, H. A., Vertino, P., Issa, J. P., Pitha, P. M., Turcoski-Corrales, S., Taub, D. D., Petrow, C. L., Baylin, S. B., and Ruscetti, F. W. (1998) Infection with human immunodeficiency virus type 1 upregulates DNA methyltransferase, resulting in *de novo* methylation of the gamma interferon (IFN-γ) promoter and subsequent downregulation of IFN-γ production. *Mol. Cell. Biol.* **18**, 5166–5177.
7. Blackwell, T. K., Kretzner, L., Blackwood, E. M., Eisenman, R. N., and Weintraub, H. (1990) Sequence-specific DNA binding by the c-Myc protein. *Science* **250**, 1149–1151.
8. Höller, M., Westin, G., Jiricny, J., and Schaffner, W. (1998) Sp1 transcription factor binds DNA and activates transcription even when the binding site is CpG methylated. *Genes Dev.* **2**, 1127–1135.

9. Ohtani, K., DeGregori, J., and Nevins, J. R. (1995) Regulation of the cyclin E gene by transcription factor E2F1. *Proc. Natl. Acad. Sci. USA.* **92**, 12146–12150.
10. Li, Y., Liu, L., Andrews, L. G., and Tollefsbol, T. O. (2009) Genistein depletes telomerase activity through cross-talk between genetic and epigenetic mechanisms. *Int. J. Cancer* **125**, 286–296.
11. Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genet.* **19**, 187–191.
12. Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389.
13. Berger, S. L. (2007) The complex language of chromatin regulation during transcription. *Nature* **447**, 407–412.
14. Kouzarides, T. (2007) Chromatin modifications and their function. javascript:AL_get(this, 'jour', 'Cell.');*Cell* **128**, 693–705.
15. Vettese-Dadey, M., Grant, P. A., Hebbes, T. R., Crane-Robinson, C., Allis, C. D., and Workman, J. L. (1996) Acetylation of histone H4 plays a primary role in enhancing transcription factor binding to nucleosomal DNA *in vitro*. *EMBO J.* **15**, 2508–2518.
16. Lee, D. Y., Hayes, J. J., Pruss, D., and Wolffe, A. P. (1993) A positive role for histone acetylation in transcription factor access to nucleosomal DNA. *Cell* **72**, 73–84.
17. Frommer, M., McDonald, L. E., Millar, D. S., Collis, C.M., Watt, F., Grigg, G. W., Molloy, P. L., and Paul, C. L. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* **89,** 1827–1831.
18. Das, P. M., Ramachandran, K., vanWert, J., and Singal, R. (2004) Chromatin immunoprecipitation assay. *Biotechniques* **37**, 961–969.
19. Weinmann, A. S., and Farnham, P. J. (2002) Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* **26**, 37–47.
20. Buck, M. J., and Lieb, J. D. (2004) ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**, 349–360.
21. Orlando, V., Strutt, H., and Paro, R. (1997) Analysis of chromatin structure by in vivo formaldehyde cross-linking. *Methods* **11**, 205–214.
22. O'Neill, L. P., and Turner, B. M. (2003) Immunoprecipitation of native chromatin: NChIP. *Methods* **31**, 76–82.
23. Olek, A., Oswald, J., and Walter, J. (1996) A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res.* **24**, 5064–5066.
24. Li, L. C. (2007) Designing PCR primer for DNA methylation mapping. *Methods Mol. Biol.* **402,** 371–384.
25. Li, L. C. (2002) Dahiya R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **18**, 1427–1431.

Chapter 19

Studying RNA–Protein Interactions In Vivo By RNA Immunoprecipitation

Luke A. Selth, Pierre Close, and Jesper Q. Svejstrup

Abstract

The crucial roles played by RNA-binding proteins in all aspects of RNA metabolism, particularly in the regulation of transcription, have become increasingly evident. Moreover, other factors that do not directly interact with RNA molecules can nevertheless function proximally to RNA polymerases and have significant effects on gene expression. RNA immunoprecipitation (RIP) is a powerful technique used to detect direct and indirect interactions between individual proteins and specific RNA molecules in vivo. Here, we describe RIP methods for both yeast and mammalian cells.

Key words: RNA immunoprecipitation, RNA-binding protein, Transcription factor, mRNA, Chromatin immunoprecipitation, Gene expression

1. Introduction

RNA immunoprecipitation (RIP) is a powerful technique for detecting the association of individual proteins with specific RNA species in vivo (1). In practice it is highly similar to chromatin immunoprecipitation (ChIP), one of the most widely used tools in molecular biology, but with some important caveats. Live cells are treated with formaldehyde to generate protein–protein, protein–DNA, and protein–RNA cross-links between molecules in close proximity. A whole-cell extract is prepared in the presence of RNase inhibitors to maintain the integrity of RNA, and the cross-linked nucleic acids are sheared by sonication to enable their solubilization. The extract is then enzymatically treated to remove DNA, and the resulting material is immunoprecipitated with an antibody against the protein of interest. RNA sequences that cross-link with this protein are selectively enriched in the immunoprecipitated sample.

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_19,

Reversal of the formaldehyde cross-linking and a second DNase treatment results in the recovery of these RNA molecules, which can be accurately quantitated by reverse transcription-PCR. The enrichment of specific sequences relative to control sequences provides information about the relative level of association of a given protein with different RNAs. This chapter describes a RIP protocol for the baker's yeast *Saccharomyces cerevisiae* and human embryonic kidney (HEK) 293T cells, but other cell types could be utilized according to the user's preferences. A flow-chart outlining the main steps of the protocol is shown in Fig. 1.

One important consideration prior to commencing RIP experiments designed to identify protein–RNA interactions is whether cross-linking is necessary. Omitting the fixing step and performing what we refer to as "native" RIP may be more appropriate for proteins that bind RNA directly and/or with high affinity, because cross-linking reduces RNA recovery rates and introduces sequence bias. Several excellent native RIP protocols have been developed for the identification and analysis of RNA-binding proteins (2–5). However, utilizing a reversible cross-linker such as formaldehyde provides many advantages. First and foremost, it enables the identification of indirect protein–RNA associations. This utility is

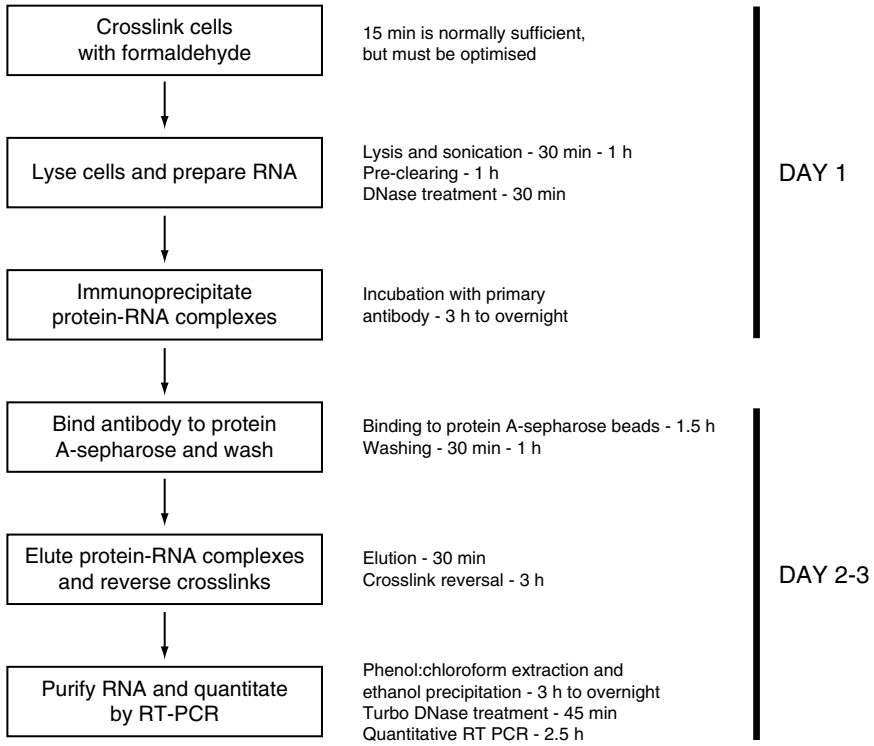


Fig. 1. Flowchart of RNA immunoprecipitation protocol. (Reproduced from (10) with permission from Cold Spring Harbor Press).

significant because proteins which do not directly bind to RNA molecules can nevertheless play crucial roles in transcription and downstream mRNA processing. One such example would be the large group of factors which interact with the RNA polymerase II (RNAPII) C-terminal domain, including chromatin remodeling complexes and histone modifying enzymes, during transcript elongation (6). Such factors can be cross-linked to specific RNA molecules via DNA or protein (e.g., RNAPII) bridges, a finding which would be indicative of a functional role in transcription of these RNAs. This is analogous to the standard ChIP assay, which can be used to study factors which do not bind DNA directly but associate with chromatin to mediate DNA-related processes. Second, cross-linking stabilizes protein–RNA interactions which can easily be lost during cell lysis and immunoprecipitation. For example, protein–RNA complexes may disassociate due to changes in solute composition or dilution effects (7). Alternatively, the binding of a protein of interest to a specific RNA ligand may be out-competed by abundant, nonspecific RNA-binding proteins that were separated from the RNA in compartmentalized cells but are available in cell extract (7). Finally, because formaldehyde inactivates cellular enzymes essentially immediately upon its addition to cells it can be used to examine "snapshots" of protein–RNA interactions at specific time points. Thus, the RIP protocol described here is likely to be more useful for kinetic analyses of events occurring on RNA in vivo than native RIP. In summary, except in the rare cases of abundant, high-affinity RNA-binding proteins, we propose that treating extracts with a reversible cross-linking agent like formaldehyde is the method of choice to characterize the role of a protein of interest in the transcription of a particular gene. However, it must always be kept in mind that the observed interaction may occur via multiple layers of cross-linked factors.

2. Materials

2.1. Yeast Cell Culture and Lysis

1. YPD (1% Bacto-yeast extract (w/v), 1% Bacto-peptone (w/v), 2% dextrose (w/v)), or appropriate yeast culture media.
2. Formaldehyde (37%). Toxic.
3. Glycine (2 M).
4. Tris-buffered saline (TBS): 140 mM NaCl, 10 mM Tris–HCl (pH 7.5).
5. FA lysis buffer: 1 mM EDTA (pH 8.0), 50 mM HEPES–KOH (pH 7.5), 140 mM NaCl, 0.1% sodium deoxycholate (w/v), 1% triton X-100 (v/v). Add 1× protease inhibitors immediately before use (for 100× protease inhibitors: Dissolve 1.42 mg leupeptin, 6.85 mg pepstatin A, 0.85 mg phenylmethylsulfonyl

fluoride, and 1.65 mg benzamidine in 50 mL 100% ethanol. Store at –20°C).

6. RNasin (Promega). Store at –20°C.
7. Zirconia/silica beads, 0.5 mm diameter (BioSpec).

2.2. Mammalian Cell Culture and Lysis

1. Dulbecco's modified Eagle's medium (DMEM) (Invitrogen/Gibco, Carlsbad, CA) supplemented with 10% fetal bovine serum (FBS) (Hyclone, Logan, UT). Store at 4°C.
2. Phosphate-buffered saline (PBS) (for 10× stock:1.37 M NaCl, 27 mM KCl, 100 mM Na_2HPO_4, 18 mM KH_2PO_4 (adjust to pH 7.4 with HCl). Autoclave before storage at room temperature. Prepare working solution by dilution of one part with nine parts water).
3. Teflon cell scrapers (Fisher).
4. Swelling buffer: 5 mM Hepes (pH 8.0), 85 mM KCl, 0.5% Nonidet P-40. Add 1× protease inhibitors immediately before use.
5. Nuclei lysis buffer: 50 mM Tris–HCl (pH 8.1), 10 mM EDTA (pH 8.0), 1% sodium dodecyl sulphate (SDS) (w/v). Add 1× protease inhibitors immediately before use.

2.3. Immunoprecipitation

1. Protein A/G Agarose (50% (v/v); Pierce). Wash and equilibrate in FA lysis buffer containing 1 mg/mL of filter-sterilized bovine serum albumin, 0.1 mg/mL yeast transfer RNA (Sigma). Add 40 U/mL immediately RNasin before use.
2. $MgCl_2$ (1 M).
3. $CaCl_2$ (2 M).
4. DNase I, RNase-free (20 mg/mL; Sigma). Store at –20°C.
5. EDTA (0.5 M, pH 8.0).
6. Antibody targeted against protein or epitope of interest and an appropriate (isotype-matched) control antibody (e.g., IgG).
7. RIP elution buffer (10 mM EDTA, 100 mM Tris–HCl (pH 8.0), 1% sodium dodecyl sulfate (w/v), RNasin (40 U/μL; Promega)). Make fresh before use.
8. FA500 buffer (EDTA (1 mM, pH 8.0), HEPES–KOH (50 mM, pH 7.5), NaCl (500 mM), sodium deoxycholate (0.1% (w/v)), triton X-100 (1% (v/v)).
9. LiCl wash (EDTA (1 mM, pH 8.0), LiCl (250 mM), Nonidet P-40 (0.5% (v/v)), sodium deoxycholate (0.1% (w/v)), Tris–HCl (10 mM, pH 8.0)).
10. TE buffer (EDTA (10 mM, pH 8.0), Tris–HCl (100 mM, pH 8.0)).
11. GELoader tips, 20-μL (Eppendorf).

2.4. Cross-Link Reversal and RNA Purification

1. NaCl (5 M).
2. Proteinase K (20 mg/mL; Roche Applied Science). Store at 4°C.
3. Water, nuclease-free (Ambion).
4. Acid-phenol:chloroform (with IAA, 25:24:1, pH 4.5; Ambion). Store at 4°C. Toxic.
5. Tubes, MaXtract, high density, 2-mL (Qiagen).
6. Ethanol (70 and 100%).
7. Sodium acetate (3 M, pH 5.2).
8. Glycogen (20 mg/mL; Roche Applied Science). Store at –20°C.
9. Turbo DNase kit (Ambion).

2.5. RNA Quantitation

1. Oligonucleotide primers (for RNA quantitation).
2. Multiscribe Reverse transcriptase (Applied Biosystems). Store at –20°C.
3. SYBR Green mix (ABsolute QPCR, ABgene).

3. Methods

3.1. Preparation of Chromatin: Yeast Cells

1. Grow 40 mL of yeast cells of interest to a density of $0.5–2 \times 10^7$ cells/mL (see Note 1) in YPD or an appropriate alternative growth medium.
2. To 20 mL of culture, add formaldehyde to a final concentration of 1% (see Note 2) and cross-link by shaking slowly on a platform for 10 min at room temperature (see Note 3). The remaining culture will be used as a control in which formaldehyde cross-linking is omitted (see Note 4).
3. Add 2 M glycine to a final concentration of 200 mM to both the cross-linked and control cultures. Incubate for an additional 5 min at room temperature with gentle shaking. The glycine stops the cross-linking by reacting with formaldehyde.
4. Centrifuge the cells at $2{,}500 \times g$ for 5 min at 4°C. Discard the supernatant into a chemical waste container.
5. Resuspend the pellet in 25 mL of ice-cold TBS. Transfer to a 50-mL centrifuge tube. Repeat the TBS wash once (see Note 5).
6. Centrifuge the cells at $2{,}500 \times g$ for 5 min at 4°C. Discard the supernatant.
7. Resuspend the cells in 0.5 mL of ice-cold FA lysis buffer. Transfer to a 1.5-mL microcentrifuge tube.
8. Pellet the cells in a benchtop centrifuge at $15{,}000 \times g$ for 30 s at 4°C. Discard the supernatant.

9. Resuspend the cell pellet in 0.75 mL of ice-cold FA lysis buffer containing 40 U/mL RNasin (see Note 6).
10. Add 500 μL of zirconia/silica beads. Lyse the cells in a cell disruptor (FastPrep, MP Biomedicals) using a speed setting of 5.5 for 30 s. Incubate the samples for 1 min on ice. Repeat four to five times for a total disruption time of 2–3 min (see Note 7).
11. Using a syringe needle, make a hole in the bottom of the 1.5-mL tube. Place the 1.5-mL tube into a 2-mL tube. Centrifuge at $2{,}000 \times g$ for 2 min at 4°C.
12. Resuspend the pellet in the 2-mL tube and then transfer the sample into two 15-mL tubes for sonication (i.e., 0.375 mL per tube). Add 0.625 mL of FA lysis buffer containing 40 U/mL RNasin to each tube. At this stage, each tube should contain extract from the equivalent of 1×10^8 cells.
13. Sonicate the sample in a BioRuptor (Diagenode, Belgium) on high ("H") for 5 min with 30 s on/off cycles. Ensure that the sample is kept cold by filling the bath with an ice/water mix prior to sonication (see Note 8).
14. Transfer the sonicated extracts into two 1.5 mL tubes. Clear by centrifugation at maximum speed in a benchtop centrifuge for 10 min at 4°C and transfer to new 1.5-mL tubes.

3.2. Preparation of Chromatin: HEK 293T Cells

1. Grow 8 × 150 mm plates of HEK 293T cells to 70–80% confluence (see Note 1) in DMEM + 10% FBS. Cells should be healthy and not density-arrested prior to cross-linking.
2. To four of the plates, add formaldehyde to a final concentration of 1% (see Note 2) directly to the tissue culture media. Cross-link by shaking slowly on a platform for 10 min at room temperature (see Note 3). The remaining plates will be used as a control in which formaldehyde cross-linking is omitted.
3. Add 2 M glycine to a final concentration of 125 mM to both the cross-linked and control plates. Incubate for an additional 5 min at room temperature. The glycine stops the cross-linking by reacting with formaldehyde.
4. Discard the media into a chemical waste container and wash the cells with 10 mL ice-cold PBS per plate. Repeat the PBS wash once (see Note 5).
5. Add 2–3 mL of ice-cold PBS to each plate and scrape the cells into a 50-mL centrifuge tube. Rinse the culture plates with 30 mL PBS and add the remaining cells to the 50-mL tube.
6. Centrifuge the cells at $500 \times g$ for 10 min at 4°C. Carefully aspirate supernatant so as to not lose cells.
7. Resuspend the pellet in 50 mL of PBS. Pellet and aspirate the supernatant as above (see Note 6).

8. Resuspend cells in 0.4 mL of Swelling Buffer (0.2 mL/1×10^7 cells).
9. Incubate on ice for 15 min with occasional flicking.
10. Transfer sample to 1.5 mL tubes and centrifuge at $2{,}500 \times g$ for 5 min at 4°C to pellet nuclei (see Note 9). Remove the supernatant with a pipette.
11. Resuspend nuclei in 0.2 mL of Nuclei Lysis Solution (0.1 mL/1×10^7 cells) containing 40 U/mL RNasin and incubate on ice for 10 min.
12. Transfer the sample into two 15-mL tubes for sonication. Dilute the extracts tenfold with FA lysis buffer containing 40 U/mL RNasin (i.e., add 0.9 mL FA lysis buffer to 0.1 mL sample in each tube). At this stage, each tube should contain extract from the equivalent of 1×10^7 cells.
13. Sonicate the sample in a BioRuptor (Diagenode, Belgium) on high ("H") for 5 min with 30 s on/off cycles. Ensure that the sample is kept cold by filling the bath with an ice/water mix prior to sonication (see Note 8).
14. Transfer the sonicated extracts into two 1.5 mL tubes. Clear by centrifugation at maximum speed in a benchtop centrifuge for 10 min at 4°C and transfer to new 1.5-mL tubes.

3.3. Immunoprecipitation

1. To each tube, add 25 µL (bead volume; i.e., 50 µL of the 50% slurry) of Protein A/G agarose. Incubate with end-over-end rotation for 1 h at 4°C to pre-clear the extract.
2. Centrifuge at $1{,}000 \times g$ for 2 min at 4°C. Carefully transfer the supernatant to a new microcentrifuge tube.
3. Adjust the extract to 25 mM $MgCl_2$ (with 1 M $MgCl_2$) and 5 mM $CaCl_2$ (with 2 M $CaCl_2$). Add 3 µL RNasin and 6 µL RNase-free DNase I (see Note 10). Incubate at 37°C for 20 min.
4. Stop the reaction by adding 0.5 M EDTA to a final concentration of 20 mM.
5. Centrifuge in a benchtop centrifuge at maximum speed for 10 min at 4°C. Transfer the extract to a new microcentrifuge tube.
6. To prepare input RNA, dilute an amount of extract equivalent to 1×10^6 yeast cells or 1×10^5 293T cells from each sample with 150 µL of RIP elution buffer. Snap-freeze inputs in liquid nitrogen or on a dry-ice/ethanol bath and store at –80°C (see Note 11).
7. Incubate one of the two DNase-treated extracts with 1–5 µg of the primary antibody against the protein or epitope of interest with end-over-end rotation overnight at 4°C (see Note 12). To the other extract, incubate with the same amount of an isotype-matched control antibody (e.g., IgG).

8. Add 25-µL (bead volume; i.e., 50 µL of the 50% slurry) of Protein A/G agarose to the sample. Incubate with end-over-end rotation for 90 min at 4°C.
9. Pellet the beads by centrifugation at 1,000×*g* for 2 min at 4°C. Remove the supernatant and resuspend in 1 mL of ice-cold FA lysis buffer. Incubate with end-over-end rotation for 5 min at 4°C.
10. Repeat spin and washing with 1 mL of ice-cold FA500 buffer.
11. Repeat spin and washing with 1 mL of ice-cold LiCl wash buffer (see Note 13).
12. Repeat spin and washing with 1 mL of ice-cold TE buffer.
13. Remove most of the supernatant from the TE pellet with a 1-mL tip. Aspirate the remaining supernatant with a GELoader tip to prevent loss of beads.
14. Add 75 µL of RIP elution buffer containing 40 U/mL RNasin. Incubate in a Thermomixer (Eppendorf) at 1,200 rpm for 10 min at 37°C.
15. Pellet the beads by centrifugation at 1,000×*g* for 2 min at room temperature. Transfer the supernatant into a new microcentrifuge tube. Repeat the elution with a fresh 75-µL aliquot of RIP elution buffer. Pool the eluates.

3.4. Cross-Link Reversal and RNA Purification

1. Add 6 µL of 5 M NaCl (to a final concentration of ~200 mM) and 20 µg proteinase K. Incubate for 1 h at 42°C and then for 1 h at 65°C (see Note 14). Do the same for the input samples (from Subheading 3.3, step 6).
2. Add 100 µL of nuclease-free water to the RIP and input samples together with an equal volume (250 µL) of acid-equilibrated phenol:chloroform. Separate the phases by centrifuging at 10,000×*g* for 3 min at room temperature in MaXtract tubes.
3. Transfer the aqueous layer to a new microcentrifuge tube. Add 25 µL of 3 M sodium acetate, 20 µg of glycogen and 625 µL of ice-cold 100% ethanol. Precipitate the RNA for 1–2 h at −80°C or overnight at −20°C.
4. Centrifuge in a benchtop centrifuge at maximum speed for 30 min at 4°C.
5. Wash the pellet with 1 mL of ice-cold 70% ethanol. Centrifuge in a benchtop centrifuge at maximum speed for 5 min at 4°C. Discard the supernatant.
6. Allow the pellet to air-dry for 5–10 min. Resuspend the precipitated RNA in 90 µL of RNase-free water.
7. Add 10 µL of TURBO DNase buffer and 1 µL of Turbo DNase (see Note 10). Incubate at 37°C for 30 min.

8. Add 10 µL DNase inactivation reagent (included in the Turbo DNase kit). Incubate with occasional agitation for 2 min at room temperature.
9. Centrifuge at $10{,}000 \times g$ for 1.5 min at room temperature. Transfer the supernatant to a new microcentrifuge tube.

3.5. RNA Quantification

1. Prepare reverse transcription reactions in a final volume of 20 µL:

ABsolute QPCR SYBR Green reagent	10 µL
RNasin	0.1 µL
Multiscribe reverse transcriptase (see Note 15)	0.2 µL
Oligonucleotide primers (see Note 16)	70 nM
RIP or input RNA	1–4 µL
H_2O	to 20 µL

2. Perform one-step quantitative reverse transcription-PCR on a Bio-Rad MyIQ iCycler (or equivalent): one cycle of 30 min at 50°C (reverse transcription); one cycle of 15 min at 95°C (DNA polymerase activation); forty cycles of 15 s at 95°C, 1 min at 60°C (amplification) (see Note 17).
3. Quantitate the PCR products using appropriate software provided with the real-time PCR machine.
4. Calculate the immunoprecipitation efficiency for a specific fragment by dividing the amount of product obtained with the immunoprecipitated RNA by the amount obtained with the input RNA. Calculate the RIP enrichment by dividing the IP efficiency of the experimental RIP by the IP efficiency of the control (e.g., IgG) RIP. Compare experimental RIP efficiencies at predicted RNA binding/association sites and control sites (see Note 16).

4. Notes

1. Extract from 1×10^8 yeast cells and 1×10^7 human cells is typically used per immunoprecipitation, although it must be noted that this amount can vary depending on the abundance of the relevant protein and the RNA transcript. Therefore, this step utilizes yeast and human cell cultures of sufficient size for approximately two RIPs, i.e., one RIP using an antibody against the protein of interest and a control RIP using a nonspecific antibody. For experiments involving the analysis of multiple factors (i.e., multiple immunoprecipitations), the volume of culture/number of plates can be scaled up according to need.

2. Keep cultures covered or work in a fume hood to avoid noxious formaldehyde fumes.
3. It is strongly advised that an initial time course of cross-linking (5 min to 1 h) and/or titration of formaldehyde concentration (0.1–1%) be performed for each new factor studied, since excessive cross-linking can reduce cell lysis efficiency, introduce sequence biases, increase background, and cause a reduction in the availability of epitopes/changes in epitopes for antibody binding. Conversely, suboptimal cross-linking may lead to incomplete fixation and, subsequently, fewer immunoprecipitated RNA–protein complexes. It is also important to note that indirect protein–RNA interactions may require increased cross-linking times and/or formaldehyde concentrations.
4. It is important to simultaneously process a sample from cells not treated with formaldehyde, since many RNA-binding proteins have the propensity to gain interactions in cell extract that do not occur in cells (8). Thus, such a control will verify that the interaction is occurring in vivo and is not a post-lysis artifact.
5. Following the TBS/PBS washes, the cells can be stored on ice for a few hours while other samples are collected so that all samples can be processed simultaneously.
6. At this stage, the cells can be frozen in liquid nitrogen or a dry ice/ethanol bath and stored indefinitely at −80°C.
7. While we use a FastPrep FP120A Instrument (Q-BIOgene) for yeast disruption because of its ease of use, reproducibility and capacity to process 12 samples simultaneously, alternative methods for cell lysis (e.g., a multivortexer with acid-washed glass beads) are suitable.
8. It is important to shear nucleic acids to an average length of 400–500 bases, since longer fragments increase background signals. If a different sonication device is used, determine the conditions necessary to achieve the desired level of nucleic acid shearing empirically. Moreover, optimization of the sonication step is likely to be required for different mammalian cell types.
9. The protocol outlined here for HEK293 cells involves separation of nuclei from the other cellular material. It is therefore designed for the immunoprecipitiation of nuclear proteins and subsequent characterization of nuclear protein–RNA complexes. For characterization of cytoplasmic protein–RNA complexes, preparation of whole cell or cytoplasmic lysates is required; protocols describing this are available (1, 7).
10. Because the initial DNase treatment (see Subheading 3.3, step 3) often does not result in complete removal of DNA, it must be complemented with a subsequent Turbo DNase treatment following reversal of cross-links (see Subheading 3.4, step 7). We have found that, for RIP from yeast cells, this dual DNase

treatment is necessary to prevent significant levels of DNA contamination and resultant decreased RIP signal:input ratios. For human cells, the Turbo DNase treatment alone is often sufficient.

11. These will be processed along with the RIPs in Subheading 3.4, step 1.
12. The actual amount of antibody required must be determined empirically and can vary considerably. Moreover, the immunoprecipitation conditions can be modified (e.g., time, temperature, salt concentration, presence of detergents) if desired or necessary.

    As in ChIP, antibody quality in RIP is extremely important. Polyclonal antibodies are preferable for both techniques, since different antibodies in the polyclonal population will reduce the probability that all specific epitopes will be masked by the process of cross-linking. If the protein target of interest is tagged, use of a well-defined antibody against this tag often improves results. In this case, the ideal experimental set-up would be to use the same antibody in two extracts that only differ in whether or not the target protein carries an epitope tag (Fig. 2b), as opposed to the "specific antibody" versus "nonspecific antibody" strategy described in this protocol.
13. For many polyclonal antibodies, the more stringent washes in Subheading 3.3, steps 10–11 result in a cleaner signal, whereas less stringent washes frequently lead to an unacceptably high background. For some antibodies (e.g., monoclonal against peptide epitopes) repeated washes with FA lysis buffer, a less stringent approach, might be more appropriate. High background signals in control samples may be overcome by increasing the salt

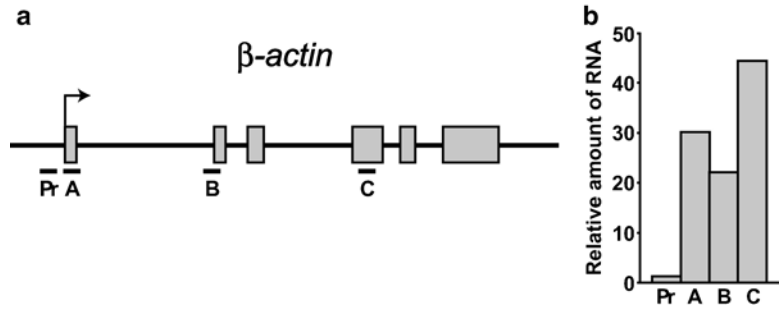


Fig. 2. Analysis of complex formation between hnRNP A1 and β-actin mRNA in HEK 293T cells by RIP. (**a**) The human β-actin gene and the position of the primers used for quantitation of immunoprecipitated RNA (Pr, promoter; A, exon 1; B, overlapping intron 1/exon 2; C, exon 4). (**b**) RIP of β-actin mRNA from cells expressing hnRNP A1-Flag. hnRNP A1-Flag was immunoprecipitated with an antibody against the Flag tag (M2; Sigma). Signals were corrected for input and are expressed as fold increase over the negative control (untagged cells).

and/or detergent concentrations in the IP and/or wash steps, or increasing the number and duration of the washes.

14. Incubation at 42°C allows for efficient proteinase K digestion of cross-linked polypeptides, while the 65°C incubation reverses the formaldehyde cross-links.
15. It is crucial to include duplicate reactions lacking reverse transcriptase to ensure that the final product results from RNA rather than contaminating DNA.
16. The use of appropriate control primers targeted to a predicted non-binding site, in addition to the experimental primers, is required to assess levels of contaminating RNA. For example, primers immediately up- or downstream of transcriptional start and stop sites are often useful controls (Fig. 2). Alternatively, if mRNA-binding proteins are studied, the control primers might be designed to detect specific tRNAs or rRNAs. Similarly, if cotranscriptional association of proteins with pre-mRNA is studied, primers across an intron–exon junction can be used. Because splicing occurs cotranscriptionally (and introns are exclusively nuclear), this makes it possible to study nuclear-specific RNA interactions (9).
17. These PCR conditions are generally appropriate for most reactions, although the annealing temperature might have to be adjusted if the melting temperature of the primers is substantially above or below 55°C.

References

1. Niranjanakumari, S., Lasda, E., Brazas, R., and Garcia-Blanco, M. A. (2002) Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions in vivo, *Methods 26*, 182–190.
2. Hieronymus, H., and Silver, P. A. (2003) Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery, *Nat Genet 33*, 155–161.
3. Hurt, E., Luo, M. J., Rother, S., Reed, R., and Strasser, K. (2004) Cotranscriptional recruitment of the serine-arginine-rich (SR)-like proteins Gbp2 and Hrb1 to nascent mRNA via the TREX complex, *Proc Natl Acad Sci USA 101*, 1858–1862.
4. Keene, J. D., Komisarow, J. M., and Friedersdorf, M. B. (2006) RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts, *Nat Protoc 1*, 302–307.
5. Tenenbaum, S. A., Lager, P. J., Carson, C. C., and Keene, J. D. (2002) Ribonomics: identifying mRNA subsets in mRNP complexes using antibodies to RNA-binding proteins and genomic arrays, *Methods 26*, 191–198.
6. Buratowski, S. (2009) Progression through the RNA polymerase II CTD cycle, *Mol Cell 36*, 541–546.
7. Conrad, N. K. (2008) Chapter 15. Co-immunoprecipitation techniques for assessing RNA-protein interactions in vivo, *Methods Enzymol 449*, 317–342.
8. Mili, S., and Steitz, J. A. (2004) Evidence for reassociation of RNA-binding proteins after cell lysis: implications for the interpretation of immunoprecipitation analyses, *RNA 10*, 1692–1694.
9. Gilbert, C., Kristjuhan, A., Winkler, G. S., and Svejstrup, J. Q. (2004) Elongator interactions with nascent mRNA revealed by RNA immunoprecipitation, *Mol Cell 14*, 457–464.
10. Selth, L. A., Gilbert, C., and Svejstrup, J. Q. (2009) RNA immunoprecipitation to determine RNA-protein associations in vivo, *CSH Protoc 2009*, pdb prot5234.

Chapter 20

Using ChIP-Seq Technology to Generate High-Resolution Profiles of Histone Modifications

Henriette O'Geen, Lorigail Echipare, and Peggy J. Farnham

Abstract

The dynamic modification of DNA and histones plays a key role in transcriptional regulation through altering the packaging of DNA and modifying the nucleosome surface. These chromatin states, also referred to as the epigenome, are distinctive for different tissues, developmental stages, and disease states and can also be altered by environmental influences. New technologies allow the genome-wide visualization of the information encoded in the epigenome. For example, the chromatin immunoprecipitation (ChIP) assay allows investigators to characterize DNA–protein interactions in vivo. ChIP followed by hybridization to microarrays (ChIP-chip) or by high-throughput sequencing (ChIP-seq) are both powerful tools to identify genome-wide profiles of transcription factors, histone modifications, DNA methylation, and nucleosome positioning. ChIP-seq technology, which can now interrogate the entire human genome at high resolution with only one lane of sequencing, has recently surpassed ChIP-chip technology for epigenomic analyses. Importantly, for the study of primary cells and tissues, epigenetic profiles can be generated using as little as 1 μg of chromatin. In this chapter, we describe in detail the steps involved in performing ChIP assays (with a focus on characterizing histone modifications in primary cells) either manually or using the IP-Star ChIP robot, followed by a detailed protocol to prepare successful libraries for Illumina sequencing. Critical quality control checkpoints are discussed. Although not a focus of this chapter, we also point the reader to several methods by which massive ChIP-seq data sets can be analyzed to extract the tremendous information contained within.

Key words: Chromatin immunoprecipitation, ChIP-seq, Next generation sequencing, Epigenomics, Histone modifications, IP-Star, ChIP robot

1. Introduction

Although the genetic information encoded in our DNA plays a major role in specifying our individual phenotypes, it is becoming increasingly clear that epigenetic information is also an important contributor to our mental and physical attributes (1–8). Our epigenome is defined as methylated DNA and modified histone

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_20,

proteins (around which both methylated and unmethylated DNA are wrapped). DNA methylation and histone modifications undergo global changes during transitions in developmental states and in diseases such as cancer and therefore are major contributors to the dynamic nature of chromatin. Histone modifications such as acetylation (e.g., acetylation of lysine 9 of histone H3 which is called H3K9Ac) are typically associated with open and accessible chromatin regions, while histone methylation can be associated with either open or compacted (also referred to as heterochromatic) chromatin regions, depending on the specific histone amino acid that is methylated (9–12). For example, mono- or trimethylation of lysine 4 on histone H3 (H3K4me1 or H3K4me3) and trimethylation of histone H3 on lysine 36 (H3K36me3) are associated with open chromatin. However, each of these marks represents a unique category of open chromatin, with H3K4me3 marking gene promoter regions, H3K4me1 marking transcriptional enhancers, and H3K36me3 marking transcribed regions of the genome. In contrast, trimethylation of lysines 9 and 27 on histone H3 (H3K9me3 and H3K27me3, respectively) is associated with compacted chromatin regions resulting in repression of target genes. Although both of these modifications mark repressive chromatin, H3K9me3 and H3K27me3 regulate distinct sets of target genes; H3K27me3 predominantly represses homeobox transcription factors and H3K9me3 predominantly targets the genes of zinc finger transcription factors (13). Knowing the genome-wide pattern of single histone modifications, such as the six marks described above, provides a great deal of information about cell identity and disease state (10, 14–17). Accordingly, these six marks have been selected to provide "roadmaps" of the epigenomic profiles of primary cells by the Roadmap Epigenome Mapping Centers (http://www.roadmapepigenomics.org/). However, it is also becoming increasingly clear that different combinations of histone marks can provide even more detailed information. For example, the presence of both the open chromatin mark H3K4me3 and the compacted chromatin mark H3K9me3 at a promoter can identify imprinted genes (18). We are just beginning to understand the interrelationships between specific histone modifications and transcriptional regulation and more insights will certainly be forthcoming with the analysis of more and more epigenetic profiles.

Currently, the method of choice to study the epigenome is the chromatin immunoprecipitation (ChIP) assay. To perform a ChIP assay, proteins such as histones or transcription factors are covalently crosslinked to their genomic DNA substrates in living cells. This provides an opportunity to take a snapshot of histone or other protein–DNA interactions in a given cell type, in cells taken at different developmental stages, or in cells altered by disease. After isolation and fragmentation of chromatin, the protein–DNA complexes are captured using antibodies specific to the histone or transcription factor of interest. After reversal of crosslinks, the

ChIP DNA is then purified and analyzed by either hybridization to microarrays (ChIP-chip) or by high-throughput sequencing (ChIP-seq). While multiple DNA microarrays are needed to cover the entire human genome, resulting in high costs for comprehensive studies, ChIP-seq offers the possibility to interrogate the entire genome in one sequencing run. Therefore, ChIP-seq has generally replaced ChIP-chip for comprehensive epigenomic studies. To date there are four high-throughput sequencing platforms that have been used for ChIP-seq (see ref. 19 for review): (1) 454 Genome Sequencer FLX from Roche (20, 21), (2) Genome Analyzer GA2 from Illumina (9, 18, 22), (3) Sequencing by Oligo Ligation/Detection (SOLiD) from Applied Biosystems (23, 24), and (4) true Single Molecular Sequencing (tSMS) by Helicos (19, 25). Unlike the first three methods that use an amplification step, the Helicos platform sequences unamplified DNA. Another difference among the platforms include read length; the Illumina, SOLiD, and Helicos platforms produce short read lengths of on average 35–50 bp whereas the 454 platform produces longer reads of 200–400 bp. Most ChIP-seq studies to date have used the Illumina sequencing platform (sometimes called Solexa sequencing). Therefore, this chapter describes in detail the steps needed to prepare ChIP samples and libraries for high-throughput sequencing using the Illumina GA2 platform and includes descriptions of quality control steps necessary to ensure a successful ChIP-seq experiment (see Fig. 1).

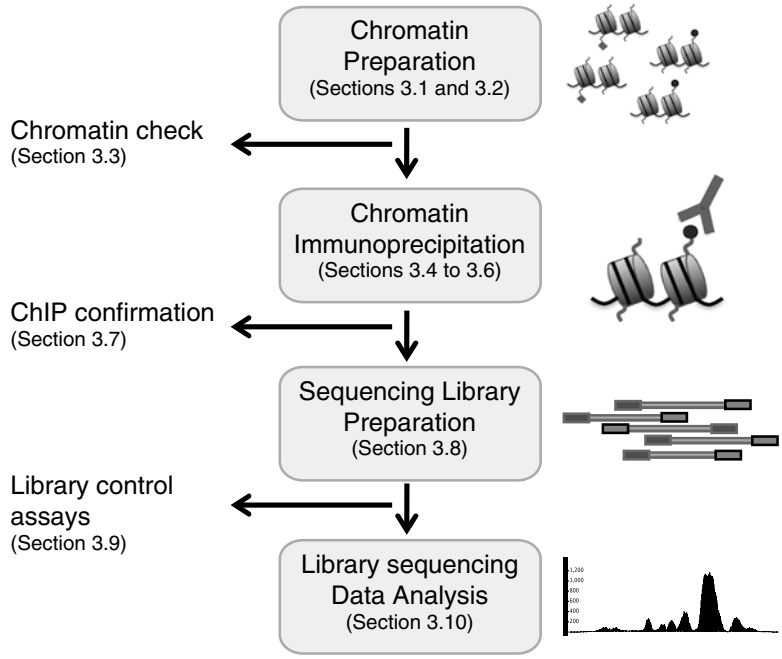


Fig. 1. ChIP-seq diagram summarizing the main experimental steps and quality check points.

2. Materials

2.1. Crosslinking Reagents

1. Crosslinking reagent: formaldehyde solution (37% w/w).
2. Stopping reagent: glycine (electrophoresis grade).
3. Wash solution: phosphate-buffered saline (PBS).

2.2. Chromatin Preparation Reagents

1. Protease inhibitor stock solutions (store in small aliquots at –20°C): 10 mg/ml aprotinin (in water), 10 mg/ml leupeptin (in water), 100 mM PMSF (in isopropanol).
2. Cell lysis buffer (store at room temperature): 5 mM PIPES pH 8, 85 mM KCl. Add igepal fresh each time to give a final concentration of 1% (10 µl/ml). Warm buffer in 37°C water bath and vortex briefly to help mixing of igepal. After mixing has occurred, place buffer containing igepal on ice to allow solution to cool down and then add protease inhibitors [PMSF (10 µl/ml f.c.), aprotinin (1 µl/ml f.c.), and leupeptin (1 µl/ml f.c.)].
3. Nuclei lysis buffer (store at room temperature): 50 mM Tris–HCl pH 8, 10 mM EDTA, and 1% (w/v) SDS. Place buffer on ice right before use to avoid precipitation of SDS and add protease inhibitors [PMSF (10 µl/ml f.c.), aprotinin (1 µl/ml f.c.), and leupeptin (1 µl/ml f.c.)] just prior to use.
4. Bioruptor UCD-200 (Diagenode) or equivalent is used for sonication.

2.3. Chromatin Check Reagents

1. Elution buffer (store at room temperature): 50 mM $NaHCO_3$, 1% (w/v) SDS.
2. DNase-free RNase A (Fermentas; 10 mg/ml).
3. QIAquick PCR purification kit (QIAGEN).
4. A NanoDrop 1000 is used to determine the concentration of double-stranded DNA samples. This instrument is invaluable for measuring low DNA concentrations (e.g., 10 ng/µl) and for small sample volumes (as little as 1 µl of sample can be measured).

2.4. Chromatin Immunoprecipitation Reagents

Note, the reagents listed in steps 2 and 3 of Subheading 2.4 are required for manual ChIP assays; for automated ChIP assays, use the Auto ChIP kit for the IP-Star (Diagenode).

1. ChIP grade antibodies specific for the following six histone modifications:
   (a) H3K4me3: Anti-Tri-Methyl-Histone H3 (Lys4) (C42D8) rabbit monoclonal antibody (CST #9751S).
   (b) H3K9ac: Anti-acetyl-Histone H3 (Lys9) rabbit antibody (Millipore #07-352).

(c) H3K27me3: Anti-Tri-Methyl-Histone H3 (Lys27) (C36B11) rabbit monoclonal antibody (CST #9733S).

(d) H3K9me3: Anti-Tri-Methyl-Histone H3 (Lys9) rabbit antibody (CST #9754S).

(e) H3K36me3: Anti-Tri-Methyl-Histone H3 (Lys36) rabbit antibody (CST #9763S).

(f) H3K4me1: Anti-Mono-Methyl-Histone H3 (Lys4) rabbit antibody (Diagenode #pAb-037-050).

2. Protease inhibitor stock solutions (store in small aliquots at –20°C): 10 mg/ml aprotinin (in water), 10 mg/ml leupeptin (in water), 100 mM PMSF (in isopropanol).
3. IP dilution buffer (store at 4°C): 50 mM Tris–HCl pH 7.4, 150 mM NaCl, 1% (v/v) igepal, 0.25% (w/v) deoxycholic acid, 1 mM EDTA pH 8. Add protease inhibitors [PMSF (10 µl/ml f.c.), aprotinin (1 µl/ml f.c.), and leupeptin (1 µl/ml f.c.)] just prior to use. This buffer is used to adjust the salt and SDS concentrations for the immuno-precipitation step.

2.5. Capture the Antibody/Chromatin Complexes and to Reverse Crosslinks

Note, the reagents listed below are required for manual ChIP assays; for automated ChIP assays, use the Auto ChIP kit for the IP-Star (Diagenode).

1. Magnetic protein G beads (Cell Signaling Technology) and magnetic rack. Do not use magnetic beads that have been blocked with foreign DNA, such as herring sperm or salmon sperm DNA. This may result in sequencing of the blocking DNA, resulting in lower quality ChIP-seq data. Although protein G binds antibodies from a variety of species (rabbit, mouse, goat, etc.) with high affinity, magnetic protein A beads can be used if desired.
2. IP wash buffer 1 (store at 4°C): same as IP dilution buffer, but without protease inhibitors.
3. IP wash buffer 2 (store at room temperature): 100 mM Tris–HCl pH 9, 500 mM LiCl, 1% (v/v) igepal, 1% (w/v) deoxycholic acid.
4. IP wash buffer 3 (store at room temperature): 100 mM Tris–HCl pH 9, 500 mM LiCl, 150 mM NaCl, 1% (v/v) igepal, 1% (w/v) deoxycholic acid.
5. Elution buffer (store at room temperature): 50 mM $NaHCO_3$, 1% (w/v) SDS.
6. 5 M NaCl.

2.6. DNA Purification

1. DNase-free RNase A (Fermentas; 10 mg/ml).
2. QIAquick PCR purification kit (QIAGEN).

2.7. ChIP Confirmation

1. SYBR-Green qPCR mix, such as SYBR Green JumpStart Taq ReadyMix (SIGMA).
2. Positive and negative control primer sets (see Note 1).

2.8. Sequencing Library Preparation

1. End-It DNA END Repair Kit (Epicentre).
2. Klenow (3′–5′ exo minus) (NEB; 5,000 U/ml).
3. 100 mM dATP.
4. LigaFast DNA ligase (Promega; 3 U/μl).
5. Oligo-only kits for single end or paired end read sequencing are available from Illumina (#FC-102-1003 and PE-102-1003, respectively). Alternatively, adapter oligos and PCR primers compatible with Illumina sequencing can be purchased elsewhere; HPLC purification is recommended. The paired end DNA oligonucleotides are more universal since the resulting library can be sequenced with either single end or paired end sequencing primers. The following stock solutions are prepared: 15 μM Paired End Adapter Oligo mix, 25 μM Paired End PCR primer 1.01, and 25 μM Paired End PCR primer 2.01.

   Paired End DNA oligonucleotide sequences (Oligonucleotide sequences© 2006 and 2008 Illumina, Inc. All rights reserved).

   PE Adapters

   5′ P-GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG

   5′ ACACTCTTTCCCTACACGACGCTCTTCCGATCT

   PE PCR Primer 1.01

   5′ AATGATACGGCGACCACCGAGATCTACACTCTTTCC CTACACGACGCTCTTCCGATCT

   PE PCR Primer 2.01

   5′ CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATT CCTGCTGAACCGCTCTTCCGATCT
6. 2% Agarose precast E-Gel® (Invitrogen #G501802), loading dye such as TrackIt™ Cyan/Orange Loading Buffer (Invitrogen #10482-028), 50 or 100 bp DNA markers (e.g., TrackIt™ 50 bp DNA Ladder, Invitrogen #10488-043).
7. QIAquick Gel Extraction Kit (QIAGEN #28704).
8. Phusion DNA polymerase (NEB #F531).
9. QIAquick PCR purification kit (QIAGEN #28104) and MinElute PCR purification kit (QIAGEN #28004).
10. Agencourt AMPure system (Beckman Coulter Genomics #A29152).

2.9. Library Control Assay

1. DNA High Sensitivity Kit (Agilent Cat# 5067–4626) for use with the Agilent 2100 Bioanalyzer.

2. SYBR-Green qPCR mix, such as SYBR Green JumpStart Taq ReadyMix (SIGMA #S4438).
3. Positive and negative control primer sets (see Note 1).

2.10. Sequencing Reagents These reagents will be supplied by the sequencing facility.

3. Methods

3.1. Preparation of Crosslinked Cells

1. Cell cultures should be healthy and not density-arrested prior to crosslinking. For primary cells or tissues, the samples can be snap frozen in liquid nitrogen immediately after collection or snap frozen after crosslinking. The amount of cells needed for ChIP-seq will vary depending on the antibody used and the abundance of the histone mark of interest. In general between 100,000 and 500,000 cells are used per histone antibody. In a chemical hood, prepare 1% formaldehyde solution in PBS and add directly to frozen cell pellet. Resuspend cell pellet by pipetting up and down. Alternatively for cultured cells, add formaldehyde (37% stock) directly to tissue culture media to a final concentration of 1%.
2. Rotate primary cells in a tightly closed tube or rock cultured cells on a shaking platform for 10 min at room temperature. Do not crosslink for longer periods since this may cause cells to form aggregates that do not sonicate efficiently.
3. Stop crosslinking reaction by adding glycine to a final concentration of 0.125 M. We use a 10× (1.25 M) stock solution. Continue to rotate/rock at room temperature for 5 min.
4. For primary cells or other cells crosslinked in suspension, centrifuge cells at 430 rcf for 5 min at 4°C, discard the solution, wash the pellet twice with ice-cold 1× PBS (mix by pipetting, pellet cells by centrifugation at 430 rcf for 5 min at 4°C and discard wash solution). For adherent cells, pour off media and rinse plates twice with ice-cold 1× PBS and pour off wash solution. Using a cell scraper, transfer adherent cells from the culture dish to a 15-ml conical tube on ice. Centrifuge the crosslinked cells at 430 rcf for 5 min at 4°C. It is important to carefully aspirate supernatants so as to not lose cells. Note: media containing formaldehyde should be treated as hazardous waste.
5. Cells may be used immediately for a chromatin preparation or snap frozen in liquid nitrogen and stored at –80°C.

3.2. Preparation of Chromatin

1. If using frozen crosslinked cells thaw them on ice; keep all cells and chromatin samples on ice at all times. Prepare the cell lysis buffer (1 ml cell lysis buffer per 1×10^7 cells): add Igepal

(10 µl per ml cell lysis buffer, agitate at 37°C to dissolve, cool on ice), then add protease inhibitors [PMSF (10 µl/ml), aprotinin (1 µl/ml), and leupeptin (1 µl/ml)]. Resuspend cell pellet in freshly prepared ice-cold cell lysis buffer by pipetting. The final volume of cell lysis buffer should be sufficient so that there are no clumps of cells. Incubate on ice for 15 min.

2. Homogenize cells using a glass dounce homogenizer (type B) to break open the cells and release nuclei. Homogenize cells on ice with 20 strokes. Omit this step when processing less than one million cells.
3. Centrifuge cells at 430 rcf for 5 min at 4°C.
4. Discard the supernatant and resuspend the nuclear pellet in nuclei lysis (NL) buffer plus protease inhibitors. Be careful not to use too much NL buffer as it may lead to dilute chromatin; we suggest ~20 µl/10^6 cells. Incubate on ice for 30 min.
5. An optional flash-freezing step may help break open nuclei more efficiently. This step is critical if the homogenizing in step 2 is omitted. After incubation of nuclei in NL buffer for 30 min, flash freeze samples in liquid nitrogen, thaw at room temperature (once thawed, immediately transfer to ice; do not allow samples to warm up to room temperature), and proceed to sonication.
6. Sonicate cells in a coldroom and/or on ice to achieve average chromatin length of 200–500 bp (see Note 2). Larger chromatin fragments can negatively influence data quality and can lead to failure of the ChIP-seq experiment. Therefore, before processing large quantities of cells, sonication conditions should be optimized for each cell type (see Note 3).
7. Transfer sonicated samples into an Eppendorf tube and centrifuge using a microcentrifuge at 10,000 rcf for 10 min at 4°C. Carefully transfer the supernatant (sonicated chromatin) to a new tube while avoiding cell debris. Keep sonicated chromatin at 4°C while performing quantification and determining chromatin size; then proceed with the ChIP assays (see Note 4).

3.3. Determination of Chromatin Size and Concentration

1. Take an aliquot of chromatin sample from Subheading 3.2, step 7 prepared above. A typical size determination uses chromatin from 100,000 to 200,000 cells (see Note 5).
2. Add ChIP elution buffer to a total volume of 100 µl and then 12 µl 5 M NaCl to give a final salt concentration of 0.54 M. Boil samples in a water bath for 20 min to reverse crosslinks.
3. Allow sample to cool down, add 1 µl DNase-free RNase (10 mg/ml), and incubate for 20 min at 37°C. This step is important because the presence of RNA results in false estimation of chromatin size.

4. Purify DNA using a PCR purification kit, elute DNA in 25 μl water. Measure chromatin concentration by NanoDrop and calculate the chromatin yield (see Note 6).
5. Run remaining chromatin on a 1.2% agarose gel to visualize average size. If the chromatin is larger than ~600 bp, adjust the sonication conditions by adding more pulses and repeat steps 1–5.

3.4. Chromatin Immunoprecipitation

ChIP is usually done within the same day as the chromatin preparation to avoid any concern about the quality of chromatin (see Note 4). The steps detailed in Subheadings 3.4 and 3.5 are for manual ChIP assays. However, ChIP reactions can be automated using a ChIP robot (e.g., the IP-Star from Diagenode) (see Note 7).

1. Measure volume of chromatin and divide chromatin as needed. The amount of chromatin needed for each ChIP reaction varies depending on the histone modification. For histone marks covering a small portion of the genome displaying sharp peaks, such as H3K4me3, we typically use 1 μg chromatin. Based on our experience, we prefer to use 5 μg chromatin for spreading histone marks such as H3K9me3 or H3K36me3 that cover large portions of the genome.
2. Optional: An IgG negative control sample can be included along with experimental antibodies. However, oftentimes chromatin cannot be spared for a control ChIP when using small amounts of primary cells. In this case, one can rely on regions bound by the activating histone marks as negative controls for the repressive histone marks and vice versa (Fig. 2a).

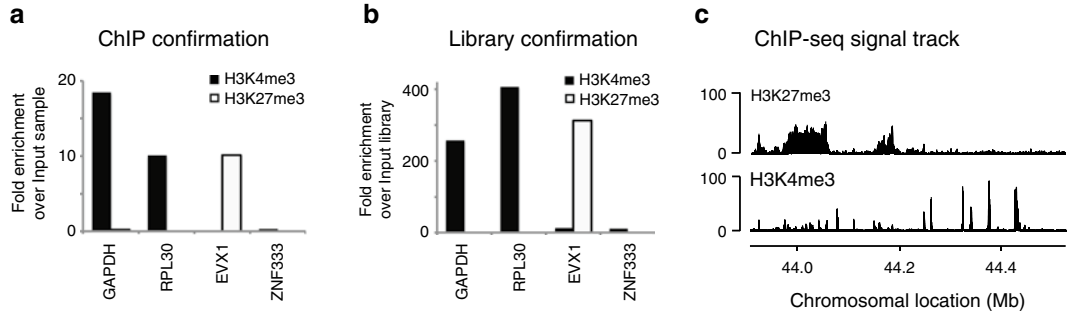


Fig. 2. ChIP-seq experiments using the IP-Star ChIP robot. ChIP assays using antibodies specific for H3K4me3 and H3K27me3 were performed using 1 μg chromatin from Ntera2 cells and the IP-Star ChIP robot (Diagenode). Libraries were prepared as outlined in Subheading 3.8. Quantitative PCR confirms specific enrichment over input in (**a**) the ChIP samples and (**b**) the ChIP-seq libraries. *Black bars* represent H3K4me3 enrichment and *white bars* represent H3K27me3 enrichment. Primer sets used are shown on the *x*-axis. GAPDH and RPL30 are positive control primer sets for H3K4me3 and negative control primer sets for H3K27me3. EVX1 was used as a positive primer set for H3K27me3 and a negative primer set for H3K4me3. The ZNF333 primer set was used as a negative control for both histone marks. (**c**) ChIP-seq binding patterns for H3K4me3 and H3K27me3 obtained from samples prepared using the IP-Star are shown for a region on chromosome 17 encompassing the HOXB gene cluster. The peak height is plotted along the *y*-axis; chromosomal coordinates (hg18 coordinates) are shown on the *x*-axis. Samples were sequenced at the DNA Technologies Core at UC Davis (http://genomecenter.ucdavis.edu/dna_technologies/).

3. Save volume corresponding to 500 ng of chromatin to prepare an input sample (often also referred to as total DNA). Store the reserved amount at −20°C until the next day and then reverse the crosslinks in the input chromatin at the same time as the crosslinks in the ChIP samples are reversed.
4. Dilute chromatin fivefold with ice-cold IP Dilution buffer (1 volume chromatin and 4 volumes IP Dilution buffer) containing protease inhibitors.
5. Add an antibody specific to the histone mark of interest to capture the protein/chromatin complexes (see Note 8). Although antibody amounts are determined empirically, we typically use between 1 and 5 μg antibody per ChIP assay. Always record catalogue number and lot number of antibodies used.
6. Incubate 8–16 h on a rotating platform at 4°C.

3.5. Capture of Antibody/Chromatin Complexes and Reversal of Crosslinks

Step 1 is carried out at 4°C, whereas steps 2–9 are carried out at room temperature.

1. Add 15 μl magnetic protein G beads to each ChIP sample ranging from 1 to 5 μg chromatin starting material and incubate on a rotating platform for 2 h at 4°C.
2. At room temperature, allow beads to settle for 1 min in a magnetic separation rack. Carefully remove the supernatant without disturbing magnetic beads.
3. Wash magnetic beads two times with IP Dilution buffer (take tubes out of magnetic rack and mix by pipetting). Efficient washing is critical to reduce background. Avoid cross contamination of samples and loss of magnetic beads.
4. Wash magnetic beads two times with IP wash buffer 2 (take tubes out of magnetic rack and mix by pipetting). Discard all wash solution after final wash.
5. Wash once with the higher stringency IP wash buffer 3. Discard wash solutions.
6. Elute antibody/chromatin complexes by adding 100 μl elution buffer per ChIP sample. Shake samples on vortexer for 30 min.
7. Allow beads to settle for 1 min in a magnetic separation rack. Carefully transfer the supernatant containing antibody/chromatin complexes to a siliconized tube.
8. Add 12 μl of 5 M NaCl per 100 μl elution buffer mix to give a final concentration of 0.54 M NaCl.
9. At this point, thaw the input sample from the previous day (Subheading 3.4, step 3). Dilute 1 volume input sample with 4 volumes ChIP elution buffer (e.g., add 80 μl ChIP elution

buffer to 20 µl input sample). Add 12 µl of 5 M NaCl per 100 µl elution buffer mix.

10. Incubate all samples in a 67°C water bath overnight to reverse formaldehyde crosslinks.

3.6. DNA Purification

1. Allow samples to cool, add 1 µl of RNaseA; incubate at 37°C for 20 min.
2. Purify DNA with a PCR clean up kit, one column per sample. Elute each sample with 40 µl EB buffer.
3. Assess ChIP enrichments by quantitative PCR (qPCR) before proceeding to preparation of Solexa libraries.

3.7. ChIP Confirmation

Enrichment of histone marks in the ChIP samples are determined by quantitative real-time PCR (qPCR). The input sample is diluted with EB to give a final concentration of 2 ng/µl and serves as a reference. Prepare a master reaction mix for each library with triplicate reactions per primer set. Add extra reagents for 10% of the total number of reagents to account for loss of volume. Add 14 µl of reaction mix to each PCR reaction well. Add 2 µl primer mix to each well.

Recipe for one reaction:

1 µl	Undiluted ChIP sample or diluted Input sample (2 ng/µl)
4.5 µl	Nuclease-free H20
7.5 µl	2× SYBR Green mix (containing polymerase)
2 µl	5 µM target primer mix (containing both Forward and Reverse primers)
15 µl	Total reaction volume

Amplify using the following PCR protocol:

3 min at 95°C.

40 cycles of 30 s at 95°C, then 30 s at 60°C.

Include a 70–95°C melting curve at the end of the qPCR program, reading all points or every 0.2°C.

Analyze the qPCR results by first manually determining the cycle threshold for each reaction across the plate within the linear range of the amplification curve. Calculate the average cycle threshold for each triplicate reaction of each sample. The relative DNA amount is then calculated for any given primer set as 2 to the power of the cycle threshold (cT) difference between input chromatin and ChIP samples, where cT is the average value.

$$\text{Relative DNA amount} = 2^{(\text{cT input} - \text{cT sample})}$$

The enrichment is then calculated by comparing relative enrichment for the target and a negative control. This is accomplished by dividing the relative DNA amount of each sample for a target primer

set by the corresponding value for a negative control primer set. The resulting quotient represents the fold enrichment. The fold enrichment will vary depending on the histone marks as well as the location of the chosen target primer set (see Fig. 2a for an example).

3.8. Preparation of the Sequencing Library

The library protocol is based on the Illumina Sample Preparation Kit for Genomic DNA with some modifications. This protocol describes the preparation of libraries from ChIP DNA that are compatible with the Illumina sequencing platforms. Libraries are prepared from the ChIP sample as well as matching input DNA from the same cell type (see Note 9). Boiled chromatin samples should not be used since single-stranded DNA will lower the library preparation efficiency.

3.8.1. End-Repair

End-repair is performed using the "End-It DNA End Repair Kit" from Epicentre. This step ensures that all DNA fragments are converted to 5′-phosphorylated blunt-ended DNA. The entire ChIP DNA volume from Subheading 3.6 is used. Combine and mix the following components in a siliconized Eppendorf tube:

1–34 µl	ChIP DNA from Subheading 3.6 or 200 ng input DNA
5 µl	10× End-Repair Buffer
5 µl	10 mM ATP
5 µl	2.5 mM dNTP Mix
1 µl	End-Repair Enzyme Mix
50 µl	Total reaction volume

Incubate at room temperature for 45 min and purify DNA using a PCR purification kit (such as QIAquick PCR purification kit), elute in 34 µl EB buffer.

3.8.2. Addition of an "A" Base to the 3′ End of DNA Fragments

Before starting, prepare stocks of 1 mM dATP from 100 mM dATP stock (e.g., add 5 µl of 100 mM dATP to 495 µl sterile RNase DNase free water), and store aliquots of 11 µl at −20°C. Once thawed, 1 mM dATP solution should not be refrozen. Combine and mix the following components in PCR tubes:

34 µl	DNA from Subheading 3.8.1
5 µl	10× Klenow buffer
10 µl	1 mM dATP
1 µl	Klenow fragment (3′–5′ exo minus)
50 µl	Total reaction volume

Incubate for exactly 30 min at 37°C using a PCR machine. Purify DNA using a PCR purification kit (such as MinElute PCR purification kit), elute in 12 µl EB buffer.

3.8.3. Ligation of Adapters to DNA Fragments

The Paired End Adapter Oligo mix is diluted 1:10 in water before use to adjust for the small quantity of ChIP DNA.

Combine and mix the following components in a siliconized Eppendorf tube:

12 µl	DNA from Subheading 3.8.2
15 µl	2× DNA ligase buffer
1 µl	1:10 dilution of PE Adapter Oligo mix
2 µl	LigaFast DNA ligase
30 µl	Total reaction volume

Incubate for 15 min at room temperature. Purify DNA using a PCR purification kit (such as QIAquick PCR purification kit), elute in 19 µl EB buffer.

3.8.4. Size Selection of DNA Fragments

Size selection of the sample ensures removal of unused adapters and selection of proper fragment size for amplification and sequencing (see Note 10). We use precast agarose gels to minimize risk of contamination.

1. Dilute 10 µl of 6× Cyan/Orange with 50 µl EB buffer to obtain 1× Cyan/Orange buffer dye. Add 1 µl of 1× Cyan/Orange buffer dye to eluted DNA from Subheading 3.8.3.
2. Prerun e-gels according to manufacturer's instructions.
3. Load 20 µl of appropriately diluted 50 or 100 bp DNA ladder (500 ng ladder per well) in one well per gel. Skip at least one well between marker and samples to avoid contamination.
4. Load 20 µl DNA with dye in each well. Skip at least one well between samples to avoid contamination.
5. Load 20 µl of EB buffer in each of the empty wells.
6. Run the gel until desired size separation is achieved (30 min for e-gels).
7. Take a gel picture to visualize sample.
8. Size select samples: using a fresh razor blade, excise two gel pieces of 200–400 bp and 400–600 bp (see Note 11). The DNA concentrations may be too low for the sample to be visible by eye; in this case, use markers as a guide. Keep exposure to UV light to a minimum to reduce DNA damage. Alternatively, a non-UV transilluminator can be used. Gel slizes can be stored at –20°C.
9. Solubilize gels at room temperature using the QIAgen gel extraction buffer by shaking on the vortexer for 30 min.
10. Purify on one QIAquick column using QIAquick PCR Purification Kit. Elute in 25 µl EB buffer.

3.8.5. Amplification of Adapter-Modified DNA Fragments and Gel Purification

Because we make libraries from both the small (200–400 bp) and the big (400–600 bp) size selected DNA fragments from Subheading 3.8.4, we prepare two amplification reactions (and two libraries) per ChIP sample. We also prepare a 200–400-bp and a 400–600-bp size-selected input library. Therefore, if all six histone modifications are analyzed, there will be 14 amplification reactions and 14 libraries (12 ChIP libraries and 2 input libraries). For these reactions, dilute Paired End primers 1:4 with sterile water.

Combine and mix the following components in PCR tubes:

23 µl	DNA from step 2
25 µl	2× Phusion DNA polymerase
1 µl	Paired End PCR primer 1.01 (1:4 dilution)
1 µl	Paired End PCR primer 2.01 (1:4 dilution)
50 µl	Total reaction volume

Amplify using the following PCR protocol:

30 s at 98°C.

15 cycles: 10 s at 98°C, 30 s at 65°C, 30 s at 72°C.

5 min at 72°C.

Hold at 4°C.

Perform 15 cycles of amplification.

3.8.6. Library Purification

Purify library samples from Subheading 3.8.5 using the Agencourt AMPure system following manufacturer's instructions.

1. Mix Ampure beads thoroughly before addition.
2. Add 90 µl of Ampure beads to each 50 µl DNA sample from step 5. Pipette several times to ensure proper mixing.
3. Use magnetic rack to separate bead–DNA complexes and discard the supernatant. Allow beads to settle, this may take several minutes.
4. Wash bead–DNA complexes using 70% ethanol without disturbing the beads. Leave the tube in the magnetic rack and add 200 µl of 70% ethanol. After 30 s, discard 70% ethanol by pipetting.
5. Repeat wash one more time for a total of two washes.
6. Allow beads to air dry for 10–20 min.
7. Add 30 µl of EB buffer and elute DNA on the vortexer for 30 min.
8. Place tubes back in magnetic rack to collect DNA, transfer liquid to siliconized Eppendorf tubes. Store libraries at –20°C.

3.9. Library Quality Control Assays

3.9.1. Library Quantification

The constructed libraries are assessed on a Bioanalyzer using the DNA High Sensitivity chip. The High Sensitivity DNA chip allows sizing and quantification of DNA samples in the single-digit pg/μl concentration range; the sequencing flow cell is loaded according to the library concentration (check with your sequencing facility for their requirements). The Bioanalyzer also allows visualization of possible adapter contamination; adapter dimers are visible as a sharp peak at approximately 120 bp. Libraries having large adapter dimer peaks should not be sequenced (contaminating adapter dimers can be removed by an additional gel size selection step as described in 3.8.4). For libraries that are not quantifiable, five additional cycles can be performed as in Subheading 3.8.5. However, over-amplification should be avoided to reduce the risk of PCR artifacts.

3.9.2. Library Enrichment Confirmation

To verify that a ChIP library has maintained a specific enrichment of target sites, perform qPCR on the ChIP-seq library using both positive targets and negative control primer pairs. The input library serves as a control to normalize the qPCR data to determine the relative enrichment of a given target (see Note 9).

A. Real-time quantitative PCR (qPCR)

Analyze the ChIP-seq sample as well as the appropriately sized input library for reference. Prepare a master reaction mix for each library with triplicate reactions per primer set. Add extra reagents for 10% of the total number of reagents to account for loss of volume. Add 15 μl of reaction mix to each PCR reaction well. Add 2 μl of primer mix to each well.

2 μl	2 ng library from Subheading 3.8.6
3.5 μl	Nuclease-free H_2O
7.5 μl	2× SYBR Green mix (containing polymerase)
2 μl	5 μM target primer mix
15 μl	Total reaction volume

Amplify using the following PCR protocol:

3 min at 95°C.

40 cycles of 30 s at 95°C then 30 s at 60°C.

Include a 70–95°C melting curve at the end of the qPCR program, reading all points or every 0.2°C.

B. Determine enrichment

Analyze library enrichments by qPCR as described in Subheading 3.7. The enrichment values vary depending on histone modification and placement of primers (see Fig. 2b, for an example). If enrichment is acceptable, then the sample can be provided to a sequencing facility (e.g., http://genomecenter.ucdavis.edu/dna_technologies/) for high-throughput sequencing; do not proceed with sequencing unless the positive targets are at least 20-fold enriched.

3.10. Library Sequencing and Data Analysis

After the sequencing is performed, the short tags (~25–50 nts) are mapped to the human genome, the tags that map uniquely to only one location in the genome are selected, the unique tags are extended to the average size of the library fragments (~200 nt), and then the extended fragments are grouped into consecutive bins running the length of each chromosome. The binned data can be visualized using the USCS browser (http://www.genome.ucsc.edu/) or the Affymetrix Integrated Genome Browser (http://www.affymetrix.com/partners_programs/-programs/developer/tools/download_igb.affx; see Fig. 2c). Target sites can be identified using a variety of peak calling methods (26–33). Most peak calling programs account for binding pattern in peak shape, but algorithms need to be adjusted for spreading histone marks covering larger regions of the genome. Sole-search is a peak-calling program that was initially developed to identify transcription factor binding sites, which typically display peaks (34). Applying this type of program to spreading histone marks such as H3K9me3 produces a large number of small peaks rather than identifying a binding region. Sole-search version 2 has been modified to address this issue and offers the choice of using the peak or histone method for peak calling (Blahnik et al., in preparation). We note that the histone method can be used to call peaks for both types of binding; it correctly identifies sharp peaks for H3K4me3, and also calls broad regions occupied by H3K9me3 (Fig. 3).

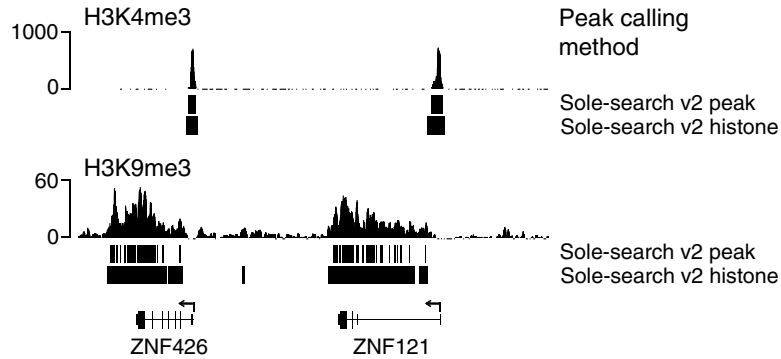


Fig. 3. Comparison of target identification for histone modifications having peak-like vs. spreading binding patterns. Libraries were prepared from ChIP samples using an antibody for H3K4me3 (K562 chromatin) and an antibody for H3K9me3 (Ntera2 chromatin) and were sequenced using the Illumina platform. As expected, sharp H3K4me3 peaks are observed proximal to the transcription start sites, whereas H3K9me3 covers larger chromatin regions; two zinc finger genes on chromosome 19 are shown to illustrate the different binding patterns (the number of tags is plotted along the *y*-axis and the hg19 chromosomal coordinates are shown on the *x*-axis). Targets were identified with Sole-search version 2 using the peak method as well as the histone method of the program; targets identified using both methods are depicted underneath each ChIP-seq signal track. The peak method (alpha value 0.01; FDR 0.001) works well for H3K4me3 but not for H3K9me3, whereas the histone method (alpha value 0.0001; FDR 0.01) can identify the sharp peaks in the H3K4me3 dataset as well as the broad binding regions of the H3K9me3 dataset.

The number of reads required to identify all sites bound by a particular histone mark depends on the characteristics of this mark. Similar to most site-specific transcription factors, the binding patterns of certain modified histones (H3K9ac and H3K4me3) can be identified as sharp peaks. For analysis of these modifications throughout the human genome, ten million sequenced tags should be sufficient. With the current Illumina GA2 platform, 20–40 million sequenced reads can be obtained routinely in a single sequencing run and 75% or more of the reads can be mapped to the human genome. Thus, one lane is usually sufficient to identify the regions bound by H3K9ac and H3K4me3. However, reads should ideally come from two independent ChIP samples, with the binding sites identified in each replicate having at least a 60% overlap. Therefore, a minimum of two lanes of sequencing (one lane each from two independent ChIP assays) are usually performed for each of the histone modifications. As noted above, certain histone modifications do not have a peak-like binding pattern, but instead spread over large chromatin regions. Spreading histone marks require more reads to identify significant enrichment over background. For spreading marks such as H3K36me3, H3K4me1, or H3K27me3, 20–40 million reads may be required. H3K9me3 is not only a spreading mark but it is also present on repetitive regions of the genome (such as centromeres). For this mark, up to 50% of the reads may map to more than one place in the human genome and are discarded from analysis. Thus, up to 40–80 million sequenced reads may be required to achieve 20–40 million mapped reads (see Fig. 4). For the Illumina GA2 machine this translates to ~2–3 lanes of sequencing. Fortunately, the newest Illumina technology (HiSeq) will greatly increase the number of reads/lane. However, HiSeq will generally provide more reads/lane than is required for many site-specific factors or certain modified histones. Therefore, multiplexing and barcoding of ChIP-seq libraries will become necessary to ensure the most cost-effective sequencing strategy (35).

4. Notes

1. Use the primer design program Primer3 (36) or another suitable program to design the target and control primers, making the product length 90–150 bp. If possible, design primers for at least two positive targets as well as for two negative control regions. Before testing the ChIP sample or library, it is important to determine that the primers work for the SYBR Green-based real-time PCR assay. To do this, use input DNA and run a melting curve following the real-time PCR reaction conditions and view the dissociation curve to ensure that the desired amplicon was detected, as seen by a single peak.

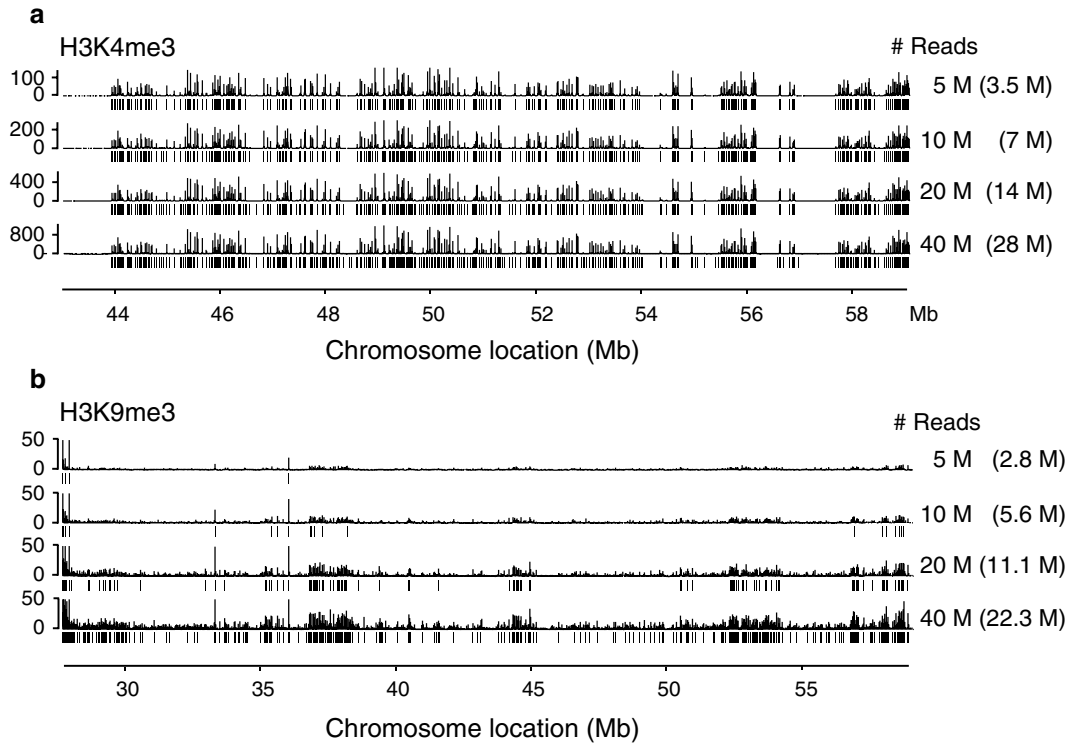


Fig. 4. Required read number varies depending on the histone mark. The ChIP-seq binding patterns from 5, 10, 20, and 40 million (M) sequenced reads for the libraries described in Fig. 3 are shown for a region on chromosome 19 (hg19 coordinates). The number of sequences that uniquely mapped to the genome is given in *brackets* and is representative for a typical ChIP-seq experiment. Peak height is plotted along the *y*-axis and the chromosomal location is shown on the *x*-axis. Called peaks are indicated as *black bars* below each track. While ten million reads are sufficient to identify most H3K4me3 sites in the human genome, more reads are required for the spreading mark H3K9me3. For example, in the region shown, only one peak is called with five million reads and only 20 peaks are identified with ten million reads; appropriate peak calling is achieved when ~40 million reads are analyzed.

2. Chromatin can be sheared by sonication or digested by micrococcal nuclease. Although both methods work well, this protocol is based on sonication. An alternative method using micrococcal nuclease is available from Cell Signaling Technology (http://www.cellsignal.com).
3. Sonication conditions should be optimized for each cell type and chromatin size determined before processing large quantities of cells, especially if the cells are collected from patients. We use the BioRuptor UCD-200 (Diagenode) on high setting for sonication. Wear hearing protection! Volumes between 0.5 and 2 ml are sonicated in 15 ml tubes, volumes between 0.1 and 0.3 ml are sonicated in 1.5 ml Eppendorf tubes, and volumes between 10 and 100 μl are sonicated in 0.5 ml tubes. The pulse duration, intensity, and number will vary depending on the sonicator, the extent of cross-linking, and cell type.

Ideally, the least amount of input energy that gives satisfactory fragmentation should be used. We commonly sonicate 20–30 min (pulses of 30 s at setting high, with 1.5 min pauses in between pulses).

4. It is recommended to prepare chromatin and set up ChIP assays on the same day for best results. If necessary, unused chromatin can be snap frozen in liquid nitrogen and stored at –80°C for use at a later time.
5. In some cases the cell number is extremely limited. As an alternative to determining chromatin yield before the ChIP assay, chromatin can be isolated after incubation with magnetic protein G beads in Subheading 3.4, step 2. Do not discard the supernatant, but reverse crosslinks and purify DNA instead. This sample can substitute for the Input sample.
6. A typical chromatin yield ranges from 5 to 10 μg chromatin per one million tissue culture cells. In our experience, primary cells often give lower chromatin yields than cultured cells. If the yield drops below 1 μg chromatin per one million cells, the ChIP reaction mixture becomes too dilute and the ChIP quality may be significantly compromised.
7. While manual ChIP assays are very reproducible when performed by experienced investigators, the IP-Star® (Diagenode) offers an alternative by automating Subheadings 3.4 and 3.5. The ChIP robot can perform up to 16 ChIP assays at a time, requiring only manual set-up of the tubes and solutions. Automation of the ChIP assay can enhance consistency of ChIP results from one experiment to the other (especially for novice researchers) or between different users. We use the IP-Star in combination with the Auto ChIP kit (Diagenode) following the manufacturer's instructions. Typically, an automated ChIP assay starts with 1–5 μg chromatin (see Fig. 2 for PCR analysis of ChIPs performed using 1 μg of chromatin and the IP-Star); if higher chromatin amounts are required, it may be best to perform manual ChIP assays. Chromatin concentration is important since the chromatin sample has to be diluted to a total reaction volume of 100 μL with Buffer A (Diagenode) to ensure appropriate dilution of the SDS that is present in the nuclei lysis buffer. Therefore, the chromatin sample used per ChIP should not exceed 20 μL to allow fivefold dilution with buffer A; more dilute chromatin (higher volumes) may decrease ChIP enrichment and are thus not desirable. Volumes of antibody and Protein G magnetic beads used per assay are the same as described in the manual ChIP protocol. The main difference in the manual and the automated protocol is that in the IP-Star, protein G magnetic beads are coated with antibody before proceeding to 8 h of IP reaction.

8. The biggest limitations to histone ChIP assays are the specificity of the antibody and the variability between different lot numbers of the same antibody. Companies have acknowledged this problem and some are making an effort to provide antibody specificity information. However, the number of ChIP validated antibodies is still very small and it is extremely important to test histone antibodies for cross reactivity with other histone marks. Antibody specificity can be tested using histone dot blots (each dot on the membrane contains a histone peptide carrying a specific modification(s)). The dot blot can be probed with the histone antibody of interest using a Western blotting protocol. Antibodies with high specificity will only recognize one dot corresponding to the histone modification(s) that was used to raise the antibody. Antibody efficiency can vary significantly between different batches, resulting in variation of the quality of the resultant ChIP-seq data. It is therefore important to record antibody details, such as catalog number, lot number, batch of affinity purification, etc. and to test each new antibody batch before performing a ChIP assay.
9. An input library is also critical for determining a baseline genome for the identification of binding sites. It is important that the input library size matches the size of the ChIP-seq library. For each cell type, 20–40 million sequenced tags of an input library are required. The same input library used to determine enrichment can be used for sequencing.
10. Gel size selection is the most variable step among protocols used by different investigators. For example, some protocols incorporate a size selection of the ChIP sample, whereas others size select after adapter ligation or after amplification (some protocols size select at more than one of these steps). We have chosen to perform size selection only once at the step after adapter ligation for three reasons: (1) each size selection results in loss of sample, so we limit the protocol to only one size selection step, (2) selecting after adapter ligation allows the removal of unincorporated adapters from the amplification reaction (which would not be possible if size selection was done after the ChIP assay but before adapter ligation), and (3) by selecting prior to PCR, we only amplify ChIP fragments of the proper length. This is especially important when the sonicated sample contained a significant fraction of fragments bigger than 500 bp (which can sometimes occur with certain cell types).
11. We prepare two libraries for each sample because the efficiency of sonication varies for different chromatin regions. Compacted chromatin sonicates less efficiently than accessible chromatin regions. The smaller portion of sonicated chromatin (200–400 bp) contains more promoter fragments, while fragments from compacted chromatin are overrepresented in the bigger

fragment (400–600 bp). We proceed with both sized fragments through the library-making process and determine enrichment of positive controls by qPCR in both the large and small libraries for each sample. We typically find that the smaller sized library gives the highest enrichment for histone marks of open chromatin (e.g., H3K9Ac, H3K4me3, H3K4me1, and H3K36me3) whereas the larger sized library gives the highest enrichment for compacted chromatin (e.g., H3K9me3 or H3K27me3).

References

1. Crews, D. Epigenetics, brain, behavior, and the environment, *Hormones (Athens)* **9**, 41–50.
2. Grayson, D. R., Chen, Y., Dong, E., Kundakovic, M., and Guidotti, A. (2009) From trans-methylation to cytosine methylation: evolution of the methylation hypothesis of schizophrenia, *Epigenetics* **4**, 144–149.
3. Iwamoto, K., and Kato, T. (2009) Epigenetic profiling in schizophrenia and major mental disorders, *Neuropsychobiology* **60**, 5–11.
4. Jiang, Y., Langley, B., Lubin, F. D., Renthal, W., Wood, M. A., Yasui, D. H., Kumar, A., Nestler, E. J., Akbarian, S., and Beckel-Mitchener, A. C. (2008) Epigenetics in the nervous system, *J Neurosci* **28**, 11753–11759.
5. McCarthy, M. M., Auger, A. P., Bale, T. L., De Vries, G. J., Dunn, G. A., Forger, N. G., Murray, E. K., Nugent, B. M., Schwarz, J. M., and Wilson, M. E. (2009) The epigenetics of sex differences in the brain, *J Neurosci* **29**, 12815–12823.
6. Jones, P. A., and Baylin, S. B. (2007) The epigenomics of cancer, *Cell* **128**, 683–692.
7. Ellis, L., Atadja, P. W., and Johnstone, R. W. (2009) Epigenetics in cancer: targeting chromatin modifications, *Mol Cancer Ther* **8**, 1409–1420.
8. Lund, A. H., and van Lohuizen, M. (2004) Epigenetics and cancer, *Genes Dev* **18**, 2315–2335.
9. Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome, *Cell* **129**, 823–837.
10. Kouzarides, T. (2007) Chromatin modifications and their function, *Cell* **128**, 693–705.
11. Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., Kellis, M., and Ren, B. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression, *Nature* **459**, 108–112.
12. Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Calcar, S. V., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007) Distinct predictive chromatin signatures of transcriptional promoters and enhancers in the human genome, *Nature Genetics* **published online Feb 4, 2007**.
13. O'Geen, H., Squazzo, S. L., Iyengar, S., Blahnik, K., Rinn, J. L., Chang, H. Y., Green, R., and Farnham, P. J. (2007) Genome-Wide Analysis of KAP1 Binding Suggests Autoregulation of KRAB-ZNFs, *PLoS Genet* **3**, e89.
14. Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., and Chang, H. Y. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs, *Cell* **129**, 1311–1323.
15. Chen, L., and Daley, G. Q. (2008) Molecular basis of pluripotency, *Hum Mol Genet* **17**, R23–27.
16. Hemberger, M., Dean, W., and Reik, W. (2009) Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal, *Nat Rev Mol Cell Biol* **10**, 526–537.
17. Lanzuolo, C., and Orlando, V. (2007) The function of the epigenome in cell reprogramming, *Cell Mol Life Sci* **64**, 1043–1062.
18. Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E. S., and Bernstein, B. E. (2007) Genome-wide maps of

chromatin state in pluripotent and lineage-committed cells, *Nature* **448**, 553–560.

19. Park, P. J. (2009) ChIP–seq: advantages and challenges of a maturing technology, *Nature Rev Genet*, 669–680.
20. Wei, C. L., Wu, Q., Vega, V. B., Chiu, K. P., Ng, P., Zhang, T., Shahab, A., Yong, H. C., Fu, Y., Weng, Z., Liu, J., Zhao, X. D., Chew, J. L., Lee, Y. L., Kuznetsov, V. A., Sung, W. K., Miller, L. D., Lim, B., Liu, E. T., Yu, Q., Ng, H. H., and Ruan, Y. (2006) A global map of p53 transcription-factor binding sites in the human genome, *Cell* **124**, 207–219.
21. Bhinge, A. A., Kim, J., Euskirchen, G. M., Snyder, M., and Iyer, V. R. (2007) Mapping the chromosomal targets of STAT1 by Sequence Tag Analysis of Genomic Enrichment (STAGE), *Genome Res* **17**, 910–916.
22. Hawkins, R. D., Hon, G. C., Lee, L. K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L. E., Kuan, S., Luu, Y., Klugman, S., Antosiewicz-Bourget, J., Ye, Z., Espinoza, C., Agarwahl, S., Shen, L., Ruotti, V., Wang, W., Stewart, R., Thomson, J. A., Ecker, J. R., and Ren, B. (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells, *Cell Stem Cell* **6**, 479–491.
23. Motallebipour, M., Ameur, A., Reddy Bysani, M. S., Patra, K., Wallerman, O., Mangion, J., Barker, M. A., McKernan, K. J., Komorowski, J., and Wadelius, C. (2009) Differential binding and co-binding pattern of FOXA1 and FOXA3 and their relation to H3K4me3 in HepG2 cells revealed by ChIP-seq, *Genome Biol* **10**, R129.
24. Koerber, R. T., Rhee, H. S., Jiang, C., and Pugh, B. F. (2009) Interaction of transcriptional regulators with specific nucleosomes across the Saccharomyces genome, *Mol Cell* **35**, 889–902.
25. Goren, A., Ozsolak, F., Shoresh, N., Ku, M., Adli, M., Hart, C., Gymrek, M., Zuk, O., Regev, A., Milos, P. M., and Bernstein, B. E. (2010) Chromatin profiling by directly sequencing small quantities of immunoprecipitated DNA, *Nat Methods* **7**, 47–49.
26. Fejes, A. P., Robertson, G., Bilenky, M., Varhol, R., Bainbridge, M., and Jones, S. J. M. (2008) FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology, *Bioinformatics* **24**, 1729–1730.
27. Xu, H., Wei, C.-L., Lin, F., and Sung, W.-K. (2008) An HMM approach to genome-wide identification of differential histone modification sites from ChIP-seq data, *Bioinformatics* **24**, 2344–2349.
28. Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S. (2008) Model-based analysis of ChIP-Seq (MACS), *Genome Biology* **9**, R137.
29. Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B. (2007) Genome-wide mapping of in vivo protein-DNA interactions., *Science* **316**, 1497–1502.
30. Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T., Euskirchen, G., Bernier, B., Varhol, R., Delaney, A., Thiessen, N., Griffith, O. L., He, A., Marra, M., Snyder, M., and Jones, S. (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing, *Nat Methods* **4**, 1–7.
31. Jothi, R., Cuddapah, S., Barski, A., Cui, K., and Zhao, K. (2008) Genome-wide identification of in vivo protein-DNA binding sites from ChIP-seq data, *Nucleic Acids Res* **36**, 5221–5231.
32. Rozowsky, J., Euskirchen, G., Auerbach, R. K., Zhang, Z. D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M., and Gerstein, M. B. (2009) PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls, *Nat Biotechnol* **27**, 66–75.
33. Valouev, A., Johnson, D. S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R. M., and Sidow, A. (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-seq data, *Nature Methods* **5**, 829–834.
34. Blahnik, K. R., Dou, L., O'Geen, H., McPhillips, T., Xu, X., Cao, A. R., Iyengar, S., Nicolet, C. M., Ludaescher, B., Korf, I., and Farnham, P. J. (2010) Sole-search: An integrated analysis program for peak detection and functional annotation using ChIP-seq data, *Nucleic Acids Res* **38**, e13.
35. Lefrancois, P., Euskirchen, G. M., Auerbach, R. K., Rozowsky, J., Gibson, T., Yellman, C. M., Gerstein, M., and Snyder, M. (2009) Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing, *BMC Genomics* **10**, 37.
36. Rozen, S., and Skaletsky, H. J. (2000) Primer3 on the WWW for general users and for biologist programmers, in Bioinformatics Methods and Protocols: Methods in Molecular Biology (Krawetz, S., and Misener, S., Eds.), pp 365–386, Humana Press, Totowa, NJ.

Chapter 21

Mapping Open Chromatin with Formaldehyde-Assisted Isolation of Regulatory Elements

Takao Nammo, Santiago A. Rodríguez-Seguí, and Jorge Ferrer

Abstract

Noncoding regulatory genomic elements are central for cellular function, differentiation, and disease, but remain poorly characterized. FAIRE (formaldehyde-assisted isolation of regulatory elements) has emerged as a simple method to identify and analyze active regulatory sequences based on their decreased nucleosomal content. More recently FAIRE was combined with high-throughput sequencing (FAIRE-seq) to locate tissue-specific regulatory elements at a genome scale in purified human pancreatic islets. Here we describe the implementation of the FAIRE method in human pancreatic islet cells.

Key words: FAIRE, Gene transcription, Open chromatin, Diabetes, Epigenetics, High-throughput sequencing, Pancreatic islets

1. Introduction

The dissection of functional noncoding sequence elements is currently one of the most important challenges in genome biology. Such sequence elements are believed to instruct genomic regulatory programs that underlie cellular phenotypes, and are likely to play an important role in disease mechanisms. The analysis of functional regulatory sequences can thus provide an opportunity to increase our understanding of cellular function, differentiation, and disease.

Genomic regions that are directly involved in transcriptional regulatory functions can be discriminated because they exhibit distinct structural chromatin features. Studies of sequences that contain binding sites for *cis*-regulatory proteins have thus revealed either local nucleosome eviction, or altered nucleosome–DNA interactions.

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_21,

This correlation could result from an increased ability of DNA binding factors to access their cognate sites at sequences that have decreased intrinsic binding affinities for nucleosomes, or because transcription factors promote local depletion of nucleosomes (1, 2).

Several methods have been developed to identify open chromatin regions. Some exploit the increased accessibility of DNA in open chromatin to nucleases such as DNAse I or restriction enzymes (3, 4). Another method, named FAIRE (formaldehyde-assisted isolation of regulatory elements), exploits the fact that upon exposure to formaldehyde, crosslinks are formed between intimately interacting nucleosomal histones and DNA (5–7). In this method, formaldehyde-treated cells are first lysed and sonicated. DNA fragments that are not bound by nucleosomes do not establish nucleosomal-DNA crosslinks and are thus preferentially recovered from the aqueous phase of a phenol–chloroform reaction (5, 6). FAIRE has now been employed by diverse investigators to study open chromatin in a variety of eukaryotic cells, including yeast, protozoan parasites, and mammalian cells (5, 6, 8–10). A recent study (11) combined FAIRE with high-throughput sequencing (FAIRE-Seq) in purified human pancreatic islets, which are clusters of endocrine cells that produce insulin and other polypeptide hormones. This study verified that FAIRE indeed provides a signature of active regulatory elements in a primary tissue; it accordingly revealed prominent FAIRE enrichment in active promoters, as well as in predicted and known long range regulatory elements (11). The same study also uncovered thousands of clusters of tissue-selective open chromatin sites, many of which appear to be unexpectedly broad regulatory domains (11).

FAIRE is easy to perform, and is applicable to very small amounts of primary tissue. It can therefore be readily employed to study chromatin states linked to lineage-specific differentiation processes in living organisms or during in vitro differentiation protocols. It can also be employed to identify chromatin changes in human disease samples, or to understand how sequence variation impacts open chromatin states. Here, we describe a detailed implementation of the FAIRE protocol in isolated human pancreatic islets. Adaptations of this method are generally applicable to cultured cell lines and primary tissue fragments.

2. Materials

2.1. Tissue Culture

1. Dithizone stock solution. Add 10 mg dithizone (Sigma, St. Louis, MO) to 2 mL dimethyl sulfoxide (DMSO) (Sigma). Store the solution at −20°C.
2. RPMI 1640 (Lonza, Basel, Switzerland) supplemented with 10% heat-inactivated fetal calf serum (FCS, Lonza) and penicillin/streptomycin 100 units/mL (Lonza).

3. Hanks' balanced salt solution (Lonza) and phosphate-buffered saline (PBS, Lonza).

2.2. FAIRE

1. 37% Formaldehyde (Calbiochem, Darmstadt, Germany).
2. 2 M Glycine (Sigma).
3. Protease inhibitor cocktail (Roche, Basel, Switzerland).
4. Dry ice and ethanol.
5. Lysis buffer (2% Triton X-100, 1% SDS, 100 mM NaCl, 10 mM Tris–HCl, pH 8.0, 1 mM EDTA). Add 20 µL protease inhibitor cocktail (Roche) to 1,000 µL of Lysis buffer freshly before use.
6. TE buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA, pH 8.0).
7. 25-gauge needles.
8. 2 mL screw cap tubes (one per sample).
9. Glass beads 0.5 mm (Biospec, Bartlesville, OK).
10. Conical 75-mm tall polypropylene tube (Sarstedt #57.512, Nümbrecht, Germany).
11. Sodium acetate 3 M, pH 5.2.
12. Ethanol 70% (store at –20°C), ethanol 95% (store at –20°C), ethanol 100% (store at –20°C).
13. Glycogen (Invitrogen, San Diego, CA).
14. Proteinase K 20 mg/mL (Roche).
15. RNase A 10 mg/mL (Qiagen, Hilden, Germany).
16. Tris-saturated phenol:chloroform:isoamyl alcohol (25:24:1) (USB, Cleveland, OH), chloroform (Sigma).
17. 5 M NaCl.
18. 0.5 M EDTA pH 8.0.
19. QIAquick PCR Purification Kit (Qiagen).

3. Methods

The analysis of primary human samples for functional genomic studies entails considerable experimental variability due to unavoidable differences in donors' history and tissue procurement procedures. Human pancreatic endocrine islets are a typical example. They are obtained from organ donors after a lengthy experimental procedure that involves collagenase digestion and gradient purification that separates islets from the more abundant exocrine tissue (12). To mitigate environmentally induced variation, we culture islets in suspension under uniform conditions for 3 days prior to performing FAIRE.

The FAIRE method described here is based on that developed by Lieb and colleagues (5–7) and has been adapted for purified primary human islets, which are essentially tissue fragments of typically ~500–2,000 cells that can be cultured in suspension. Adequate fixation of human islets requires slightly longer times of exposure to formaldehyde than cells grown in monolayer. Fixation conditions for other types of primary tissue samples may need to be optimized empirically. Proper sonication of primary tissues is also challenging. It is not only crucial to ensure that most chromatin fragments are smaller than 1 kb, but also that fragment length is uniform across samples that need to be compared. Our protocol overcomes difficulties in the sonication of isolated islets, and can assist efforts to optimize sonication of other primary tissues.

3.1. Tissue Culture

1. Prior to culture, islets are quantified and assessed for purity using dithizone, which binds to zinc ions present in ß-cells and thus permits discrimination from pancreatic exocrine contaminants. Dilute 200 µL dithizone solution with 800 µL Hanks' balanced salt solution, centrifuge at maximum speed at room temperature for 5 min. After uniform dispersion of islets, an aliquot of 400 µL is taken and transferred to a 1.6-mL eppendorf tube (a larger volume is needed if there are not at least 50 islets). Let the islets sediment, then add 900 µL diluted dithizone solution to this, and incubate for 15 min at 37°C. Islets are next rinsed with PBS twice, and then examined under a microscope. Islet number in this aliquot is extrapolated to the total islet count based on the total culture volume. These islets can be spun, snap frozen, and stored at −80°C for subsequent extraction of unfixed DNA that can be employed as the input control (see below).
2. To place islets in culture after isolation or transportation, they are first transferred to 50-mL Falcon tubes in a tissue culture hood, and spun at 500 rpm ($45 \times g$) for 1 min in a benchtop centrifuge at room temperature. After discarding the supernatant, 10 mL of the culture medium is added for every ~2,000 islets. Islets are very gently resuspended, 10 mL medium is transferred to each 100 mm bacterial culture dish, which are then placed in culture at 37°C for 3 days.

3.2. Fixation

1. After culture, collect islets in the center of the culture dish by very slow constant anticlockwise swirling of the dish. Use a pipette to remove as much debris and nonislet material as possible.
2. Aspirate islets from two culture dishes (~4,000 islets) and transfer to a 50-mL Falcon tube. Spin at 500 rpm ($45 \times g$) for 1 min in a benchtop centrifuge at RT.
3. Remove the most supernatant, and transfer all islets into a single new 15-mL tube.

4. Rinse the islets with 10 mL PBS and spin at 500 rpm (45 × *g*) for 1 min. Remove the supernatant. Repeat this twice (see Note 1).
5. After removing supernatant add PBS up to 10 mL.
6. Add 278 μL 37% formaldehyde (final concentration 1%) and fix cells shaking gently for 10 min at room temperature (see Note 2).
7. Stop fixation by adding 685 μL 2 M glycine (final concentration 125 mM) and incubate for 5 min shaking gently at room temperature.
8. Centrifuge islets at 500 rpm (45 × *g*) for 1 min at room temperature.
9. Remove the supernatant, resuspend the islets in 1 mL ice-cold PBS. Centrifuge at 500 rpm for 1 min at 4°C. Repeat this step twice.
10. Remove the supernatant, snap-freeze fixed islets in liquid nitrogen or a dry ice/ethanol bath, and store the samples at −80°C or liquid nitrogen. We have stored chromatin samples at −80°C for over 6 months without any apparent deterioration of FAIRE enrichment.

3.3. Sonication

1. Allow a tube containing ~4,000 fixed islets to briefly thaw on ice and add 1 mL Lysis buffer with protease inhibitor cocktail. Pipette the resuspended islets into a 2-mL screw-top tube. Add 1 mL of 0.5-mm glass beads to each tube.
2. Lyse cells by vortexing vigorously for 1 min. Repeat this ten times, placing tubes on ice 1 min. between each vortexing session.
3. Open the tube cap, insert a 25-gauge needle in the bottom of the tube, allow the extract to drip into a conical 75-mm tall tube on ice. Add 0.7 mL Lysis buffer to the beads, mix gently, and allow that to drip into the same tube. Aim for ~1.5–1.7 mL total lysate volume.
4. Label Eppendorf tubes for checking DNA sonication (e.g., 0, 5, 10 cycles). Take a 25–50 μL aliquot for cycle 0. Prepare a dry ice/ethanol bath.
5. Clean the sonicator probe by squirting water, ethanol, and then water again. Adjust sonication settings as follows (Branson 450D sonifier with tapered microtip (#101-148-062)): 30”-1” ON/0.5” OFF; amplitude 15% (set manually).
6. Sonicate for ten cycles, placing the tube in the dry ice bath for 2 s between each cycle (see Note 3). Take 25–50 μL aliquots at fifth and tenth cycles.
7. To assess the efficiency of sonication, add 150–175 μL TE buffer (to make 200 μL) to the 0, 5, 10 cycle samples. Spin at full speed in the microfuge at 4°C for 5 min to clear the extract of

debris and unlysed cells. Transfer the supernatants to new tubes. Add 1 μL RNase A (10 mg/mL) to each tube, tap the tubes lightly to mix, and incubate at 65°C for 20 min.

8. Add 2 μL Proteinase K 20 mg/mL, tap the tube to mix, and incubate at 65°C for 3 h.
9. Add 160 μL TE to make 360 μL total volume.
10. Add phenol–chlorofom solution (450 μL), vortex to mix. Spin at room temperature at maximum speed in a microfuge for 5 min. Recover the aqueous phase and transfer to new tubes.
11. Add chloroform (450 μL), vortex to mix. Spin at room temperature at maximum speed for 5 min. Recover the aqueous phase and transfer to new tubes.
12. Add 1/10 volume of sodium acetate (3 M, pH 5.2), 2.5× volume ice-cold 100% ethanol, and 1 μL glycogen (5 mg/mL). Invert several times and keep at −20°C for >15 min.
13. Spin at 4°C at maximum speed for 20 min and remove the supernatants.
14. Rinse with 500 μL ice-cold 70% ethanol. Spin at 4°C at maximum speed for 5 min and remove the supernatants.
15. Speed-vac ~10 min. at room temperature. Do not allow the pellet to overdry. Resuspend in 10 μL dH_2O. Incubate at 37°C for 30 min.
16. Run 5–8 μL of the resuspended DNA on a 1% agarose gel in 0.5× TBE buffer (110 V, 1 h).
17. After electrophoresis, check the DNA size. If the chromatin has the intended size at ten cycles (see Note 4), aliquot 300 μL chromatin samples in labeled eppendorf tubes, snap freeze and store at −80°C, or use fresh. If fragments are too large, sonicate for 2–5 further cycles, check DNA quality again, and then store chromatin aliquots and/or use fresh.
18. Aliquot 75 μL of chromatin extract in one tube and label it as fixed chromatin input.

3.4. Phenol–Chloroform Extraction

1. Take 300 μL sonicated extract, add TE buffer to make a total volume of 500 μL. Spin for 10 min at 4°C at full speed to clear debris. Recover the supernatant and transfer to a new tube.
2. Add an equal volume of phenol–chloroform solution to this supernatant. Vortex. Spin at top speed in microfuge for 5 min. Recover the supernatant and transfer to a new tube.
3. Add an equal volume of phenol–chloroform solution, vortex, and spin as in the previous step. Recover the supernatant ensuring that the interphase content is not collected and transfer to a new tube.
4. Add an equal volume of chloroform, vortex, spin as in the previous step, and recover the supernatant.

5. Add 1/10 volume sodium acetate (3 M, pH 5.2). Mix. (Optional: add 1 μL of 20 mg/mL Glycogen). Add 2.5× volume ice-cold 95% ethanol. Mix. Place at −20°C for >1 h.
6. Centrifuge at full speed for 20 min at 4°C. Remove the supernatant, add ice-cold 70% ethanol, spin for 5 min at full speed. Remove the supernatant and speed-vac at room temperature. Be careful not to overdry the pellet.
7. Resuspend in 50 μL of TE buffer and add 1 μL RNase A (10 mg/mL). Flick the tube and incubate at 37°C for >30 min.
8. DNA concentration can be quantified using a fluorometry-based method such as Qubit (Invitrogen) (see Note 5).

3.5. Preparation of Input DNA

1. Place the sonicated extract labeled as input on ice.
2. Add TE to make 300 μL. Add 1.5 μL RNase A (10 mg/mL). Incubate at 65°C for >30 min.
3. Add 4.5 μL Proteinase K (20 mg/mL). Incubate at 65°C for 5 h to O/N (see Note 6).
4. Purify DNA with a spin column such as the QIAquick PCR Purification Kit (QIAGEN) following the manufacturer's instructions. Elute DNA in 50 μL TE buffer.

3.6. Overview of FAIRE DNA Analysis

1. The analysis of FAIRE-enriched DNA is in many ways analogous to that of chromatin immunoprecipitation experiments, and can therefore be performed using either locus-specific assays (11), hybridization to miroarrays (5), or high-throughput sequencing (11). Several general features of this analysis are nevertheless distinct in FAIRE as outlined below.
2. FAIRE can be analyzed with semiquantitative PCR assays that target specific genomic sites. As in chromatin immunoprecipitation experiments, this involves designing oligonucleotides for the experimental target site as well as for negative control sites that presumably do not harbor open chromatin. FAIRE DNA is then assayed for enrichment at experimental and control sites relative to input DNA. There are, however, important caveats to bear in mind when performing this type of analysis. First, without a priori knowledge of where the FAIRE enrichment sites are located in the cell of interest, any PCR assay can easily fail to detect an open chromatin site even if it encompasses an open chromatin site. This can occur, for example, if one of the two oligonucleotides targets a sequence that flanks a true open chromatin site but is intimately bound by nucleosomes; DNA fragments containing this flanking sequence will not be retained in the aqueous phase of the phenol extraction, and therefore the PCR assay will not detect FAIRE enrichment. Another critical issue is the selection of appropriate negative control regions. For unknown reasons FAIRE-seq experiments have

shown detectable differences in "background" DNA between different broad genomic *loci*. This difference can occur between two regions that do not necessarily harbor any discrete open chromatin sites, which warrant the use of local negative control regions for PCR assays. These limitations can theoretically be overcome by designing a tile of overlapping short amplimers across a region of interest. In general, however, target-specific assays in FAIRE are ideally suited for interrogating previously characterized open chromatin sites.

3. High-throughput short-read sequencing provides a very powerful means of assessing FAIRE-enriched DNA at a genomic scale. Library construction, sequencing, and alignment can be performed analogously to chromatin immunoprecipitation sequencing experiments. Some noteworthy differences are nonetheless also applicable here. First, the fraction of the genome that exhibits open chromatin is higher than that enriched in typical transcription factor chromatin immunoprecipitations, and furthermore the assay inherently leads to the recovery of a low level of DNA from nonenriched regions. These factors determine that a high sequencing depth is required to attain a reliable read coverage at any given site. As a guideline, >50 million high-quality mapped 36-nucleotide reads from a single sample can provide a highly informative genome-scale open chromatin map in human tissue. A second related issue is that FAIRE does not divide a cellular genome into discrete closed regions and open chromatin peaks. Instead, this method provides a diverse range of enrichment values at different sites, and different sites can exhibit considerable length heterogeneity. Calling enriched sites thus provides a greater challenge than in typical transcription factor binding ChIP-seq experiments. A previous report of FAIRE-seq in human islets employed F-seq (13), a method that calculates a read density probability for each base, and identifies enriched regions of varying lengths that exceed different threshold standard deviations above a background mean. In general, different epigenomic features require different types of algorithms for optimal detection, and any additional candidate algorithms need to be tested to assess their performance in FAIRE-seq datasets.

4. Notes

1. Do not expose islets to unnecessary physical stress, such as inverting or vigorous shaking of tubes. Vigorous pipetting of fixed islets will cause significant loss of the sample due to adherence to the plastic surface.

2. These fixation conditions are adequate for FAIRE in human islets, but need to be optimized for other tissues or cell types. The same conditions are valid for chromatin immunoprecipitation assays for histone modifications.
3. Keep the probe at the highest possible point that does not lead to foaming and in a well-centered position, avoiding high-pitched sounds. The number of cycles varies according to several parameters, including fixation time and number of cells, and needs to be determined empirically by checking the DNA size ranges after ten cycles. Because primary human samples can vary considerably in ways that affect sonication efficiency, we generally check sonication for all samples to ensure that we obtain consistent sonication lengths in different samples.
4. We aim for fragment sizes ranging from 200 to 1,000 bp. When assessing this by simple inspection, it is worth remembering that ethidium bromide stains larger fragments more intensely than smaller DNA fragments. The optimal DNA size is empirical, yet different size ranges can yield different results for both FAIRE and chromatin immunoprecipitation experiments. It is therefore critical to ensure that the size range is similar across samples that are to be compared with each other at some point.
5. Avoid using spectrophotometric methods as they can significantly overestimate DNA in FAIRE samples. FAIRE-enriched DNA can also be both quantified and assessed for size range using the Agilent 2100 Bioanalyzer and Agilent High Sensitivity DNA reagents (Agilent Technologies, Palo Alto, CA).
6. It is important to ensure that the removal of crosslinks is exhaustive, otherwise "open" chromatin may theoretically be preferentially recovered when decrosslinking is insufficient in the input sample. This may influence any downstream assays that use input as a reference. Using unfixed DNA is a valid control for sequence analysis of FAIRE DNA.

Acknowledgments

The authors would like to thank Jason D. Lieb and Paul G. Giresi for advice and support in the adaptation of the FAIRE protocol for human islets. Work in the authors' laboratory has been funded by Ministerio de Ciencia e Innovación and Centro de Investigación Biomédica en Red (CIBER) de Diabetes y Enfermedades Metabólicas.

References

1. Henikoff S.(2008) Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat Rev Genet* **9**,15–26.
2. Segal E, Widom J.(2009) What controls nucleosome positions? *Trends Genet* **25**, 335–343.
3. Almer A, Horz W.(1986) Nuclease hypersensitive regions with adjacent positioned nucleosomes mark the gene boundaries of the PHO5/PHO3 locus in yeast. *EMBO J* **5**,2681–2687.
4. Elgin SC.(1988) The formation and function of DNase I hypersensitive sites in the process of gene activation. *J Biol Chem* **263**, 19259–19262.
5. Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD.(2007) FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17**,877–885.
6. Nagy PL, Cleary ML, Brown PO, Lieb JD.(2003) Genomewide demarcation of RNA polymerase II transcription units revealed by physical fractionation of chromatin. *Proc Natl Acad Sci U S A* **100**,6364–6369.
7. Giresi PG, Lieb JD.(2009) Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). *Methods* **48**,233–239.
8. Ponts N, Harris EY, Prudhomme J et al.(2010) Nucleosome landscape and control of transcription in the human malaria parasite. *Genome Res* **20**,228–238.
9. Eeckhoute J, Lupien M, Meyer CA et al.(2009) Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. *Genome Res* **19**,372–380.
10. Berchowitz LE, Hanlon SE, Lieb JD, Copenhaver GP.(2009) A positive but complex association between meiotic double-strand break hotspots and open chromatin in Saccharomyces cerevisiae. *Genome Res* **19**,2245–2257.
11. Gaulton KJ, Nammo T, Pasquali L et al.(2010) A map of open chromatin in human pancreatic islets. *Nat Genet* **42**,255–259.
12. Street CN, Lakey JR, Shapiro AM et al.(2004) Islet graft assessment in the Edmonton Protocol: implications for predicting long-term clinical outcome. *Diabetes* **53**, 3107–3114.
13. Boyle AP, Guinney J, Crawford GE, Furey TS.(2008) F-Seq: a feature density estimator for high-throughput sequence tags. *Bioinformatics* **24**,2537–2538.

Chapter 22

Inhibition of Histone Deacetylases

Yi Huang, Patrick G. Shaw, and Nancy E. Davidson

Abstract

Lysine acetylation of histones is one of the major epigenetic regulators of chromatin conformation and gene expression. The dynamic nature of histone acetylation is determined by the counterbalancing activity of histone acetyltransferase and histone deacetylase (HDAC) enzymes. Acetylation of histones is generally associated with open and transcriptionally active chromatin, whereas the activity of HDACs leads to histone deacetylation, condensation of chromatin, and inhibition of transcription. Aberrant silencing of tumor suppressors and other genes has been found in different types of cancer. Abnormal activity of HDACs has been implicated in tumorigenesis and therefore considerable effort has been put into the development of HDAC inhibitors as a means of modifying histone acetylation status and reexpressing aberrantly silenced tumor suppressor genes. This has led to the generation of a number of structurally diverse compounds that can effectively inhibit HDAC activity, thus altering chromatin structure in cancer cells. This unit discusses the methods and recent technological developments with respect to the studies of HDAC inhibition in cancer.

Key words: Histone deacetylases, Histone acetyltransferases, HDAC inhibitors, Epigenetic gene silencing, Chromatin remodeling

1. Introduction

Epigenetic modifications refer to heritable and reversible changes in chromatin structure that are not due to alterations in primary DNA sequence (1, 2). The biochemical modifications that dictate epigenetic changes include methylation of cytosine residues in CpG dinucleotides and posttranslational modifications of the histone tails such as acetylation, methylation, phosphorylation, ubiquitination, sumoylation, and ADP-ribosylation (3, 4). Among these modifications, histone acetylation is one of the major regulators of chromatin conformation and gene expression. The histone deacetylase (HDAC) family is divided into zinc-dependent (class I, IIa, IIb,

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_22,

Table 1
Human histone deacetylase subunits

Class	Subunit	Zinc-dependent	NAD+-dependent
I	HDAC 1, 2, 3, 8	Yes	No
IIa	HDAC 4, 5, 7, 9	Yes	No
IIb	HDAC 6, 10	Yes	No
III	SIRT 1, 2, 3, 4, 5, 6, 7	No	Yes
IV	HDAC 11	Yes	No

HDAC histone deacetylase, *SIRT* sirtuin

and IV of which there are 11 subtype enzymes) and zinc-independent enzymes (class III, also called sirtuins), which require NAD^+ for their catalytic activities (Table 1) (5). Acetylation by histone acetyltransferases (HATs) is generally associated with transcriptionally active chromatin (euchromatin) and activity of HDACs typically leads to chromatin condensation and inhibition of transcription (heterochromatin). Over the past decade, a number of HDAC inhibitors have been rationally designed and developed. These HDAC inhibitors have been examined for their ability to alter chromatin structure and reexpress aberrantly silenced genes which is associated with growth inhibition and apoptosis in cancer cells (6, 7). The field of HDAC inhibitors is moving rapidly into a new stage of development that has now started to produce success in the clinic, particularly in the field of cancer therapy. Based on their chemical structures, HDAC inhibitors are divided into four groups: hydroxamic acids, cyclic tetrapeptides, short-chain fatty acids, and benzamides (Table 2). Most of the HDAC inhibitors developed so far are nonselective reagents and among the most potent inhibitors are those that have been designed to target primarily the zinc cofactor of the enzyme active site and exhibit their effects in nano- or micromolar levels (8, 9). However, several recent studies revealed some unique features of class IIa HDAC biochemistry and demonstrated unexpected selectivity of HDAC inhibitors presumed to be nonselective (10, 11).

The efficacy of HDAC inhibitors such as TSA (trichostatin A), SAHA (vorinostat), romidepsin (FK-228), LBH589 (Panobinostat), PDX101 (Belinostat), and MS-275 (Entinostat) as antitumor agents has been demonstrated in a wide range of cancer cell lines as well as in animal models (12–15). Two pharmaceutical HDAC inhibitors, SAHA and romidepsin, have already been approved by the US-FDA for the clinical treatment of cutaneous T-cell lymphoma (CTCL). A number of other promising HDAC inhibitors are currently under evaluation in advanced clinical trials. The exact mechanisms through which HDAC inhibitors mediate anticancer

Table 2
Characteristics of some HDAC inhibitors in clinical trials

Class	Compound	Targeted HDACs	Clinical trial stage
Hydroxamic acid	SAHA (Vorinostat)	Class I, II, IV	USFDA approved for CTCL
	LBH-589 (Panobinostat)	Class I, II, IV	Phase III
	PXD-101 (Belinostat)	Class I, II, IV	Phase II
	ITF2357	Class I, II	Phase I
Cyclic peptide	Romidepsin (FK/228)	Class I, II	USFDA approved for CTCL
Short-chain fatty acid	Valproic acid	Class I HDAC 1	Phase II
	Phenylbutyrate	Class I	Phase I, II
Benzamide	MS-275 (Entinostat)	Class I HDAC 1, 2, 3	Phase II
	MGC0103	Class I, 11	Phase II

CTCL cutaneous T-cell lymphoma, *HDAC* histone deacetylase, *SAHA* suberoylanilide hydroxamic acid, *USFDA* US Food and Drug Administration

activity have not been fully elucidated. One model suggests that HDAC inhibitor-induced hyperacetylation of histones activates tumor-suppressor genes and represses oncogenes, thus activating intrinsic apoptotic pathways (16, 17). For example, in ER-negative breast cancer cells, inhibition of HDAC activity by specific HDAC inhibitors reactivates aberrantly silenced estrogen receptor alpha (ERα) and progesterone receptor (PR) gene expression (18–21). Pruitt et al., demonstrated that the inhibition of class III HDAC SIRT1 using a pharmacologic inhibitor, splitomicin, or siRNA reactivates epigenetically silenced *SFRP1*, *SFRP2, E-cadherin*, and *CRBP1* genes in human breast and colon cancer cells despite full retention of DNA hypermethylation at promoters of reactivated genes (22). A recent study demonstrated that HDAC inhibitors induce cellular senescence through downregulation of polycomb group genes, suggesting that HDAC activity is important for self-renewal of human multipotent stem cells (MSCs) (23). In addition, a growing field of mass spectrometry-based proteomic techniques have identified several nonhistone proteins whose lysine acetylation is directly regulated by HDACs (24–26). These studies suggest that HDAC inhibitors can also affect diverse pathways in the cell and have recently been used to predict lysine acetylation motifs (27). A broader discussion of mass spectrometry techniques used in epigenetic research can be found in volume 593, Chapter 13 of this series (28).

Studies investigating the effects of HDAC inhibitors on chromatin and gene transcription generally involve measuring the alterations of histone acetylation levels or expression of genes and

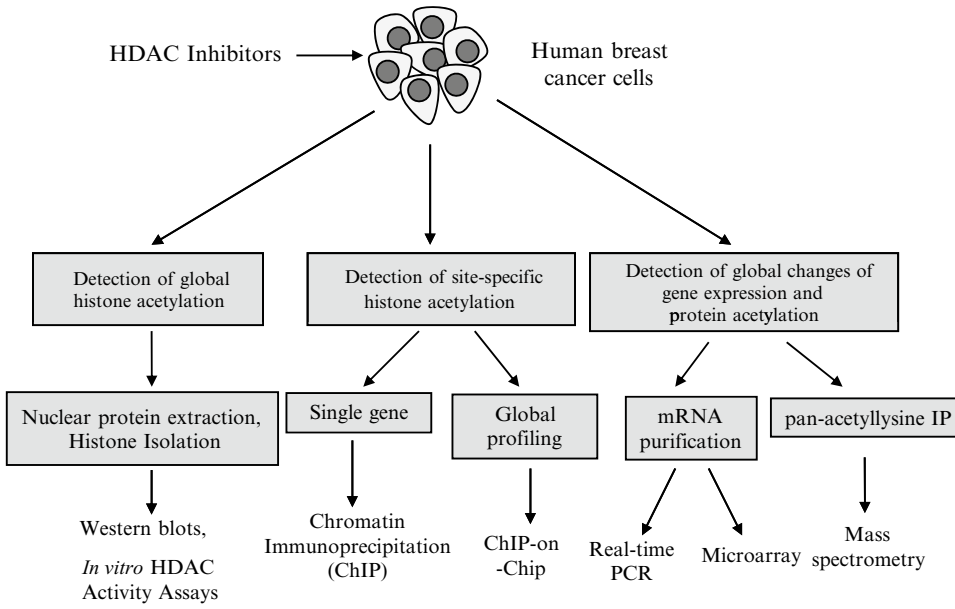


Fig. 1. Techniques for studying HDAC inhibition in human breast cancer. After human breast cancer cells are treated with HDAC inhibitors, immunological detection methods such as Western blot or immunochemistry can be used to determine the level of histone acetylation using specific antibodies against histone H3 or H4 or specific histone lysine residues such as AcH3K9, AcH3K27, and AcH4K20. Chromatin immunoprecipitation (ChIP) is used to determine the interaction of site-specific acetylated histones with promoters of genes of interest in breast cancer cells after HDAC inhibitor treatment. DNA sequences bound to a particular acetylated histone or nonhistone protein can be isolated by ChIP and these fragments can subsequently be hybridized to a DNA microarray (such as a tiling array). This so-called ChIP-on-chip technology allows the determination of acetylated histone binding occupancy throughout the genome of the cancer cell. Quantitative-PCR is able to precisely measure the specific gene expression changes in the presence of HDAC inhibitor treatment. Microarray-based gene expression profiling can be used to identify genes whose expression is altered by HDAC inhibitor treatment. Several mass spectrometry (MS)-based proteomic methods exist to quantitatively analyze proteins that are hyperacetylated after treatment with HDAC inhibitors as well as determine isoform specific occupancy of histone modifications.

gene products associated with acetylated histones induced by drug (Fig. 1). Although histones are often enriched prior to analysis using a classical acid extraction protocol (29–31), we find that nuclear extraction using a kit-based method described here is faster and typically sufficient. Additionally, more efficient methods for histone isolation which preserve more labile modifications such as phosphorylation have recently been developed (32).

As part of an effort to define the "histone code" of variable histone tail modifications, a significant area of research focuses on the detection of specific acetylated lysines of histones. Immunological detection (Western blots or immunochemistry) has become the method of choice to determine histone acetylation in cancer cells as a result of the growing availability of site-specific, histone family-specific, or pan-acetylation antibodies. While recent developments in mass spectrometry enable accurate quantification of isoform-specific histone modifications (33), the existing array of specific antibodies remains indispensable for the analysis of gene expression in concert with histone acetylation.

A range of methods is now available for assessing gene regulation relevant to histone acetylation in a gene-specific or genome-wide manner. The relationship between a gene of interest and site-specific histone acetylation can be analyzed by chromatin immunoprecipitation (ChIP) using antibodies (preferably monoclonal) to identify specifically modified histones bound to DNA. To globally assess genes correlated with specific histone acetylation sites, the ChIP-on-chip method (ChIP in combination with microarray) is used. This strategy has been extensively described in other recent editions of this collection (34–36). Quantitative-PCR has been widely used to quantify specific gene expression changes, and microarray expression analysis has been successfully used in our laboratory to measure the global gene expression after HDAC inhibitor treatment (37). The methods and protocols for the analysis of the cellular effects of histone deacetylase inhibition in human breast cancer cells are described below.

2. Materials

2.1. Cell Culture, HDAC Inhibitors

1. Phosphate-buffered saline (PBS), pH 7.2.
2. Dulbecco's modified Eagle's medium (DMEM, Mediatech) containing 5% fetal bovine serum (Mediatech) used for culture of MDA-MB-231 human breast cancer cells.
3. Trypsin/EDTA solution: 0.05% trypsin/0.53 mM EDTA (Mediatech).
4. HDAC inhibitors: SAHA (suberoylanilide hydroxamic acid, vorinostat, Cayman); TSA (7-[4-(dimethylamino) phenyl]-*N*-hydroxy-4,6-dimethyl-7-oxo-2,4-heptadienamide; Sigma); MS275 (Selleck Chemicals); belinostat (PXD101, Selleck Chemicals); and panobinostat (LBH589, Selleck Chemicals). These HDAC inhibitors are dissolved in 100% DMSO and stored at –20°C (see Note 1).

2.2. Histone Isolation

1. PBS, pH 7.2.
2. Lysis buffer A: 10 mM Tris–HCl, 1 mM $MgCl_2$, 5 mM butyrate, 1% Triton X-100.
3. Buffer B: 0.25 M sucrose, 3 mM $CaCl_2$, 10 mM Tris–HCl, and 5 mM butyrate.
4. Sulfuric acid.
5. Acetone.

2.3. Nuclear Protein Extraction

1. NE-PER® nuclear and cytoplasmic extraction reagents (Thermo Scientific/Pierce).

2. Protease inhibitors: benzamidine 250 mg/ml; aprotinin 2 mg/ml; leupeptin 2 mg/ml; PMSF (phenlymethlysulfonyl flouride) 0.2 M.

2.4. Western Blots and Antibodies

1. Tris–HCl SDS-PAGE precast gels, gel running and protein transfer apparatus, and PVDF membrane.

2.4.1. Chemiluminescence Detection

2. ECL plus Western blotting detection system (GE Healthcare).

2.4.2. Odyssey Quantitative Fluorescence Detection

1. Infrared dye 800CW goat anti-mouse secondary antibody and infrared dye 680 goat anti-rabbit secondary antibody (Li-COR).
2. Odyssey blocking buffer (Li-COR).
3. Tween 20, PBS buffer, methanol, and SDS.
4. The Odyssey® infrared imaging system (Li-COR).

2.4.3. Histone Antibodies

Rabbit anti-acetyl-histone H3 polyclonal IgG (Millipore), Rabbit anti-acetyl-histone H4 polyclonal IgG (Millipore), Rabbit anti-acetyl-histone H3 (Lys 9) polyclonal IgG (Millipore), Rabbit anti-acetyl-histone H3 (Lys 27) polyclonal IgG (Millipore), Rabbit anti-histone H3 polyclonal IgG (Abcam), Goat anti-rabbit immunoglobulin/HRP (DAKO).

2.5. In vitro HDAC Activity Assay

1. HDAC assay buffer: 50 mM Tris–Cl, pH 8.0, 137 mM NaCl, 2.7 mM KCl, 1 mM $MgCl_2$.
2. HDAC substrate (Calbiochem/EMD).
3. HDAC developer concentrate (20×) (Calbiochem/EMD).
4. Deacetylated standard (Calbiochem/EMD).
5. 96-Well microplates.
6. Fluorimeter.

2.6. RNA Extraction, cDNA Synthesis, PCR

1. RNA extraction: TRIzol® reagent (Invitrogen).
2. First-strand cDNA synthesis: $Oligo(dT)_{12–18}$, dNTP Mix (dATP, dGTP, dCTP, and dTTP), 5× first-strand buffer, DTT, RNaseOUT™, M-MLV reverse transcriptase (Invitrogen).
3. PCR reagent: JumpStart™ Taq ready mix (Sigma).
4. Real-Time PCR reagents: SYBR green or Taqman® (Applied Biosystems).
5. Applied Biosystems Real-Time PCR system.

2.7. Chromatin Immunoprecipitation

1. 37% Formaldehyde.
2. SDS lysis buffer: 1% SDS, 10 mM EDTA pH 8.0, 50 mM Tris–HCl pH 8.1.
3. ChIP diluent buffer: 0.01% SDS, 1.1% Triton-X 100, 1.2 mM EDTA pH 8.0, 16.7 mM Tris–HCl pH 8.1, 167 mM NaCl.

4. Low salt buffer: 0.1% SDS, 1.0% Triton X-100, 2 mM EDTA pH 8.0, 20 mM Tris–HCl pH 8.1, 150 mM NaCl.
5. High salt buffer: 0.1% SDS, 1.0% Triton X-100, 2 mM EDTA pH 8.0, 20 mM Tris–HCl pH 8.1, 500 mM NaCl.
6. LiCl immune complex: 0.25 M LiCl, 1% NP40, 1% deoxycholate, 1% SDS, 1 mM EDTA pH 8.0, 10 mM Tris–HCl pH 8.1.
7. Elution buffer: 1% SDS, 0.1 M $NaHCO_3$.
8. Protein A agarose slurry/Salmon Sperm DNA (Millipore).
9. Protein A agarose beads (Millipore).
10. Magnetic separator.
11. Sonifier (Branson).

3. Methods

3.1. Histone Extraction and Isolation

Histone proteins from HDAC inhibitor-treated human breast cancer cells are isolated according to previously published method (29).

1. Treat human breast cancer cells with HDAC inhibitors.
2. Harvest cells by scraping or trypsinization and centrifugation at 500 × *g* for 5 min and wash cells three times in ice-cold PBS.
3. Remove the supernatant and resuspend in 1 ml lysis buffer A with protease inhibitor and transfer to a 1.5-ml eppendorf tube on ice for 30 min.
4. Centrifuge for 15 min at 16,000 × *g* at 4°C.
5. Resuspend the pellet in 250 µl buffer B with protease inhibitor.
6. Add 11 µl 3.8N H_2SO_4 to make final concentration of 0.4N and leave the tube at 4°C overnight.
7. Centrifuge at 16,000 × *g* for 15 min at 4°C and transfer the supernatant to a new tube and precipitate with 10× cold acetone at –20°C.
8. Centrifuge 16,000 × *g* 15 min at 4°C and wash the pellet with cold acetone containing 0.2% H_2SO_4.
9. Dry pellet and dissolve pellet in ddH_2O.

3.2. Nuclear Protein Extraction (See Note 2)

This protocol is derived from the published protocol from manufacturer (Thermo Scientific/Pierce).

1. Isolate approximately 20 µl packed cell volume of breast cancer cells treated with HDAC inhibitors or vehicle control by centrifugation at 500 × *g* for 2–3 min.
2. Add 200 µl of ice-cold CER I reagent with 1× protease inhibitor mixture.

3. Vortex vigorously for 15 s to fully resuspend the cell pellet and incubate the samples on ice for 10 min.
4. Add 11 µl of ice-cold CER II.
5. Vortex the tube for 5 s on the highest setting. Incubate tube on ice for 1 min.
6. Vortex the tube for 5 s and centrifuge the tube for 5 min at 16,000 × *g*.
7. Immediately transfer the supernatant (cytoplasmic extract) fraction to a clean prechilled tube and resuspend the insoluble pellet fraction containing nuclei in 100 µl of ice-cold NER with 1× protease inhibitor mixture.
8. Vortex for 15 s every 10 min, for a total of 40 min.
9. Centrifuge the tube at 16,000 × *g* for 10 min and transfer the supernatant (nuclear extract) fraction to a clean prechilled tube.
10. Store all extracts at –80°C until use.

3.3. Western Blot Analysis of Histone Acetylation

1. Treat cells with HDAC inhibitors.
2. Extract nuclear proteins or histone proteins as described above in Subheadings 3.1 and 3.2.

3.3.1. Chemiluminescence Western Blot

3. Dilute equal amounts of proteins in 2× SDS loading buffer, and boil at 95°C for 5 min.
4. Separate nuclear extract or histones on SDS-PAGE gels and transfer them onto PVDF membrane according to appropriate protocols.
5. Block blot in TBST with 5% nonfat dry milk for 2 h at room temperature or 4°C overnight.
6. Add the primary antibodies against acetylated histone proteins at a dilution of 1:2,000, and incubate the membranes with the diluted antibody at room temperature for 2 h.
7. Wash the membrane with TBST on a shaker with the revolution at 40–50 rpm for 3 × 10 min.
8. Incubate blot with secondary antibody (rabbit) at a concentration of 1:3,000 in TBST containing 5% nonfat dry milk at room temperature for 1.5 h.
9. Wash the blot three times for 10 min in TBST and once for 5 min in 1× TBS, and rinse with water.
10. Visualize the acetylated histones with the ECL kit.

3.3.2. Odyssey Western Blot (See Note 3)

1. Protein sample preparation, gel running and PVDF membrane transfer should be performed using standard blotting procedures as described in Subheading 3.3.1.
2. Place membrane in Odyssey blocking buffer (without Tween 20) for at least 1 h with gentle shaking at room temperature.

3. Incubate blot with primary antibodies in Odyssey blocking buffer with 0.1% Tween 20 for 2 h at room temperature or overnight at 4°C.
4. Rinse membrane with 1× PBST (0.2% Tween 20).
5. Incubate blot with infrared dye secondary antibody at a concentration of 1:5,000 in Odyssey blocking buffer with 0.1% Tween 20 and 0.01% SDS at room temperature for 60 min. Protect membrane from light during incubation.
6. Rinse membrane with 1× PBST (0.1% Tween 20). The membrane can be scanned wet or dry.

3.4. HDAC Activity Assay

In vitro HDAC activity assays are performed using the Calbiochem® HDAC activity assay kit according to the manufacturer's instructions. This method is an assay system to measure HDAC activity in whole cell or nuclear extracts, immunoprecipitates, or purified recombinant human HDACs. In this unit, we describe the method of in vitro measurement of nuclear HDAC activity using the peptide substrate comprising an acetylated side chain lysine residue and a bound fluorescent group (Calbiochem/EMD Chemical).

1. Treat cells with HDAC inhibitors.
2. Extract nuclear proteins using the methods as described in Subheading 3.2.
3. Prepare the standard curve of deacetylation. Optimize the concentration ranges of deaceylated standard.
4. Add HDAC assay buffer to appropriate wells of the 96-well plate.
5. Add diluted cell nuclear extract or other HDAC containing samples to designated wells in triplicate.
6. Add HDAC substrate to each well containing nuclear extract or "no enzyme control."
7. Mix thoroughly and incubate the plate at room temperature (~25°C) for 10–15 min.
8. Read samples in a fluorimeter at an excitation wavelength of 350–380 nm and an emission wavelength of 440–460 nm (see Note 4).

3.5. RT-PCR to Detect the Reexpression of Epigenetically Silenced Genes by HDAC Inhibitors in Human Breast Cancer Cells

Previous studies in our laboratory showed that pharmacological inhibition of histone deacetylation resulted in the expression of functional ERα mRNA and protein in ER negative human breast cancer cells (18–20). Here, we describe the method to detect the reactivation of ERα mRNA by SAHA treatment using RT-PCR in ER negative MDA-MB-231 cells. The RT-PCR primers and conditions for some other genes reactivated by HDAC inhibitors in human breast cancer cells are summarized in Table 3.

Table 3
RT-PCR primers and conditions for genes reactivated by HDAC inhibitors in human breast cancer cells

Gene	Primers	Annealing temperature (°C)	HDAC inhibitor	Reference
ERα	S: GCACCCTGAAGTCTCTGGAA AS: TGGCTAAAGTGGTGCATGAT	55	TSA Scriptaid LBH SAHA	(18–21)
PR	S: TCATTACCTCAGAAGATTTG TTTAATC AS: TGATCTATGCAGGACTAGACAA	60	TSA Scriptaid	(18–21)
SFRP1	S: GGCCCATCTACCCGTGTCG AS: GATGGCCTCAGATTTCAACTCGT	60	Splitomicin	(22)
SFRP2	S: AAGCCTGCAAAAATAAAAATGATG AS: TGTAAATGGTCTTGCTCTTGGTCT	53	Splitomicin	(22)
E-cadherin	S: CCGCCGGCGTCTGTAGGAA AS: AGGGCTCTTTGACCACCGCTCTC	57	Splitomicin	(22)
CRBP1	S: CATCCGCACGCTGAGCACTTTTAG AS: CACGCCCCTCCTTCTCACCCTTCT	58	Splitomicin	(22)

ERα estrogen receptor alpha, *PR* progesterone receptor, *SFRP* secreted frizzled-related protein, *CRPB1* cellular retinol binding protein 1

1. Treat ER negative human breast cancer cells (MDA-MB-231) with 1–10 μM SAHA for 24 h.
2. Rinse cells with ice-cold PBS and lyse cells directly by adding 1 ml of TRIZOL reagent (Invitrogen) per 10-cm² dish and scraping with cell scraper.
3. Add 0.2 ml of chloroform per 1 ml of TRIzol reagent. Vortex samples for 15 s and incubate them at room temperature for 2–3 min.
4. Centrifuge the samples at 12,000 × *g* for 15 min at 4°C and transfer upper aqueous phase containing RNA into nuclease-free tubes.
5. Add 0.5 ml of isopropyl alcohol per 1 ml of TRIZOL reagent to precipitate the RNA and centrifuge at 12,000 × *g* for 10 min at 4°C.
6. Wash the RNA pellet with 75% ethanol, centrifuge, and dissolve RNA in 100 μl DEPC-treated water. Measure the RNA concentrations.
7. First strand cDNA is synthesized by mixing 3 μg total RNA with 1 μl oligo $(dT)_{12-18}$ (500 μg/ml), and 1 μl of 10 mM dNTP mix (Invitrogen), then adding sterile ddH_2O to 12 μl.

8. Heat mixture to 65°C for 5 min and quickly chill on ice.
9. Add 4 µl 5× first strand buffer, 2 µl 0.1 M DTT and 1 µl RNaseOUT™ Recombinant Ribonuclease Inhibitor (40 U/µl) (Invitrogen), mix contents of the tube gently and incubate at 37°C for 2 min.
10. Add 1 µl (200 U) of M-MLV reverse transcriptase (Invitrogen) and incubate 50 min at 37°C followed by heating at 70°C for 15 min to inactivate the reaction.
11. Conventional PCR was performed in cDNA samples as previously described (20) using the following primers: ERα S: GCACCCTGAAGTCTCTGGAA; AS: TGGCTAAAGTGG TGCATGAT; Actin S: ACCATGGATGATGATATCGC; AS: ACATGGCTGGGGTGTTGAAG. Amplified products are analyzed on 1.5% agarose gels.

3.6. Chromatin Immunoprecipitation to Analyze Changes in Regulatory Chromatin Marks by HDAC Inhibitors at the Specific Gene Promoters

ChIP is a powerful tool to study protein–DNA interaction in HDAC inhibitor treated cells. This technique can map minute-by-minute changes of histone acetylation at a single promoter, or over the entire genome by using advanced ChIP on DNA microarray technology (ChIP-on-chip) (38). The protocol and reagents for ChIP used in our laboratory are described below as recommended by the manufacturer's instructions (Millipore/Upstate).

1. Treat 2×10^6 human breast cancer cells with 1–10 µM HDAC inhibitors for 24 h.
2. Crosslink DNA and proteins by adding 37% formaldehyde directly to growth media to a final concentration of 1%, and gently shake dishes to mix, and incubate at room temperature for 10 min.
3. Add glycine to a final concentration of 0.125 M to quench crosslinking reactions.
4. Wash cells with cold PBS containing 1× protease inhibitor and scrape cells from each dish into a microcentrifuge tube.
5. Spin at $500 \times g$ at 4°C for 2–5 min to pellet cells.
6. Resuspend cell pellet in 200 µl of SDS lysis buffer containing 1× protease inhibitor.
7. Sonicate cell lysate to shear DNA to ~200–1,000 bp in length (see Note 5). In between pulses, let samples sit on ice for at least 2 min.
8. Centrifuge samples at $14{,}000 \times g$ at 4°C for 10 min and transfer 200 µl of the sonicated cell supernatant to a new microcentrifuge tube.
9. Dilute the sonicated cell supernatant tenfold in 1.8 ml ChIP dilution buffer with 1× protease inhibitor.
10. Preclear the 2-ml diluted cell supernatant with 80 µl of salmon sperm DNA/Protein A Agarose beads (50% slurry) for 30 min at 4°C with agitation.

11. Separate beads by brief centrifugation and transfer the supernatant to a fresh tube (if using magnetic beads, beads are separated on magnetic rack).
12. Remove 10 μl (1%) of the supernatant as input and save at 4°C.
13. Collect the supernatant by aliquoting 1 ml into fresh microfuge tubes.
14. Add the immunoprecipitating antibodies against acetylated H3, H4, or specific lysine residues on histone tail to the supernatant fraction and incubate overnight at 4°C with rotation. Rabbit immunoglobulin G (IgG) and H3 antibodies are used for negative and quantitative controls, respectively (see Note 6).
15. Add 60 μl of Protein A Agarose beads (Millipore) and mix for 1 h at 4°C with rotation.
16. Separate beads by brief centrifugation or magnetic rack and remove the supernatant fraction.
17. Wash the beads once in 1 ml of low salt immune complex wash buffer (Millipore), once in high salt immune complex wash buffer (Millipore), once in LiCl Immune complex wash buffer (Millipore), and twice in TE buffer.
18. Add 100 μl of elution buffer (20% SDS, 1 M $NaHCO_3$) to each tube containing the antibody/bead complex. Mix and incubate at room temperature for 15 min. Transfer the eluate to fresh tube and wash the beads with 250 μl ChIP elution buffer. Repeat the wash and pool the eluates.
19. Add 20 μl 5 M NaCl to the pooled eluates and reverse cross-links at least 4 h at 65°C.
20. Add 1 μl of RNase A and incubate for 30 min at 37°C. Add 4 μl 0.5 M EDTA, 8 μl 1 M Tris–HCl and 1 μl Proteinase K, and incubate at 45°C for 1 h.
21. Extract the samples with phenol/chloroform (1:1), ethanol-precipitate DNA in the presence of 20 μg of glycogen, wash with 70% ethanol, and dissolve in 50 μl TE.
22. IP DNA can be further analyzed by quantitative Real-Time PCR to quantify alteration in acetylated histone marks at the promoter region of gene of interest (see Note 7). DNA immunoprecipitated by H3 antibody is used for normalization.

4. Notes

1. HDAC inhibitors should be dissolved in DMSO first for maximum solubility and then diluted in aqueous buffer of choice. Since some HDAC inhibitors, such as vorinostat, are unstable in aqueous media, we replace the drug-containing medium every day if the length of treatment time is more than 24 h.

2. This method has been successfully used to extract nuclear proteins and examine the expression level of nuclear histone proteins (39). The CER I reagent from cytoplasmic/nuclear protein extraction kit induces swelling of the cell leading to stress on the cellular membrane and the CER II reagent lyses the cell membrane, allowing cytoplasmic proteins to be collected. The NER reagent is then used to extract nuclear proteins from the intact nucleus. EDTA-free protease inhibitors can be used for the extraction agents. However, protease inhibitors that contain alcohols should be avoided.
3. By use of fluorescence-labeled antibodies rather than enzyme labels, Odyssey infrared image system quantitatively detects protein expression on the Western blot with wide linear dynamic range that cannot be achieved by conventional chemiluminescence. With two detection channels, multiple separate targets can be probed in the same experiment. Therefore, quantification accuracy is improved when the second channel is used for loading normalization. This method has been used in our recent study to quantify drug-induced changes in global histone marks (39).
4. This method has been successfully used with preparations of all class I and II HDACs and class III HDAC SIRT1. It is necessary to use a potent HDAC inhibitor, such as TSA, as a positive control in experiments. The exact concentration range of the deaceylated standard, substrate and inhibitors should be carefully optimized and determined.
5. It is important to optimize conditions for shearing crosslinked DNA to 200–1,000 bp in length. These conditions vary with different cell types, cell density, and the specific sonication equipment setting including the power output, duty cycle, and number of pulses. During the sonication, keep all the samples on ice to avoid the occurrence of protein denaturation.
6. It is possible that an anti-acetylated histone antibody used in ChIP will not recognize the epitope of the antigen in fixed chromatin. In such case, choose an antibody with higher affinity that has been validated as suitable for ChIP. It is important to use a negative control in every ChIP experiment (such as IgG) to detect nonspecific binding. If polyclonal antibodies are used, a control using unimmunized sera from the same species should be included.
7. ChIP Primers for silenced genes reactivated with HDAC inhibition in human breast cancer cells: ERα, forward: TGAACCGTCCGCAGCTCAAGATC and reverse: GTCTGACCGTAGACCTGCGCGTTG (19); SFRP1, forward: AGCCGCGTCTGGTTCTAGT and reverse: GGAGGCTGCAGGGCTG; E-cadherin, forward: TAGAGGGTCACCGCG TCTATG and reverse: GGGTGCGTGGCTGCAGCCAGG (22).

Acknowledgments

This work was funded by NIH SPORE grant CA88843 (to N.E.D.) and the Breast Cancer Research Foundation (to N.E.D.).

References

1. Feinberg, A. P. and Tycko, B. (2004) The history of cancer epigenetics. *Nat Rev Cancer.* **4**, 143–153.
2. Jones, P. A. and Baylin, S. B. (2007) The epigenomics of cancer. *Cell* **128**, 683–692.
3. Jenuwein, T. and Allis, C. D. (2001) Translating the histone code. *Science* **293**, 1074–1080.
4. Ho, L. and Crabtree, G. R. (2010) Chromatin remodelling during development. *Nature* **463**, 474–484.
5. Minucci, S. and Pelicci, P. G. (2006) Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. *Nat. Rev. Cancer* **6**, 38–51.
6. Marson, C. M. (2009) Histone deacetylase inhibitors: design, structure-activity relationships and therapeutic implications for cancer. *Anticancer Agents Med Chem.* **9**, 661–692.
7. Khan, O. and La Thangue, N. B. (2008) Drug Insight: histone deacetylase inhibitor-based therapies for cutaneous T-cell lymphomas. *Nat. Clin. Pract. Oncol.* **5**, 714–726.
8. Marks, P. A., Richon, V. M., Miller, T. and Kelly, W. K. (2004) Histone deacetylase inhibitors. *Adv.Cancer. Res.* **91**, 137–168.
9. Ficner, R. (2009) Novel structural insights into class I and II histone deacetylases. *Curr Top. Med. Chem.* **9**, 235–240.
10. Bradner, J. E., West, N., Grachan, M. L., Greenberg, E. F., Haggarty, S. J., Warnow, T. and Mazitschek, R. (2010) Chemical phylogenetics of histone deacetylases. *Nat. Chem. Biol.* **6**, 238–243.
11. Schuetz, A., Min, J., Allali-Hassani, A., Schapira, M., Shuen, M., Loppnau, P., Mazitschek, R., Kwiatkowski, N. P., Lewis, T. A., Maglathin, R. L., McLean, T. H., Bochkarev, A., Plotnikov, A. N., Vedadi, M. and Arrowsmith, C. H. (2008) Human HDAC7 harbors a class IIa histone deacetylase-specific zinc binding motif and cryptic deacetylase activity. *J. Biol. Chem.* **283**, 11355–11363.
12. Stimson, L., Wood, V., Khan, O., Fotheringham, S. and La Thangue, N. B. (2009) HDAC inhibitor-based therapies and haematological malignancy. *Ann Oncol.* **20**, 1293–1302.
13. Marks, P. A. and Breslow, R. (2007) Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. *Nat. Biotechnol.* **25**, 84–90.
14. Marks, P. A. and Xu, W. S. (2009) Histone deacetylase inhibitors: Potential in cancer therapy. *J. Cell. Biochem.* **107**, 600–608.
15. Campas-Moya, C. (2009) Romidepsin for the treatment of cutaneous T-cell lymphoma. *Drugs Today (Barc)* **45**, 787–795.
16. Nebbioso, A., Clarke, N., Voltz, E., Germain, E., Ambrosino, C., Bontempo, P., Alvarez, R., Schiavone, E. M., Ferrara, F., Bresciani, F., Weisz, A., de Lera, A. R., Gronemeyer, H. and Altucci, L. (2005) Tumor-selective action of HDAC inhibitors involves TRAIL induction in acute myeloid leukemia cells. *Nat. Med.* **11**, 77–84.
17. Peart, M. J., Smyth, G. K., van Laar, R. K., Bowtell, D. D., Richon, V. M., Marks, P. A., Holloway, A. J. and Johnstone, R. W. (2005) Identification and functional significance of genes regulated by structurally different histone deacetylase inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3697–3702.
18. Keen, J. C., Yan, L., Mack, K. M., Pettit, C., Smith, D., Sharma, D. and Davidson, N. E. (2003) A novel histone deacetylase inhibitor, scriptaid, enhances expression of functional estrogen receptor alpha (ER) in ER negative human breast cancer cells in combination with 5-aza 2'-deoxycytidine. *Breast Cancer Res. Treat.* **81**, 177–186.
19. Zhou, Q., Atadja, P. and Davidson, N. E. (2007) Histone deacetylase inhibitor LBH589 reactivates silenced estrogen receptor alpha (ER) gene expression without loss of DNA hypermethylation. *Cancer Biol. Ther.* **6**, 64–69.
20. Yang, X., Ferguson, A. T., Nass, S. J., Phillips, D. L., Butash, K. A., Wang, S. M., Herman, J. G. and Davidson, N. E. (2000) Transcriptional activation of estrogen receptor alpha in human breast cancer cells by histone deacetylase inhibition. *Cancer Res.* **60**, 6890–6894.
21. Sharma, D., Saxena, N. K., Davidson, N. E. and Vertino, P. M. (2006) Restoration of tamoxifen sensitivity in estrogen receptor-negative

breast cancer cells: tamoxifen-bound reactivated ER recruits distinctive corepressor complexes. *Cancer Res.* **66**, 6370–6378.

22. Pruitt, K., Zinn, R. L., Ohm, J. E., McGarvey, K. M., Kang, S. H., Watkins, D. N., Herman, J. G. and Baylin, S. B. (2006) Inhibition of SIRT1 reactivates silenced cancer genes without loss of promoter DNA hypermethylation. *PLoS Genet.* **2**, e40.
23. Jung, J. W., Lee, S., Seo, M. S., Park, S. B., Kurtz, A., Kang, S. K. and Kang, K. S. (2010) Histone deacetylase controls adult stem cell aging by balancing the expression of polycomb genes and jumonji domain containing 3. *Cell Mol. Life Sci.* **67**, 1165–1176.
24. Zhou, Q., Chaerkady, R., Shaw, P. G., Kensler, T. W., Pandey, A. and Davidson, N. E. (2010) Screening for therapeutic targets of vorinostat by SILAC-based proteomic analysis in human breast cancer cells. *Proteomics* **10**, 1029–1039.
25. Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L., Grishin, N. V., White, M., Yang, X. J. and Zhao, Y. (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell.* **23**, 607–618.
26. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V. and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840.
27. Gnad, F., Ren, S., Choudhary, C., Cox, J. and Mann, M. (2010) Predicting Posttranslational Lysine Acetylation Using Support Vector Machines. *Bioinformatics* **26**,1666–1668.
28. Beck, H. C. Mass spectrometry in epigenetic research. Methods Mol Biol. **593**, 263–282
29. Yoshida, M., Kijima, M., Akita, M. and Beppu, T. (1990) Potent and specific inhibition of mammalian histone deacetylase both in vivo and in vitro by trichostatin A. *J. Biol. Chem.* **265**, 17174–17179.
30. Cousens, L. S., Gallwitz, D. and Alberts, B. M. (1979) Different accessibilities in chromatin to histone acetylase. *J. Biol. Chem.* **254**, 1716–1723.
31. Jackson, V., Shires, A., Chalkley, R. and Granner, D. K. (1975) Studies on highly metabolically active acetylation and phosphorylation of histones. *J. Biol. Chem.* **250**, 4856–4863.
32. Rodriguez-Collazo, P., Leuba, S. H. and Zlatanova, J. (2009) Robust methods for purification of histones from cultured mammalian cells with the preservation of their native modifications. *Nucleic Acids Res.* **37**, e81
33. Young, N. L., DiMaggio, P. A., Plazas-Mayorca, M. D., Baliban, R. C., Floudas, C. A. and Garcia, B. A. (2009) High throughput characterization of combinatorial histone codes. *Mol. Cell. Proteomics.* **8**, 2266–2284.
34. Tong, Y. and Falk, J. (2009) Genome-wide analysis for protein-DNA interaction: ChIP-chip. *Methods Mol Biol.* **590**, 235–251.
35. Reimer, J. J. and Turck, F. (2010) Genome-wide mapping of protein-DNA interaction by chromatin immunoprecipitation and DNA microarray hybridization (ChIP-chip). Part A: ChIP-chip molecular methods. *Methods Mol Biol.* **631**, 139–160.
36. Gobel, U., Reimer, J. and Turck, F. (2010) Genome-wide mapping of protein-DNA interaction by chromatin immunoprecipitation and DNA microarray hybridization (ChIP-chip). Part B: ChIP-chip data analysis. *Methods Mol. Biol.* **631**, 161–184.
37. Keen, J. C., Garrett-Mayer, E., Pettit, C., Mack, K. M., Manning, J., Herman, J. G. and Davidson, N. E. (2004) Epigenetic regulation of protein phosphatase 2A (PP2A), lymphotactin (XCL1) and estrogen receptor alpha (ER) expression in human breast cancer cells. *Cancer Biol. Ther.* **3**, 1304–1312.
38. Weinmann, A. S. and Farnham, P. J. (2002) Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. *Methods* **26**, 37–47.
39. Huang, Y., Greene, E., Murray Stewart, T., Goodwin, A. C., Baylin, S. B., Woster, P. M. and Casero, R. A., Jr. (2007) Inhibition of lysine-specific demethylase 1 by polyamine analogues results in reexpression of aberrantly silenced genes. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 8023–8028.

Chapter 23

Computational Methods for Epigenetic Analysis: The Protocol of Computational Analysis for Modified Methylation-Specific Digital Karyotyping Based on Massively Parallel Sequencing

Jian Li, Qian Zhao, and Lars Bolund

Abstract

Massively parallel sequencing technology opens new possibilities for epigenetic research. Many methods have been developed based on the new sequencing platforms, allowing an ultra-deep mapping of epigenetic variants in a fast and cost-effective way. However, handling millions of short reads produced by these sequencing platforms is a huge challenge for many laboratories. Thus, there is a need for the development of accurate and fast computational tools for epigenetic studies in the new era of genomic sequencing.

Modified methylation-specific digital karyotyping (MMSDK) is an improved method for genome-wide DNA methylation profiling based on the combination of traditional MSDK and Illumina/Solexa sequencing. Here, we introduce our computational tools used in the MMSDK analysis process from the experimental design to statistical analysis. We have developed a mapping process based on the in silico simulation of combined enzyme cutting and tag extraction of the reference genome. Subsequently, the 20–21 nucleotides (nt) long tags obtained by sequencing are mapped to the simulated library using an open source software Mapping and Assembly with Qualities. Our computational methods include trimming, annotation, normalization, and counting the reads to obtain digital DNA methylation profiles. We present the complete protocol and discuss some important issues that should be considered by readers, such as handling of repeat sequences, SNPs, and normalization. The core part of this protocol (mapping and annotation of tags) is suitable for any tag profiling-based methods, and it could also be modified to analyze results from other types of epigenetic studies based on massively parallel sequencing.

Key words: Computational methods, Massively parallel sequencing, MMSDK

1. Introduction

The advent of massively parallel sequencing platforms such as 454 Life Sciences (Roche) (1), Illumina/Solexa (2), and Applied Biosystems SOLiD (3) opens opportunities for the development of a variety of methods for epigenetic studies. These novel sequencing platforms allow ultra-deep sequencing for the mapping of epigenetic

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5_23, © Springer Science+Business Media, LLC 2011

information. However, various bioinformatic challenges arise from the analysis of numerous short reads and data amounts that are several magnitudes higher than that traditionally obtained by Sanger sequencing. Thus, there is an urgent need for fast and accurate computational methods to overcome this bottleneck and realize the full potential of the improved sequencing technology.

DNA methylation is the most widely studied epigenetic mechanism known to correlate with gene expression and genomic stability. Currently, there are three main approaches that are compatible with massively parallel sequencing for genome-wide mapping of DNA methylation information: (1) endonuclease digestion; (2) affinity enrichment; and (3) bisulphate conversion (4). All the above methods involve a three-stage analysis process (see Fig. 1a): Firstly, the epigenetic information is biochemically converted into genetic information, through the treatment of genomic DNA with one of the three methylation-dependent procedures. Subsequently, massively parallel sequencing technology is applied. Finally, computational algorithms are used to infer the DNA methylation information from numerous sequenced reads (4, 5). Mapping and annotation of numerous short reads to the reference genome is the central issue for the computational methods. The challenge comes not only from the requirement of highly efficient algorithms but also from the need of accuracy (6). Here, we have chosen the endonuclease digestion-based method, modified methylation-specific digital karyotyping (MMSDK) (7), as an example to introduce our computational methods for mapping DNA methylation in the human genome.

MMSDK is based on traditional MSDK, but the tags are analyzed by Illumina/Solexa sequencing (7) (see the left side of Fig. 1b and Note 1). Briefly, genomic DNA is tandemly digested by a methylation-sensitive enzyme (mapping enzyme, e.g., MluI (cutting site: ACGCGT), total 21,309 recognition sites in the human genome) and a fragmenting enzyme (e.g., NlaIII (cutting site: ACTG), total 13,787,676 recognition sites in the human genome). The resulting DNA fragments are linked with designed adaptors that contain the recognition sequence of a tagging enzyme (MmeI). MmeI generates 20–21 nt tags from the fragments. The tags are subsequently linked with another adaptor. As both ends of the tags have been linked with recommended sequencing adaptors, the tags can directly be amplified with PCR followed by single-end Illumina/Solexa sequencing (7). In subsequent computational analysis, the sequence of the tags can easily be derived through the identification and removal of adaptor sequences. Similar to long serial analysis of gene expression (LongSAGE) (8), 20–21 nt tags isolated from genomic DNA contain sufficient information to allow mapping of each tag sequence to the genomic locus from which it was derived. To increase confidence and efficiency of mapping, we developed a mapping process based on the simulation of

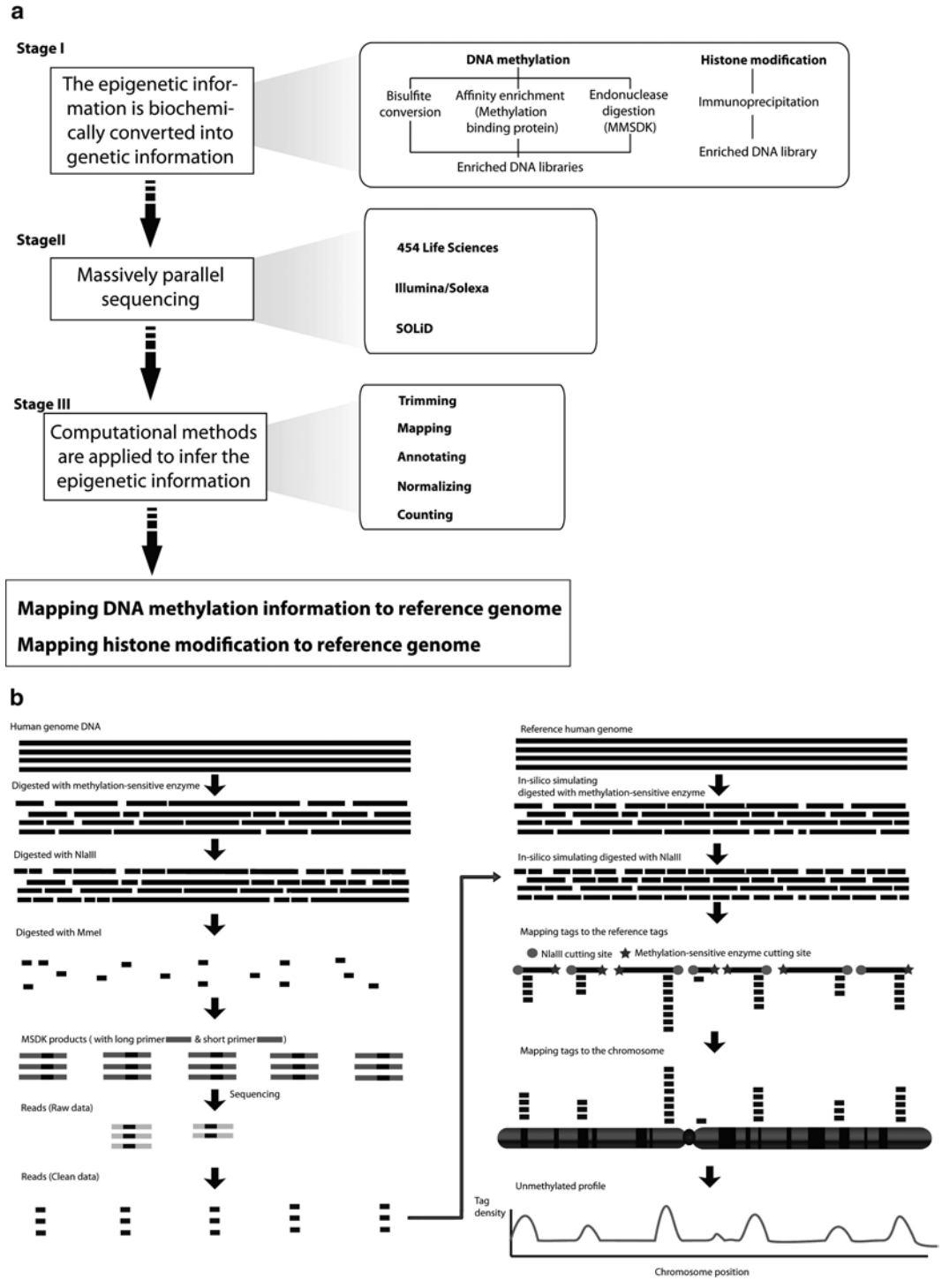


Fig. 1. (**a**) The schematic outline of computational analysis for epigenetic study. (**b**) Schematic workflow of the mapping strategy for MMSDK.

the combined enzyme digestion of the human reference genome to generate a mapping reference library for a given set of methylation-sensitive, fragmenting and tagging enzymes (see the right side of Fig. 1b). Subsequently, the open source software MAQ (Mapping and Assembly with Qualities) was used to map all observed tags to this in silico simulated reference library (6, 9). If SNPs, sequencing errors, and repeat sequences are properly handled, the tags with unique sequences can be identified as the corresponding genomic elements based on the alignment with the annotated reference genome. The tags with repeat sequences can also be classified to their corresponding repeat families based on the alignment with a repeat sequence database. The mapped tags are ordered sequentially by genomic position, and groups of tags are counted to quantify the extent of methylation along each chromosome.

The core part of this protocol (mapping and annotation of tags) is useful for any tag profiling-based methods such as digital gene expression and digital karyotyping, thereby providing a direct analysis of the relationship between DNA methylation, gene expression, and DNA copy number variation. The protocol could be also modified for the analysis of the results obtained from other massively parallel sequencing-based epigenetic approaches such as ChIP-seq (10, 11), bisulfite methylome sequencing (12), and methyl-DNA immunoprecipitation (MeDIP) (13). Thus, our work provides useful computational tools to bridge the gap between epigenetics and the application of massively parallel sequencing technology.

2. Materials

2.1. Biological Materials

In this protocol, we take two human cell lines (named A and B) as samples. Genomic DNA is isolated from the two cell lines as previously described (7).

2.2. Computational Equipment

We suggest a 64-bit machine for the computations (see Note 2). A Linux system is also recommended since the computational methods introduced here are developed based on this type of system.

3. Methods

3.1. Construction of MMSDK Tag Libraries

Genomic DNA is digested with a methylation-sensitive mapping enzyme. Here, we use MluI (New England Biolabs) as an example. MluI has two CpG sites in its recognition sequence, ACGCGT, and therefore its recognition site is preferentially located in CpG islands (CGIs). Digested DNA is ligated to biotinylated linkers and fragmented by NlaIII (New England Biolabs) cleavage. Since MluI

is only able to cut unmethylated regions, binding of DNA fragments to streptavidin-conjugated magnetic beads will separate the unmethylated fragments from the methylated. Bound DNA is linked to an adaptor (N) that contains a MmeI restriction enzyme recognition site. Subsequently, the linked DNA fragment is digested with MmeI (New England Biolabs) that produces short sequence tags (20–21 nt, due to enzyme cut floating). The cut ends of the tags are linked with another adaptor (P7), followed by PCR amplification by 15 cycles (7) (see the left side of Fig. 1b).

3.2. Massively Parallel Sequencing Using Illumina/Solexa 1 G Genome Analyzer

Clusters are generated from single-end DNA fragments using Illumina Cluster Generation (Illumina) and Sequencing-By-Synthesis (SBS) strategy is performed using 1 G Genome Analyzer (Illumina) according to the manufacturer's instructions (7). All reagents for the sequencing process are purchased from Illumina Inc. The libraries are sequenced using the single-end read protocol of Illumina Genome 1 G Analyzer. Base calling is performed by the Illumina instrument software.

3.3. Computational Protocol of MMSDK

3.3.1. The Option of Methylation-Sensitive Restriction Enzymes

Cutting of the human reference genome by the most commonly used methylation-sensitive restriction enzymes was simulated and, the resulting genomic information is provided in Table 1. Readers could select the methylation-sensitive restriction enzymes which depends on the resolution requirement of their experiments: if readers want to get higher resolution of methylation profile for a given sample, we recommend the enzyme that has more recognition sites in the human genome, but that would require large sequencing amount. Otherwise, insufficient sequencing would lead to low coverage; if readers want to sequence more samples with relatively small sequencing amount, we recommend the enzyme with rare recognition sites, by which readers can use bar code (index) system to sequence more than one sample in a single lane. Alternatively, readers can get a high sequencing depth for each recognition site without using bar code system. Thus, readers can make a cost benefit analysis and choice of enzymes for their experimental design based on the information given in Table 1. Since we take MluI enzyme as an example in this chapter, the detailed information on the human reference genome digested by MluI is shown in Fig. 2, by which the distributions of the distances of the recognition site of MluI in the human genome and CGIs have been illustrated.

3.3.2. Identifying and Trimming Reads (Tags)

1. 35 nt single-end reads are generated by the Illumina/Solexa sequencer.
2. According to the experimental design, 16–17 nt tags together with the neighboring 4 nt (the overhang of the adaptor N that is complementary to the NlaIII recognition sequence) are eligible

Table 1
The human genome digested by the commonly used methylation-sensitive restriction enzyme

Enzymes	Recognition	Nr recongnition sites	Nr CGI involved	% CGI invovled	Nr fragments (CGI)
AciI	C′CGC	2,041,173	27,008	95.68	331,220
HpaII	C′CGG	2,321,290	26,975	95.57	298,094
HhaI	GCG′C	1,612,632	26,539	94.02	288,170
HinP1I	G′CGC	1,612,632	26,539	94.02	288,170
BstUI	CG′CG	693,646	25,919	91.83	243,065
HpyCH4IV	A′CGT	2,167,391	17,955	63.61	58,209
SmaI	CCC′GGG	378,855	16,724	59.25	60,040
TspMI	C′CCGGG	378,855	16,724	59.25	60,040
KasI	G′GCGCC	231,709	15,398	54.55	49,658
NarI	GG′CGCC	231,709	15,398	54.55	49,658
SfoI	GGC′GCC	231,709	15,398	54.55	49,658
SacII	CCGC′GG	66,312	15,359	54.41	51,775
NaeI	GCC′GGC	124,438	14,416	51.07	43,549
NgoMIV	G′CCGGC	124,438	14,416	51.07	43,549
EagI	C′GGCCG	89,339	14,179	50.23	44,705
BssHII	G′CGCGC	59,946	13,390	47.44	43,981
AfeI	AGC′GCT	109,855	6,608	23.41	15,263
FspI	TGC′GCA	88,653	6,103	21.62	14,033
PmlI	CAC′GTG	278,675	5,461	19.35	12,818
NotI	GC′GGCCGC	9,776	4,982	17.65	11,527
PaeR7I	C′TCGAG	121,324	4,530	16.05	10,041
BmgBI	CAC′GTC	136,268	4,169	14.77	9,344
AatII	GACGT′C	70,665	3,843	13.62	8,576
ZraI	GAC′GTC	70,665	3,843	13.62	8,576
FseI	GGCCGG′CC	13,412	2,876	10.19	6,208
NruI	TCG′CGA	14,791	2,739	9.70	5,860

(continued)

Table 1 (continued)

Enzymes	Recognition	Nr recongnition sites	Nr CGI involved	% CGI invovled	Nr fragments (CGI)
AscI	GG′CGCGCC	4,696	2,619	9.28	5,631
MluI	A′CGCGT	21,309	2,432	8.62	5,345
AgeI	A′CCGGT	53,417	1,923	6.81	4,024
PvuI	CGAT′CG	12,448	1,557	5.52	3,222

The column "Enzyme" displays commonly used methylation-sensitive restriction enzymes
The column "Recognition" shows the recognition sequence for each enzyme
The column "Nr recognition sites" displays the total number of recognition sites for each enzyme in the human genome
The column "Nr CGI involved" displays the total number of CpG islands that can be theoretically investigated by a given enzyme in the human genome
The column "% CGI involved" displays the percentage of CpG islands that can be theoretically investigated by a given enzyme in the human genome
The column "Nr fragments (CGI)" shows the total number of fragments that can be theoretically obtained after the digestion by a given enzyme in the human genome

to be mapped to the reference for deciphering the methylation status. However, the length of reads is 35 nt, which means that there is a 18–19 nt tail at the end of each read belonging to the sequencing adaptor P7. We coded a dynamic programming algorithm implemented in Perl to trim the 18–19 nt tail. The scores are defined as +1 for base hit and −1 for mismatch. The gaps are not allowed in the alignment. The alignment is performed, and the score is calculated for each alignment (see Fig. 3). Finally, the matched region with highest score will be trimmed by removing 18–19 nt adaptor sequence as shown in Fig. 3.

3.3.3. Construction of a Simulated MluI Reference Library

We developed an analysis pipeline for tag alignment based on the principle of MMSDK. The whole strategy of computational analysis is presented in Fig. 1b. There are 21,309 MluI sites in the human genome to be identified by our in silico simulation. According to Repeatmasker (14), only 353 (1.66%) of the MluI sites are located within repeat sequences. Figure 4 shows the definition of the MluI reference library. We generated a MluI reference library by combined enzyme digestion (MluI and NlaIII) in silico. Subsequently, we use this simulated library instead of the whole human genome for MMSDK tag mapping. The usage of the simulated library allows us to map tags more accurately and efficiently, because most of high-confidence mapped tags (mapping quality score more than 20 by MAQ) have low accuracy when using the whole human genome (only about 20% of them are unique).

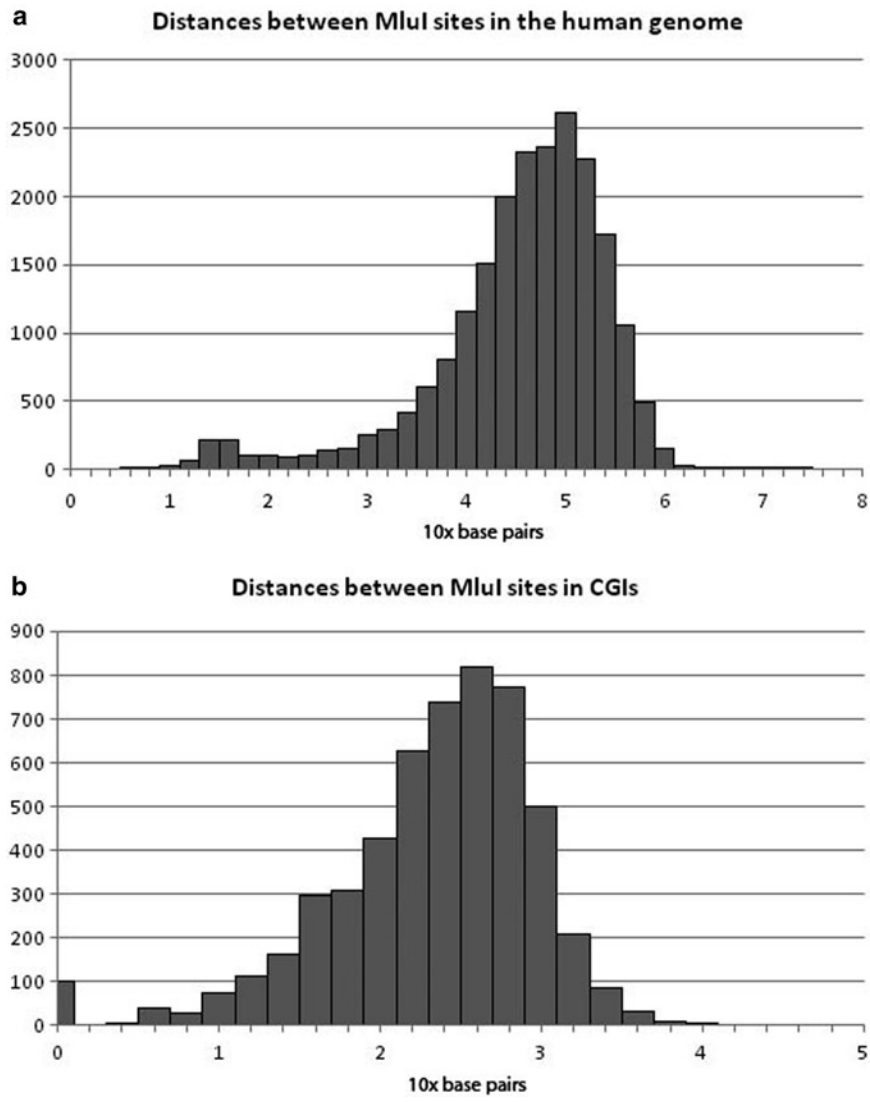


Fig. 2. MluI cutting of the human genome. (**a**) Distribution of the distances of neighboring MluI recognition sites in the human genome (the *x*-axis presents the distance, shown in the scale of $\log_{10}$ of the number of base pairs; the *y*-axis presents the count of the DNA fragments). (**b**) Distribution of the distances of neighboring MluI recognition sites in the CpG islands (the *x*-axis shown in the scale of $\log_{10}$ of the number of base pairs; the *y*-axis presents the count of the DNA fragments).

However, by using our simulated human genome as reference, the unique mapping tags increased to more than 70%. Furthermore, a smaller simulated human genome could improve the mapping efficiency. Therefore, using simulated human genome could increase the confidence and efficiency of mapping.

In a MluI reference library, one end of each reference should be CATG (the recognition site of fragmenting enzyme-NlaIII) and the other end should be ACGCGT (the recognition site of

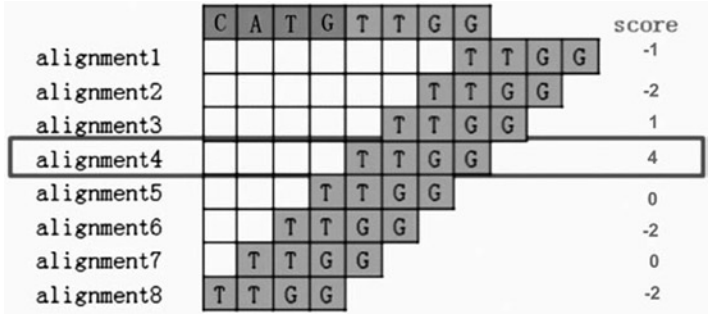


Fig. 3. The process of identifying the adaptor sequence. The red (*dark grey*) block (CATG) represents the adaptor sequence and the green (*light grey*) block (TTGG) represents reads from MMSDK generated by Illumina/Solexa sequencing. The highest score is 4 and this alignment is used to trim the sequence by removing the adaptor tail.

methylation-sensitive restriction enzyme-mapping enzyme-MluI). In Fig. 4a, b, two different situations are illustrated, when at least one ACGCGT site exists between two nearby CATG sites. In Fig. 4a, the two clusters of MMSDK tags would be available to estimate the DNA methylation status of one and the same MluI site between the two CATG sites. In Fig. 4b, the two clusters of MMSDK tags would represent two different MluI sites next to each CATG site. Figure 4a, b represents the situation when MluI sites are unmethylated so that the MMSDK tags would be produced and identified by sequencing. If a MluI site is methylated (Fig. 4c), the corresponding MMSDK tags would not exist in the MMSDK library. Thus, the count of the tags representing a particular MluI site is a measure of its degree of methylation in the genome. The smaller number of mapped tags counts the higher degree of methylation of the site in question. Some blind spots cannot be scanned by MMSDK as shown in Fig. 4d. The red (middle) ACGCGT site, which is between two other ACGCGT sites without any nearby CATG site, could never be identified with the present strategy using MluI and NlaIII. However, we found only 369 MluI sites lacking neighboring NlaIII sites.

We coded a Perl script to generate the MluI reference library (see *Additional file*) and all other related Perl scripts described in this chapter are also provided in *Additional file*.

1. At first, we constructed the library of simulated MluI digested DNA fragments by splitting the human genome at all potential cutting sites of MluI (i.e., all sites containing the sequence "ACGCGT" in the human genome).
2. Subsequently, we performed the second digestion simulation by splitting the above library at all cutting sites of NlaIII (i.e., the sequence of "CATG").
3. Finally, only the DNA fragments containing ACGCGT and CATG at each end were selected as the MluI reference library for the following mapping procedure.

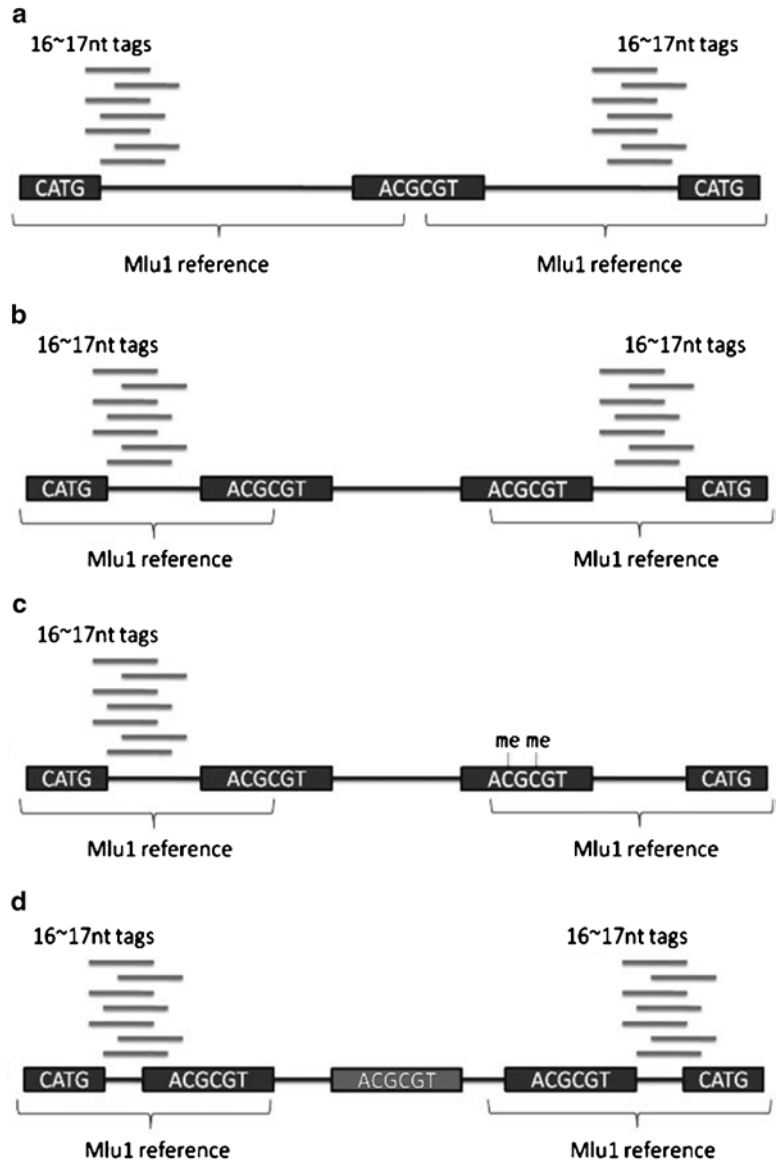


Fig. 4. Principles of the simulated MluI reference library.

3.3.4. Mapping Tags to the Simulated MluI Reference Library

We used MAQ (Mapping and Assembly with Qualities) to align the MMSDK tags to the simulated MluI reference library (see Note 3). MAQ is an open source software to build mapping assemblies from short reads generated by massively parallel sequencing platforms (see Note 4). The detailed information about the commands and the manual of MAQ can be found in the references (6, 9). Here, we just give a brief description.

1. Downloaded MAQ from the website in ref. 9.
2. Install and start according to the manual in ref. 9.

3. Convert reference sequences to the binary fasta (BFA) format (here the reference is the simulated MluI reference library), and convert reads to binary fastq (BFQ) format (9) (see Note 5).
4. Use MAQ to align reads in the BFQ format to the reference in the BFA format. Build the mapping assemblies, based on statistics from the resulting alignment, using the corresponding commands of MAQ (9) (see Note 6).
5. Use MAQ to extract and display consensus sequences and qualities (quality score) as well as a list of SNPs (9) (see Note 7).
6. Remove all the tags which were mapped to the simulated MluI reference library with the quality score < 20. According to the manual of MAQ, tags with mapping quality 20 should have an error rate of 1% and tags with a mapping quality higher than 20 should have an even lower alignment error rate on average. SNPs and sequencing errors are considered in the calculation of this mapping quality score. In another words, we considered the mapped tags with quality score ≥ 20 as uniquely and accurately aligned. Thus, these tags were eligible to calculate the tag counts for each MluI reference library (see Note 8).

3.3.5. Identifying Differentially Methylated Regions

MMSDK can be used to identify differentially methylated regions (DMRs) between different samples and create a DNA methylation profile for each independent sample.

1. Using the described protocol, we generated 5,432,906 and 5,636,928 tags from the sample A and sample B, respectively.
2. To compare the methylation profiles for the two samples, we need to perform a normalization process for the both cell lines. We assumed that total number of sequenced tags obtained from both cell lines was a constant number. Based on this hypothesis, we normalized the total number of sequenced (raw) tags from both cell lines to 1 M tags. If, for example, one MluI reference site contained 100 eligible tags in sample A, we regarded 100/5.432906 as the normalized tag count for sample A. Similarly, if the MluI reference site contained 100 eligible tags in sample B, the normalized tag count was 100/5.636928.
3. The normalized tag number represents the comparative methylation level that can be used to identify the DMRs between the sample A and sample B (see Note 9).
4. DMRs are identified by means of pair-wise comparison of the normalized libraries between the two samples. A *Z*-score test was used to calculate the *P*-value when estimating the significance for each MluI site in the pair-wise comparison.
5. A false discovery rate (FDR) control was added for the correction of the *P*-values to address the multiple comparisons problem inherent in high-throughput techniques (see Note 10).

6. Based on the FDR adjusted *P*-values, the DMRs were identified in the comparison of the two samples.

3.3.6. Annotating Tags with the Human Genomic Information

Gene core promoter regions and CpG island (CGI) regions are the most common genomic components for DNA methylation analysis, because their methylation states are affecting the related gene expression (see Note 11). In addition, Lister Ryan demonstrated a positive correlation between the methylation level in the gene body and the gene expression level (15). Thus, the DNA methylation status of the gene body is also important information that needs to be investigated. The DNA methylation state in regions of repeat sequences plays a key role in maintaining genomic stability and oncogenesis (16). Therefore, it is also valuable to measure DNA methylation levels of repeat regions and, especially, of transposon elements.

1. The human genome sequence and mapping information (Santa Cruz human genome assembly (hg18), March 2006) was downloaded from the University of California, Santa Cruz Genome Bioinformatics Site (17). The corresponding gene annotation was downloaded from the National Center for Biotechnology Information "The Reference Sequence (RefSeq)" database (18). The definition and classification of repeat sequences in this protocol are based on the information of the downloaded database from Repeatmasker (14). We defined gene core promoters as the regions located in the upstream 2 kb from transcript starting sites (TSS) and the first exon. We adopted the same criteria as used by the UCSC Genome Browser for the definition of CGIs (17) (see Note 11).
2. In our MMSDK analysis, we annotated the DMRs based on the genomic information of the simulated MluI reference library, and we used the methylation status of each MluI site to represent the corresponding genomic region in which this MluI site is located.

3.3.7. The Statistical Analyses

We recommend the readers to use R statistical tools (R 2.1.1 package) to perform further statistical analysis. We have provided some commonly applied statistical commands that could be used in MMSDK analysis (see *Additional file*).

4. Notes

1. DNA samples are generally derived from mixtures of cell populations with heterogeneous DNA methylation profiles. Our present method provides an average measurement across the sampled DNA molecules for particular loci (methylation-sensitive restriction enzyme sites).

2. Our Perl script and MAQ commands are best compiled and used on a 64-bit machine because it frequently uses 64-bit integers (9). Compiling as a 32-bit executable will work but the speed will be affected (9).
3. We applied the MAQ method to map the MMSDK tags to the simulated MluI reference library. MAQ is an open source software to map short DNA reads to reference genomes, with a special emphasis on mapping quality (6). The manual and commands are provided in the MAQ website (9). Readers can get more detailed information from the above website.
4. MAQ is specifically designed for the Illumina/Solexa sequencer to achieve fast aligning of short reads to a reference genome. It also has preliminary functionality to handle SOLiD data. However, MAQ also has some limitations: MAQ cannot perform de novo assembly and cannot align reads longer than 63 bp such as the reads from capillary sequencing or 454 Life Sciences instruments (Roche) (9).
5. Since the format of raw reads generated by Illumina/Solexa is different from the one required by MAQ, users need to first convert the format prior to using MAQ. The detailed information is available in the MAQ website (9).
6. For small-scale datasets (e.g., one lane reads), readers are recommended to run Easyrun script to decrease analysis time.
7. Interestingly, the methylation state of CpG dinucleotides can be associated with linked SNPs (19). Therefore, sequence-based technologies including MMSDK that can provide allele-specific DNA methylation information are preferred to allele-agnostic methylation assays (19).
8. An important caveat is that the process of mapping tags to a reference genome can bias the analysis toward genomic regions with unique and complex sequence patterns. This is because short sequencing reads that (partially) overlap with low-complexity regions or with interspersed repeats stand a higher chance of being discarded due to lack of unique genomic alignment (5). Repeat region, a very dense SNP region, rearrangements and amplification may affect the mapping result, and there was no efficient computational way to solve these problems so far. Readers need to consider about this issue when they run their own data. However, by using our computational method, most of the ambiguous mapping result could be avoided. First, the mapping quality generated by MAQ has already considered the repeat region, the base quality of the read, and mismatch. And we only focused on those tags with high mapping quality score. In fact, our result shows that over 80% of all the tags are high-confidence mapped tags. Thus to a certain extent, these unambiguous data are representative of all the

sequencing tags; Second, we constructed simulated combined enzyme digested human genome instead of the whole human genome as mapping reference. Therefore, mapping to the much smaller reference could avoid many regions that may along with these problems. Third, few of some enzyme recognition sites (e.g., ACGCGT, 1.66%) were identified in repeat region.

9. Enzyme-based analyses and enrichment techniques both rely on the relative enrichment or depletion of regions of the genome. The occurrence of aneuploidy, such as in cancer cells, can lead to inaccuracies in methylation measurements, if the influence of copy-number variations is not appropriately controlled (4). To address this problem, a control produced from the experimental sample, such as DNA artificially methylated by M.SssI (20) or digested with MspI (21) or other approaches serving for revealing DNA copy number alterations (such as array CGH, digital karyotyping, and low coverage genomic sequencing) can be used to normalize copy number variations. Bisulphite-based methods are less susceptible to copy-number variation.
10. *Z*-score test was used to calculate the *P*-value for each MluI site in the genome by means of pair-wise comparison. FDR was calculated for the correction of the *P*-values to address the multiple comparisons. Briefly, the calculation procedure is (22): (1) Order *P*-values of pair-wise comparisons in increasing order and denote them by $P(1)$, $P(2)$, …, $P(m)$; (2) For each of these *P*-values, its FDR = $P(k) \times m/k$.
11. The definitions of promoters are somewhat unclear. However, several alternative prediction methods for annotating functional promoters have been developed and evaluated (23), which substantially improves the accuracy of promoter annotation (23). CpG island prediction to some extent overlaps with promoter prediction because the majority of promoters in mammalian genomes co-localize with CpG islands. However, CpG islands play a more general role as mediators of open chromatin structure, and they also overlap with enhancers and other regulatory elements (5). CpG islands are often predicted solely based on their GC content, CpG frequencies, and region length (24), and multiple variants of the original definition are in use (25). The definition of CGIs in this protocol is based on the criteria of UCSC (GC content >50%, ratio of the observed CpGs to the expected CpGs >0.6, length >200 bp). The readers can make a choice among the prediction methods to define promoters and CGIs according to the purpose of their studies.

Acknowledgments

This work is supported by the project "Molecular Tools for Optimal Personalized Treatment of Breast Cancer" under the auspices of Sino-Danish Breast Cancer Research Centre, financed by the Danish "Grundforskningsfonden" and "The Natural Science Foundation of China." Support was also received from the project "Molecular Tools for Optimal Personalized Treatment of Colorectal Cancer" (FØSU), Danish Centre for Translational Breast Cancer Research (DCTB), "A Race Against Breast Cancer" (ARABC), the OAK foundation and "Den Bøhmske Fond."

References

1. Margulies, M., Egholm, M., Altman, W. E., et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. **437**, 376–380.
2. Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. **456**, 53–59.
3. Shendure, J., Porreca, G. J., Reppas, N. B., et al. (2005) Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*. **309**, 1728–1732.
4. Laird, P. W. (2010) Principles and challenges of genome-wide DNA methylation analysis. *Nat Rev Genet*. **11**, 191–203.
5. Bock, C., Lengauer, T. (2008) Computational epigenetics. *Bioinformatics*. **24**, 1–10.
6. Li, H., Ruan, J., Durbin, R. (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome research*. **18**, 1851–1858.
7. Li, J., Gao, F., Li, N., et al. (2009) An improved method for genome wide DNA methylation profiling correlated to transcription and genomic instability in two breast cancer cell lines. *BMC Genomics*. **10**, 223.
8. Hanriot, L., Keime, C., Gay, N., et al. (2008) A combination of LongSAGE with Solexa sequencing is well suited to explore the depth and the complexity of transcriptome. *BMC Genomics*. **9**, 418.
9. MAQ (Mapping and Assembly with Qualities). (Accessed at http://maq.sourceforge.net.)
10. Barski, A., Cuddapah, S., Cui, K., et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*. **129**, 823–837.
11. Mikkelsen, T. S., Ku, M., Jaffe, D. B., et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. **448**, 553–560.
12. Lister, R., O'Malley, R. C., Tonti-Filippini, J., et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*. **133**, 523–536.
13. Pomraning, K. R., Smith, K. M., Freitag, M. (2009) Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods*. **47**, 142–150.
14. Yao, D., Ehrlich, M., Henis, Y. I., Leof, E. B. (2002) Transforming growth factor-beta receptors interact with AP2 by direct binding to beta2 subunit. *Mol Biol Cell*. **13**, 4001–4012.
15. Lister, R., Pelizzola, M., Dowen, R. H., et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. **462**, 315–322.
16. Ehrlich, M. (2002) DNA methylation in cancer: too much, but also too little. *Oncogene*. **21**, 5400–5413.
17. Zuckermann, A., Bohdjalian, A., Deviatko, E., et al. (2002) The University of Vienna experience in heart transplantation. *Clin Transpl*. 229–242.
18. The Reference Sequence (RefSeq). (Accessed at http://www.ncbi.nlm.nih.gov/RefSeq.)
19. Kerkel, K., Spadola, A., Yuan, E., et al. (2008) Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet*. **40**, 904–908.
20. Gal-Yam, E. N., Egger, G., Iniguez, L., et al. (2008) Frequent switching of Polycomb repressive marks and DNA hypermethylation in the

PC3 prostate cancer cell line. *Proc Natl Acad Sci U S A.* **105**, 12979–12984.

21. Oda, M., Glass, J. L., Thompson, R. F., et al. (2009) High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. *Nucleic Acids Res.* **37**, 3829–3839.
22. Benjamini, Y., Hochberg, Y. (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met.* **57**, 289–300.
23. Trinklein, N. D., Karaoz, U., Wu, J., et al. (2007) Integrated analysis of experimental data sets reveals many novel promoters in 1% of the human genome. *Genome Res.* **17**, 720–731.
24. Gardiner-Garden, M., Frommer, M. (1987) CpG islands in vertebrate genomes. *J Mol Biol.* **196**, 261–282.
25. Irizarry, R. A., Wu, H., Feinberg, A. P. (2009) A species-generalized probabilistic model-based definition of CpG islands. *Mamm Genome.* **20**, 674–680.

INDEX

Trygve O. Tollefsbol (ed.), *Epigenetics Protocols: Second Edition*, Methods in Molecular Biology, vol. 791,
DOI 10.1007/978-1-61779-316-5, © Springer Science+Business Media, LLC 2011

D

E

F

G

H

Q

R

S

T