



HHS Public Access

Author manuscript

Methods Mol Biol. Author manuscript; available in PMC 2022 January 01.

Published in final edited form as:

Methods Mol Biol. 2021 ; 2198: 301–310. doi:10.1007/978-1-0716-0876-0_23.

Genome-Wide Mapping of DNA Methylation 5mC by Methylated DNA Immunoprecipitation (MeDIP)-Sequencing

Millissia Ben Maamar, Ingrid Sadler-Riggelman, Daniel Beck, Michael K. Skinner

Center for Reproductive Biology, School of Biological Sciences, Washington State University, Pullman, WA, USA.

Abstract

Methylated DNA immunoprecipitation is a large scale purification technique. It enables the isolation of methylated DNA fragments for subsequent locus-specific or genome-wide analysis. Here we describe an immunoprecipitation protocol using a monoclonal mouse anti 5-methylcytosine antibody followed by next-generation sequencing (MeDIP-Seq).

Keywords

DNA methylation; Methylated DNA immunoprecipitation; Epigenetics; Immunoprecipitation; Next generation sequencing; Review

1 Introduction

DNA methylation is one of the key epigenetic mechanisms; methylation of cytosines, for example, can mediate epigenetic gene regulation [1, 2]. DNA methylation is found in both eukaryotic and prokaryotic organisms. However, in prokaryotes methylation can occur on cytosine and adenine, whereas in multicellular organisms it seems to be restricted to cytosine [3]. In mammals, a DNA methylation mark occurs through the transfer of a methyl group from a donor S-adenosylmethionine to a cytosine at position C5 which produces a new nucleotide 5-methylcytosine (5mC), predominantly in the context of 5'-C-phosphate-G-3' (CpG) dinucleotides [4]. In the mammalian genome, most of the CpG sites are methylated thus contributing to altered gene expression, the condensation of chromatin (heterochromatin), silencing of transposable elements, and X-chromosome inactivation [1].

The methylated DNA immunoprecipitation (MeDIP) method is based on using methylcytosine antibodies to specifically isolate methylated DNA fragments for subsequent locus-specific or genome-wide analysis. Anti-5mC (or anti-5hmC) antibodies are used. The anti-5mC antibodies recognize 5mC independently of surrounding DNA sequence. An alternate procedure uses a methyl-CpG-binding domains (MBDs) of the methyl-CpG-binding proteins (MBPs) which recognize the ^mCpG sites of different densities or within different sequencing contexts. The MeDIP procedure is biased to low density CpG desert sites and the MBP procedure biased to high density island sites [5]. The bisulfite sequencing is also biased to higher density CpG regions. Therefore, each procedure is useful but focused

on different regions of the genome. Since the majority of the genomes of most organisms (e.g., mammals) are low density CpG (>90%), the MeDIP method provides one of the best genome-wide analysis of DNA methylation.

The method described in this chapter is the MeDIP protocol followed by next generation sequencing for an MeDIP-Seq procedure. The MeDIP uses an anti-5mC mouse monoclonal antibody (type IgG) and requires genomic DNA shearing and denaturation steps for efficient binding of the short target DNA fragments by the antibody. Then, the methyl DNA–antibody complexes are precipitated using magnetic beads which bind to the antibody and the resulting complexes are then thoroughly washed to remove unmethylated DNA fragments. Protein digestion and DNA purification steps follow. As a result, the immunoprecipitated DNA fraction is enriched with the methylated fragments for comparison between groups to identify differential DNA methylation regions (DMRs). This can include control versus exposure groups or any parameter to compare. The MeDIP has been developed and optimized [6, 7] and the MeDIP-Seq optimized [7] as described below.

2 Materials

Prepare all solutions using ultrapure water. Prepare and store all reagents at room temperature (unless indicated otherwise). Diligently follow all waste disposal regulations as determined by your institution.

Buffers

1. 1× TE: 10 mM Tris–HCl, pH 7.5, 1 mM EDTA.
2. 1.5% agarose gel: 200 ml 0.5× TBE, 3 g of agarose, 7 µl of 10 mg/ml ethidium bromide.
3. 5× IP Buffer: 100 mM Na-Phosphate pH 7.0, 5 M NaCl, 2.5% Triton X-100 (Sigma Aldrich T-9284), MilliQ water. Use a 0.2 µm filter to sterilize. Store at 4 °C.
4. Washing Buffer: 1× PBS (Ca²⁺ and Mg²⁺ free) with 0.1% BSA and 2 mM EDTA. Store at 4 °C.
5. Digestion Buffer: 1 M Tris–HCl, pH 8.0, 0.5 M EDTA, 10% SDS, MilliQ-water. Use a 0.2 µm filter to sterilize. Store at 4 °C.

3 Methods

3.1 DNA Shearing

1. Extract genomic DNA according to the cells or tissues studied.
2. Sonicate purified genomic DNA using the Covaris M220 Focused Ultra-Sonicator (*see Note 1*).

¹Other sonication devices can be used and will result in equally usable fragmentation. Examples are Bioruptor by Diagenode.

3. Dilute 6 µg genomic DNA into 130 µl (final volume) 1× TE Buffer and pipet into the appropriate Covaris tube.
4. Set Covaris to 300 bp fragment size program (*see* Note 2).
5. Run program for each tube.
6. Run 5 or 10 µl (around 400–500 ng) of sheared genomic DNA and 10 µl DNA ladder on 1.5% agarose gel to verify fragments size. Both “self-poured” gels or using precast gels like for Thermo’s eGEL apparatus will work. Unsonicated DNA (100–200 ng) can be run on the same gel as a comparison (*See* Note 3).

3.2 Antibody Addition

1. Measure volume of sonicated DNA after gel run and dilute it with 1× TE Buffer to 400 µl.
2. Heat-denature in dry bath heating block for 10 min at 95 °C and immediately cool on ice for 10 min.
3. While keeping the sample cold, add 100 µl of cold 5× IP and 4–5 µg (*see* Note 4) of antibody (monoclonal mouse anti 5-methyl-cytidine) to the denatured sonicated DNA. Incubate the DNA–antibody mixture overnight on a rotator (e.g., Paddle Tube Revolver by Fisher at speed setting 10) at 4 °C (*See* Note 5).

3.3 Bind Beads to DNA–Antibody Mixture

1. Prewash magnetic beads (We use Dynabeads M-280 sheep anti-mouse IgG) (*see* Note 6) as follows: Thoroughly resuspend the beads by pipetting up and down or rotating (*see* Note 7). They need to be in a homogeneous solution and if processing multiple samples they need to be resuspended frequently since they settle fast.
2. Transfer needed total volume (50 µl per sample, *see* Note 8) to a centrifuge tube, add the same volume of Washing Buffer (at least 1 ml) and resuspend.
3. Place the tube in a magnetic rack (e.g., DynaMag-2, Thermo Fisher) for 1–2 min and discard supernatant (*See* Note 9).

²-If the Covaris programs that were preinstalled by the manufacturer do not give satisfactory results, parameters, like treatment time or peak incident power can be adjusted.

³-Genomic DNA is randomly sheared by sonication to generate fragments between 300 and 1000 bp. Genomic DNA can also be fragmented with restriction enzymes like Alu I, but it is not recommended for unbiased sequencing studies. The sonication efficiency varies with DNA concentration, sonicator settings and size and quality of the sonication instrument used, therefore it is recommended to check the size of the sheared DNA to ensure reproducible sonication between experiments.

⁴-As in all antibody experiments, it is necessary to make a dose curve to determine what amount of antibody works best for your experimental setting. It also depends on the amount of DNA in the experiment how much antibody to use since there might not always be 6 µg of DNA available for the MeDIP. Four to five micrograms is a guideline and needs to be adjusted according to your specific experiment.

⁵-Rotate the tubes at a low enough speed to prevent foaming but still ensure thorough mixing.

⁶-Both Dynabeads anti mouse IgG as well as Protein G magnetic beads work well. In our comparisons the anti-IgG worked slightly better. Although we use magnetic beads, it is also possible to use Protein A/G agarose beads. The reason we prefer magnetic beads is that the washing steps can be done more efficiently.

⁷-Vortexing may damage these beads so should be avoided unless done at a low setting.

⁸-The volume of beads that works best in a certain experimental setting should also be determined in a dose curve.

⁹-Remove the supernatant carefully when the solution turns clear. This will take about 1–2 min; some magnets work faster than others. Be careful as to not disturb the beads on the magnetic rack.

4. Resuspend beads in 1 ml Washing Buffer and incubate for 1 min on ice. Put tube on magnet for 1–2 min and discard supernatant.
5. Remove the tube from the magnetic rack and resuspend washed beads in the same volume of 1× IP Buffer as the initial volume of beads.
6. Add 50 µl of beads to the 500 µl of DNA–antibody mixture from **step 2**.
7. Incubate for 2 h on a rotating platform at 4 °C.

3.4 DNA–Antibody–Bead Mixture Washing

1. After the 2 h incubation wash beads three times with 1× IP Buffer as follows: Place tube in magnetic rack for 1–2 min and discard supernatant. Remove tube from magnetic rack. Add 1 ml of cold 1× IP Buffer. Mix by inverting tube or gently vortexing (*See Note 7*). Incubate tube for 1 min on ice. Place tube in magnetic rack for 1–2 min and discard the supernatant. Repeat twice for a total of three washes.
2. Resuspend the beads in 250 µl Digestion Buffer.
3. Add 3.5 µl Proteinase K (20 mg/ml) to the resuspended beads.
4. Incubate for 2–3 h on a rotating platform at 55 °C (*See Note 10*).

3.5 DNA Purification

1. Remove parafilm and add 250 µl phenol–chloroform–isoamyl alcohol to each tube. Vortex for 30 s and centrifuge at 14,000 × *g* for 5 min at room temperature. Remove the aqueous supernatant and transfer it to a fresh microcentrifuge tube.

Danger: Phenol as well as chloroform should be handled only in a chemical hood, since both are toxic. Also, gloves need to be worn. Dispose of properly.
2. Add 250 µl of chloroform to the supernatant from **step 1**. Vortex briefly and centrifuge at 14,000 × *g* for 5 min at room temperature. Remove the aqueous supernatant and transfer it to a fresh microcentrifuge tube.
3. Add 2 µl of the coprecipitant GlycoBlue (20 mg/ml, Life Technologies) and mix well.
4. Add 20 µl 5 M NaCl and then 500 µl of 100% ethanol. Mix well.
5. Precipitate in –20 °C freezer for 1 h to overnight.
6. Centrifuge at 14,000 × *g* for 20 min at 4 °C. Carefully remove the supernatant while not disturbing the blue pellet.
7. Wash once to twice with 1 ml 70% ethanol by incubating at –20 °C for 10 min then spinning again for 10 min. Discard supernatant. Then spin again briefly to

¹⁰Make sure to seal the lids well so they do not leak. USA Scientific SealRite tubes or Eppendorf Safe-Lock tubes sealed with parafilm usually work well.

collect residual liquid to bottom of tube and remove all the liquid with gel loading or other fine pipette tip.

8. Air-dry the samples on bench (*See Note 11*).
9. Resuspend in 25 μ l of nuclease-free water.
10. Measure the DNA concentration (*See Note 12*).

3.6 Libraries for Next Generation Sequencing

1. Since the DNA fragments retrieved through MeDIP are single stranded, the first step needs to produce double stranded DNA for library preparation. The method we use and which has worked well for us, is using a library preparation kit designed for RNA which includes a step where double-stranded DNA is produced (for example: NEB's NEBNext Ultra II RNA Library Prep Kit, E7770).
2. Use between 10 and 1000 ng (*see Note 13*) of single-stranded DNA fragments and anneal 10 ng/ μ l random hexamer primers to the sample by heating it in a thermal cycler to 95 °C, then cooling immediately on ice.
3. Perform second strand DNA synthesis, which if using the NEB kit as mentioned above would be step 1.4.
4. Follow the rest of the manufacturer's protocol to receive libraries for next generation sequencing (*see Note 14*).
5. Determine the library yield with the help of Qubit High Sensitivity dsDNA Kit (Qubit™ dsDNA HS Kit, Q32851, Thermo Fisher), then perform quality control for fragment size range and concentration/molarity on the Bioanalyzer (Agilent).
6. Libraries are sequenced on the sequencers available to you, for example Illumina HiSeq (*see Note 15*).

Methods controls: *See Notes 16 and 17*.

4 Data Analysis

Raw data resulting from the next generation sequencing (NGS) step is generally one or two files for each sample. Typically these files are FASTQ formatted, however, other formats are

¹¹It usually takes 5 min to dry the samples.

¹²Since the resulting DNA is single stranded, we use the ssDNA kit for the Qubit (*Qubit ssDNA Assay Kit, Q10212, Thermo Fisher*).

¹³This is the amount that can be used in the kit for one reaction per the manufacturer, but based on MeDIP yields it will probably be on the lower end of the spectrum.

¹⁴If more than one library is run in one lane for sequencing, different index primers have to be attached to each sample to be able to differentiate between the libraries for sequence analysis. For example NEB: NEBNext Multiplex Oligos, E7335. Oligos can of course also be synthesized according to your specifications.

¹⁵NGS sequencing kits are specific for the type of sequencer you use, so make sure you work with compatible kits.

¹⁶Methylation efficiency control: Remove input DNA from sample then "spike" with methylated and nonmethylated control DNA. Sonicate. Keep part of it at -20 °C as non-IP control, then perform immunoprecipitation as for the sample. Use specific primers to determine the amount of methylated versus nonmethylated DNA after and before IP.

¹⁷A no-antibody control should be run in addition to the sample, where the control is treated the same way as the sample, except that no antibody is added. This should not yield any methylated DNA fragments after the experiment showing that the beads themselves will not bind DNA nonspecifically. Nonspecific binding does not seem to be a big issue with magnetic beads.

possible depending on the sequencing platform used. The data analysis is broken into three parts: data verification and quality control, differential methylation analysis, and final result processing and summary. Code to perform the analyses described below is available at <https://github.com/skinnerlab/MeDIP-seq>.

4.1 Data Verification and QC

1. Ensure data integrity by verifying the raw sequencing files on the analysis server are identical to the ones from the sequencer. This can be done using any file checksum such as MD5.
2. Prior to any analysis the raw data should be backed up to a secure location. Backup files should also be verified using file checksums.
3. Examine raw data quality using FastQC [8]. This tool generates several summary plots and tables that help determine the quality of the raw data. Anomalous results may indicate resequencing or more stringent quality filtering is required.
4. Clean and filter raw reads using Trimmomatic [9] to remove adapters and low quality bases (*see* Note 18).

4.2 Differential Methylation Analysis

1. Libraries are sequenced on the sequencers available to you, for example Illumina HiSeq (*see* Note 15).
2. MeDIP-seq analysis requires a reference genome. An appropriate genome should be selected and downloaded. NCBI is a good source for the genome files (*see* Note 19).
3. The reference genome is needed in two forms. Use the *bowtie2-build* command to generate index files that will be used during the mapping step. Create an R BSgenome [10] package of the reference genome using the *forgeBSgenomeDataPkg* function. This function is part of the BSgenome R package.
4. For each sample, map the cleaned reads to the reference genome using Bowtie 2 [11]. Default parameters can be used. This mapping produces SAM formatted files (*see* Note 20).
5. Convert the SAM files to a sorted BAM format using the SAMtools [12] utility. This is accomplished using *samtools view* followed by *samtools sort*. The original SAM files can be deleted to conserve disk space.
6. Using the R packages MEDIPS [13] and edgeR [14], perform the differential methylation analysis. Read in the sorted BAM files for each sample using the

¹⁸.A cleaning and filtering step is often performed by the sequencing lab. Repeating this step is often unnecessary, but ensures consistent data quality.

¹⁹.The genome files are typically in FASTA format and include a single file for each chromosome. Unassembled genomes can also be used; however, concatenation of the contigs/scaffolds is often required.

²⁰.The logfile produced by the Bowtie 2 mapping should be examined closely to look for potential problems. A low mapping percent generally indicates problems with the input data quality or with the selected reference genome.

MEDIPS.createSet function. Identify the samples in each treatment group. Use the *MEDIPS.meth* function to perform the differential analysis for each genomic window (see Note 21). This analysis will result in a large table containing p -values and other information for each genomic window.

7. The differential analysis result table is next processed to identify DMR. Preliminary DMR are identified by selecting all genomic windows that meet a preselected p -value threshold. Both the raw edgeR p -value and the FDR adjusted p -value can be used. Merge multiple neighboring preliminary DMR together into a single DMR. This is done by extending preliminary DMR edges until there is no genomic window within 1000 base pairs with a p -value less than 0.1. These are arbitrarily selected thresholds that seem to work well. DMR can be additionally filtered by the log fold change in expression.

4.3 Final Result Processing and Summary

1. Calculate CpG density, length, and other desired DMR attributes using the reference genome.
2. Figures such as histograms of p -values for all genomic windows, principal component analysis (PCA) plots using sample read depths, and sample dendrograms can be helpful for diagnosing problems with the underlying samples.
3. Optionally, annotate DMRs by looking for nearby genes using the biomaRt [15] R package. It may be necessary to annotate the DMRs in another manner (such as BLAST) if there is not an appropriate Biomart database.
4. DMR can be plotted by chromosome to determine if they are distributed genome wide or are concentrated in certain genomic regions.

References

1. Bird A (2002) DNA methylation patterns and epigenetic memory. *Genes Dev* 16(1):6–21. 10.1101/gad.947102 [PubMed: 11782440]
2. Reamon-Buettner SM, Borlak J (2007) A new paradigm in toxicology and teratology: altering gene activity in the absence of DNA sequence variation. *Reprod Toxicol* 24(1):20–30. 10.1016/j.reprotox.2007.05.002 [PubMed: 17596910]
3. Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 31(2):89–97. 10.1016/j.tibs.2005.12.008 [PubMed: 16403636]
4. Lister R, Ecker JR (2009) Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res* 19(6):959–966. 10.1101/gr.083451.108 [PubMed: 19273618]
5. Nair SS, Coolen MW, Stirzaker C, Song JZ, Statham AL, Strbenac D, Robinson MD, Clark SJ (2011) Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. *Epigenetics* 6(1):34–44. 10.4161/epi.6.1.13313 [PubMed: 20818161]
6. Guerrero-Bosagna C, Settles M, Lucker B, Skinner M (2010) Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm epigenome. *PLoS One* 5(9):1–17. 10.1371/journal.pone.0013100

²¹.The p .adj parameter should generally be set to “fdr” to perform the FDR p -value multiple testing adjustment.

7. Beck D, Sadler-Riggleman I, Skinner MK (2017) Generational comparisons (F1 versus F3) of vinclozolin induced epigenetic transgenerational inheritance of sperm differential DNA methylation regions (epimutations) using MeDIP-Seq. *Environment Epigenet* 3(3):1–12. 10.1093/eep/dvx016
8. Andrews S (2016) FastQC—a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
9. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120. 10.1093/bioinformatics/btu170 [PubMed: 24695404]
10. Pagès H (2018) BSgenome: software infrastructure for efficient representation of full genomes and their SNPs. R package version 1.48.0
11. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359. 10.1038/nmeth.1923 [PubMed: 22388286]
12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079. 10.1093/bioinformatics/btp352 [PubMed: 19505943]
13. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L (2014) MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. *Bioinformatics* 30(2):284–286. 10.1093/bioinformatics/btt650 [PubMed: 24227674]
14. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–140. 10.1093/bioinformatics/btp616 [PubMed: 19910308]
15. Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 4(8):1184–1191. 10.1038/nprot.2009.97 [PubMed: 19617889]