Chapter 16

Determining DNA Methylation Profiles Using Sequencing

Suhua Feng, Liudmilla Rubbi, Steven E. Jacobsen, and Matteo Pellegrini

Abstract

Cytosine methylation is an epigenetic mark that has a significant impact on the regulation of transcription and replication of DNA. DNA methylation patterns are highly conserved across cell divisions and are therefore highly heritable. Furthermore, in multicellular organisms, DNA methylation patterning is a key determinant of cellular differentiation and tissue-specific expression patterns. Lastly, DNA demethylases can affect global levels of DNA methylation during specific stages of development. Bisulfite sequencing is considered the gold standard for measuring the methylation state of cytosines. Sodium bisulfite converts unmethylated cytosines to uracils (which after PCR are converted to thymines), while leaving methylated cytosines unconverted. By mapping bisulfite treated DNA back to the original reference genome, it is then possible to determine the methylation state of individual cytosines. With the advent of next-generation sequencers during the past few years, it is now possible to determine the methylation state of an entire genome. Here, we describe in detail two protocols for preparing bisulfite treated libraries, which may be sequenced using Illumina GAII sequencers. The first of these uses premethylated adapters, which are not affected by bisulfite treatments, while the second uses a two-stage adapter strategy and does not require premethylation of the adapters. We also describe the specialized protocol for mapping bisulfite converted reads. These approaches allow one to determine the methylation state of each cytosine in the genome.

Key words: DNA methylation, Epigenetics, Next-generation sequencing, Whole-genome methylome

1. Introduction

The methylation state of cytosines in the genome has a profound effect on many biological processes. Most organisms contain maintenance DNA methyltransferases that can preserve the methylation state of CpG dinucleotides during cell division (1). A second class of maintenance methyltransferases are also commonly found, which can often methylate cytosines not in CpG dinucleotides (2, 3). These enzymes are found in a wide variety

Young Min Kwon and Steven C. Ricke (eds.), *High-Throughput Next Generation Sequencing: Methods and Applications*, Methods in Molecular Biology, vol. 733, DOI 10.1007/978-1-61779-089-8_16,

of organisms ranging from plants to animals, and from multicellular organisms to single-celled ones. Although DNA methylation is found very frequently, certain organisms such as *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Caenorhabditis elegans* have lost the ability to methylate DNA.

For those organisms that methylate their DNA, this process is usually essential to their survival, and mutants of DNA methyltransferases often lead to nonviable strains (3). The reason for this is that DNA methylation plays a critical role in patterning the transcriptional state of a cell, and DNA methylation mutants typically have aberrant development. By modifying the DNA methylation state of cytosines, it is possible to suppress the transcription of transposable and other repetitive elements, thus constraining unregulated growth of genome sizes (4, 5). Similarly, aberrant methylation of promoters can lead to transcriptional suppression of the downstream genes, as is often seen in diseases such as cancer (5).

Because of the importance of DNA methylation in regulating fundamental biological processes, it is of interest to measure the methylation state of all the cytosines in a genome. Conventional DNA sequencing techniques cannot distinguish between methylated and unmethylated cytosines. It is therefore common to utilize sodium bisulfite to convert unmethylated cytosines to uracils, while leaving the methylated cytosines unconverted. By mapping the converted DNA back to the reference DNA, it is possible to determine the methylation state of each of the cytosines simply by counting the number of cytosines and thymines that align to that position.

While the sequencing of bisulfite-converted DNA is often considered the gold standard for determining the methylation state of DNA, in the past this approach was usually applied to a limited number of loci. However, with the advent of relatively inexpensive high-throughput sequencing, it is now possible to shotgun-sequence an entire bisulfite converted genome, thus enabling genome-wide measurements of cytosine methylation (6–8). Here, we present two protocols for generating libraries of bisulfite-treated DNA, for sequencing on Illumina sequencers. The first of these protocols uses premethylated adapters that are not affected by the bisulfite conversion step (Fig. 1). This protocol has the advantage of requiring fewer overall steps to prepare the library and fewer amplification cycles, thus limiting potential sequence composition biases, and has to date been used more frequently in published papers. The second protocol proceeds in two steps and does not require that the standard library adapters be premethylated (Fig. 2). The advantage of the second approach is that it leads to the sequencing of both the converted DNA strands and its reverse complement, and therefore maintains relatively high frequencies of cytosines in a library even when the

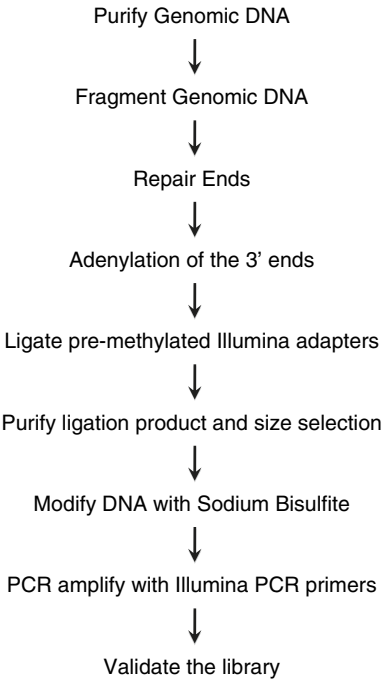


Fig. 1. Protocol I: library generation using premethylated adapters. (Adapted from Illumina protocol.).

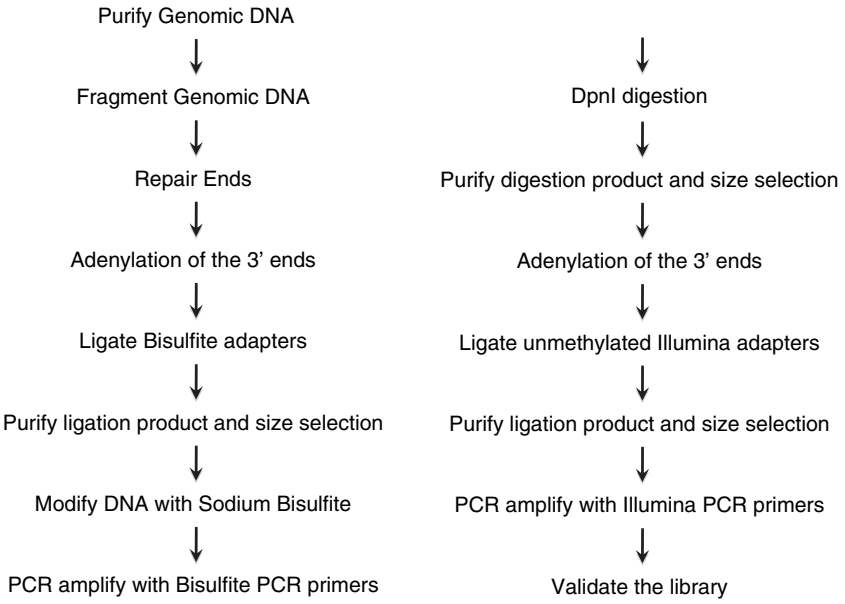


Fig. 2. Protocol II: library generation using unmethylated adapters. (Adapted from Millipore and Illumina protocols.).

genome is mostly unmethylated. This feature potentially improves base calling (which is problematic when very few cytosines are present in a library), but because it requires two rounds of amplification, it can lead to greater sequence composition biases.

Once the library has been prepared and sequenced, the next step requires that the converted reads be aligned to the genome. This results in a nontraditional alignment in which thymines in reads can align to either thymines or cytosines in the genome. To accomplish this, we align a three-letter version of reads to a three-letter genome, and appropriately score correct versus incorrect cytosine to thymine transitions. We briefly describe this approach along with a software we have developed to conduct these alignments.

2. Materials (see Note 1)

2.1. Protocol I: Library Generation Using Premethylated Adapter

2.1.1. DNA Sample Preparation

1. HMB buffer: 25 mM Tris–HCl, pH 7.6, 0.44 M sucrose, 10 mM $MgCl_2$, 0.1% Triton X-100, 10 mM β-mercaptoethanol, 2 mM spermine, 1 mM PMSF, 1 μg/ml pepstatin, 1× EDTA–free protease inhibitors (Roche). Make fresh and keep at 4°C.
2. Miracloth filter (Calbiochem).
3. HBB buffer: 25 mM Tris–HCl, pH 7.6, 0.44 M sucrose, 10 mM $MgCl_2$, 0.1% Triton X-100, 10 mM β-mercaptoethanol. Make fresh and keep at 4°C.
4. 40%/60% percoll gradient: Make 40% and 60% percoll by mixing percoll with HBB. Add 10 ml of 60% percoll into a 30-ml centrifuge tube first and then carefully lay 10 ml of 40% percoll on top of the 60% percoll. Make fresh.
5. Resuspension buffer: 50 mM Tris–HCl (pH 7.5), 20 μg/μl Proteinase K (Roche). Make fresh.
6. Lysis Buffer: 50 mM Tris–HCl (pH 7.5), 15 mM EDTA, 1.5% SDS. Make fresh.
7. QIAquick PCR Purification Kit (Qiagen).
8. Biorupter sonicator (Diagenode).

2.1.2. End Repair

1. T4 DNA polymerase (3 U/μl, New England Biolabs).
2. Klenow DNA polymerase (5 U/μl, New England Biolabs).
3. T4 polynucleotide kinase (10 U/μl, New England Biolabs).
4. dNTP mix (10 mM each, New England Biolabs).
5. 10× T4 DNA ligase buffer with 10 mM ATP (New England Biolabs).

2.1.3. Adenylation

1. Klenow fragment $3' \rightarrow 5'$ exo minus (5 U/μl, New England Biolabs).

2. 10× Klenow buffer (Buffer 2, New England Biolabs).
3. 1 mM dATP (New England Biolabs).

2.1.4. Ligation with Premethylated Illumina Adapters

1. Quick T4 DNA ligase (2,000 U/μl) and 2× Quick ligation buffer (New England Biolabs).
2. Premethylated Adapters single-end (Illumina):
   (a) *Premethylated Illumina Adapter 1* (all the cytosines in this oligo are 5-methylated; ask the oligo synthesis company to incorporate 5-methylated cytosines instead of regular unmethylated cytosines during the synthesis of this oligo):

   5'-GATCGGAAGAGCTCGTATGCCGTCTTCTGCTTG-3'

   ↑

   with 5' phosphate

   (b) *Premethylated Illumina Adapter 2* (all the cytosines in this oligo are 5-methylated; ask the oligo synthesis company to incorporate 5-methylated cytosines instead of regular unmethylated cytosines during the synthesis of this oligo):

   5'- ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3'

   ↑

   no 5' phosphate
3. 2% certified low range ultra agarose (Bio-Rad).
4. QIAquick column (Qiagen).

2.1.5. Bisulfite Treatment

1. CpGenome DNA Modification Kit (Millipore).
2. 3 M NaOH (freshly prepared prior to each use). Dissolve 1 g of NaOH pellet in 8.3 ml of water. Use appropriate caution when manipulating this caustic base.
3. Millipore reagent I – Urea solution (freshly prepared prior to each use): use appropriate caution when handling this reagent, as it is irritating to the respiratory system and skin.
4. Millipore reagent II (freshly prepared): excess reagent can be stored in a foil-wrapped container at 2–8°C in the dark for up to 6 weeks.
5. 70% ETOH. Store at –20°C.
6. 20 mM NaOH/90% EtOH (freshly prepared): to prepare 1 ml of this solution, combine 900 μl of 100% EtOH, 93.4 μl of water, and 6.6 μl of 3 M NaOH.
7. 90% EtOH. Store at –20°C.
8. TE buffer. Store at room temperature.

2.1.6. PCR Amplification of the Library

1. Pfu Turbo Cx polymerase and 10× Pfu Turbo Cx buffer (Agilent).
2. dNTPs (2.5 mM each, New England Biolabs).
3. PCR single-end Primers 1.1 and 2.1 (Illumina):
   (a) Illumina PCR primer 1.1:

   5'-
   AATGATACGGCGACCACCGAGATCTACACTCTTT
   CCCTACACGACGCTCTTCCGATCT-3'

   (b) Illumina PCR primer 2.1:

   5'- CAAGCAGAAGACGGCATACGAGCTCTTCCGATCT-3'
4. PCR Purification Kit (Qiagen).

2.1.7. Qualitative and Quantitative Controls of the Library

1. 2% certified low range ultra agarose gel (Bio-Rad).
2. PCR4Blunt-TOPO vector (Invitrogen).
3. Library resuspension Buffer: EB Buffer (Qiagen), 0.1% Tween-20.

2.2. Protocol II: Library Generation Using Unmethylated Adapters

Materials for the following six steps are described in the above sections: *DNA Sample Preparation* (Subheading 2.1.1), *End Repair* (Subheading 2.1.2), *Adenylation* (Subheading 2.1.3), *Bisulfite Treatment* (Subheading 2.1.5), *PCR Amplification of the Library* (Subheading 2.1.6), and *Quantitative and Quantitative Controls of the Library* (Subheading 2.1.7)

2.2.1. Ligation with Bisulfite Adapters

1. Adapters ligation: Quick T4 DNA ligase (2,000 U/μl) and 2× Quick ligation buffer (New England Biolabs)
2. Bisulfite Adapters.
   (a) Bisulfite Adapter 1:

   5'-AGTTATTCCGGACTGTCGAAGCTGAAGTGATC^{m}CGT -3'

   no 5' phosphate — 5-methylated

   (b) Bisulfite Adapter 2:

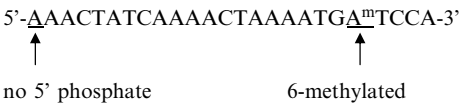
   with 5' phosphate

   5'-CGGATC^{m}ACTTCAGCTTCGACAGTCCGGAAT-3'

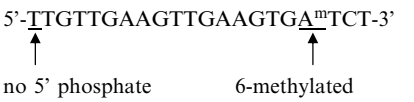
   5-methylated
3. 2% certified low range ultra agarose (Bio-Rad).
4. QIAquick column (Qiagen).

2.2.2. PCR Amplification of the Bisulfite Treated DNA

1. Pfu Turbo Cx polymerase and 10× Pfu Turbo Cx buffer (Agilent).
2. dNTPs (2.5 mM each, New England Biolabs).
3. Bisulfite PCR Primers 1 and 2:
   (a) Bisulfite PCR Primer 1:



   (b) Bisulfite PCR primer 2:



4. 2% certified low range ultra agarose gel (Bio-Rad).

2.2.3. DpnI Digestion of the PCR Products

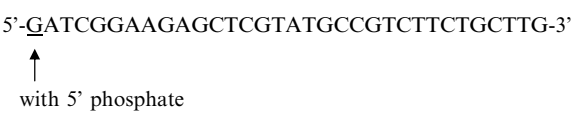
1. 10× DpnI buffer (Buffer 4, New England Biolabs), BSA (10 mg/ml, New England Biolabs), DpnI (20 U/µl, New England Biolabs).
2. QIAquick column (Qiagen).
3. 2% certified low range ultra agarose gel (Bio-Rad).

2.2.4. Adenylation

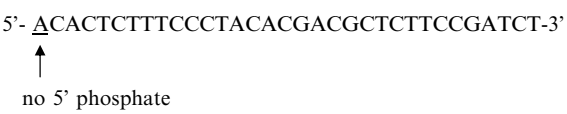
1. 10× Klenow buffer (Buffer 2 New England Biolabs) and Klenow fragment 3′→5′ exo minus (5 U/µl, New England Biolabs).
2. dATP (1 mM New England Biolabs).
3. QIAquick MinElute column (Qiagen).

2.2.5. Ligation with Unmethylated Illumina Adapters

1. Quick T4 DNA ligase (2,000 U/µl) and 2× Quick ligation buffer (New England Biolabs).
2. Unmethylated Adapters single-end (Illumina):
   (a) Unmethylated Illumina Adapter 1:



   (b) Unmethylated Illumina Adapter 2:



3. 2% certified low range ultra agarose gel (Bio-Rad).
4. QIAquick column (Qiagen).

3. Methods (see Note 1)

3.1. Protocol I: Bisulfite Library Generation Using Premethylated Adapters

Genomic DNA can be purified from different sources (e.g., cells, frozen tissue) using various standard DNA purification protocols. It is important to ensure that the DNA sample to be processed is highly pure. In this chapter, we describe the procedure to obtain highly pure DNA from *Arabidopsis thaliana*.

3.1.1. DNA Sample Preparation

1. Ground 1 g of plant tissues into a fine powder in liquid nitrogen and homogenize three times in 10 ml of HBM buffer.
2. Filter through two layers of Miracloth filters (Calbiochem). Centrifuge the homogenate (3,000–3,500 rpm [1,000–1,500 × *g*] for 5 min at 4°C, in an SS-34 rotor), and resuspend the pellet in 5 ml of HBB buffer.
3. To isolate the nuclei load the sample onto 20 ml of 40%/60% percoll gradient HBB and centrifuge for 1,500–2,000 rpm [300–500 × *g*] for 30 min at 4°C without applying the brake. Wash twice with 10 ml of HBB buffer and resuspend in 500 µl of 50 mM Tris–HCl (pH 7.5) containing 20 µl (final µg 20 µg/µl) of Proteinase K (Roche) and incubate at room temperature for 30 min. Add 1 ml of Lysis buffer to lyse the nuclei. Purify DNA by phenol/chloroform extraction and ethanol precipitation. Resuspend DNA in totally 200 µl of Qiagen EB buffer to complete.
4. To fragment the genomic DNA, sonicate 5 µg of DNA (in 100–300 µl solution) (see Note 2), in a 1.5-ml tube, with Biorupter sonicator (Diagenode) for four cycles of 15 min each (30 s "on" and 30 s "off", output H). In between cycles, samples are kept on ice. The DNA fragments are next purified with QIAquick column from QIAquick PCR purification Kit (Qiagen) and eluted in 35 µl of EB buffer (Qiagen) (see Note 3).

3.1.2. End Repair

1. To repair, blunt, and phosphorylate ends ("End Repair reaction"), the DNA fragments are subsequently treated with a mixture of T4 DNA polymerase, *Escherichia coli* DNA polymerase I Klenow fragment, and T4 polynucleotide kinase.

   Set up the End-Repair reaction as follows:

35 µl	DNA from previous step
40 µl	Water
10 µl	10× T4 DNA ligase buffer (with 10 mM ATP)
4 µl	dNTP mix (10 mM each)
5 µl	T4 DNA polymerase (3 U/µl)

1 µl	Klenow DNA polymerase (5 U/µl)
5 µl	T4 PNK (10 U/µl)
100 µl	Total reaction volume
Incubate for 30 min at 20°C.	

2. Purify DNA with QIAquick column (Qiagen) and elute in 32 µl of EB buffer.

3.1.3. Adenylation

1. To add a single "A" base to the 3' end ("Adenylation reaction"), incubate the repaired DNA fragments from previous step with Klenow exo-fragment (3'→5' exo⁻).

   Set up the "Adenylation reaction" as follows:

32 µl	DNA (from previous step)
5 µl	10× Klenow buffer
10 µl	dATP (1 mM)
3 µl	Klenow fragment 3′→5′ exo minus (5 U/µl)
50 µl	Total reaction volume
Incubate for 30 min at 37°C.	

2. Purify DNA with one QIAquick MinElute column (Qiagen) and elute in 14 µl of EB buffer.

3.1.4. Ligation with Premethylated Illumina Adapters

1. To ligate the DNA fragments with the adapters incubate the adenylated fragments from previous step with premethylated, single-end, adapters (Illumina).

   Set up the Adapter Ligation reaction as follows:

14 µl	DNA (from previous step)
30 µl	2× Quick ligation buffer
10 µl	Premethylated Illumina adapters (single-ended) (anneal adapter 1 and adapter 2 at 1:1 molar ratio, adjust the concentration of the adapter duplex so that there is a 10:1 molar ratio of adapter duplex to DNA insert)
6 µl	Quick T4 DNA ligase (2,000 U/µl)
60 µl	Total reaction volume
Incubate for 15 min at room temperature.	

2. Purify DNA with one QIAquick column and elute in 30 µl of Qiagen EB buffer.
3. Run DNA from previous step on 2% certified low range ultra agarose gel (Biorad) for 1 h at 100 V. Excise fragments

ranging from 150 to 300 bp (or according to specific needs, see Note 4) and extract DNA with one QIAquick column. Elute DNA in 30 µL of Qiagen EB buffer.

3.1.5. Bisulfite Treatment

At this step, adapter-ligated DNA is ready for bisulfite treatment. Several commercial bisulfite treatment kits are available. In the following steps, the CpGenome DNA Modification Kit (Millipore) is used. The protocol described here follows the manufacturer's instructions and previously described method (9).

1. Transfer 1 µg of the adapter-ligated DNA to a 1.5–2.0-ml screwcap microcentrifuge tube. Bring the total volume to 100 µl with water (add 2 µl of Millipore Reagent IV if less than 1 µg of DNA is used). Add 7 µl of freshly prepared 3 M NaOH, mix and incubate at 55°C in a heat block for 20 min.
2. Prepare the Millipore Reagent I-Urea solution: for each sample to be modified, dissolve 0.227 g of Millipore Reagent I (warm bottle to room temperature before opening) into 0.464 ml of water. Adjust the pH by adding 20 µl of 3 M NaOH and mix. Check the pH with the pH indicator paper and the pH should be 5.0. Add 0.22 g of urea to 450 µl of this solution. The final volume should be around 650 µl.
3. Add 650 µl of the freshly prepared Millipore Reagent I-Urea solution to the denatured DNA (from step 2), vortex, and incubate at 55°C in a water bath for 24 h. Protect the samples from light exposure.
4. Prepare Millipore Reagent II: add 1 µl of β-mercaptoethanol to 20 ml of deionized water. For each sample to be modified, add 750 µl of this solution to 1.35 g of Millipore Reagent II (warm to room temperature before opening). Mix well until completely dissolved.
5. Resuspend Millipore Reagent III by vortexing vigorously and pipeting up and down for ten times to disperse any remaining clumps. Add 5 µl of well-suspended Millipore Reagent III to the DNA solutions in the tubes. Add 750 µl of freshly prepared Millipore Reagent II, mix briefly, and incubate at room temperature for 15 min.
6. Spin for 30 s at 5,000 × *g* to pellet the Reagent III, discard supernatant (a small white pellet should be visible). Add 1 ml of 70% EtOH, vortex, spin 5,000 × *g* for 30 s, and discard the supernatant. Repeat this step twice. After the last wash has been removed, centrifuge at top speed for 2 min, and remove all the remaining supernatant with a plastic pipette tip.
7. Add 50 µl of the freshly prepared 20 mM NaOH/90% EtOH solution to the samples, vortex briefly, and incubate at room temperature for 10 min.

8. Spin for 30 s at 5,000 × *g*. Add 1 ml of 90% EtOH, vortex to wash the pellet, and again to remove the supernatant. Repeat this step once. Spin at top speed for 3 min, remove all the remaining supernatant with a pipette tip, and let dry for 30 min at room temperature.
9. Add 40 µl of TE, incubate at 55°C for 20 min to elute DNA, centrifuge at top speed for 3 min, and transfer the supernatant to a new tube. Proceed to PCR or store as aliquots at –80°C. Avoid repeated thawing and refreezing.

3.1.6. PCR Amplification of the Library

1. Enrich Adapters-ligated and Bisulfite-modified DNA fragments by PCR reactions.

   Set up four PCR reactions for each sample as follows:

2.5 µl	DNA
5.0 µl	10× PfuTurbo Cx buffer
4.0 µl	dNTPs (2.5 mM each)
1 µl	Illumina PCR primer 1.1 (single-end)
1 µl	Illumina PCR primer 2.1 (single-end)
0.5 µl	PfuTurbo Cx polymerase
36 µl	Water
50 µl	Total reaction volume

   Amplify using the following PCR protocol: 2 min at 98°C. 15 cycles of 10 s at 98°C and 90 s at 60°C, 10 min at 60°C. Hold at 4°C.
2. Follow the instructions in the QIAquick PCR Purification Kit (Qiagen) to purify with one QIAquick column, elute in 30 µl of EB buffer. This is the final library.

3.1.7. Qualitative and Quantitative Controls of the Library

1. Load 10% of the volume of the library (5 µl) on 2% agarose gel to check whether the size range is as expected (see Note 5). It should be slightly larger in size than the size-range excised during the gel purification step (since PCR primers add ~25 bp to the length of the product).
2. Clone 4 µl of the library into a sequencing vector. (e.g., pCR-4Blunt-TOPO, Invitrogen) (see Note 6). Sequence individual clones by conventional Sanger sequencing.
3. Determine the concentration of the library by measuring its absorbance at 260 nm. To determine the molar concentration of the library, examine the gel image (from step 1) and estimate the median size of the library smear. Multiply this size by 650 (the molecular mass of a base pair) to get the molecular weight of the library. Use this number to calculate

the molar concentration. Make 10-nM aliquots of library in Qiagen EB buffer containing 0.1% Tween-20 for Illumina/Solexa sequencing and store at −20°C.

3.1.8. Aligning Bisulfite-Converted Reads

Sodium bisulfite converts unmethylated cytosines to uracils, while leaving methylated cytosines unconverted. After the steps of PCR, uracils are converted to thymines, which is the base associated with unmethylated cytosines in the final reads. When mapping these converted reads back to the reference genome, we cannot therefore use a standard alignment tool, as this would consider the alignment of a thymine in a read to a cytosine in the genome as a mismatch.

To circumvent this limitation, we conduct the alignments in three-letter space, where all cytosines are converted to thymines (or vice versa) in both the reads and the genome. Typically, we restrict our attention to only those reads that map to a unique position in the genome in this three-letter space, although it is possible to include more sophisticated treatments on nonuniquely mapping reads. This approach allows us to correctly align thymines in reads to cytosines in the genome, but also permits incorrect alignments of cytosines in reads to thymines in the genome. Consequently, after the three-letter alignment is complete, we reconvert our sequence to four-letter space, and penalize all alignments that contain read cytosines aligned to genome thymines.

We have implemented a version of this protocol that utilizes a fast short read aligner, Bowtie, to perform the three-letter alignments. Our wrapper around this software then checks for incorrect conversions and penalizes those alignments. The software may be downloaded from http://pellegrini.mcdb.ucla.edu/BS_Seeker/BS_Seeker.html.

3.2. Protocol II: Bisulfite Library Generation Using Unmethylated Adapters

The protocols for the following three steps are described in the above sections: *DNA Sample Preparation* (Subheading 3.1.1), *End Repair* (Subheading 3.1.2), and *Adenylation* (Subheading 3.1.3).

3.2.1. Ligation with Bisulfite Adapters

1. To ligate the DNA fragments with the bisulfite adapters, incubate the adenylated fragments (from previous step) with the Bisulfite Adapters.

   Set up the Adapter Ligation reaction as follows:

14 µl	DNA (from previous step)
30 µl	2× Quick ligation buffer
10 µl	Bisulfite Adapters (anneal adapter 1 and adapter 2 at 1:1 molar ratio, adjust the concentration of the adapter duplex so that there is a 10:1 molar ratio of adapter duplex to DNA insert)

6 µl	Quick T4 DNA ligase (2,000 U/µl)
60 µl	Total reaction volume
Incubate for 15 min at room temperature.	

2. Purify DNA with one QIAquick column and elute in 30 µl of Qiagen EB buffer.
3. Run DNA from previous step on 2% certified low range ultra agarose gel (Biorad) for 1 h at 100 V. Excise fragments ranging from 100 bp and up, purify using three QIAquick columns, and elute DNA from each column with 30 µL of EB buffer.
4. The protocol for *Bisulfite Treatment* is described in Subheading 3.1.5.

3.2.2. PCR Amplification of the Bisulfite Treated DNA

1. Enrich Adapters-ligated and Bisulfite-modified DNA fragments by PCR reactions.

   Set up eight PCR reactions for each sample as follows:

2.5 µl	DNA
5.0 µl	10× PfuTurbo Cx buffer
4.0 µl	dNTPs (2.5 mM each)
1 µl	Bisulfite PCR primer 1
1 µl	Bisulfite PCR primer 2
0.5 µl	Pfu Turbo Cx polymerase
36 µl	Water
50 µl	Total reaction volume

   Use the following PCR conditions: 2 min at 98°C. Eight cycles of 10 s denaturation at 94°C, 30 s annealing at temperatures from 55°C to 52°C (2 cycles at each temperature) and 4 min extension at 60°C. Then, seven cycles of 10 s denaturation at 94°C, 30 s annealing at 51°C, and 4 min extension at 60°C.

   Finally, 10 min at 60°C and hold PCR samples at 4°C.
2. Purify PCR product with two QIAquick columns, elute each column with 30 µl of Qiagen EB buffer and combine.
3. Run 5 µl of purified of the purified PCR product on 2% certified low range ultra agarose gel (Bio-Rad) to check the DNA size.

3.2.3. DpnI Digestion of the PCR Products

1. The DpnI enzyme digests GATC site with 6-methylated adenosine, which is formed immediately flanking the library DNA insert after PCR amplification with bisulfite PCR

primers (previous step). Thus, DpnI digestion will cleave the bisulfite adapters off PCR products, which are ligated to unmethylated Illumina adapter in subsequent steps.

Set up the DpnI digestion as follows:

55 µl	(DNA from previous step)
12 µl	10× DpnI Buffer
1.2 µl	BSA (10 mg/ml)
12 µl	DpnI (20 U/µl)
39.8 µl	Water
120 µl	Total reaction volume
Incubate overnight at 37°C.	

2. Purify the digestion products with one QIAquick column and elute in 30 µl of Qiagen EB buffer.
3. Run DNA from previous step on 2% certified low range ultra agarose gel (Bio-Rad) for 1 h at 100 V. Excise DNA fragments from 40 bp and up, purify DNA with four QIAquick MinElute columns, and elute each column with 16 µl of Qiagen EB buffer and combine.

3.2.4. Adenylation

1. Addition of an "A" to 3′ end of DNA fragments, "adenylation reaction":

   Set up the "Adenylation reaction" as follows:

64 µl	DNA from previous step
20 µl	10× Klenow buffer
10 µl	dATP (1 mM)
6 µl	Klenow fragment 3′→5′ exo minus (5 U/µl)
100 µl	Total reaction volume
Incubate at 37°C for 30 min.	

2. Purify DNA with one QIAquick MinElute column and elute in 14 µl of Qiagen EB buffer.

3.2.5. Ligation with Unmethylated Illumina Adapters

1. Ligation with unmethylated single-end adapters (Illumina).

   Set up the Adapter Ligation reaction as follows:

14 µl	DNA from previous step
30 µl	2× Quick ligation buffer
10 µl	Unmethylated Illumina adapters (single-end) (anneal adapter 1 and adapter 2 at 1:1 molar ratio, adjust the concentration of the adapter duplex so that there is a 10:1 molar ratio of adapter duplex to DNA insert)

6 µl	Quick T4 DNA ligase (2,000 U/µl)
60 µl	Total reaction volume
Incubate for 15 min at room temperature.	

2. Purify DNA with one QIAquick column and elute in 30 µl of Qiagen EB buffer.
3. Run DNA from previous step on 2% certified low range ultra agarose gel (Biorad) for 1 h at 100 V. Excise fragments ranging from 150 to 300 bp (or sizes suitable for specific need) and extract DNA with one QIAquick column. Elute DNA in 30 µL of Qiagen EB buffer.
4. The protocols for *PCR Amplification of the Library* and *Qualitative and Quantitative Controls of the Library* are described in Subheadings 3.1.6 and 3.1.7, respectively.

3.2.6. Aligning Bisulfite-Converted Reads

When constructing libraries using this second protocol, reads may be observed in either of four forms. This second protocol generates a forward read (+FW) from the Watson strand, the reverse complement (+RC) of +FW, a forward read (–FW) from the Crick strand, and the reverse complement (–RC) of –FW. The FW reads all start with the tag "TCTGT" (the remnant of the DPNI-digested first tag of the protocol), while the RC reads contain the tag "TCCAT". The tags are used to determine whether the read is of type FW or RC, after which the tag is removed from the read for alignment to the genome. The alignment program BS Seeker first converts all Cs to Ts on FW reads and both strands of the reference genome so that the subsequent mapping is performed using only three letters, A, T, G. Similarly, G/A conversion is performed on RC reads and both strands of the reverse complement of the reference genome. Then, it uses Bowtie to map the C/T converted FW reads to the C/T converted Watson and Crick strands, and the G/A converted RC reads to the two G/A converted reverse complements of the Watson and Crick strands. Reads that do not have a tag are treated as if they could be both FW and RC reads. During each of the four runs of Bowtie, the mapped positions for each read are recorded. After all the runs of Bowtie are complete, only unique alignments are retained; such alignments as those that have no other hits with the same or fewer mismatches in the three-letter alignment (between the converted read and the converted genomic sequence). Finally, we calculate the number of mismatches. For this calculation, we consider a read T that aligns to a genomic C as a match, while a read C that aligns to a genomic C is considered a mismatch. Similarly, when aligning RC reads, a read A that aligns to a genomic G is considered a match, while a read G that aligns to a genomic A is considered a mismatch. Low-quality alignments with the number of mismatches larger than the user-defined value are discarded.

4. Notes

1. Wear gloves at all times. Use clean equipment. Try to separate equipment and reagents for library generation from the rest of the lab.
2. Follow the yield of each step by Nanodrop.
3. When eluting QIAquick columns, make sure to add Qiagen buffer EB onto the center of the membrane of the columns.
4. During size selection, leave at least one blank well between samples to avoid cross-contamination. Make sure to change razor blades between samples.
5. If primer dimers are seen after PCR, it is very likely that the library is not good. In this case, it is not recommended to just purify away the primer dimers; instead, it is better to repeat the library construction procedures.
6. The 5′ ends of the library molecules are not phosphorylated and therefore require a phosphorylated vector for cloning.

References

1. Goll, M.G. and T.H. Bestor (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem.* **74**, 481–514.
2. Chan, S.W., Zilberman, D., Xie, Z., Johansen, L. K., Carrington, J. C., and Jacobsen, S. E. (2004) RNA silencing genes control *de novo* DNA methylation. *Science* **303**, 1336.
3. Jones, P.A. and Takai, D. (2001) The role of DNA methylation in mammalian epigenetics. *Science* **293**, 1068–1070.
4. Martienssen, R.A. and V. Colot (2001) DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293**, 1070–1074.
5. McCabe, M. T., Brandes, J. C., and Vertino, P. M. (2009) Cancer DNA methylation: molecular mechanisms and clinical implications. *Clin. Cancer Res.* **15**, 3927–3937.
6. Cokus, S.J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., et al., (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* **452**, 215–219.
7. 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.
8. Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322.
9. Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S., and Jaenisch, R. (2005) Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877.