



EpiGnome™ Methyl-Seq Kit: a novel post-bisulfite conversion library prep method for methylation analysis

Epigenomics is increasingly becoming an important field of research, and the ability to detect and quantify DNA methylation accurately is now critical for numerous fields of study, including disease biology and gene expression. The differential reactivities of methylated and nonmethylated cytosines in DNA with sodium bisulfite forms the basis for their identification in the genome by sequencing. Current whole-genome bisulfite sequencing methods require substantial amounts of starting DNA to compensate for the loss due to bisulfite-mediated DNA degradation. To address issues with sample preparation, we have developed the EpiGnome™ Methyl-Seq Kit, which utilizes a novel pre-library bisulfite conversion scheme to prepare whole-genome bisulfite sequencing libraries with no sample loss and with only 50 ng of input genomic DNA.

DNA methylation is not evenly distributed in the mammalian genome. In human somatic cells approximately 60%–80% of all CpGs (~1% of total DNA bases) are methylated. Numerous methods are available to investigate methylation patterns, including those that are enrichment based (e.g., as methylated DNA immunoprecipitation sequencing (MeDIP-seq)), restriction enzyme based (e.g., methylation-sensitive restriction enzyme sequencing (MRE-seq)) and bisulfite based (e.g., reduced-representation bisulfite sequencing (RRBS) and whole-genome bisulfite sequencing (WGBS)). Of these, the only method that captures the complete methylome of a given sample is WGBS, as only through this approach—in which unmethylated cytosine residues are converted to uracil—can genome-wide analysis of 5-methylcytosines be achieved. However, a major challenge in WGBS is the degradation of DNA that occurs during bisulfite conversion under the conditions required for complete conversion. Typically, ~90% of input DNA is degraded, something that is especially problematic when only limited starting amounts are available. Additionally, regions that are rich in unmethylated cytosines are more sensitive to strand breaks. As a consequence, a majority of DNA fragments contained in di-tagged NGS DNA libraries treated with bisulfite “post-library construction” can be rendered inactive due to strand breaks in the DNA sequence flanked by the adapter sequences. These mono-tagged templates are then excluded during library enrichment, resulting in incomplete coverage and bias when performing whole-genome bisulfite sequencing.

Here, we describe a novel library construction method, called EpiGnome™ Methyl-Seq, for preparing sequencing libraries from bisulfite-converted genomic DNA. This “post-bisulfite conversion” library construction method uses bisulfite-treated single-stranded DNA as template for the subsequent addition of adapter sequences required for cluster generation and sequencing. Thus, single-stranded DNA fragments independent of size and position of strand breaks remain viable templates for library construction, eliminating the loss of fragments and the selection bias associated with a post-library-construction bisulfite conversion strategy. EpiGnome does not require Covaris ultrasonication or an enzyme for fragmentation, or methylated adaptors, both of which are features of existing bisulfite sequencing library prep method workflows.

Method overview

With the EpiGnome™ Methyl-Seq Kit, bisulfite-treated single-stranded DNA (ssDNA) is randomly primed using a polymerase able to read uracil nucleotides to synthesize DNA strands containing a specific sequence tag (**Fig. 1**). The 3' ends of the newly synthesized DNA strands are then selectively tagged with a second specific sequence tag using a patented procedure, resulting in di-tagged DNA molecules with known sequence tags at their 5' and 3' ends. The di-tagged DNA is enriched in PCR, resulting in double-stranded DNA (dsDNA) with the appropriate sequences required for sequencing on any Illumina platform.

EpiGnome requires only 50 ng input gDNA

50 ng of Coriell's lymphoblastoid gDNA (GM12878) was treated with ZYMO's bisulfite conversion kit (EZ DNA Methylation-Lightning Kit)

Anupama Khanna, Agata Czyz & Fraz Syed

Epicentre®, an Illumina® company, Madison, Wisconsin, USA.
Correspondence should be addressed to F.S. (e-mail: fsyed@illumina.com).

APPLICATION NOTES

according to the manufacturers' recommendations. The resulting DNA was used to prepare WGBS libraries as described in **Figure 1**. Additionally, 50 ng of hypermethylated HeLa gDNA (NEB) was treated similarly. PCR amplification of each bisulfite-converted library was performed for 10 cycles. Libraries were sequenced on HiSeq and MiSeq, and data analysis was done using Bismark software (v0.7.2) and a custom analysis pipeline.

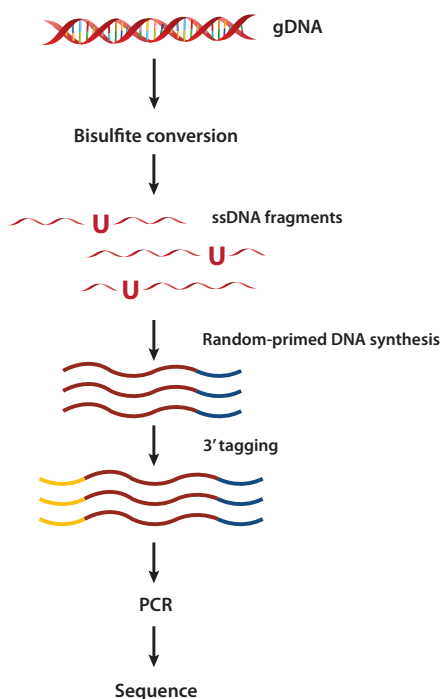


Figure 1 | Workflow for the EpiGnome™ Methyl-Seq Kit.

Sequencing metrics of HeLa and Coriell gDNA

As shown in **Table 1**, both Coriell and HeLa gDNAs have a very high mapping efficiency, with the diversity ranging from 80% to 96% depending on the methylation status of the input DNA. Total conversion of cytosines to thymines in the CpG context is also very

Table 1 | Sequencing metrics

	Coriell, 50 ng	HeLa, 50 ng
Reads % >Q30 ^a	96	96
Mapping efficiency	75.50%	70.20%
Diversity	80.50%	96.30%
Cytosines methylated in CpG context	53.50%	94.10%
Cytosines methylated in CHG context	2.50%	0.60%
Cytosines methylated in CHH context	1.20%	0.70%

^aPercentage of reads with quality score over 30.

high, ranging from above 53% to 95%, again dependent on the methylation status of the input DNA. All these libraries were enriched by 10 cycles of PCR, and the analysis was done using Bismarck v0.7.2.

Methylation calls

Approximately 52% methylated CpGs were observed in the lymphoblastoid DNA, compared to ~95% for the HeLa hypermethylated DNA, which was hypermethylated with CpG methylase. By contrast, CHG and CHH methylation levels were similar for both DNA samples. Uniform coverage across the genome was also observed (data not shown). **Figure 2** highlights CpG methylation patterns across regions of chromosome 1, showing regions of high CpG methylation (red) and low CpG methylation (blue) for GM12878 lymphoblastoid gDNA treated with bisulfite. For the hypermethylated (CpG methylase) HeLa gDNA, the majority of CpGs detected were clearly methylated. The Integrated Genome Viewer plots (IGV) shown in **Figure 2** are based on forward reads for each sample.

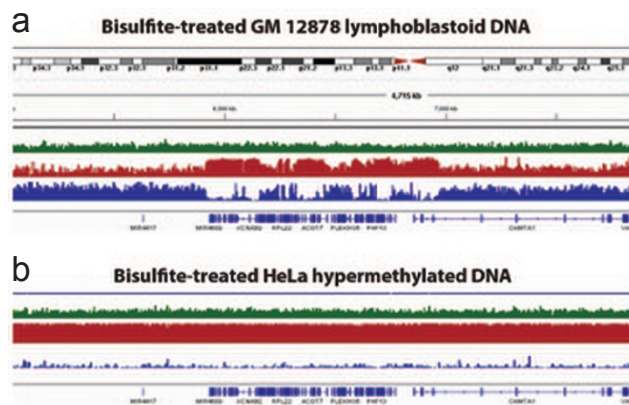


Figure 2 | CpG methylation patterns. **(a)** Example methylation across region of chromosome 1 indicates high CpG methylation (red) and regions of low CpG methylation (blue) from 50 ng inputs of GM12878 lymphoblastoid gDNA treated with bisulfite. The IGV plots shown are based on forward reads for each sample. **(b)** For the hypermethylated (CpG methylase) HeLa gDNA, the majority of CpGs detected were methylated, as seen in red. The IGV plots shown are based on forward reads for each sample.

Conclusion

The EpiGnome™ Methyl-Seq Kit is a simple, 1-day, post-bisulfite conversion library construction method that requires only 50 ng of starting gDNA. EpiGnome does not require a separate enzyme or Covaris ultrasonication for fragmentation or methylated adaptors, both of which are required for current bisulfite sequencing library prep methods. In addition, this approach also results in uniform coverage across all chromosomes and is the method of choice for performing whole-genome bisulfite sequencing.

This article was submitted to *Nature Methods* by a commercial organization and has not been peer reviewed. *Nature Methods* takes no responsibility for the accuracy or otherwise of the information provided.