EM-seq enables accurate and robust methylation detection of cell free DNA and FFPE DNA sample types

Vaishnavi Panchapakesa, Louise Williams, V.K. Chaithanya Ponnaluri, Bradley W. Langhorst, Matthew A. Campbell, Eileen T. Dimalanta, Theodore B. Davis | New England Biolabs, Inc.

INTRODUCTION

DNA methylation is one of the most important epigenetic regulatory mechanisms, and epigenomic changes are recognized as factors that influence tumor initiation, growth and progression. Advances in next generation sequencing, in particular, sample preparation, have aided large scale quantification of DNA methylation. In recent years, there has been great interest in the diagnostic applications of circulating cell-free DNA (cfDNA) and formalin-fixed, paraffin-embedded (FFPE) samples, as they represent a major source of samples in cancer research.

Bisulfite sequencing which chemically converts cytosines to uracils is the most commonly used method for DNA methylation analysis. This chemical-based conversion damages and degrades DNA, resulting in shorter insert sizes as well as introducing bias into the data. Therefore, analysis of DNA methylation from cfDNA and FFPE DNA is challenging as the DNA is typically of low quality and quantity. We have developed an enzyme-based methylation detection technology, called NEBNext® Enzymatic Methyl-seq (EM-seq™) that addresses the drawbacks of bisulfite sequencing. EM-seq minimizes damage to DNA, enabling longer insert sizes, lower duplication rates and reduced GC bias resulting in more accurate quantification of methylation in the DNA.

Using EM-seq, we profiled cfDNA and FFPE DNA from multiple tissue types. Results for these challenging DNA sample types showed that the EM-Seq libraries had similar or longer inserts, lower duplication rates, higher percentages of mapped reads and less GC bias compared to WGBS libraries. These libraries also identified a higher number of CpGs resulting in enhanced coverage across genomic features (TSS, CpG islands). In conclusion, EM-seq facilitates the generation of libraries with superior sequencing metrics resulting in reliable and robust methylation profiling for these types of challenging DNA samples.

METHODS

Sample Preparation

cfDNA or FFPE DNA + Control DNA → Ultra II End Prep/ Ligation → TET2/ Oxidation Enhancer → APOBEC → PCR Amplification → Illumina Sequencing

- cfDNA was extracted from a healthy individual using single donor human plasma (anticoagulant: Na EDTA, Innovative™ Research). QIAamp Circulating Nucleic Acid Kit was used to extract cfDNA from 5 ml of plasma. No carrier RNA was used during the extraction.
- cfDNA: 10 ng and 25 ng of cfDNA (not sheared) was combined with two sheared control DNAs: unmethylated lambda (2 ng) and CpG methylated pUC19 (0.1 ng)
- FFPE DNA was obtained commercially from Biochain. 10 or 50 ng of FFPE DNA were combined with two control DNAs: unmethylated lambda (2 ng) and CpG methylated pUC19 (0.1 ng), prior to shearing to 300bp
- DNA was end repaired and ligated to EM-seq adaptors
- 5mC and 5hmC were protected from APOBEC deamination by TET2/Oxidation Enhancer
- Cytosines were deaminated to uracils using APOBEC
- Libraries were amplified with NEBNext Q5U™ Master Mix and Unique Dual Index Primer Pairs
- Libraries were sequenced using an Illumina NovaSeq 6000, 2x100 base paired reads
- Bisulfite conversion was performed using Zymo Research EZ DNA Methylation-Gold™ kit

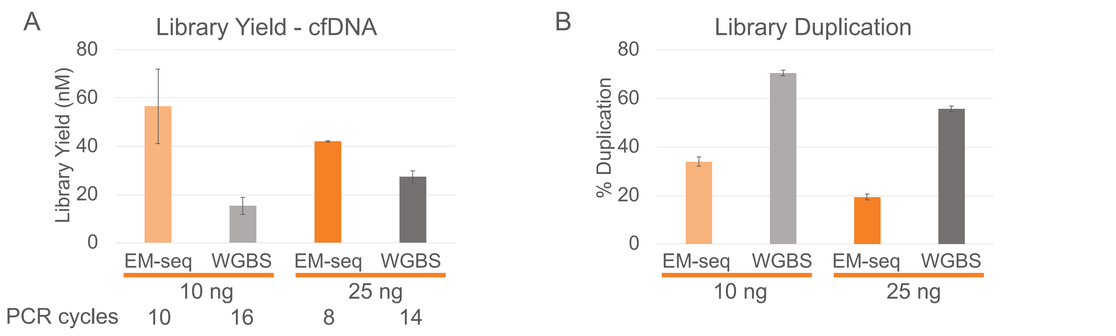
Data Analysis

Trim → Align (BWAMeth) → Deduplicate → Picard → Methyl Dackel → MethylKit

- Reads were aligned to hg38 using BWA-Meth and Methylation levels were extracted using MethylDackel
- Correlation analysis at 1x minimum coverage was performed used methylKit 1.4.0
- Picard 2.17.2 was used for determining library insert size and GC bias

RESULTS

cfDNA: Higher Quality Sequencing Data with EM-seq Libraries



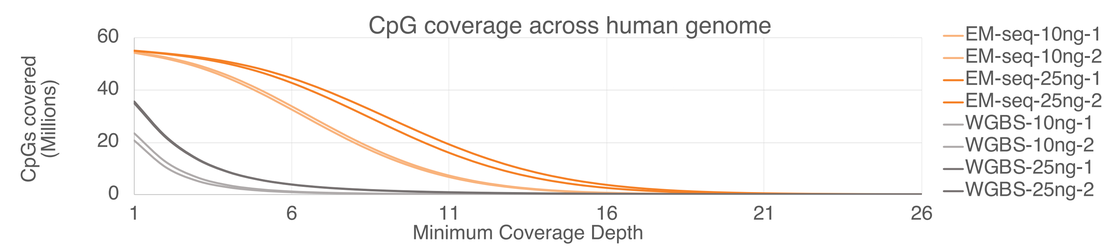
EM-seq and WGBS library quality and sequencing metrics using 10 ng and 25 ng cfDNA. Each library was sequenced using the Illumina NovaSeq 6000. (A) EM-seq libraries have higher yield using fewer PCR cycles compared to WGBS. (B) Library duplication percentages are lower for EM-seq. (C) Insert size distribution is similar between EM-seq and WGBS libraries for cfDNA samples. (D) EM-seq libraries show more even GC coverage distribution than bisulfite libraries. The bisulfite libraries are AT rich and have lower GC coverage.

		EM-seq	WGBS
% methylation (10 ng)	CpG	76. ± 0.42	77.80 ± 0.14
	CHG	0.95 ± 0.07	0.35 ± 0.07
	CHH	0.90 ± 0.14	0.35 ± 0.07
% methylation (25 ng)	CpG	76.45 ± 0.07	78.7 ± 0.14
	CHG	0.75 ± 0.07	0.60 ± 0.14
	CHH	0.75 ± 0.07	0.60 ± 0.14

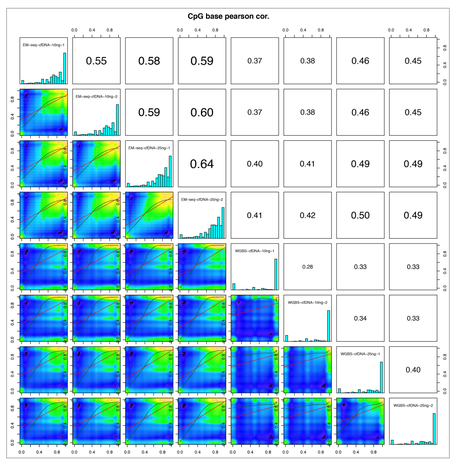
The percentage methylation for 10 ng and 25 ng cfDNA in CpG, CHG and CHH contexts. For cfDNA, CpG methylation levels are similar for all libraries. For unmethylated lambda control DNA: <1% methylated cytosines in CpG, CHG and CHH contexts were detected for all libraries (data not shown).

RESULTS

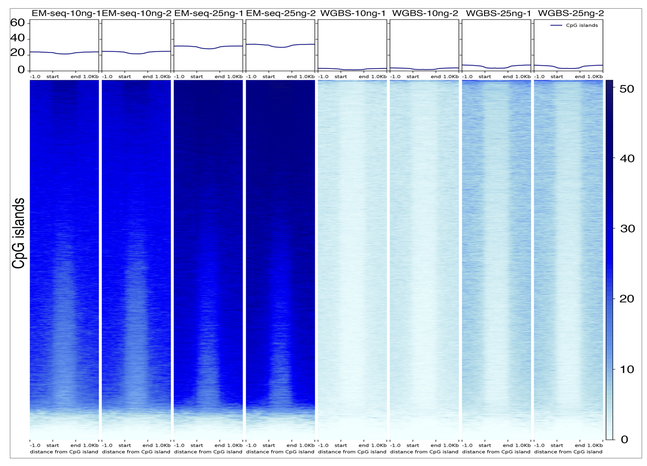
cfDNA: EM-seq Libraries are Superior



CpG Coverage at different coverage depths. Top and bottom strand CpGs were counted independently, yielding a maximum of 56 million possible CpG sites. EM-seq libraries identified more unique CpGs than bisulfite libraries for 10 ng and 25 ng inputs. EM-seq libraries have a higher number of CpGs covered with a coverage depths between 6x and 15x providing more usable data.

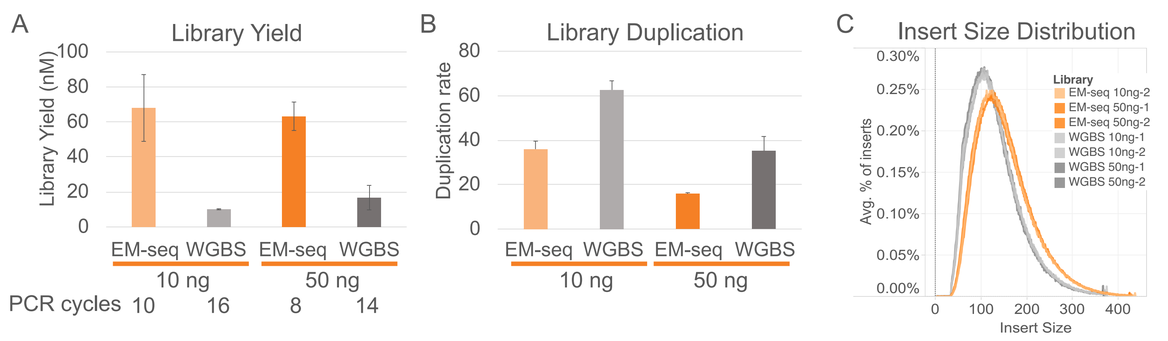


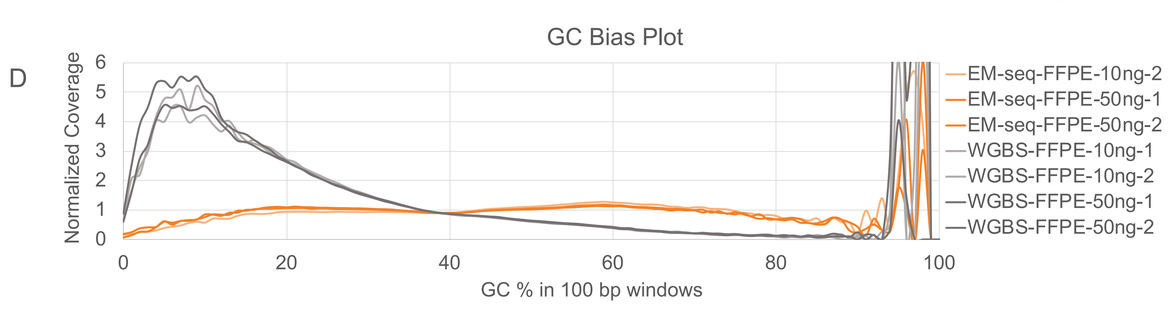
Pearson's correlations were plotted using MethylKit for 10 ng and 25 ng EM-seq and WGBS libraries at 1x minimum coverage (8 million CpGs common to all libraries).



CpG coverage across CpG islands are represented as heatmap (+/- 1kb from the start and end sites). Dark blue indicates high coverage and light blue/white indicate little or no coverage. The heatmaps show that EM-seq has higher coverage at all DNA inputs.

FFPE DNA: Higher Quality Sequencing Data with EM-seq Libraries

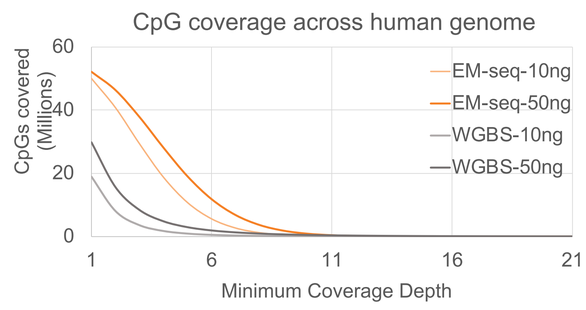




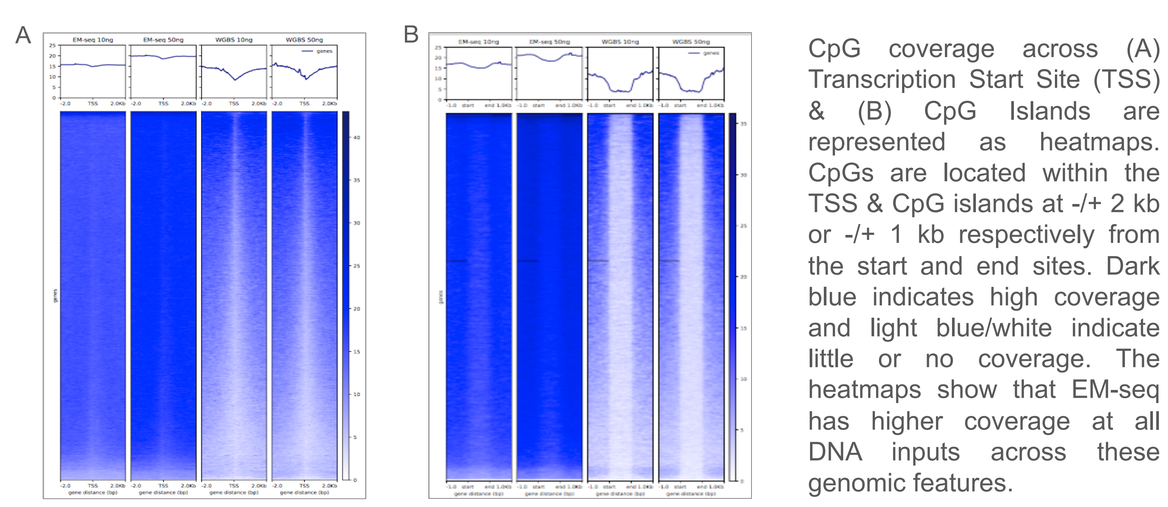
EM-seq and WGBS metrics from 10 ng and 50 ng FFPE DNA. Each library was sequenced using the Illumina NovaSeq 6000. (A) EM-seq libraries have higher yield using fewer PCR cycles compared to WGBS. (B) Library duplication percentages are lower for EM-seq. (C) EM-seq libraries have larger insert sizes compared to WGBS.(D) EM-seq libraries show more even GC coverage distribution than bisulfite libraries. The bisulfite libraries are AT rich and have lower GC coverage.

		EM-seq	WGBS
% methylation (10 ng)	CpG	75.65 ± 0.35	78.1 ± 0.14
	CHG	0.4 ± 0.00	0.5 ± 0.00
	CHH	0.4 ± 0.00	0.3 ± 0.00
% methylation (50 ng)	CpG	76.75 ± 0.07	78.3 ± 0.28
	CHG	0.3 ± 0.00	0.5 ± 0.00
	CHH	0.6 ± 0.00	0.6 ± 0.00

The percentage methylation for 10 ng and 50 ng FFPE DNA in CpG, CHG and CHH contexts. CpG methylation levels are similar for all libraries. Unmethylated Lambda: <1% methylated Cs in CpG, CHG and CHH were detected for all libraries (data not shown).



CpG Coverage at different coverage depths. Top and bottom strand CpGs were counted independently, yielding a maximum of 56 million possible CpG sites. EM-seq libraries identified more unique CpGs than bisulfite libraries for 10 ng and 50 ng inputs.



CpG coverage across (A) Transcription Start Site (TSS) & (B) CpG Islands are represented as heatmaps. CpGs are located within the TSS & CpG islands at -/+ 2 kb or -/+ 1 kb respectively from the start and end sites. Dark blue indicates high coverage and light blue/white indicate little or no coverage. The heatmaps show that EM-seq has higher coverage at all DNA inputs across these genomic features.

CONCLUSIONS

Identification of CpGs within cfDNA and FFPE DNA using the EM-seq method is robust compared to whole genome bisulfite sequencing.

- Higher library yields with less PCR cycles
- Less GC bias
- Lower percent duplication
- Detects more CpGs with fewer reads

Provides new method to evaluate low input cfDNA with higher concordance between the replicates for accurate methylation-based biomarker detection

We thank Laurie Mazzola and Danielle Fuchs from the NEB Sequencing Core for their assistance.