High-throughput methyl-binding domain-based DNA enrichment and sequencing

Kaylinnette Pinet, Anagha Kadam, Laura Blum, Matthew Angel, Matthew Campbell, Bradley W. Langhorst, V K Chaithanya Ponnaluri, Keerthana Krishnan

New England Biolabs, Ipswich, MA, USA

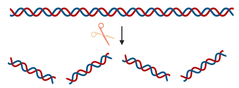
Introduction

High-throughput methylated DNA enrichment can be an important tool to support high throughput and large-scale epigenetic studies, especially when paired with information-rich sequencing analyses. An established approach for selectively binding methylated DNA involves the fusion of a methyl-CpG binding domain from human MBD2 to the Fc tail of human IgG1 (MBD2-Fc) and the coupling of this fusion protein to paramagnetic hydrophilic protein A beads (MBC2-Fc/Protein A Magnetic Bead). This stable complex will selectively bind double-stranded methylated CpG containing DNA, whether it is intact gDNA or fragmented gDNA (<1 kb) from single or mixed-genome samples. Here we describe an optimized protocol for the enrichment of methylated DNA from mixed genome samples and cell free-DNA (cfDNA), upstream of sequencing.

Methods

We employed a novel, streamlined, and high-throughput workflow for MBD2-Fc/Protein A Magnetic Bead-based enrichment relative to the standard approach. Mixed genome control samples were used to track enriched DNA yields as a measure of methylated-DNA binding capacity. We utilized standard DNA sequencing and enzymatic methylation sequencing to assess binding specificity. Finally, we applied this novel methylated DNA enrichment workflow to cfDNA samples, to evaluate epigenetic marker enrichment applications.

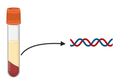
DNA Sample Types



Sheared control gDNA mixes
(7:3 Human : *E. coli* gDNA mix)

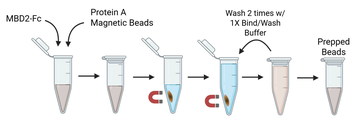
OR

Cell-free DNA
(Human blood serum
+/- methylated ct-DNA spike-in)

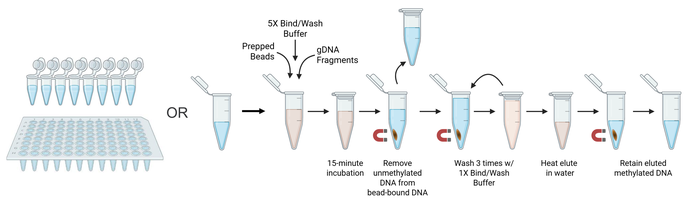


Methylated CpG DNA Enrichment Workflow[1]

1 MBD2-Fc/Protein A magnetic bead preparation



2 High OR Low throughput (HT/LT) meCpG DNA binding



Library Preparation, Sequencing, and Analysis

3 Assess meCpG binding specificity of control gDNA with NEBNext® Enzymatic Methyl-seq v2 library prep

End Repair/ dA-Tailing → Adaptor Ligation → Clean Up → Oxidation of 5mC & 5hmC → Clean Up → Deamination of C to U → PCR Amplification → Clean Up

4 Evaluate DNA enrichment of gDNA and cfDNA with NEBNext® UII DNA library prep

End Repair/ dA-Tailing → Adaptor Ligation → Clean Up → PCR Amplification → Clean Up

Illumina® Sequencing

- Sequenced 2x76 or 2x100
- Down-sampled with Seqtk[2]
- Kraken2 assessment of all reads
- Ultra II DNA library reads were mapped to a composite genome (Human and *E.coli)* using Bowtie2[3]
- For % methylation values from EM-seq v2, reads were mapped to a composite reference genome with BWA-meth[4] and methylation information was extracted using MethylDackel
- Enrichment assessment from Ultra II DNA libraries was performed by looking at percent Human vs *E.coli* in bound libraries and % CpG island coverage[5]

References
1. Created in https://BioRender.com
2. H. Li, Seqtk: a fast and lightweight tool for processing FASTA or FASTQ sequences, 2013. https://github.com/lh3/seqtk/.
3. Langmead, B., Salzberg, S. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359 (2012). https://doi.org/10.1038/nmeth.1923.
4. Pedersen, B., & Eyring, E. (2014). Fast and accurate alignment of long bisulfite-seq reads. arXiv preprint arXiv:1401.1129.
5. M. Gardiner-Garden, M. Frommer, CpG Islands in vertebrate genomes, Journal of Molecular Biology, Volume 196, Issue 2, 1987, Pages 261-282, ISSN 0022-2836, https://doi.org/10.1016/0022-2836(87)90689-9.

Results

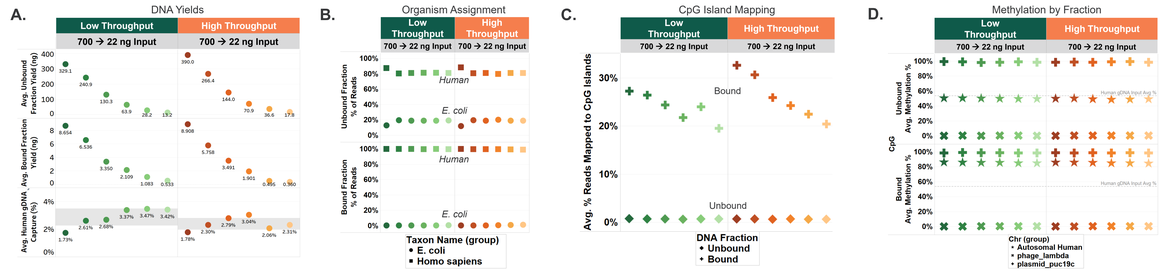


Figure 1. Percent of captured DNA and percent of CpG DNA methylation are equivalent between the low and high throughput meCpG DNA enrichment workflows from a broad input range. Reagents from the EpiMark® Methylated DNA Enrichment Kit were used on 7:3 Human NA12878 gDNA (Coriell Institute for Medical Research) and *Escherichia coli* gDNA (Lofstrand Labs Limited) mixed samples. Total DNA amounts were 700 ng, 350 ng, 175 ng, 88 ng, 44 ng, and 22 ng prior to the MBD binding reactions. Standard DNA or methyl-conversion (EM-seq v2) libraries were prepared from pre-binding, unbound, and bound DNA fractions following the NEBNext UII DNA workflow or EM-seq™ v2 workflow. For EM-seq v2 libraries, pre-sheared unmethylated lambda and CpG-methylated pUC19 spike-in controls were utilized after DNA enrichment. Libraries were sequenced on an Illumina sequencing platform. Reads for each library were aligned to a composite (*E. coli* MG1655 + human hs1 + methylation controls) reference genome using Bowtie2[3] for standard DNA libraries (10M reads) and BWA-meth[4] for EM-seq libraries (~20M reads). **A**) Average DNA yields for the unbound and bound DNA fractions from the input range tested, determined via TapeStation® (n=4 per input). Percent of input human DNA captured calculated based on bound fraction yield; standard deviation for % captured across all input amounts is shaded in grey (n=24). **B**) Percent of reads identified as human or *E.coli* DNA by Kraken2 (n=2 per input). **C**) Percent DNA mapping to human CpG islands[5] (n=2 per input) **D**) EM-seq-based percent methylation in CpG context (n=2 per input). Expected levels of methylation for unmethylated lambda (<1%), CpG methylated pUC19 (>96%) and NA12878 human gDNA (~54%). Human methylation in Bound samples is higher than the average methylation level suggesting enrichment of fragments with higher methylation levels.

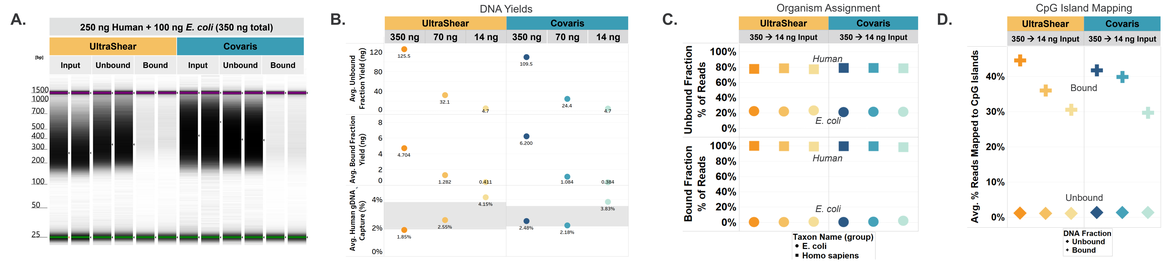


Figure 2. High specificity of meCpG binding from enzymatically and mechanically sheared mixed gDNA is demonstrated based on genomic partitioning and CpG island enrichment. Reagents from the EpiMark Methylated DNA Enrichment Kit were used on 7:3 Human NA12878 gDNA and *Escherichia coli* gDNA mixed samples. gDNA samples were either enzymatically fragmented (NEBNext Ultrashear®) or mechanically sheared (Covaris ME220). Total DNA amounts were 350 ng, 70 ng, and 14 ng prior to the high throughput MBD binding reactions. NEBNext Ultra II DNA libraries were prepared from pre-binding, unbound, and bound DNA fractions. Libraries were sequenced on an Illumina platform. 10M reads for each library were aligned to a composite reference genome using Bowtie2[3]. **A**) TapeStation® of 350 ng total input samples and post-binding fractions show expected DNA size and yield. **B**) Average DNA yields for the unbound and bound DNA fractions from the input range tested, determined via TapeStation (n=2 per input). Percent of input human DNA captured calculated based on bound fraction yield; standard deviation for % captured across all input amounts is shaded in grey (n=6). **C**) Percent of reads identified as human or *E.coli* DNA by Kraken2 (n=2 per input). **D**) Percent DNA mapping to human CpG islands[5] (n=2 per input).

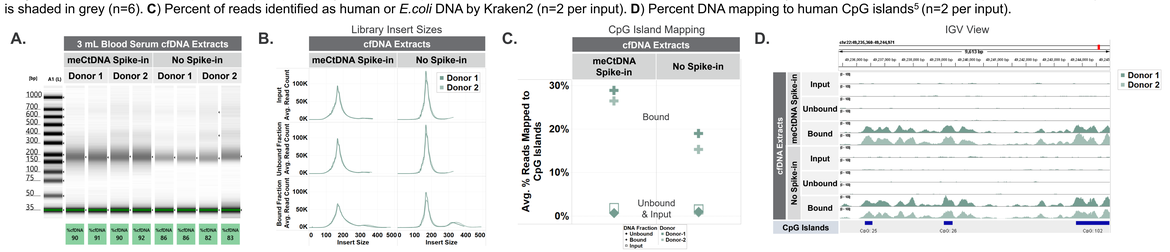


Figure 3. Epigenetic marker variability is detectable in NEB's Monarch cfDNA isolates from human blood serum. Monarch cfDNA extraction reagents and workflows were applied to human blood serum +/- 10 ng spike-in of commercially available ctDNA control material, (>90% methylated, Seraseq). Total cfDNA yields were 5 ng - 10 ng prior to enrichment. EpiMark Methylated DNA Enrichment Kit reagents were applied to isolated cfDNA. NEBNext Ultra II DNA libraries were prepared from pre-binding, unbound, and bound DNA fractions. Libraries were sequenced on an Illumina platform and 10M reads for each library were aligned to a composite reference genome using Bowtie2[3]. **A**) Extracts evaluated using TapeStation cfDNA reagents to confirm high-quality cfDNA. **B**) insert sizes distributions for each sample. **C**) Percent DNA mapping to human CpG islands[5] (n=2 per input). **D**) IGV view of 3 human CpG islands show enrichment in bound cfDNA fractions.

Conclusions

This optimized, automation-friendly, MBD-based DNA enrichment technique:

- increases the throughput of the standard MBD2-Fc/Protein A Magnetic Bead-based process
- is compatible downstream of enzymatic fragmentation using NEBNext® UltraShear which maintains methylation marks
- is compatible downstream of the automation friendly Monarch® Cell-Free DNA extraction workflow
- enhances the power of methylated DNA sequencing by reducing sequencing depth requirements for regulatory region assessment from various sample types and DNA input amounts
- facilitates the understanding of epigenetic markers and their roles in gene regulation and disease

Poster #TT054