Applications of innovative whole-genome bisulfite sequencing (WGBS) for cancer biomarker discovery in plasma liquid biopsy

Caila Ruiz[1], Hanjun Kim[1], Ryan Yancey[1], Jeffrey Bhasin[1], Mingda Jin[1], Yi Xu[1], Xiaojing Yang[1], Larry Jia[1]*

[1]Zymo Research Corporation, Irvine, CA, USA;

xyang@zymoresearch.com

Abstract

Circulating cell-free DNA (cfDNA) has been reported to contain valuable genetic and epigenetic information for the diagnosis and prognosis of cancer. Studies have shown that the blood from cancer patients contains more tumor-derived cfDNA compared to the healthy controls. DNA methylation, the most well-studied epigenetic marker, has been validated as one of the key drivers in the development of many diseases including cancer. cfDNA methylation thus holds great potential to become a biomarker that will enable early detection of cancer. However, cfDNA is also a challenging input for library preparation as cfDNA is mostly highly fragmented, and the amount of cfDNA from blood plasma is comparatively low. Two classes of library preparation protocols were evaluated for profiling DNA methylation patterns at single nucleotide resolution from ultra-low amount of blood plasma cfDNA. These methods included a bisulfite conversion-based protocol and an enzymatic-based protocol. The bisulfite conversion reaction was optimized allowing a milder treatment for less nucleic acid damage. This bisulfite conversion-based protocol also contained a further optimized adapter ligation step, allowing a simpler procedure. Circulating cfDNA were extracted from blood plasma of a healthy control, a non-small cell lung cancer patient, and an adenocarcinoma lung cancer patient. 5-ng cfDNA was used as input and the libraries were successfully prepared using both methods. Each library was sequenced to 300 – 600 million read pairs at a read length of 150 bp, enabling a median coverage of 20X per detected CpG, and 86.8% of the 30 million CpG sites in human genome were covered by at least 1 read. Unique alignment rate among all libraries were about or above 80%. The comprehensive coverage of the CpG sites across the entire human genome illustrated that the optimized bisulfite conversion reaction was compatible with fragmented cfDNA, proving the satisfactory efficacy in this simple and classic method for DNA methylation profiling of such a challenging sample type. This opens the door to apply the streamlined and simple Whole-Genome Bisulfite Sequencing (WGBS) protocol for cfDNA methylation profiling, which in turn shall facilitate further understanding in cfDNA epigenetic variations and advance the development of cancer biomarkers from the non-invasive, liquid biopsies.

Keywords: Liquid Biopsy, Cancer, Epigenetic, Whole genome bisulfite sequencing

Difficult Library Prep Made Simple

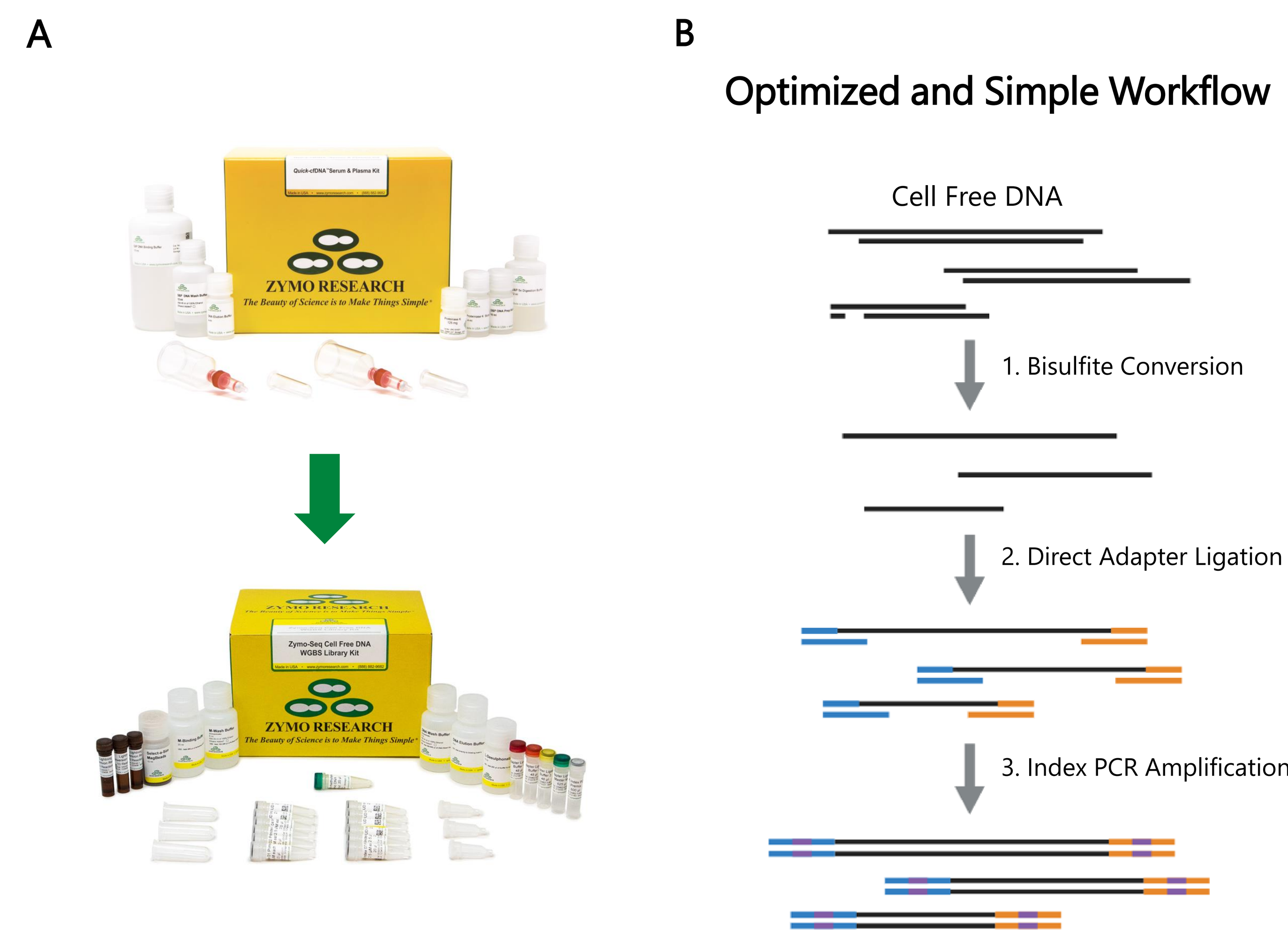


Figure 1. Zymo Research provides excellent products for methylation analysis of cfDNA from sample purification to library preparation. (A) The *Quick-cf*DNA™ Serum & Plasma Kit was used to extract and purify cfDNA from plasma samples from multiple donors. After purification, the purified cfDNA was ready for bisulfite conversion and library preparation using the Zymo-Seq™ Cell Free DNA WGBS Library Kit. (B) Overview of the Zymo-Seq™ Cell Free DNA WGBS Library Kit protocol. The cfDNA is first bisulfite converted using optimized conditions for fragmented input. Next, the innovative adapters capture and directly ligate onto any size DNA fragment, thus accurately preserving the methylation status of each terminus. Finally, the adapter ligated cfDNA is indexed and amplified via PCR, and the libraries are ready for sequencing on any Illumina instrument.

Quality Libraries from Precious cfDNA Samples

High Quality cfDNA Libraries

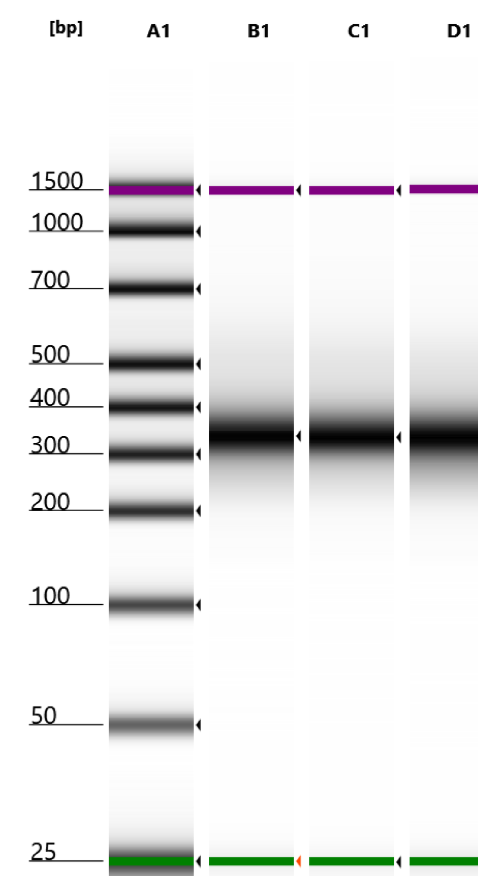
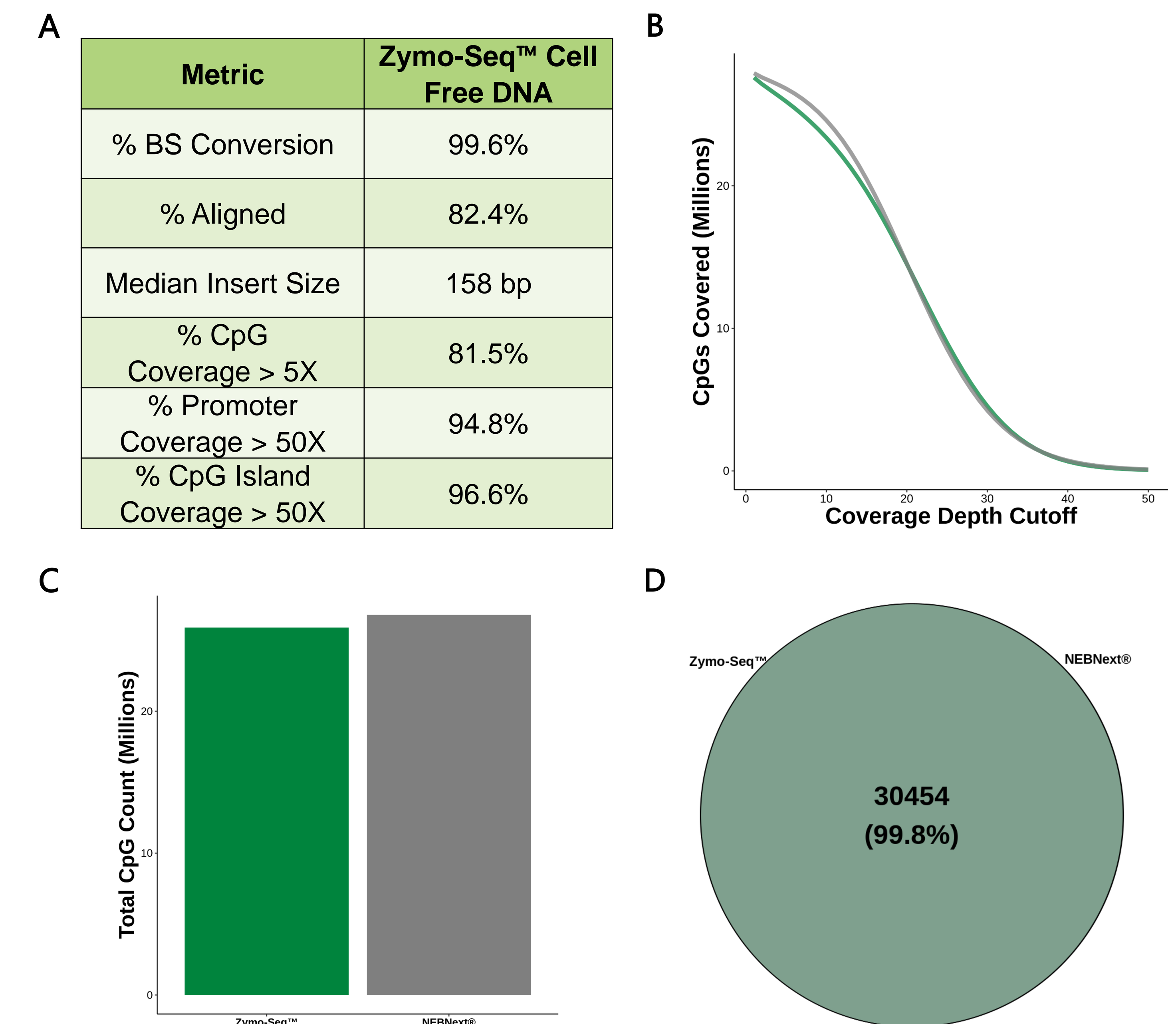


Figure 2. Zymo-Seq™ Cell Free DNA WGBS libraries prepared from multiple cfDNA samples. Agilent 4200 TapeStation HS D1000 gel of libraries prepared using cfDNA extracted from plasma of healthy and cancerous donors. A1 is the molecular weight marker. B1 was prepared from a healthy 59-year-old donor. C1 was prepared from a lung cancer NSCLC stage IV 66-year-old donor. D1 was prepared from an adenocarcinoma stage IV 69-year-old donor.

High Performance On Library Complexity And Coverage



Metric	Zymo-Seq™ Cell Free DNA
% BS Conversion	99.6%
% Aligned	82.4%
Median Insert Size	158 bp
% CpG Coverage > 5X	81.5%
% Promoter Coverage > 50X	94.8%
% CpG Island Coverage > 50X	96.6%

Figure 4. Zymo-Seq™ Cell Free DNA WGBS libraries are concordant with enzymatic methods. (A) Libraries were prepared from 5 ng input cfDNA and sequenced via NovaSeq 6000 with approximately 400M PE sequencing reads. Reads were aligned to hg38 using Bismark and methylation calling was performed with MethylDackel. Quality control metrics from the sequencing run are listed. (B) Libraries were prepared from 5 ng input cfDNA for each method. The number of CpG sites retained was determined at increasing coverage depth cutoffs. Both the enzymatic and bisulfite conversion library prep methods achieved comparable coverage depth profiles. (C) Methylation calls for each data set were filtered to a minimum 5X coverage and counted. Both libraries covered nearly 90% of all CpG loci. When genomic loci were overlapped, 95% were identical between the libraries. (D) CpG loci with at least 5X coverage were annotated with known hg38 promoters. Zymo-Seq™ and NEBNext® libraries covered ~98% of possible promoters with 99.8% agreement between the two sets. (Green: Zymo-Seq™; Grey: NEBNext®)

Accurate Methylation Calling Across the Entire Read

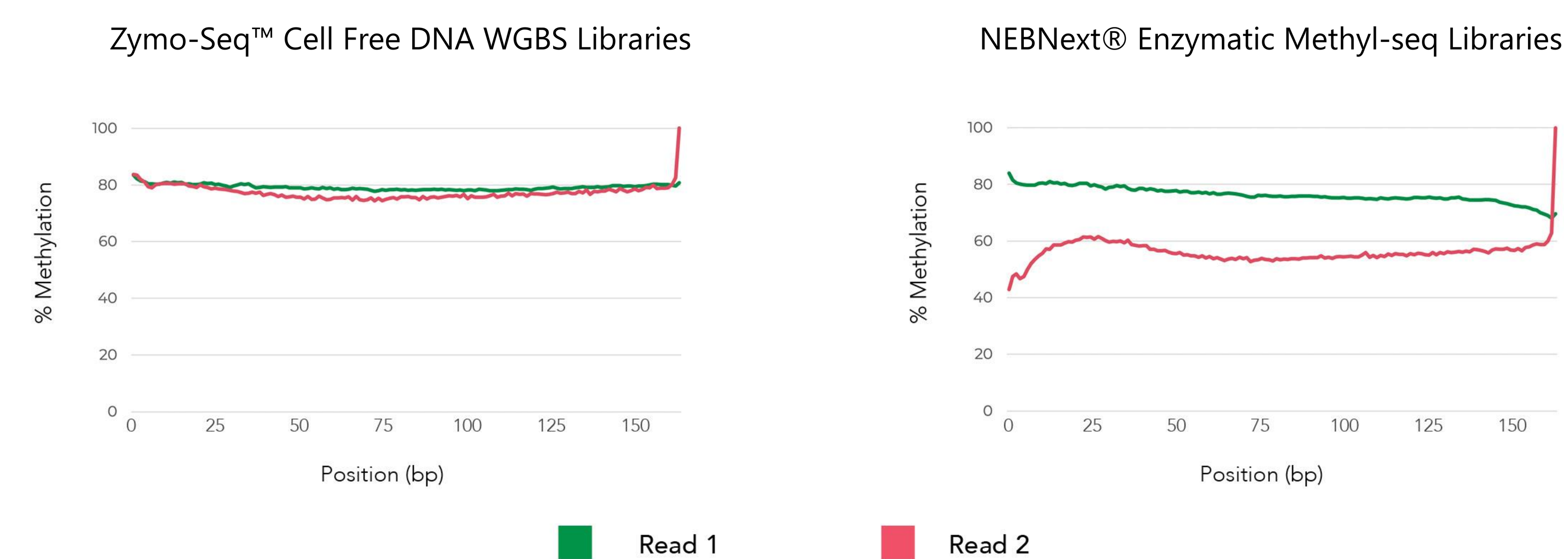


Figure 3. Zymo-Seq™ Cell Free DNA WGBS libraries minimize library preparation bias commonly found in conventional methods. Unbiased libraries will have consistent methylation levels across the entire read length. The NEBNext® Enzymatic Methyl-seq Kit protocol includes an end-repair step that incorporates artificial nucleotides to blunt damaged DNA termini, resulting in significant methylation bias on the 3' end of the DNA fragments. The Zymo-Seq™ Cell Free DNA WGBS Library Kit directly ligates the adapters, eliminating the need for end-repair and thus preserving the integrity of native methylation present on the fragment termini. The Zymo-Seq™ Cell Free DNA WGBS library (left) shows consistent CpG methylation across both Read 1 and Read 2 whereas the NEBNext® Enzymatic Methyl-seq library (right) shows significant bias. The M-Bias plots shown above were generated by plotting the average CpG methylation levels across each position of the mapped read.

Sensitive Detection of Cancer Methylation Differences

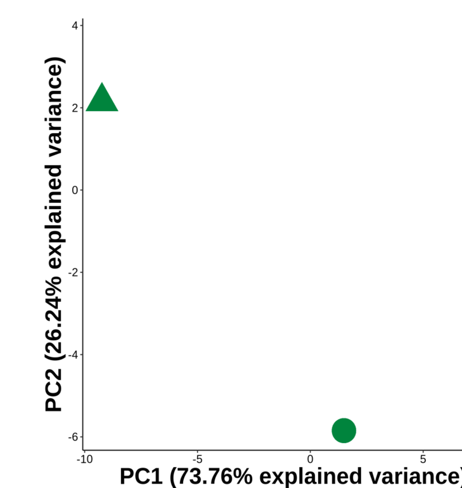


Figure 5. Cancer samples from healthy samples separate based on loci variances between datasets. Loci with at least 5X coverage were ranked by the percent methylation variance across all samples (standard deviation), and the 100 most variant sites were selected to perform Principal Component Analysis (PCA). Samples can be separated based on disease state (cancer or healthy). (▲ – Healthy; ● – Cancer)

Conclusion

The Zymo-Seq™ Cell Free DNA WGBS Library Kit produces high quality DNA methylation libraries from cell-free DNA. This streamlined and optimized WGBS library prep kit for cfDNA will facilitate fundamental research necessary for improvement and accessibility of personalized medicine. The accuracy and quality of the methylation calls demonstrate the power of free-floating DNA methylation profiling with the potential for applications in novel cancer biomarker discovery from liquid biopsies.