INTRODUCTION

DNA methylation is one of the most important epigenetic regulatory mechanisms. The ability to accurately identify 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) gives us greater insight into potential gene regulatory mechanisms. Bisulfite sequencing (BS) is traditionally used to detect methylated cytosines, however, the chemical based conversion of cytosines to uracils leads to DNA damage which subsequently translates to shorter DNA insert sizes as well as biases in the data. To overcome these limitations, we developed NEBNext® Enzymatic Methylation-Seq (EM-seq™), an enzymatic approach for detecting methylcytosine.

EM-seq and BS Illumina libraries were prepared using 10 ng to 200 ng NA12878 DNA. EM-seq libraries have longer insert sizes and less GC bias compared to bisulfite converted libraries. Global methylation levels are similar between the two methods, indicating overall detection of methylated Cs is similar. However, CpG correlation plots demonstrated higher correlations for EM-seq indicating that EM-seq libraries are more consistent than BS across replicates and input amount. GC Bias and bisulfite-degradation showed that EM-seq has more even GC distribution compared to the AT rich representation observed for BS. EM-seq libraries exhibit no even coverage allowing for a higher percentage of CpGs to be assessed and therefore leading to more consistent evaluation of methylation across key genomic features (TEs, CpG island, etc.).

There is increasing interest in the diagnostic applications of circulating cell-free DNA (cDNA). Analysis of DNA methylation from cfDNA is challenging as the DNA is typically of low quantity and quality. EM-seq and BS libraries were made using cDNA. EM-seq libraries had longer inserts, higher duplication rates, higher percentages of mapped reads and less GC bias compared to BS libraries. These libraries also identified a higher number of CpGs resulting in enhanced coverage across genomic features, such as transcription start sites (TSS) and CpG islands. EM-seq is robust and reproducible, facilitating the generation of libraries with superior sequencing metrics for these challenging DNA samples.

METHODS

Human NA12878: Higher Quality Sequencing Data with EM-seq Libraries

A

<table>
<thead>
<tr>
<th>PCR CYCLES</th>
<th>YIELD (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EM-seq</td>
<td>WGBS</td>
</tr>
<tr>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>15</td>
<td>60</td>
</tr>
<tr>
<td>20</td>
<td>50</td>
</tr>
</tbody>
</table>

B

- plots were generated using methylKit at a 1x minimum CpG coverage. EM-seq libraries have higher correlations.

(C) The number of CpGs covered for EM-seq and WGBS libraries were calculated and graphed at minimum coverage depths of 1x through 2x. (B) Genomic features detected in EM-seq and WGBS libraries. Coverage at 5x or greater is indicated. EM-seq libraries have higher coverage for all features examined. Features from WGBS’s RatLseq, the Eukaryotic Promoter, UCSC table browser and DmTm are shown. Coverages are represented with one point per region and the average coverage depth is plotted.

RESULTS

EM-seq and WGBS metrics from 10 ng, 50 ng and 200 ng NA12878 genomic DNA. Each library was sequenced using the Illumina NovaSeq 6000, 324 million, 2x150 base reads were used for methylation analysis. (A) EM-seq libraries have higher yield but require fewer PCR cycles. (B) EM-seq library insert sizes are larger than bisulfite libraries. (C) Library duplication percentages are lower for EM-seq and (D) the GC distribution of reads and less GC bias compared to BS libraries. These libraries have even coverage more than bisulfite libraries. Bisulfite libraries are AT-rich and have lower GC coverage.

cfDNA: Higher Quality Sequencing Data with EM-seq Libraries

A

- plots were generated using methylKit at a 1x minimum CpG coverage. EM-seq libraries have higher correlations.

(a) The number of CpGs covered for EM-seq and WGBS libraries were calculated and graphed at minimum coverage depths of 1x through 2x. (B) Genomic features detected in EM-seq and WGBS libraries. Coverage at 5x or greater is indicated. EM-seq libraries have higher coverage for all features examined. Features from WGBS’s RatLseq, the Eukaryotic Promoter, UCSC table browser and DmTm are shown. Coverages are represented with one point per region and the average coverage depth is plotted.

EM-seq and WGBS cfDNA library coverage across CpG islands. Heatmaps depict CpGs at +/- 14bp from the start and end sites of the CpG islands. 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.

CONCLUSIONS

Identification of CpGs using the EM-seq method is superior to whole genome bisulfite sequencing:
- Higher yields with less PCR cycles
- Larger library insert sizes
- More even base coverage
- Less GC bias
- Detects more CpGs with fewer reads

Identification of CpGs within cDNA using the EM-seq method is robust compared to whole Genome bisulfite sequencing.

Provides a new method to evaluate the low input cDNA with higher concordance between the replicates for accurate methylation based biomarker detection.

We thank the NEB sequencing core (Laura Mazzotta, Danielle Fuchs, Kristin Augugliaro and Harold Bell) for their technical assistance.