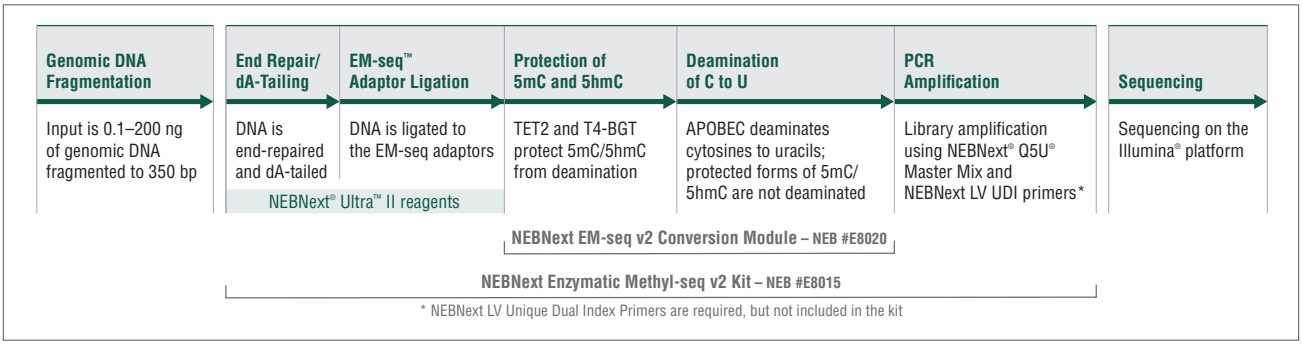
NEBNext® Enzymatic Methyl-seq v2 Kit

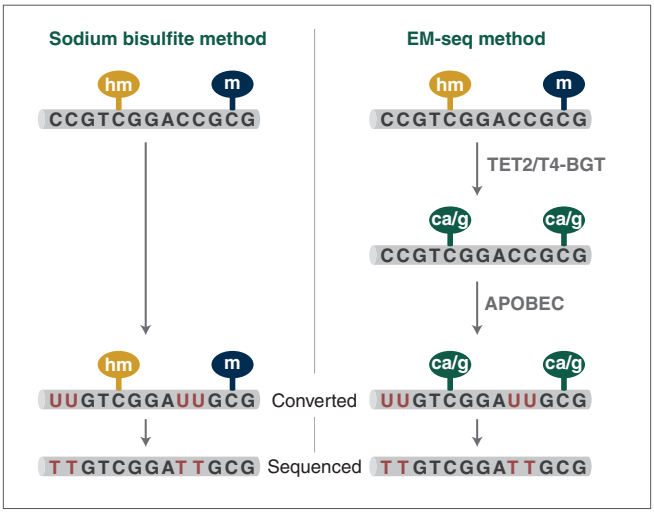
NEB #E8015

FIGURE 1: EM-seq™ v2 workflow



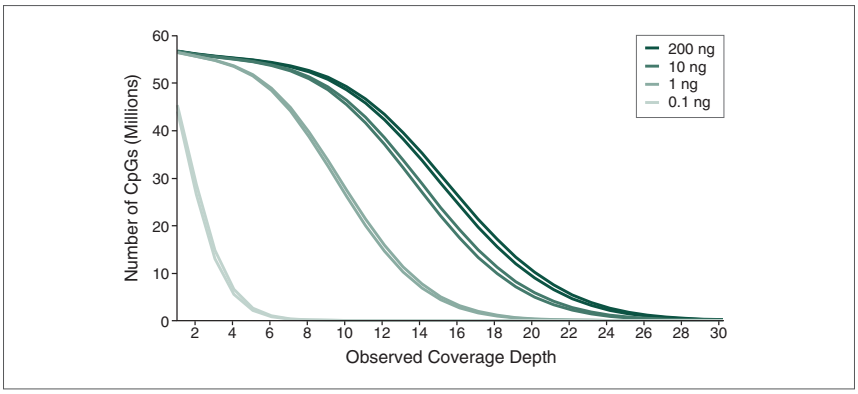
The EM-seq v2 workflow accommodates a wider input range than the original EM-seq workflow, with a 100-fold lower minimum input amount. The v2 workflow is more streamlined, has one fewer cleanup step and is 30 minutes faster. Note that NEBNext LV UDI primers are not included in the kit and are available separately.

FIGURE 2: EM-seq conversion method



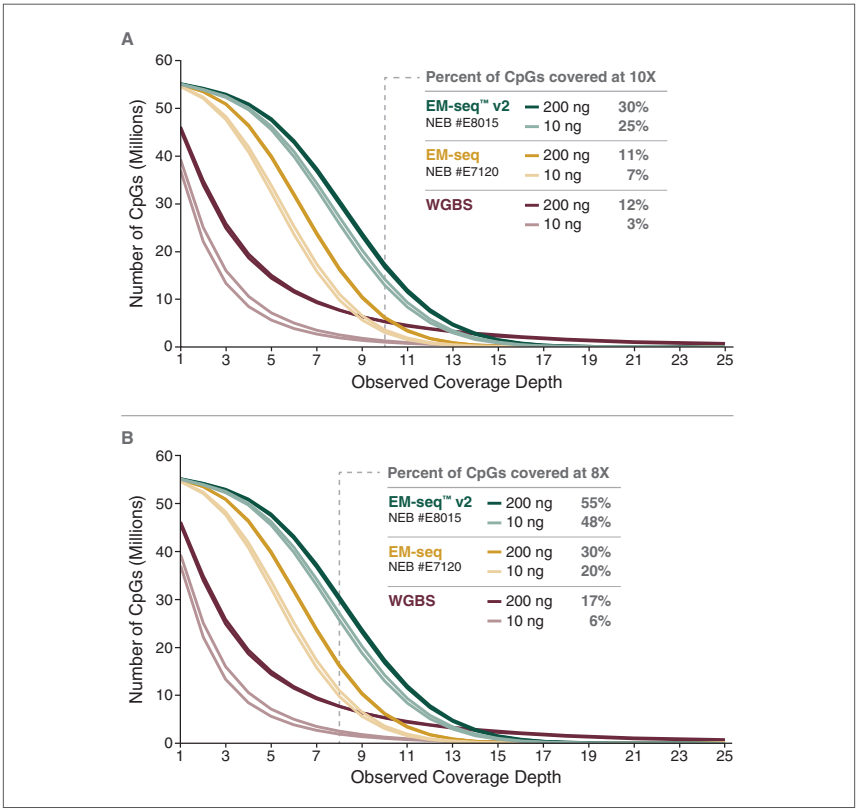
The harsh sodium bisulfite treatment deaminates unmodified cytosines to uracil. In the EM-seq workflow, 5mC and 5hmC are first protected using the enzymes TET2 and T4-BGT. Unmodified cytosines are then deaminated by the APOBEC enzyme to uracil, while the protected 5mC and 5hmC are not converted. During Illumina®sequencing, 5mCs and 5hmCs are represented as cytosine, while unmodified cytosines are represented as thymine.

FIGURE 3: **EM-seq v2 exhibits high CpG coverage across a range of inputs**



EM-seq v2 libraries were prepared from 200–0.1 ng of NA12878 DNA (sheared to 350 bp using Covaris® ME220), spiked with unmethylated lambda and CpG-methylated pUC19. Libraries were sequenced on an Illumina NovaSeq® 6000 (2 x 150 bases). Approximately 910 million reads for each library were aligned to a composite human T2T, lambda and pUC19 reference genome using bwa-meth. The T2T genome covers a maximum of 67.8 million CpGs when the top and bottom strands are counted independently. EM-seq covered over 56 million CpG sites for 200–1 ng inputs and roughly 45 million CpG sites for 0.1 ng input libraries.

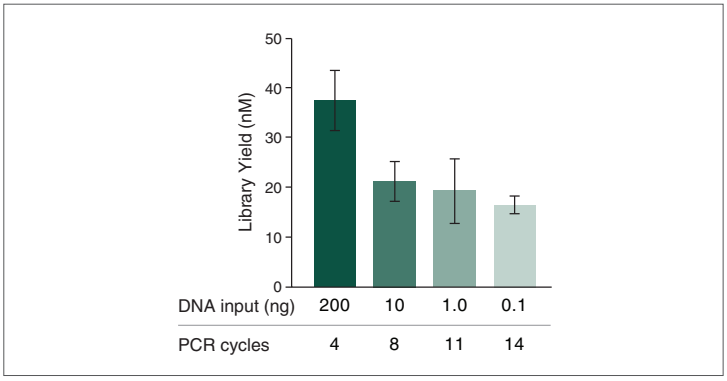
FIGURE 4: **NEBNext EM-seq v2 identifies more CpGs than WGBS and the original EM-seq, at lower sequencing coverage depth**



EM-seq v2 (NEB #E8015), EM-seq (NEB #E7120) and WGBS libraries were prepared from 200 ng and 10 ng of NA12878 DNA (sheared to ~350 bp), spiked with unmethylated lambda and CpG-methylated pUC19. Libraries were sequenced on an Illumina NovaSeq 6000. For accurate comparison of the original EM-seq and WGBS data with EM-seq v2 data, we evaluated data from approximately 625 million 100 base reads for each library aligned to a composite human T2T, lambda and pUC19 reference genome using bwa-meth.

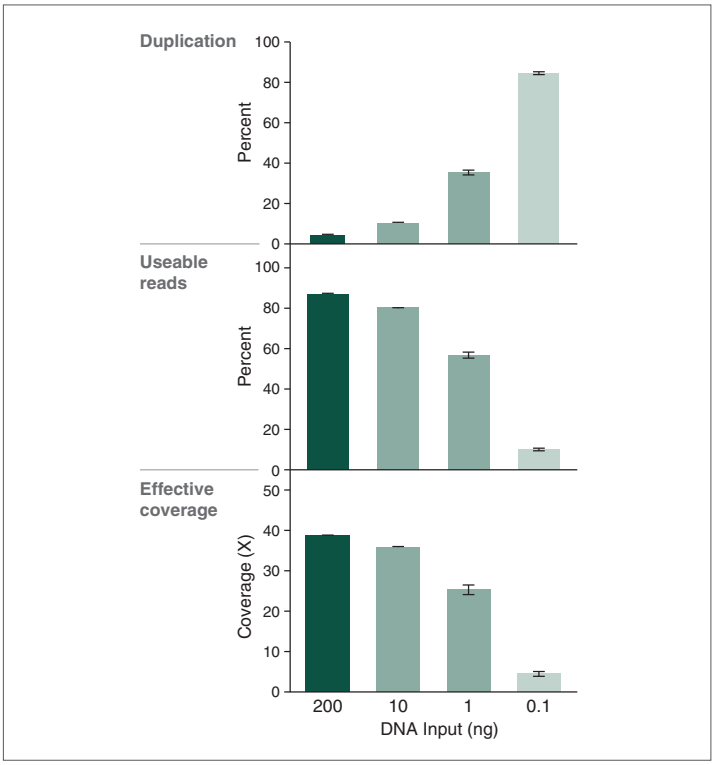
The T2T genome covers a maximum of 67.8 million CpGs when the top and bottom strands are counted independently. EM-seq v2 and EM-seq covered over 54 million CpG sites for both 200 ng and 10 ng inputs; however, WGBS libraries covered only 46 million and 39 million for 200 ng and 10 ng inputs respectively at 1X coverage. The dashed lines represent coverage of (A) 10X and (B) 8X. The table lists the percentage of CpG sites covered by different libraries at (A) 10X and (B) 8X coverage level.

FIGURE 5: **EM-seq v2 produces high library yields across a broad input range**



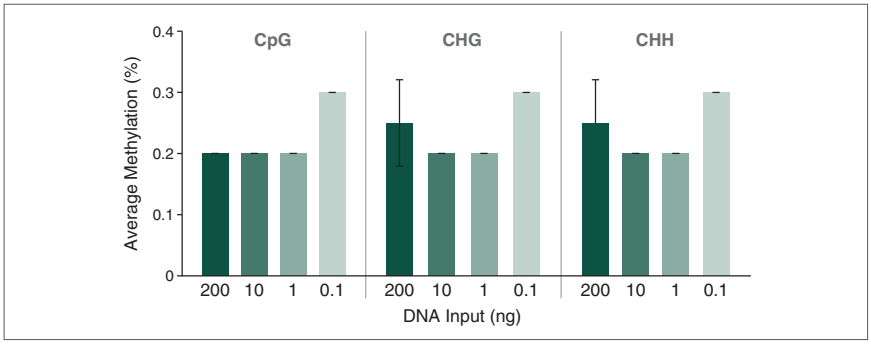
200–0.1 ng of NA12878 genomic DNA, sheared to 350 bp (Covaris® ME220) was used as input into the EM-seq v2 protocol, using the number of PCR cycles shown. Library yields were determined using the Agilent® TapeStation® with High Sensitivity D1000 reagents. Values shown are the average of two technical replicates and error bars show standard deviation. EM-seq v2 consistently produces high-yield libraries across a wide range of inputs.

FIGURE 6: **EM-seq v2 produces high quality libraries**



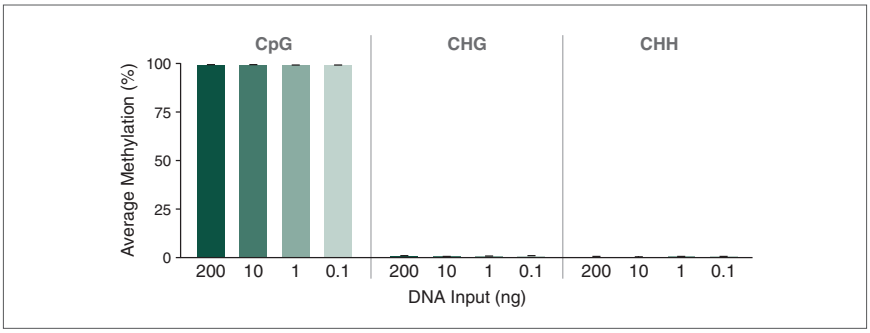
200–0.1 ng of NA12878 genomic DNA, sheared to 350 bp (Covaris ME220) was used as input into the EM-seq™ v2 protocol. Libraries were sequenced on an Illumina NovaSeq 6000 (2 x 150 bases). Sequencing metrics were calculated using approximately 910 million 150-base Illumina reads. Mapping percentage for all libraries was > 99.95%, using bwa-meth to align to a composite human T2T, lambda and pUC19 genome. Duplication: reads marked as duplicate by Picard MarkDuplicates. Usable Reads: the set of Proper-pair, MapQ > 10, primary, nonduplicate reads used in methylation calling (SAMtools view -F 0xF00 -q 10). Effective Coverage: % Usable × theoretical coverage. Theoretical coverage is calculated using the number of bases sequenced/total bases in the T2T reference.

FIGURE 7: **EM-seq v2 libraries have a high deamination efficiency**



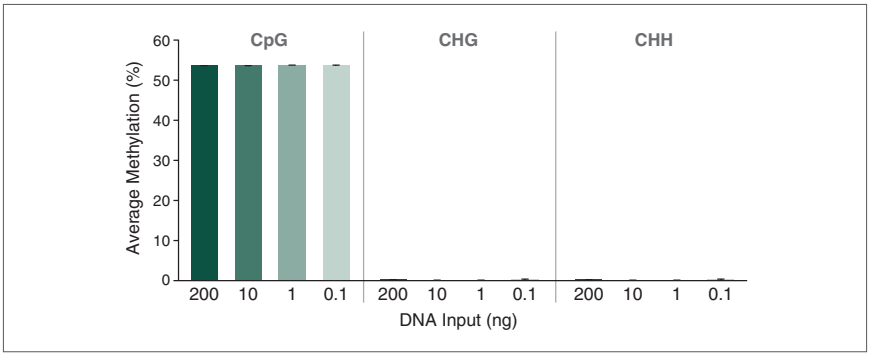
Control unmethylated lambda DNA was spiked in when preparing EM-seq v2 libraries from 200–0.1 ng of NA12878 DNA, sheared to 350 bp (Covaris ME220). Libraries were sequenced on an Illumina NovaSeq 6000 (2 x 150 bases). Approximately 910 million reads for each library were aligned to a composite human T2T, lambda and pUC19 genome using bwa-meth, and methylation information was extracted from the alignments using MethylDackel. Percent 5mC detected in each library is ≤ 0.5%, indicating a deamination efficiency of ≥ 99.5%. Values shown are the average of two technical replicates and error bars show standard deviation.

FIGURE 8: **EM-seq v2 libraries have a high protection rate**



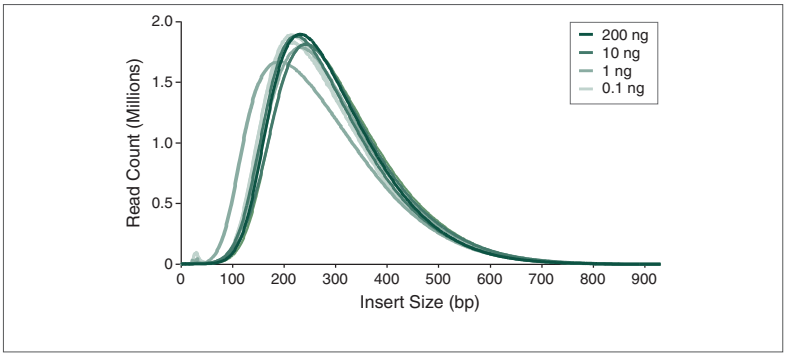
Control CpG-methylated pUC19 DNA was spiked in when preparing EM-seq v2 libraries from 200–0.1 ng of NA12878 DNA, sheared to 350 bp (Covaris ME220). Libraries were sequenced on an Illumina NovaSeq 6000 (2 x 150 bases). Approximately 910 million reads for each library were aligned to a composite human T2T, lambda and pUC19 genome using bwa-meth, and methylation information was extracted from the alignments using MethylDackel. Percent 5mC detected in CpG context for each library is ≥ 99% and < 1% for CHG and CHH contexts. Values shown are the average of two technical replicates and error bars show standard deviation.

FIGURE 9: **Methylation detected by EM-seq v2 in NA12878 genomic DNA is consistent across inputs**



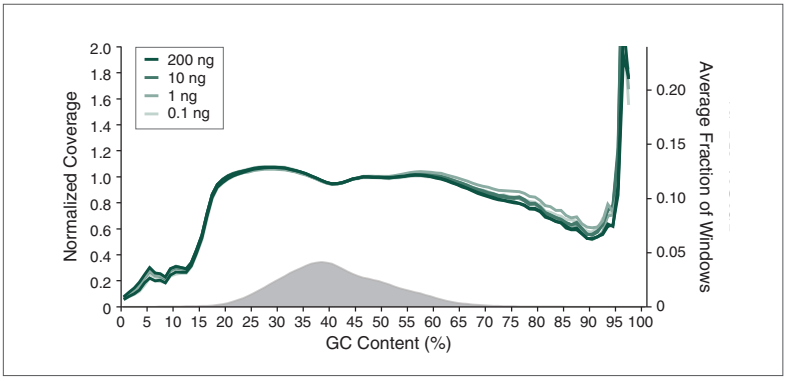
Unmethylated lambda and CpG-methylated pUC1200 were spiked in when preparing EM-seq v2 libraries from 200–0.1 ng of NA12878 DNA, sheared to 350 bp (Covaris ME220). Libraries were sequenced on an Illumina NovaSeq 6000 (2 x 150 bases). Approximately 910 million reads for each library were aligned to a composite human T2T, lambda and pUC19 genome using bwa-meth, and methylation information was extracted from the alignments using MethylDackel. Percent 5mC detected in CpG context for each library is ~ 53.5% and < 0.5% for CHG and CHH contexts. Values shown are the average of two technical replicates and error bars show standard deviation.

FIGURE 10: **EM-seq v2 libraries have a uniform insert size distribution**



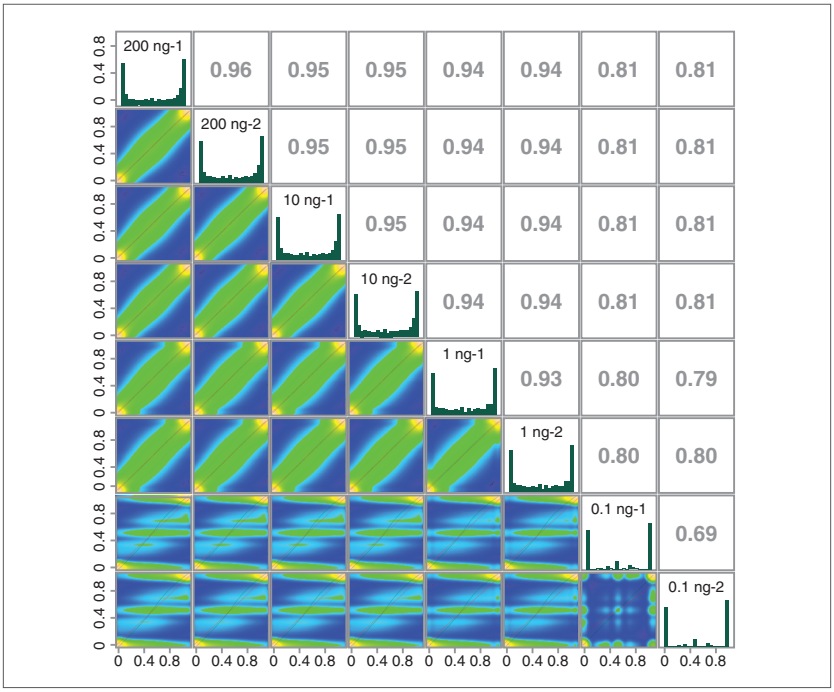
EM-seq v2 libraries were prepared from 200–0.1 ng of NA12878 DNA (sheared to 350 bp using Covaris ME220), spiked with unmethylated lambda and CpG-methylated pUC19. Libraries were sequenced on an Illumina NovaSeq 6000 (2 x 150 bases). Approximately 910 million reads for each library were aligned to a composite human T2T, lambda and pUC19 reference genome using bwa-meth. Library insert sizes were determined using Picard and the read count of each insert size was plotted. Duplicate reads were retained and used for insert size assessment. EM-seq v2 libraries have consistent insert sizes regardless of the input DNA amount.

FIGURE 11: **EM-seq v2 provides even GC coverage**



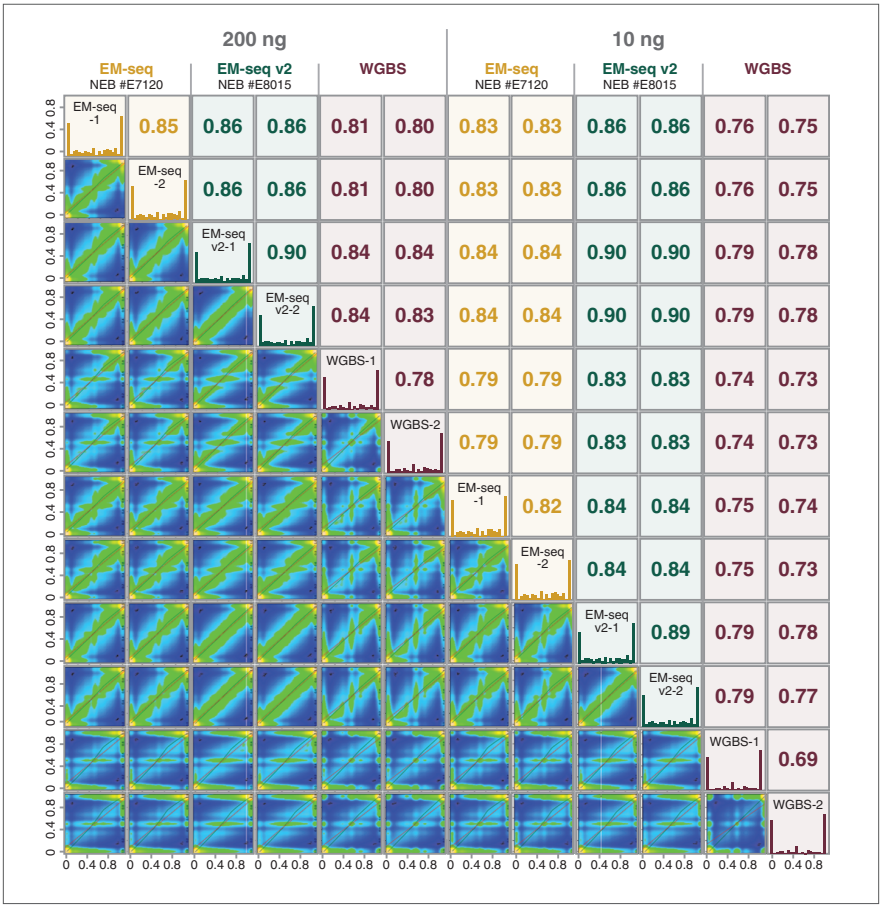
EM-seq v2 libraries were prepared from 200–0.1 ng of NA12878 DNA (sheared to 350 bp using Covaris ME220), spiked with unmethylated lambda and CpG-methylated pUC19. Libraries were sequenced on an Illumina NovaSeq 6000 (2 x 150 bases). Approximately 910 million reads for each library were aligned to a composite human T2T, lambda and pUC19 reference genome using bwa-meth. GC coverage was analyzed using Picard and the distribution of normalized coverage across different GC contents of the genome (0–100%) was plotted for reads mapping to human genome. EM-seq v2 libraries have uniform GC coverage across the input range.

FIGURE 12: **EM-seq v2 libraries are well-correlated across a range of inputs**



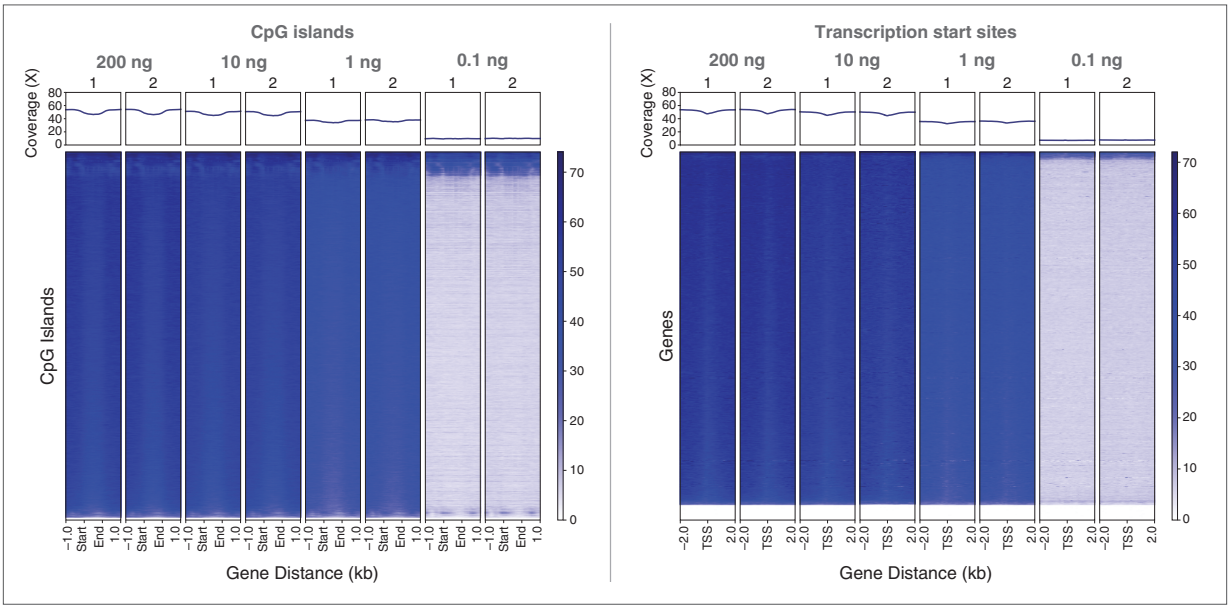
EM-seq v2 libraries were prepared from 200–0.1 ng of NA12878 DNA (sheared to 350 bp using Covaris ME220), spiked with unmethylated lambda and CpG-methylated pUC19. Libraries were sequenced on an Illumina NovaSeq 6000 (2 x 150 bases). Approximately 910 million reads for each library were aligned to a composite human T2T, lambda and pUC19 reference genome using bwa-meth, and methylation information was extracted from the alignments using MethylDackel. Pearson correlations for 200 ng, 10 ng and 1 ng input libraries were ≥ 0.93 between replicates and within inputs. Correlations for 0.1 ng were lower at 0.69. 38 million CpG sites are included in this correlation with a minimum 1X coverage cut off.

FIGURE 13: **Correlation of EM-seq v2, EM-seq and WGBS libraries**



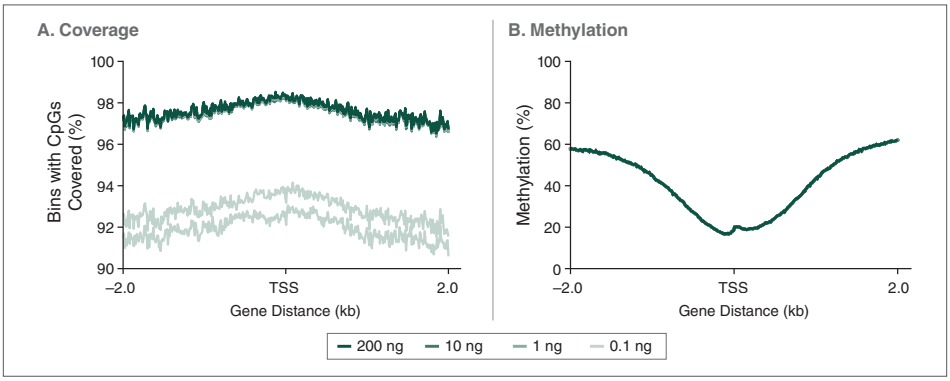
EM-seq v2 (NEB #E8015), EM-seq (NEB #E7120) and WGBS libraries were prepared from 200 ng and 10 ng of NA12878 DNA (sheared to ~350 bp), spiked with unmethylated lambda and CpG-methylated pUC19. Libraries were sequenced on an Illumina NovaSeq 6000. Approximately 625 million 100 base reads for each library were aligned to a composite human T2T, lambda and pUC19 reference genome using bwa-meth, and methylation information was extracted from the alignments using MethylDackel. Pearson correlations for 200 ng libraries for EM-seq v2 and EM-seq were > 0.8 compared to 0.78 for WGBS 200 ng libraries. In the case of 10 ng libraries pearson correlations for EM-seq v2 and EM-seq were also > 0.8 compared to 0.69 for WGBS 10 ng libraries. 24 million CpG sites are included in this correlation with a minimum 1X coverage cut off.

FIGURE 14: **EM-seq v2 libraries have even coverage across genomic features**



EM-seq v2 libraries were prepared from 200–0.1 ng of NA12878 DNA (sheared to 350 bp using Covaris ME220), spiked with unmethylated lambda and CpG-methylated pUC19. Approximately 910 million reads for each library were aligned to a composite human T2T, lambda and pUC19 reference genome using bwa-meth. Heatmaps were generated using deepTools, showing coverage across 1 kb around start and end of CpG islands and coverage across 2 kb windows around transcription start sites (TSS). EM-seq v2 libraries have even coverage across genomic features regardless of DNA input amount.

FIGURE 15: **EM-seq v2 libraries have consistent methylation across TSS**

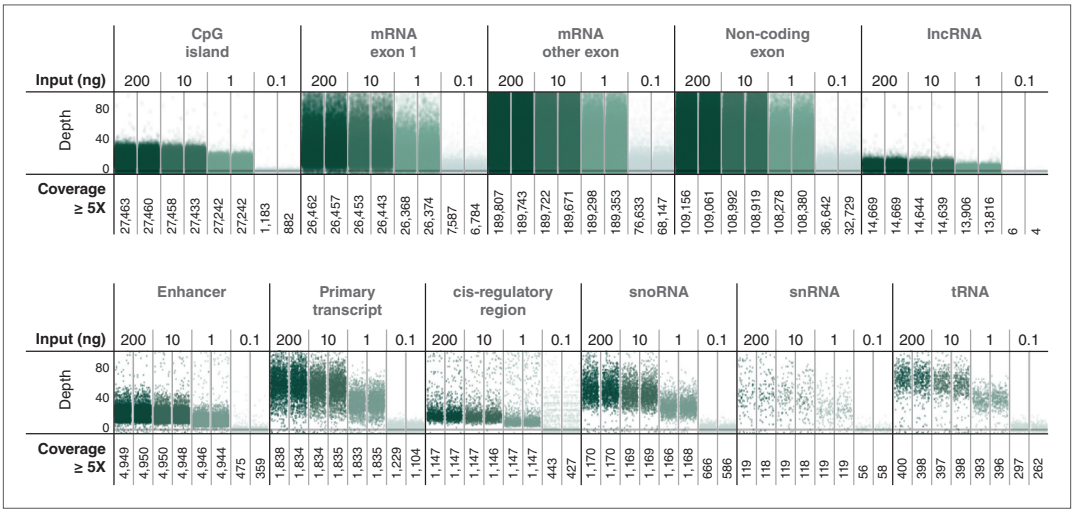


EM-seq v2 libraries were prepared from 200–0.1 ng of NA12878 DNA (sheared to 350 bp using Covaris ME220), spiked with unmethylated lambda and CpG-methylated pUC19. Libraries were sequenced on an Illumina NovaSeq 6000 (2 x 150 bases). Approximately 910 million reads for each library were aligned to a composite human T2T, lambda and pUC19 reference genome using bwa-meth, and methylation information was extracted from the alignments using MethylDackel.

A. Plot of the percentage of CpG containing bins (bin size: 10 bp) with coverage. The percentage of bins covered across 2 kb around TSS showed low variability.

B. Plot of the level of methylation observed for the same bins.

FIGURE 16: **EM-seq v2 libraries provide consistent coverage of diverse genomic feature types across inputs**



EM-seq v2 libraries were prepared from 200–0.1 ng of NA12878 DNA (sheared to 350 bp using Covaris ME220), spiked with unmethylated lambda and CpG-methylated pUC19. Libraries were sequenced on an Illumina NovaSeq 6000 (2 x 150 bases). Approximately 910 million reads for each library were aligned to a composite human T2T, lambda and pUC19 reference genome using bwa-meth. The number of features with coverage greater than 5X is indicated below each plot. Feature coverage at ≥ 5X is maintained to 1 ng and drops with 0.1 ng input. Coverage of genomic feature types are represented with one point per region with the vertical position representing the average coverage of the feature. Points are staggered horizontally to avoid excess overlapping. Feature annotations are from NCBI's RefSeq browser. CpG islands were defined based on the UCSC genome browser.

ISO 9001 Registered Quality Management | ISO 14001 Registered Environmental Management | ISO 13485 Registered Medical Devices | Certified B Corporation

www.neb.com

NEW ENGLAND Biolabs®