Supplemental Material

Enzymatic methyl sequencing detects DNA methylation at single base resolution from picograms of DNA

Romualdas Vaisvila[*], V. K. Chaithanya Ponnaluri[*], Zhiyi Sun[*], Bradley W. Langhorst, Lana Saleh, Shengxi Guan, Nan Dai, Matthew A. Campbell, Brittany S. Sexton, Katherine Marks, Mala Samaranayake, James C. Samuelson, Heidi E. Church, Esta Tamanaha, Ivan R. Corrêa Jr., Sriharsa Pradhan, Eileen T. Dimalanta, Thomas C. Evans, Jr., Louise Williams[†] and Theodore B. Davis

* These authors contributed equally to this work

[†] Correspondence should be addressed to lwilliams@neb.com

Table of Contents:

Supplemental Methods	pages 2-10
Supplemental Figures	pages 11-64
Supplemental Tables	pages 65-80

Supplemental Methods

DNA substrates

The following DNA was gifted or commercially sourced. NA12878 genomic DNA (Corriell Cell Repositories), unmethylated lambda DNA (Promega), *Arabidopsis thaliana* genomic DNA (Biochain), Mouse NIH/3T3 and human Jurkat DNA (NEB), XP12 phage DNA (Dr. Peter Weigele, NEB) and human adult normal lung FFPE (Biochain). Oligonucleotides for all assays were synthesized by IDT.

TET2 kinetic, end time point and time course reactions

Preparation of genomic DNA substrate for TET2 kinetics experiments

6 μg of genomic DNA was made up to 100 μl using 10 mM Tris, 1 mM EDTA, pH 8.0 buffer and sheared to 1500 bp using the Covaris S2 instrument. The sample was sheared using settings of 5% duty cycle, intensity 3 and 200 cycles/burst for 40 s. Sheared DNA was purified using either the Monarch PCR and DNA Cleanup kit (NEB) or QIAquick Nucleotide Removal Kit (Qiagen). The purified DNA was eluted in H_2O and used in the TET2 reactions.

TET2 end timepoint reaction

A master mix containing 50 mM Tris pH 8.0, 2 mM ATP, 1 mM DTT, 5 mM sodium ascorbate, 5 mM alpha-ketoglutarate (αKG), varying concentrations of TET2 and 2 ng/μl genomic DNA was prepared in a final volume of 50 μl. The reaction was initiated with the addition of 50 μM $FeSO_4$ and incubated at 37 °C for 60 min. The DNA sample was then treated with 0.8 Units of Proteinase K (NEB) for 60 min at 50 °C and subsequently purified using Oligo Clean & Concentrator kit (Zymo Research) and analyzed for oxidized 5mC modifications using liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS) according to the procedure and instruments described previously (Tamanaha et al. 2016).

Activity of TET2 and T4-BGT on 5mC analyzed using mass spectrometry

3 µg of XP12 genomic DNA was made up to 130 µl with 10 mM Tris, 1 mM EDTA, pH 8.0 buffer prior to shearing to 1500 bp using the Covaris S2 instrument set to 5% duty cycle, intensity 3 and 200 cycles/burst for 40 s. Sheared DNA was purified using SPRISelect beads (Beckman Coulter), and 1 µg used in the TET2/T4-BGT reactions. Briefly, 1 µg of genomic DNA was incubated in TET2 1X buffer (50 mM Tris pH 8.0, 2 mM ATP, 1 mM DTT, 5 mM sodium ascorbate, 5 mM αKG, 50 µM $FeSO_4$) with 16 µg of TET2, 40 µM UDP-glu (NEB), 10 U of T4-BGT (NEB) for up to 60 min at 37 °C. 0.8 Units of Proteinase K (NEB) was added for 30 min at 37 °C, and the DNA purified using Monarch PCR and DNA Cleanup kit (NEB). The nucleotide content of the DNA was then analyzed using LC-MS/MS (Tamanaha et al. 2016).

APOBEC3A substrate specificity and site preference analysis

The oligonucleotides used for the quantitative analysis of APOBEC3A site preference are shown in Supplemental Tables 1 and 2. The single stranded oligos shared a core sequence with either a single cytosine or a methylated cytosine preceded by either a thymine, uracil, guanine, adenine, or 5mC/cytosine (Supplemental Table 1). Each oligonucleotide was assayed as follows. 2 µM ssDNA oligonucleotide and 0.2 µM APOBEC3A in 1× reaction buffer (50 mM Bis-Tris, pH 6.0, 0.1% Triton X-100) were incubated at 37° C. Incubation times varied according to the oligonucleotide being assayed. For the site preference analysis, timepoints of 0, 1, 2, 4, 8 and 20 h were used. For substrate specificity assays, C and 5mC incubation times were 1, 2, 4, 8, 16, 32, 64, 128, and 256 min. For the oligonucleotides containing 5hmC, 5fC or 5caC, time points of 8, 16, 32, 64, 128, 256 min, 5 and 22 h were used. Reactions were quenched with 8 volumes of ethanol and the DNA was recovered by using the DNA Clean and Concentrator kit (Zymo Research). DNA samples were digested to nucleosides using a Nucleoside Digestion Mix (NEB). Global nucleoside content analysis was performed by LC-MS or LC-MS/MS as noted (see conditions below). For the site preference and substrate specificity assays, the

data points were best fitted by a single exponential equation to follow the disappearance of C, 5mC, 5hmC, 5fC, or 5caC nucleotides and the appearance of U or T (KaleidaGraph, Synergy Software).

Determining APOBEC3A activity over time

The 2 μM oligonucleotide substrates contained the same core sequence, but varied at one single position with either C, 5mC, 5hmC, 5gmC, 5fC or 5caC (Supplemental Table 2), were incubated with 25 nM of APOBEC3A for up to 23 h at 37 °C in 50 mM Bis-Tris pH 6.0, 0.1% Triton X-100. Samples were withdrawn at 1 h intervals and challenged by adding 2 μM cytosine only oligonucleotide substrate (Supplemental Table 2) for 15 min at 37 °C. Deamination was terminated by sample purification using the Oligo Clean & Concentrator Kit (Zymo Research). Deamination rates of C>U and 5mC oxidation derivatives were quantified with LC-MS (Sun et al. 2021).

Global nucleoside analysis using LC-MS and LC-MS/MS

LC-MS analysis was performed on an Agilent LC-MS System 1200 Series equipped with a G1315D diode array detector and a 6120 Single Quadrupole Mass Detector operating in positive (+ESI) and negative (-ESI) electrospray ionization modes. LC was carried out on a Waters Atlantis T3 column (4.6 × 150 mm, 3 μm) with a gradient mobile phase consisting of 10 mM aqueous ammonium acetate (pH 4.5) and methanol. The relative abundance of each nucleoside was determined by UV absorbance.

LC-MS/MS analysis was performed in duplicate by injecting digested DNA on an Agilent 1290 UHPLC equipped with a G4212A diode array detector and a 6490A Triple Quadrupole Mass Detector operating in the positive electrospray ionization mode (+ESI). UHPLC was carried out on a Waters XSelect HSS T3 XP column (2.1 × 100 mm, 2.5 μm) with the gradient mobile phase consisting of 10 mM aqueous ammonium formate (pH 4.4) and methanol. MS data acquisition was performed in the dynamic multiple reaction monitoring (DMRM) mode. Each nucleoside was identified in the extracted chromatogram

associated with its specific MS/MS transition: dC $[M+H]^+$ at m/z 228 →112, 5mC $[M+H]^+$ at m/z 242→126, 5hmC $[M+H]^+$ at m/z 258→142, 5fC $[M+H]^+$ at m/z 256→140 and 5caC $[M+H]^+$ at m/z 294→178. External calibration curves with known amounts of the nucleosides were used to calculate their ratios within the analyzed samples.

Activity of TET2, T4-BGT and APOBEC3A on NA12878 genomic DNA, analyzed using mass spectrometry

The enzymatic activity of TET2, T4-BGT and APOBEC3A on NA12878 genomic DNA was analyzed using LC-MS. Enzymes were investigated individually or in combination as follows. For individual TET2 or T4-BGT reactions or those that combined TET2 and T4-BGT in a single reaction followed by no APOBEC3A deamination: 200 ng of NA12878 DNA was made up to 50 µl with 10 mM Tris pH 8.0. The DNA was transferred to a Covaris microTUBE (Covaris) and sheared to 300 bp using the Covaris ME220 instrument. DNA was oxidized in a 50 µl reaction volume containing 50 mM Tris pH 8.0, 1 mM DTT, 5 mM Sodium-L-Ascorbate, 20 mM αKG, 2 mM ATP, 50 mM Ammonium Iron (II) sulfate hexahydrate, 0.04 mM UDP-glu (NEB), 16 µg TET2, 10 U T4-BGT (NEB). The reaction was initiated by adding Fe (II) solution to a final reaction concentration of 40 µM, but only for the TET2 containing reactions, and then incubated for 1 h at 37 °C. 0.8 U of Proteinase K (NEB) was then added and incubated for 30 min at 37 °C. The DNA was purified using 90 µl of resuspended NEBNext Sample Purification Beads according to the manufacturer's instructions and eluted in 18 µl water. 17 µl was transferred to a new tube then prepped for nucleoside digestion. DNA samples were digested to nucleosides using a Nucleoside Digestion Mix (NEB). Global nucleoside content analysis was performed by LC-MS. Reactions that used APOBEC3A either individually or after TET2 and T4-BGT treatment were set up as follows. If there was no upstream TET2 or T4-BGT reactions, 200 ng of NA12878 DNA was made up to 50 µl with 10 mM Tris pH 8.0. The DNA was transferred to a Covaris microTUBE (Covaris) and sheared to 300 bp using the Covaris ME220 instrument. The DNA was purified using 90 µl of resuspended NEBNext Sample

Purification Beads according to the manufacturer's instructions and eluted in 16 µl water. For individual APOBEC3A reactions as well as those that has upstream TET2 and T4-BGT treatments, 4 µl of formamide (Sigma-Aldrich) was added and the reactions were incubated at 85 °C for 10 min. The DNA was then deaminated in 50 mM Bis-Tris pH 6.0, 0.1% Triton X-100, 20 µg BSA (NEB) using 0.2 µg of APOBEC3A (NEB). The reaction was incubated at 37 °C for 3 h and the DNA was purified using 100 µl of resuspended NEBNext Sample Purification Beads according to the manufacturer's protocol. The sample was eluted in 18 µl water and 17 µl was transferred to a new tube then prepped for nucleoside digestion. DNA samples were digested to nucleosides using a Nucleoside Digestion Mix (NEB). Global nucleoside content analysis was performed by LC-MS. DNA for unconverted control reactions were set up as follows. 200 ng of NA12878 DNA was made up to 50 µl with 10 mM Tris pH 8.0. The DNA was transferred to a Covaris microTUBE (Covaris) and sheared to 300 bp using the Covaris ME220 instrument. The DNA was purified using 90 µl of resuspended NEBNext Sample Purification Beads according to the manufacturer's instructions and eluted in 18 µl water and 17 µl was transferred to a new tube then prepped for nucleoside digestion. DNA samples were digested to nucleosides using a Nucleoside Digestion Mix (NEB). Global nucleoside content analysis was performed by LC-MS.

PCR of long amplicons

Sodium bisulfite conversion: 200 ng of NA12878 DNA was converted using the EZ DNA Methylation-Gold Kit (Zymo Research) according to the manufacturer's instructions. Enzymatic 5mC/5hmC Conversion: 200 ng of NA12878 DNA was incubated with 16 µg of TET2 enzyme and 10 U of T4-BGT (NEB) in 40 µM UDP-glu, 50 mM Tris pH 8.0, 2 mM ATP, 1 mM DTT, 5 mM sodium ascorbate, 5 mM αKG and 50 µM Fe(II) for 30 min at 37 °C, in a final volume of 50 µl. Then, 0.8 U of Proteinase K (NEB) was added and incubated for 30 min at 37 °C. DNA was purified using a Genomic DNA Clean & Concentrator (Zymo Research). The DNA was then denatured by adding formamide to 20% (v/v) at and incubating for

10 min at 90 °C. DNA was deaminated using 0.26 µg of APOBEC3A in a 100 µl reaction volume for 3 h. 3 µl from 100 µl of enzymatically converted and 1 µl from 15 µl of bisulfite-converted DNA were PCR amplified with Q5U polymerase (NEB) and primers (Supplemental Table 4) using the following cycling conditions: 98 °C for 30 s, then 35 cycles of 98 °C for 10 s, 63 °C for 20 s, and 68 °C for 2 min. The resulting PCR products were 543, 1181, 1750, 2207 and 2945 bp. PCR amplicons were visualized on a 2% agarose gel run in 1× TBE buffer and stained with ethidium bromide.

Generation and analysis of XP12 data

EM-seq libraries were made using 50 ng NA12878 spiked with 1 ng unmethylated lambda (Promega), 1 ng XP12 and 0.05 ng CpG methylated pUC19 as described for the 10 ng - 200 ng EM-seq protocol described in the Methods section. Illumina data was analyzed as follows. The XP12 genome was divided into strand-specific 50 bp windows at 25 bp step. For each window, the number of cytosines and the average 5mC quantification by EM-seq were calculated. The EM-seq methylation quantification was plotted (Y-axis) against the cytosine density (X axis, which is also the methylation density since all the cytosines are methylated in XP12) of the XP12 genome using a boxplot in R.

Illumina data analysis

Illumina data were used to determine unique features assembled from the Eukaryotic Promoter Database (promoters), UCSC's Table Browser (CpG Islands), Dfam (repetitive elements) (Hubley et al. 2016), and NCBI's RefSeq annotation (remainder). Coverage of these features was assessed using featureCounts. Coverage of Transcription start sites (GENCODE v29 (Frankish et al. 2019)) was calculated using deepTools (Ramirez et al. 2016) after combining both technical replicates and sampling 1.3B reads for each input amount and conversion method.

bigWig coverage tabulation:

```
find . -name '*combined_reps.bam' | parallel bamCoverage -b {} -o {.}_3840_mq30.bw \
  --samFlagExclude 3840 --minMappingQuality 30
```

Coverage matrix construction:

```
computeMatrix reference-point -S *.bw  -R gencode.v29.basic.annotation.gtf \
  -a 2000 -b 2000 -o 2kb_start_codon.matrix --referencePoint TSS  -p max/2
```

Heatmap plotting:

```
plotHeatmap -m 2kb_start_codon.matrix --boxAroundHeatmaps no \
  --colorList azure,lightblue,cornflowerblue,royalblue,blue,mediumblue,darkblue,navy,midnightblue \
  --missingDataColor white --yMax 65 --yMin 0 --colorNumber 80 \
  --samplesLabel "10 ng EM-seq" "10 ng WGBS" "50 ng EM-seq" "50 ng WGBS" "200 ng EM-seq" "200 ng WGBS" \
  -o gencode.v29.basic.annotation_tss_cov_blue.pdf
```

We used a binning approach to calculate the fraction of sites covered and methylation status of sites with at least 8× coverage. Windows of +/- 2 kb around each transcription start site (TSS) were divided into 10 bp bins. We calculated the percent of bins containing at least one CpG with at least 8× coverage in comparison to all bins with at least 8× coverage. For methylation, we calculated the average methylation of CpGs with at least 8× coverage in each 10 bp bin. We assessed the number of "usable reads" using "SAMtools view -F 0xF00 -q 10 -c ${bam}"A nextflow (Tommaso et al. 2017) workflow containing the analysis detail is available in the at https://github.com/nebiolabs/EM-seq/

Due to excess control concentration in the low-input experiment, downsampling was performed to produce equivalent numbers of human reads and replicates were assessed independently before combination for final methylation assessment. Summary fractions are expressed as fractions of human reads (instead of total reads). Otherwise, software methods remained the same as for the standard protocol libraries. Software versions: MethylDackel 0.3.0, bwameth=0.2.0, bwa=0.7.17, bwakit=0.7.15, samblaster=0.1.24, sambamba=0.6.6, picard=2.18.14.

References

Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, Mudge JM, Sisu C, Wright J, Armstrong J et al. 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res* **47**: D766-D773.

Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, Smit AF, Wheeler TJ. 2016. The Dfam database of repetitive DNA families. *Nucleic Acids Res* **44**: D81-89.

Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**: W160-165 doi:10.1093/nar/gkw257.

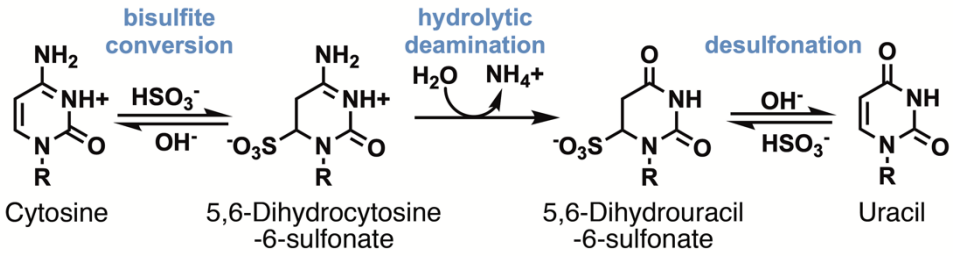
Sun Z, Vaisvila R, Hussong LM, Yan B, Baum C, Saleh L, Samaranayake M, Guan S, Dai N, Correa IR, Jr. et al. 2021. Nondestructive enzymatic deamination enables single-molecule long-read amplicon sequencing for the determination of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Genome Res* doi:10.1101/gr.265306.120.

Tamanaha E, Guan S, Marks K, Saleh L. 2016. Distributive Processing by the Iron(II)/alpha-Ketoglutarate-Dependent Catalytic Domains of the TET Enzymes Is Consistent with Epigenetic Roles for Oxidized 5-Methylcytosine Bases. *J Am Chem Soc* **138**: 9345-9348.

Tommaso PD, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. 2017. Nextflow enables reproducible computational workflows. *Nat Biotechnol* **35**: 316-319.

Supplemental Figures

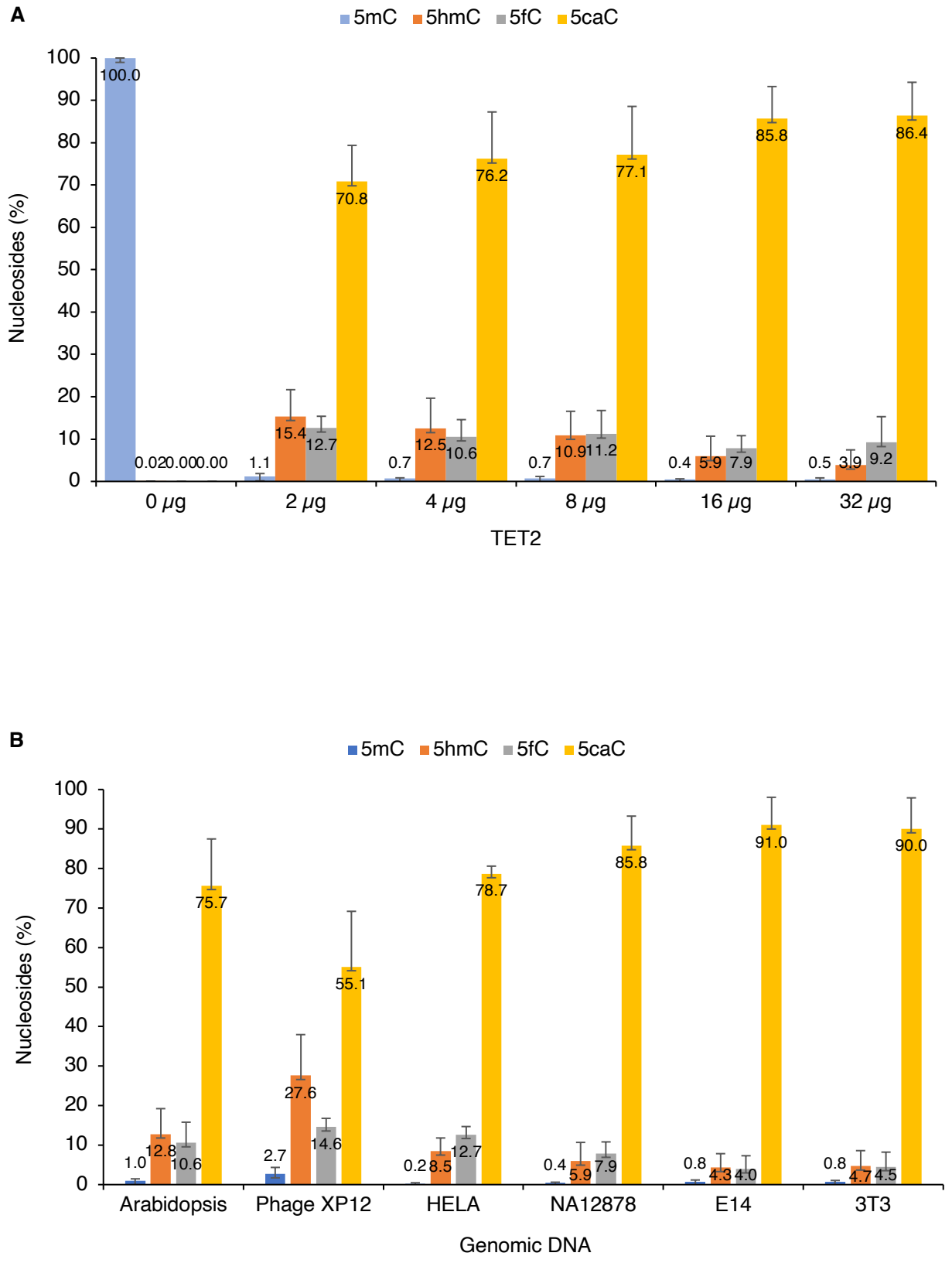
Supplemental Figure 1



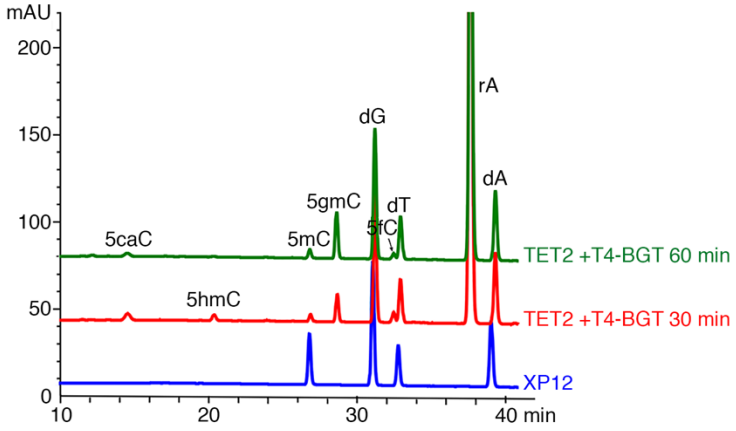
Chemistry of sodium bisulfite conversion

Three chemical steps are needed to deaminate cytosine to uracil. Bisulfite conversion forms the intermediate 5,6-Dihydrocytosine-6-sulfonate which can then undergo hydrolytic deamination to form 5,6-Dihydrouracil-6-sulfonate. Desulfonation during the final step of conversion results in the formation of uracil.

Supplemental Figure 2



C

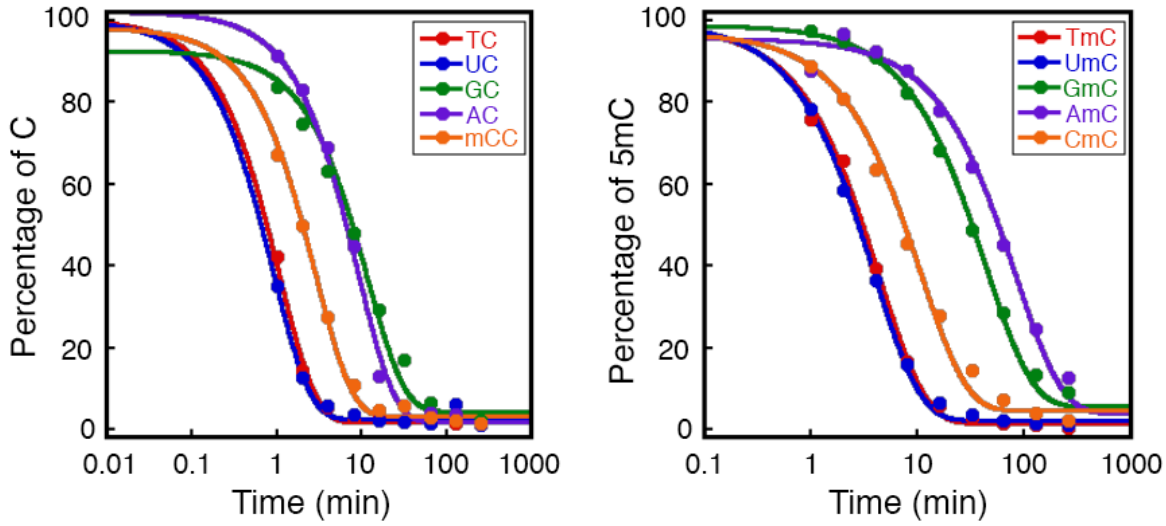


Mass spectrometry analysis of the activity of TET2 and T4-BGT on 5mC and 5hmC

(A) Titration of TET2 (0 – 32 μg) with NA12878 genomic DNA shows that for all TET2 concentrations, ≥ 99% of 5mC (blue) are oxidized to 5hmC (orange), 5fC (gray), and 5caC (yellow) with 5caC constituting approximately 70-85% of total oxidized product. (B) TET2 activity on different genomic DNAs. ≥ 99% of 5mC is converted to its oxidized forms for all genomic DNAs in the presence of 16 μg TET2 except for Phage XP12 where 5mC conversion was >95%. All reactions were performed with 200 ng genomic DNA in a total volume of 50 μl at 37 °C for 60 min. LCMS data used to generate (A) and (B) are in Supplemental File 1. (C) LC-MS spectra showing the activity of TET2 and T4-BGT on XP12 genomic DNA. rA is adenosine and is derived from the ATP added to the TET2 reaction. The blue trace shows the starting XP12 genomic DNA substrate. The red and green traces show the formation of four products, 5hmC, 5fC, 5caC, and 5gmC, by TET2 and T4-BGT at 30 and 60 min, respectively.

Supplemental Figure 3

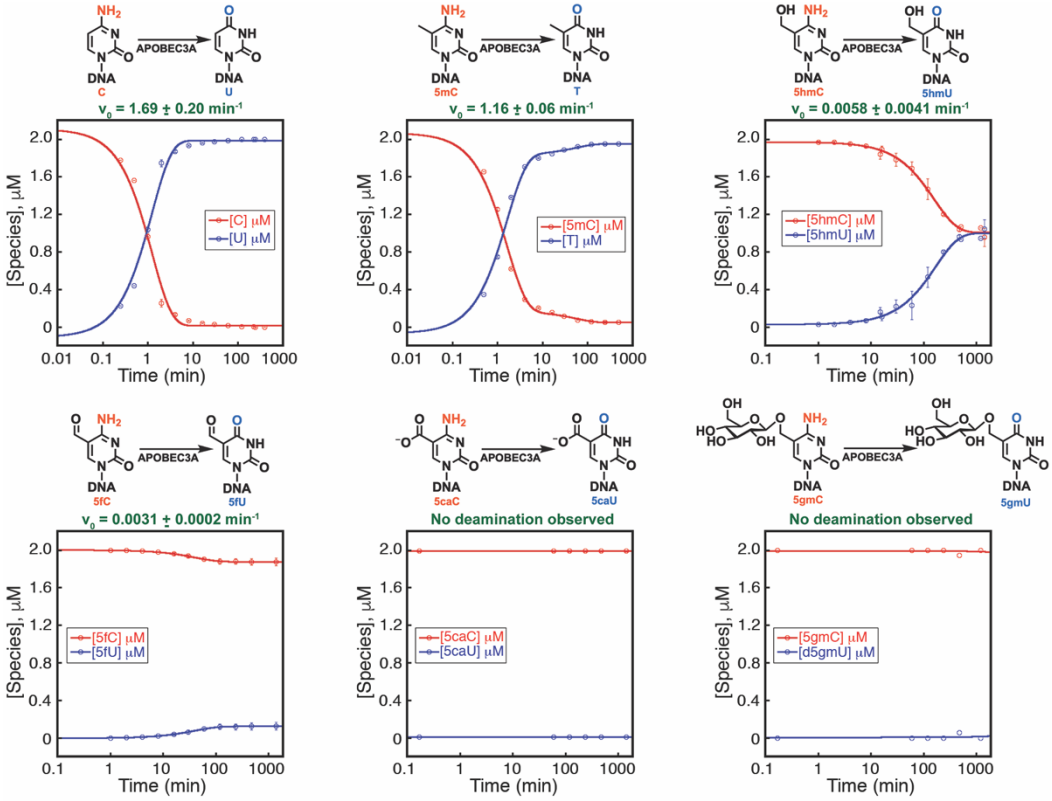
A



B

Substrate	k_{obs} (min^{-1})	Substrate	k_{obs} (min^{-1})
UC	1.1	UmC	0.26
TC	0.94	TmC	0.23
mCC	0.35	CmC	0.095
AC	0.11	GmC	0.023
GC	0.085	AmC	0.012

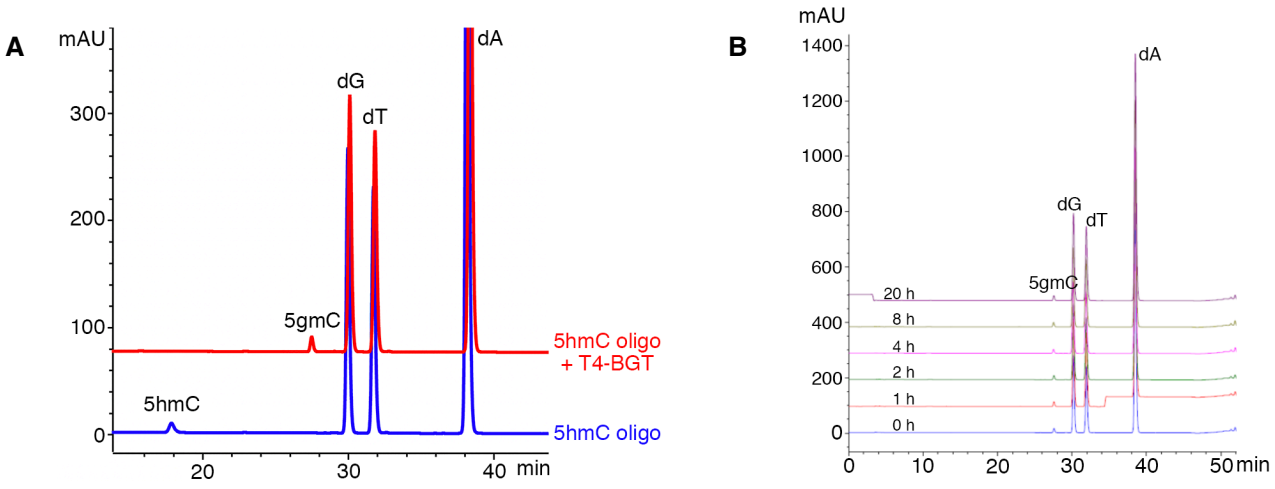
C



APOBEC3A substrate specificity

(A) Time course reactions with 0.2 μM APOBEC3A and 2 μM single-stranded oligonucleotide substrate containing either a cytosine or 5mC preceded by different bases as specified in the legend. Global nucleoside content analysis was performed by LC-MS. Oligonucleotide sequences are in Supplemental Table 1. (B) Observed rate constants (k_{obs}) for C and 5mC disappearance extracted from single exponential fits (solid traces) to the experimental data (closed circles). (C) Time course reactions of 0.2 μM APOBEC3A and 2 μM single-stranded oligonucleotides containing either cytosine, 5mC, 5hmC, 5gmC, 5fC, 5caC (Supplemental Table 2). Synthesis of 5gmC is described in Supplemental Figure 4. k_{obs} and initial rate values for the disappearance of the various C forms are extracted from single exponential fits (solid traces) to the experimental data (open circles). LC-MS was used to determine the amounts of substrate and products within each reaction and the data points underwent best fit analysis using KaleidaGraph (Synergy Software). The LCMS data used to generate (A) and (C) are in Supplemental File 1.

Supplemental Figure 4



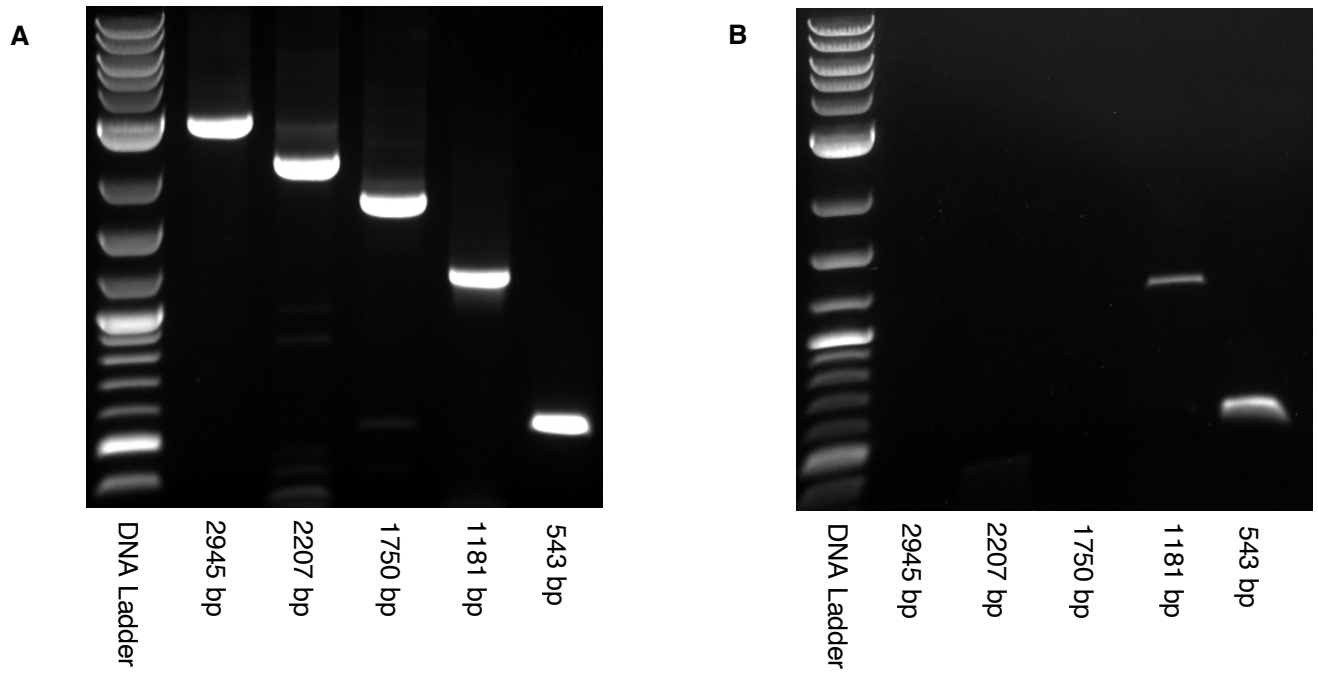
C

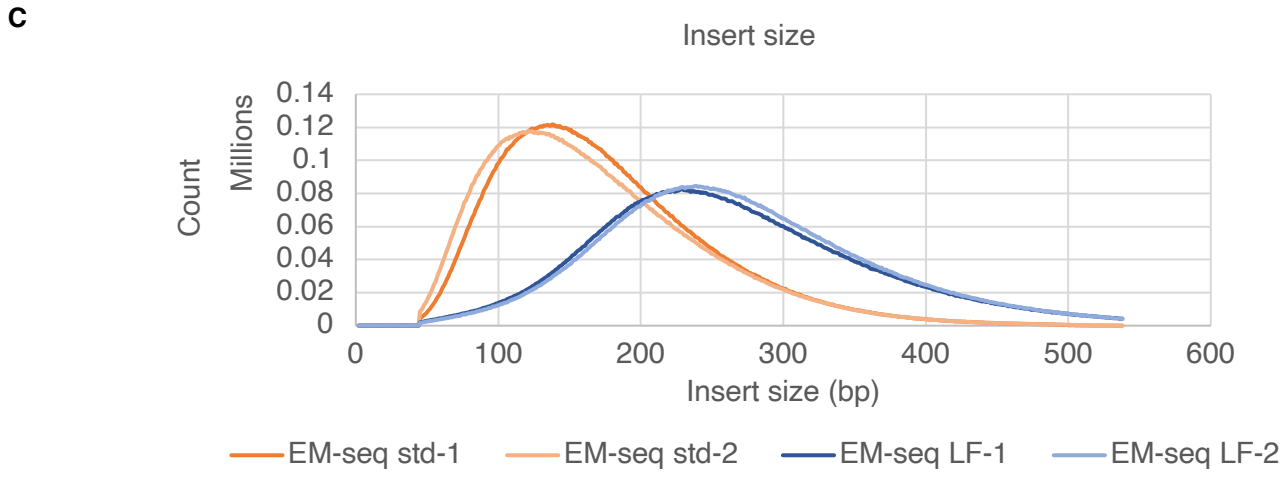
APOBEC3A incubation time (h)	5gmC	G	T	A
0	0.036	0.397	0.507	1.000
1	0.036	0.399	0.510	1.000
2	0.036	0.398	0.509	1.000
4	0.036	0.397	0.508	1.000
8	0.035	0.397	0.507	1.000
20	0.036	0.398	0.509	1.000

Generation of 5-(β-glucosyloxymethyl)cytosine single stranded DNA substrate

(A) An oligonucleotide with an internal 5-hydroxymethylcytosine (ATAAGAATAGAATGAAT/i5HydMe-dC/GTGAAATGAATATGAAATGAATAGTA) was incubated with T4-BGT and the formation of 5gmC demonstrated. (B, C) Time course showing APOBEC3A activity on the glucosylated oligonucleotide. APOBEC3A does not deaminate 5gmC. The extinction coefficient constant for 5gmC is not available, therefore the extinction coefficient for C was used instead. This oligonucleotide was used in Supplemental Figure 3C.

Supplemental Figure 5



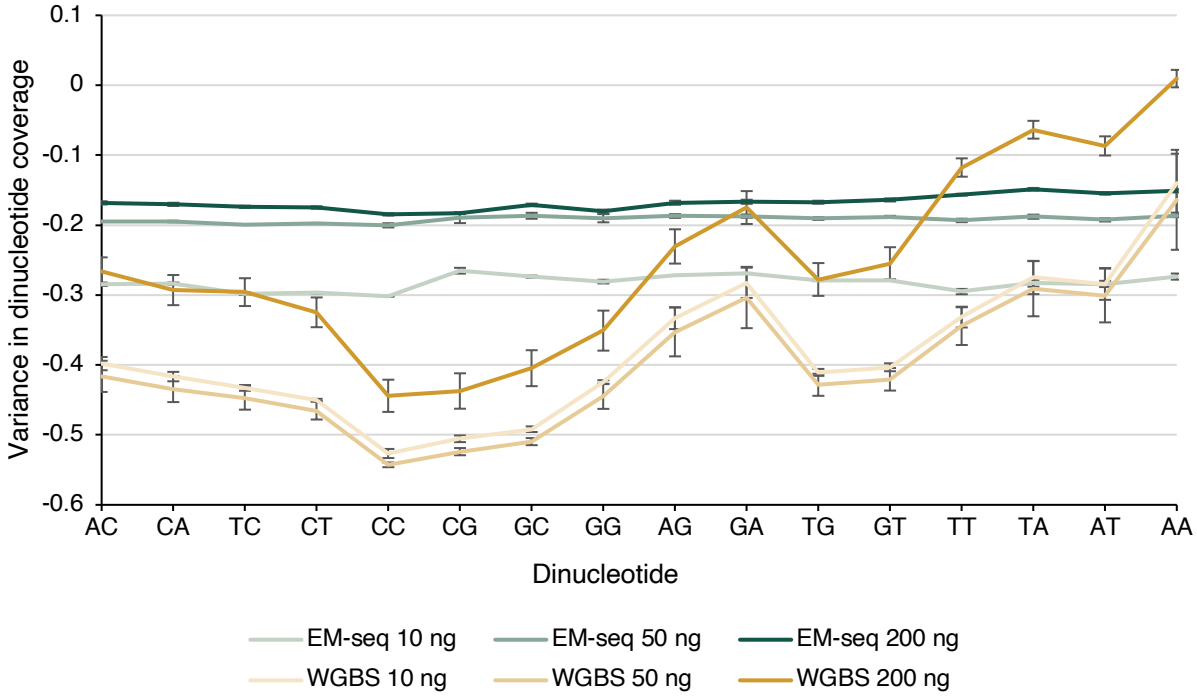


DNA integrity assessment

(A) and (B) NA12878 DNA was sheared to 5 kb prior to enzymatic conversion or bisulfite conversion. End-point PCR was performed using five amplicons ranging in size from 543 bp to 2945 bp. The PCR primers used are shown in Supplemental Table 4. The PCR products were separated using agarose gel

electrophoresis and their images are shown in (A) TET2/T4-BGT followed by APOBEC3A treated samples and (B) bisulfite treated DNA. The left lane represents 1 kb Plus DNA Ladder (NEB). (C) Larger insert Illumina libraries were generated by increasing the shearing size of the DNA to 1 kb and altering the bead clean up after PCR. The larger fragment (LF) libraries have larger inserts than the standard (std) sized EM-seq library inserts. These data indicate that DNA integrity is maintained in the enzymatically treated DNA.

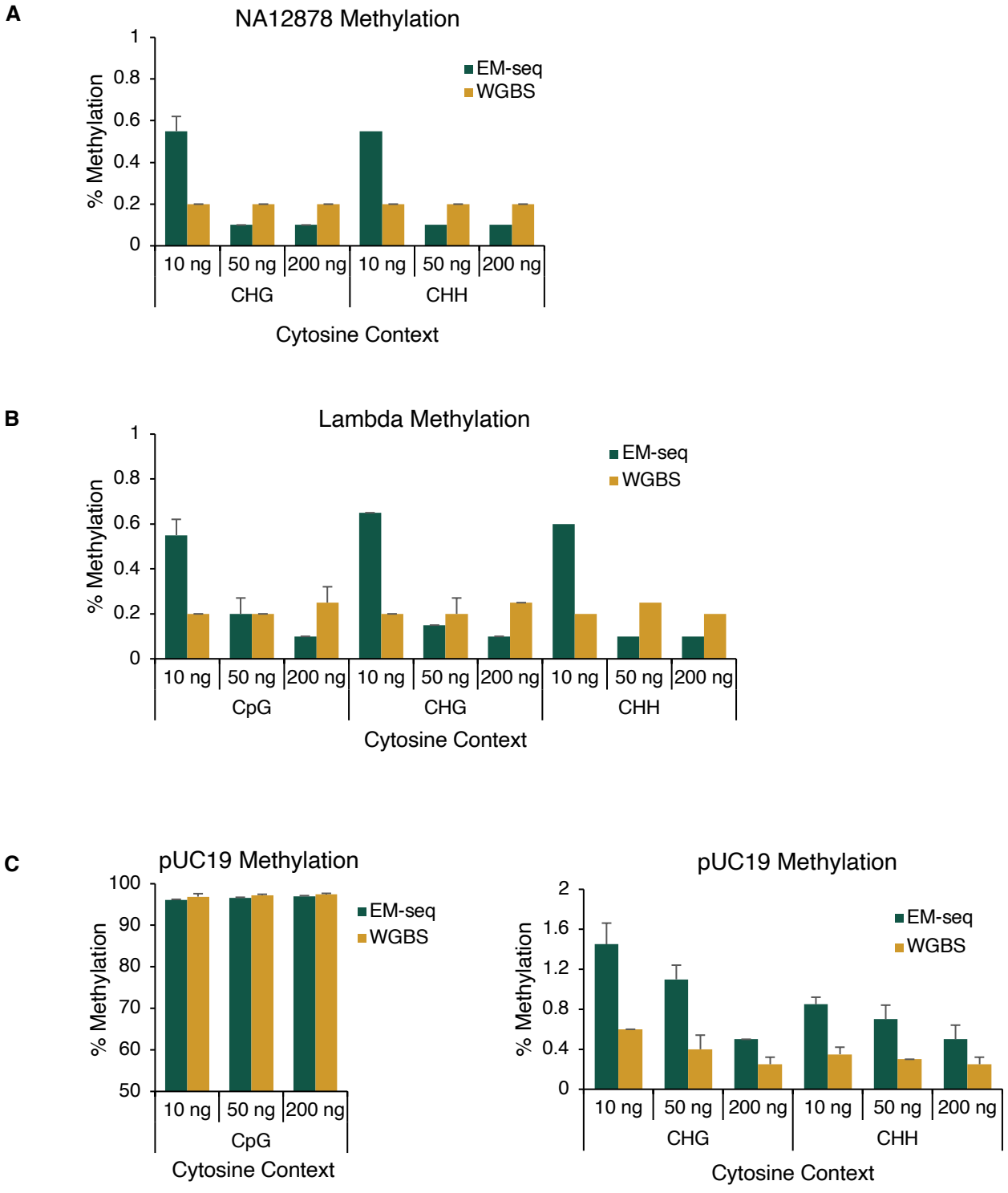
Supplemental Figure 6



Nucleotide analysis of EM-seq and bisulfite libraries.

EM-seq and WGBS libraries were made using 10 ng, 50 ng and 200 ng of NA12878 DNA. Dinucleotide distribution for the data was normalized to coverage observed in an unconverted Ultra II DNA library using the same input DNA as the EM-seq and WGBS libraries. The even distribution of dinucleotides found in EM-seq libraries provides evidence for reduced bias compared to WGBS.

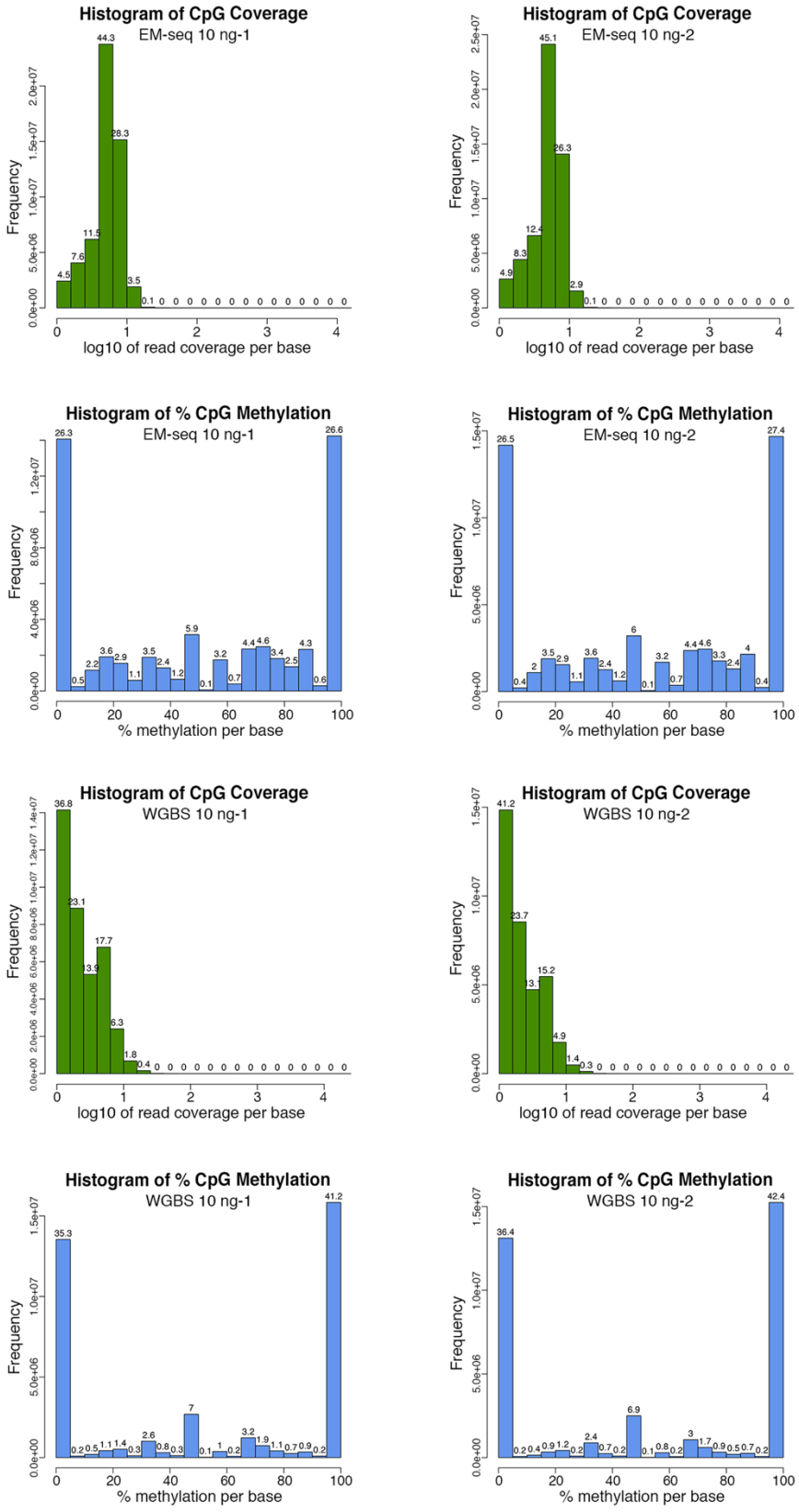
Supplemental Figure 7

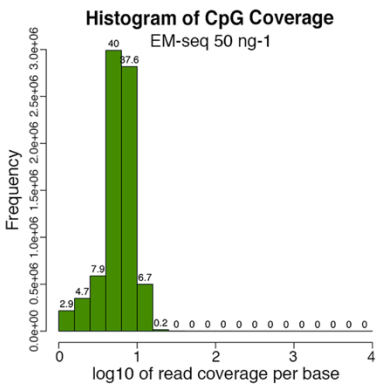


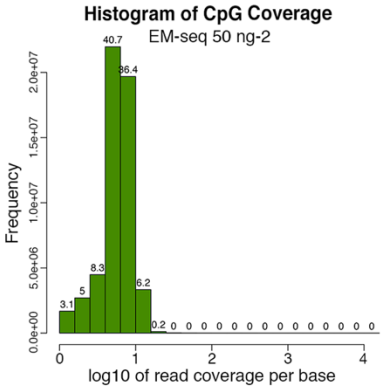
Cytosine methylation in the EM-seq and bisulfite NA12878 library internal controls

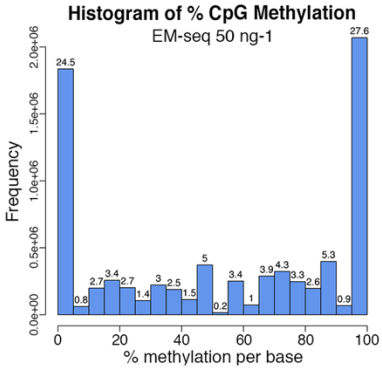
(A) Cytosine methylation of NA12878 DNA in the CHG and CHH contexts. (B) Methylation of cytosines in the unmethylated lambda controls. EM-seq and WGBS have <1% methylation in CpG, CHG and CHH contexts. (C) Methylation of cytosines in the CpG methylated pUC19 control. EM-seq and WGBS libraries have approximately 97% methylation in the CpG context and around 1% in the CHG and CHH contexts.

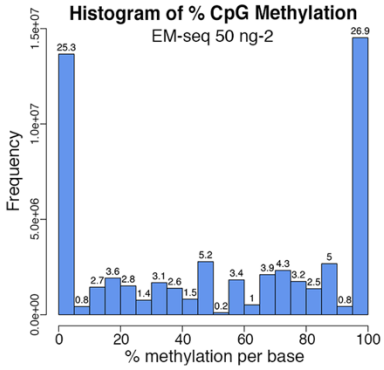
Supplemental Figure 8

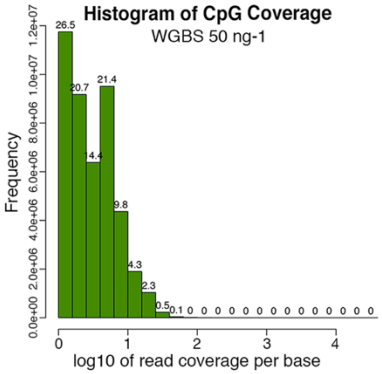


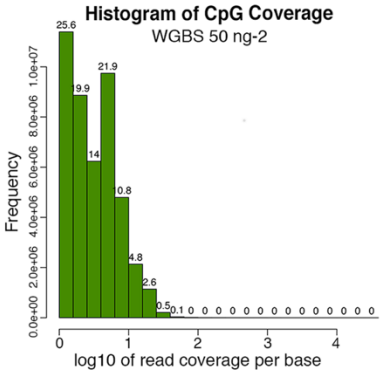

Histogram of CpG Coverage
EM-seq 50 ng-1
Frequency
log10 of read coverage per base

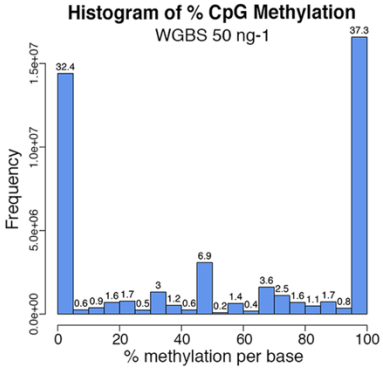

Histogram of CpG Coverage
EM-seq 50 ng-2
Frequency
log10 of read coverage per base

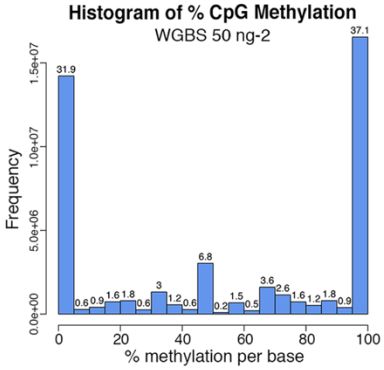

Histogram of % CpG Methylation
EM-seq 50 ng-1
Frequency
% methylation per base

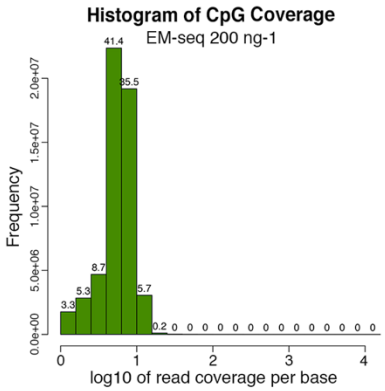

Histogram of % CpG Methylation
EM-seq 50 ng-2
Frequency
% methylation per base

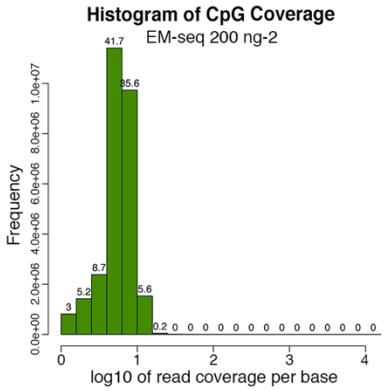

Histogram of CpG Coverage
WGBS 50 ng-1
Frequency
log10 of read coverage per base

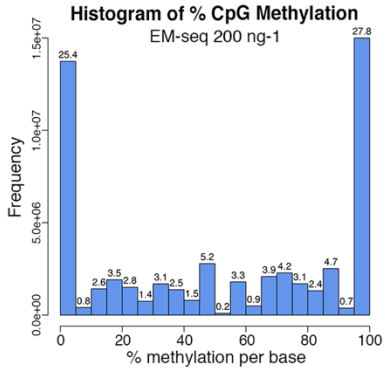

Histogram of CpG Coverage
WGBS 50 ng-2
Frequency
log10 of read coverage per base

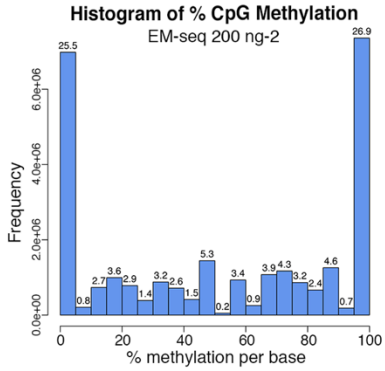

Histogram of % CpG Methylation
WGBS 50 ng-1
Frequency
% methylation per base

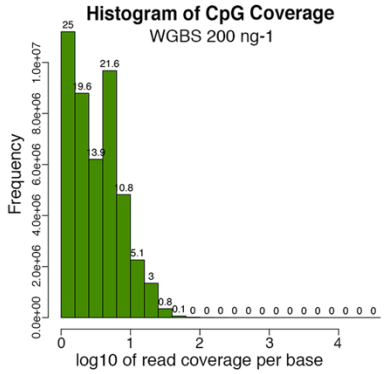

Histogram of % CpG Methylation
WGBS 50 ng-2
Frequency
% methylation per base

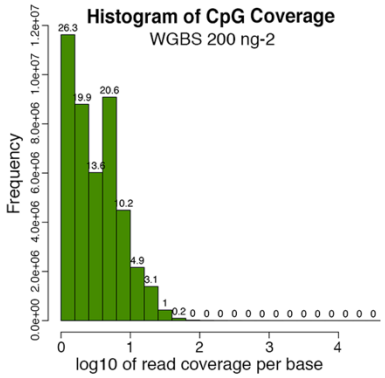

Histogram of CpG Coverage
EM-seq 200 ng-1
Frequency
log10 of read coverage per base

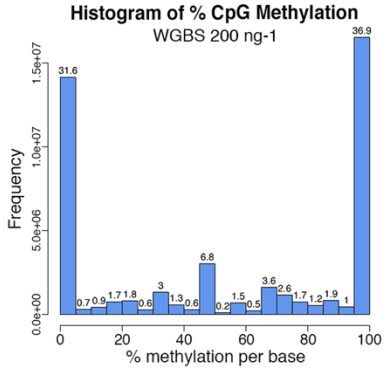

Histogram of CpG Coverage
EM-seq 200 ng-2
Frequency
log10 of read coverage per base

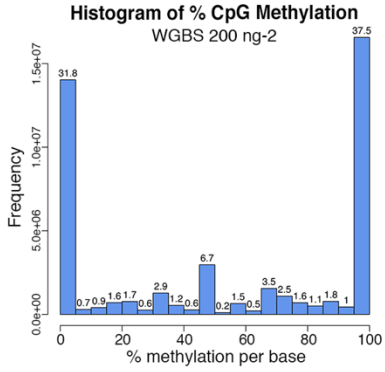

Histogram of % CpG Methylation
EM-seq 200 ng-1
Frequency
% methylation per base


Histogram of % CpG Methylation
EM-seq 200 ng-2
Frequency
% methylation per base


Histogram of CpG Coverage
WGBS 200 ng-1
Frequency
log10 of read coverage per base


Histogram of CpG Coverage
WGBS 200 ng-2
Frequency
log10 of read coverage per base

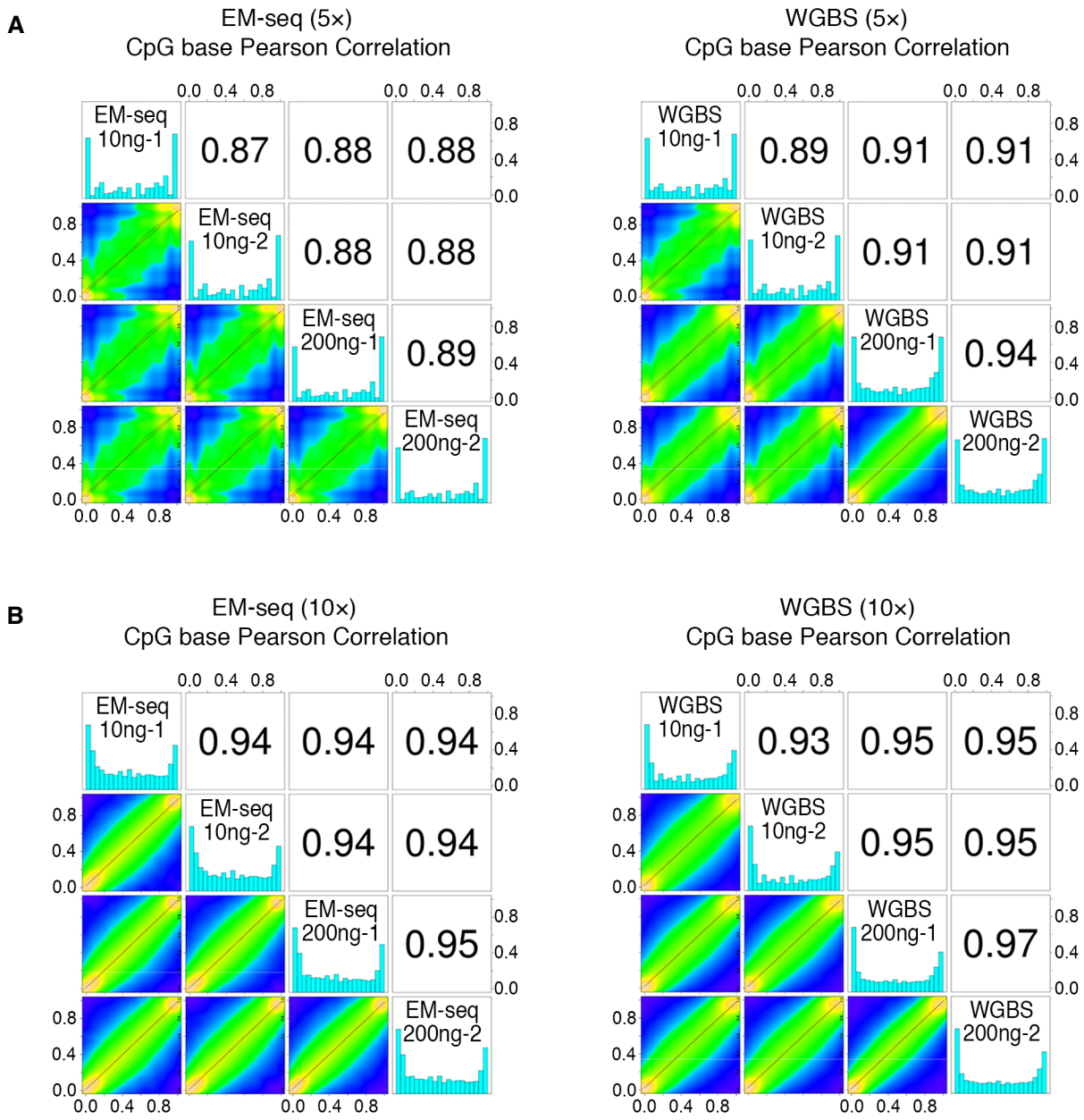

Histogram of % CpG Methylation
WGBS 200 ng-1
Frequency
% methylation per base


Histogram of % CpG Methylation
WGBS 200 ng-2
Frequency
% methylation per base

CpG coverage and methylation per base

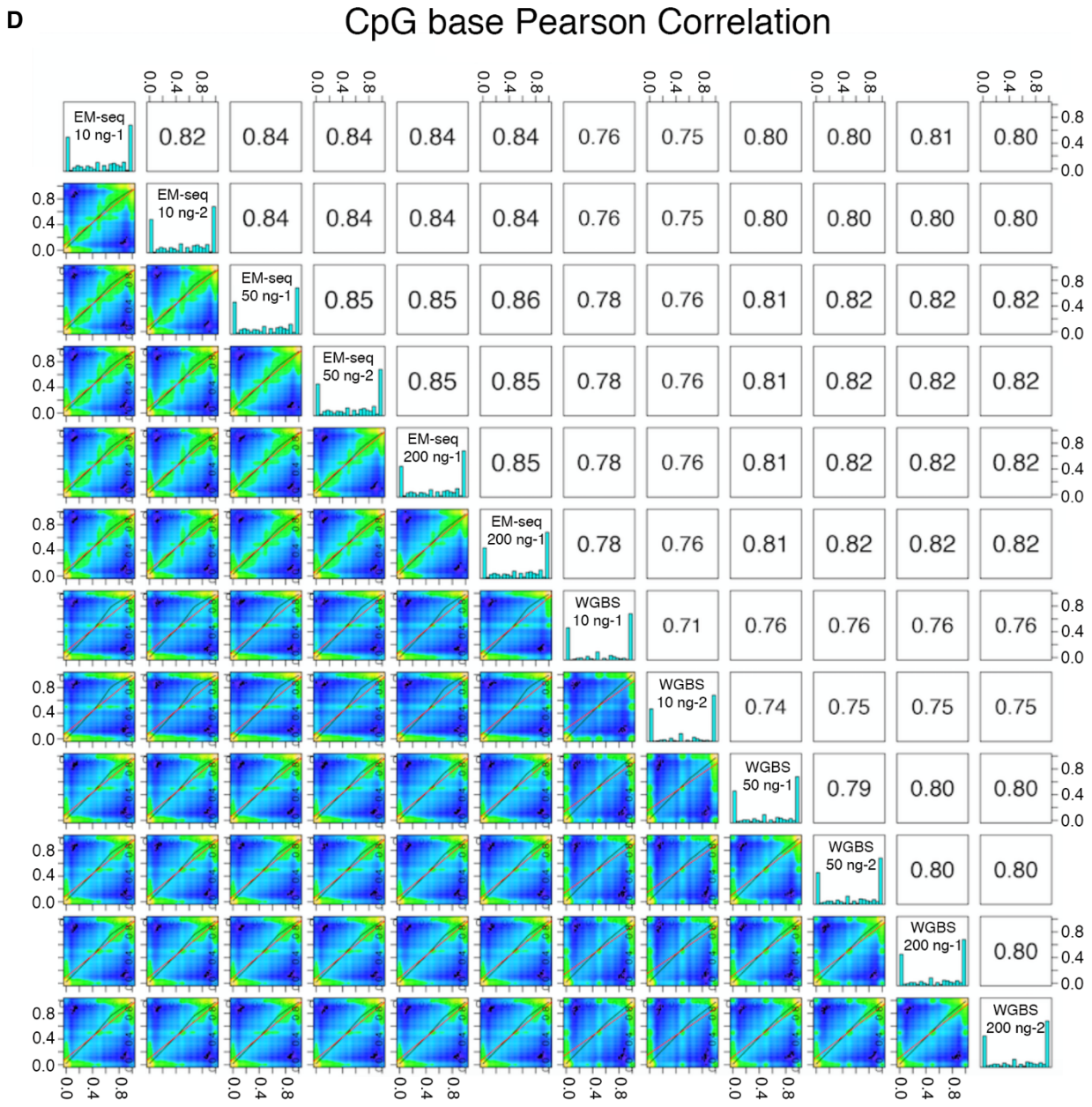
MethyKit was used to generate CpG coverage and percent methylation per base plots for data from EM-seq and bisulfite libraries made using 10, 50 and 200 ng NA12878 DNA.

Supplemental Figure 9



C

Correlation Coverage	EM-seq: CpGs used for corrlelation	WGBS: CpGs used for corrlelation
1×	53,458,888	25,958,126
5×	14,878,258	3,578,394
10×	179,991	615,413

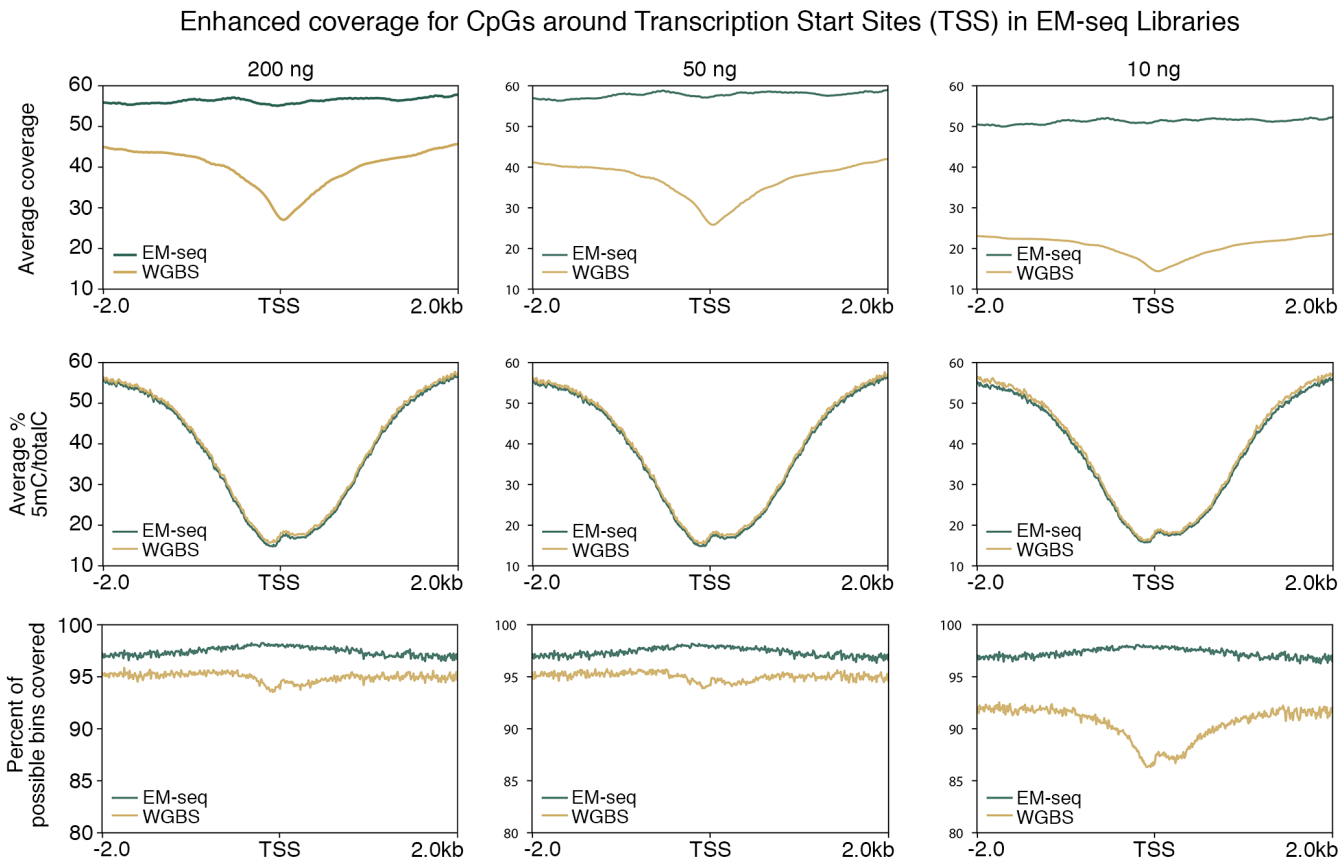


Correlations between EM-seq and bisulfite libraries

(A, B) MethylKit was used to plot correlations of data from 10 ng and 200 ng NA12878 EM-seq and bisulfite libraries at coverage depths of 5× (A) or 10× (B). (C) Different coverage thresholds result in the identification of differential numbers of CpGs between EM-seq and WGBS libraries. (D) Correlations were plotted using methylKit for 10 ng, 50 ng and 200 ng NA12878 EM-seq libraries and bisulfite

libraries at 1× minimum coverage (21 million CpGs common to all libraries). Correlations are highest for EM-seq libraries regardless of input. WGBS library correlations are lower with 10 ng inputs driving the lowest correlations and limiting the number of CpGs common to all libraries.

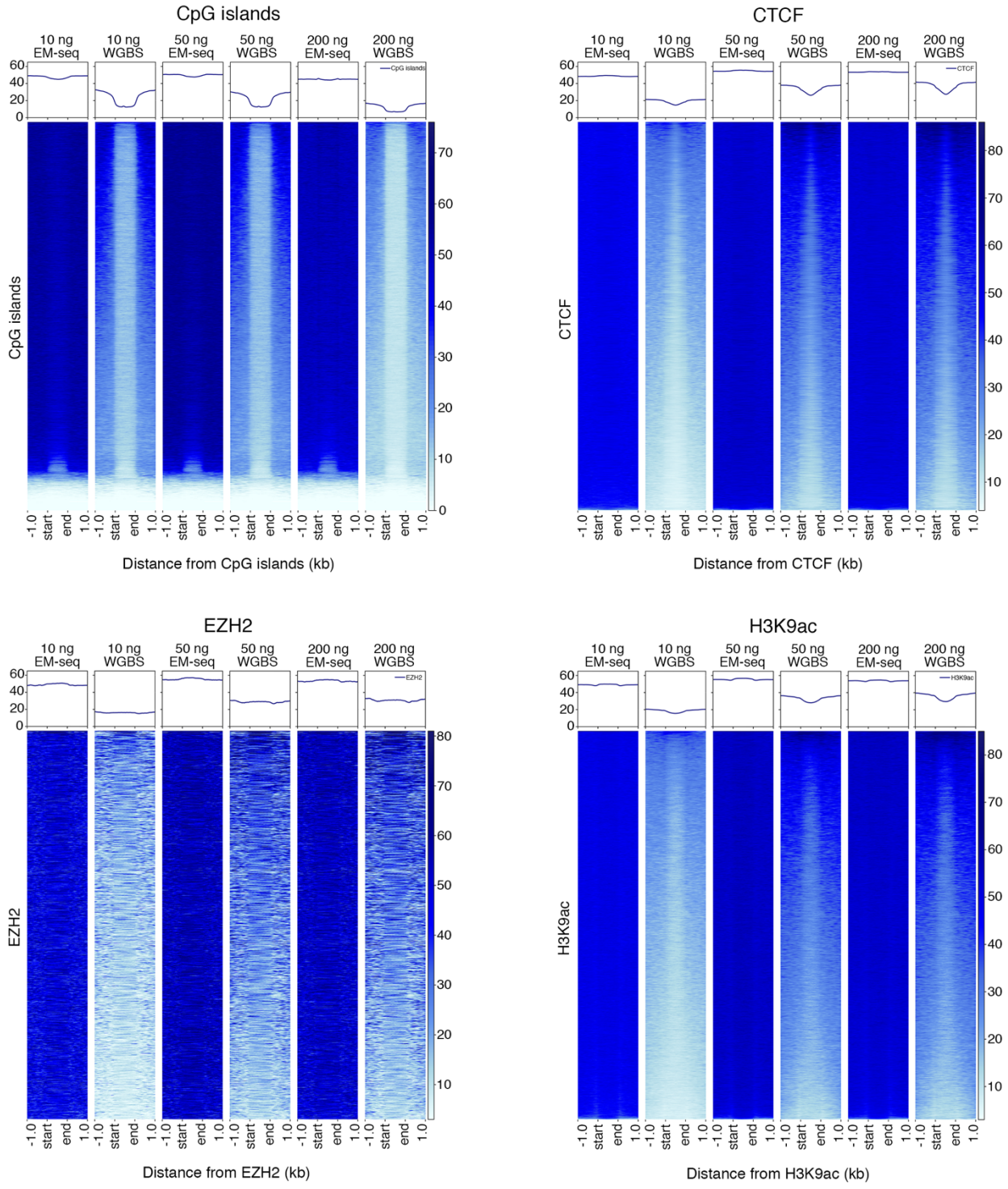
Supplemental Figure 10

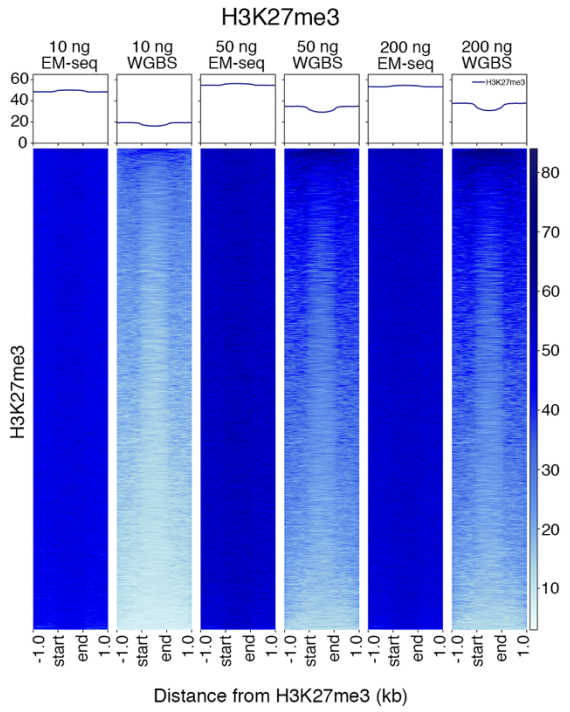


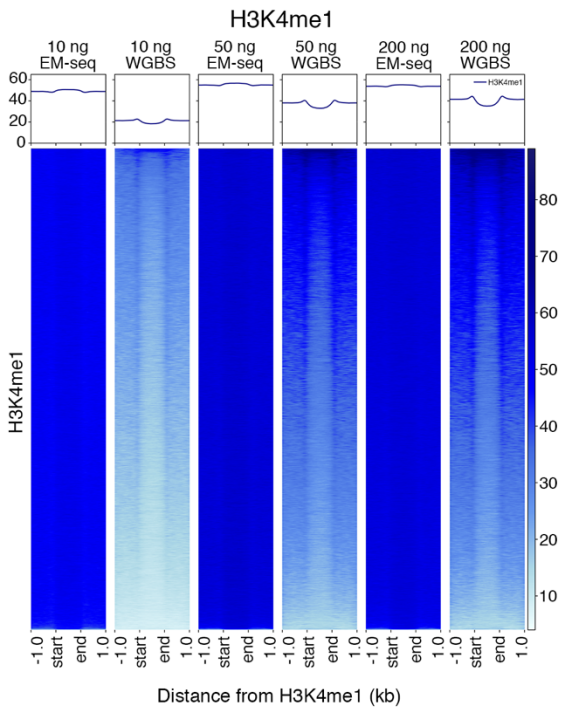
CpG coverage and methylation at the transcription start site

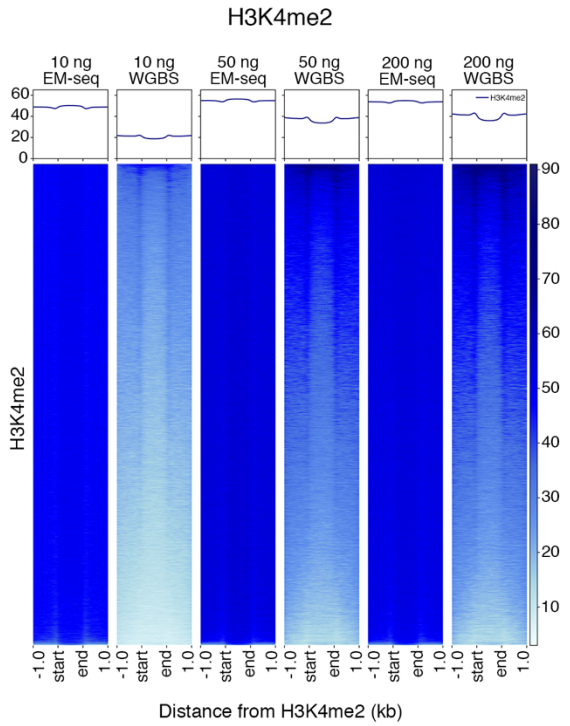
CpG coverage and methylation for EM-seq and WGBS libraries across TSS. This data had no coverage filtering before analysis.

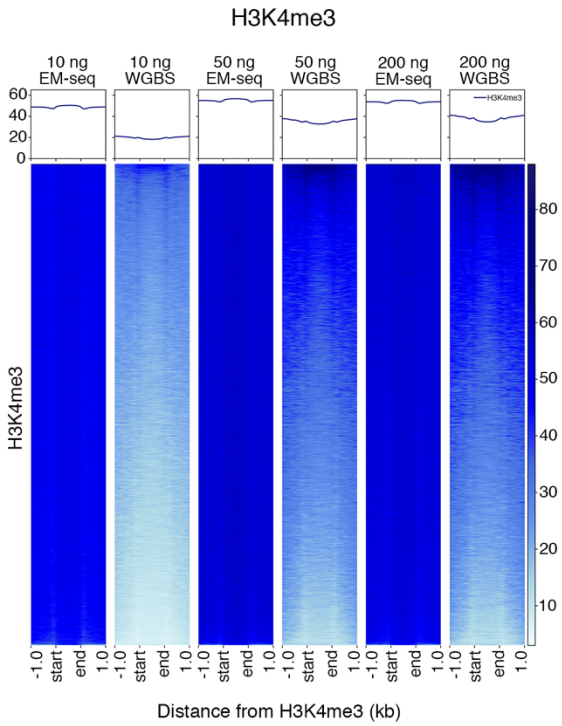
Supplemental Figure 11




H3K27me3
10 ng EM-seq
10 ng WGBS
50 ng EM-seq
50 ng WGBS
200 ng EM-seq
200 ng WGBS
H3K27me3
Distance from H3K27me3 (kb)


H3K4me1
10 ng EM-seq
10 ng WGBS
50 ng EM-seq
50 ng WGBS
200 ng EM-seq
200 ng WGBS
H3K4me1
Distance from H3K4me1 (kb)

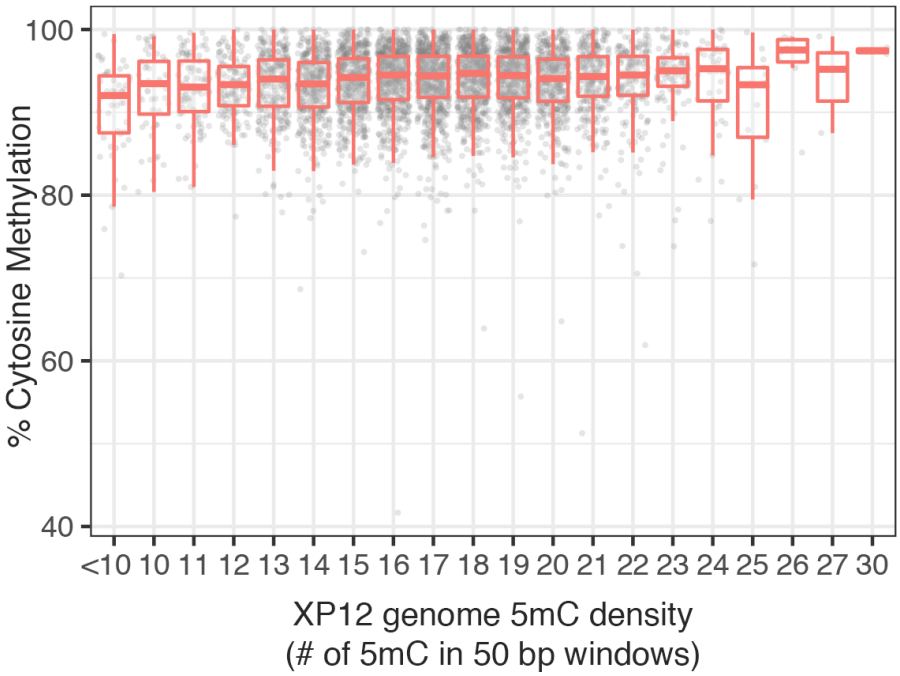

H3K4me2
10 ng EM-seq
10 ng WGBS
50 ng EM-seq
50 ng WGBS
200 ng EM-seq
200 ng WGBS
H3K4me2
Distance from H3K4me2 (kb)


H3K4me3
10 ng EM-seq
10 ng WGBS
50 ng EM-seq
50 ng WGBS
200 ng EM-seq
200 ng WGBS
H3K4me3
Distance from H3K4me3 (kb)

Even coverage of CpGs found in diverse genomic features using EM-seq

Heat maps (made using deepTools, gencode) representing CpG coverage across a range of genomic features including CTCF and EZH2 binding sites, CpG islands as well as H3K4me1, H3K4me2, H3K4me3, H3K27me3 and H3K9ac binding sites. Coverage plot include +/- 1 kb either side of the start and end of each feature.

Supplemental Figure 12

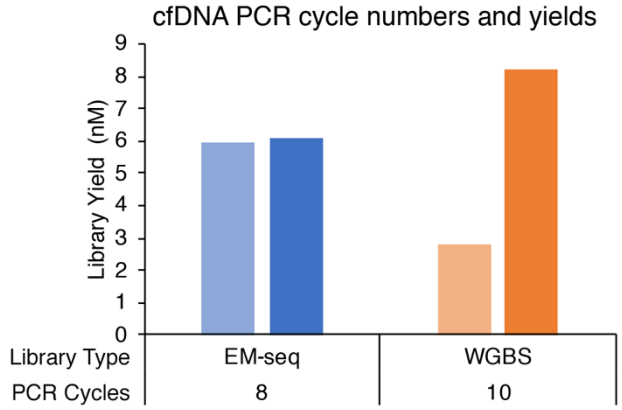


Methylation of cytosine in XP12

An EM-seq library was made using XP12 genomic DNA. The XP12 genome was divided into strand-specific 50 bp windows at 25 bp step intervals and the cytosine density plotted against methylation to investigate any reduction in observed methylation caused by the higher density of cytosines. The plot demonstrates that there are no extreme biases against high density cytosine methylated regions.

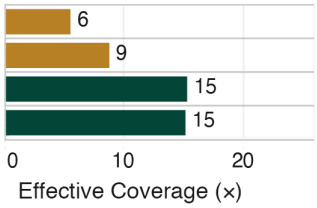
Supplemental Figure 13

A

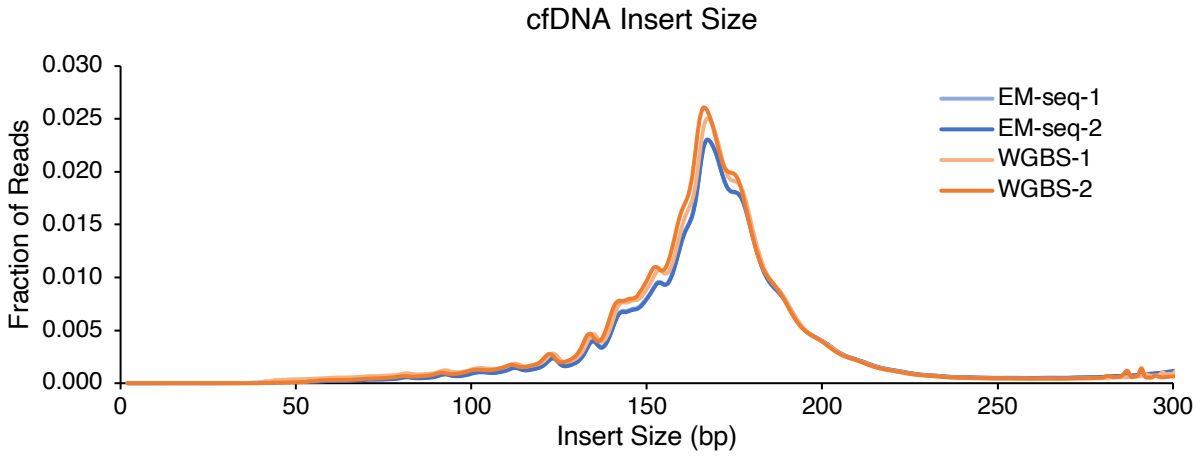


B

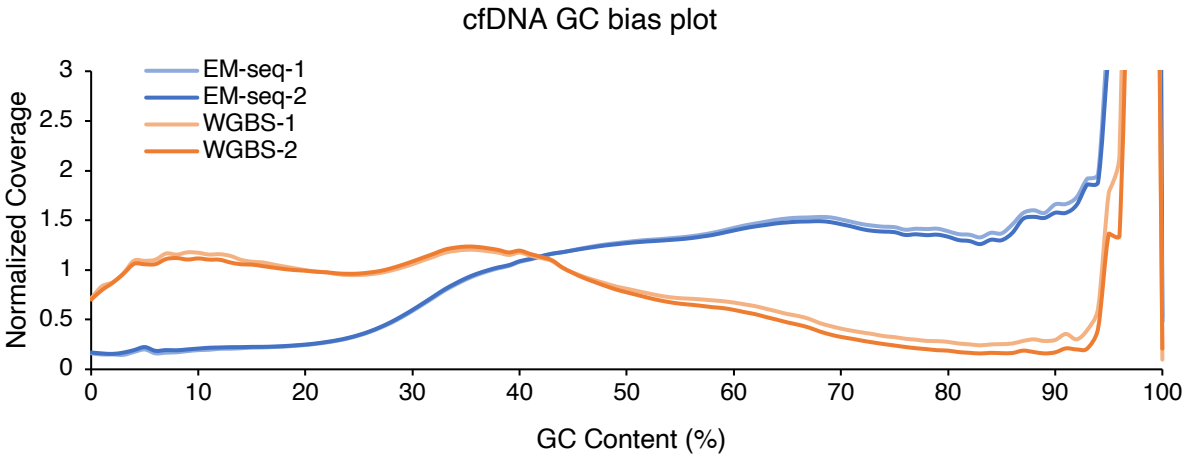
Input (ng)	Read set	Protocol	Rep.	Reads	Expected Coverage	% Mapped	% Dups	% Usable
10	cfDNA	Bisulfite (std)	1	755M	24	99.4%	71.5%	22.5%
			2	754M	24	99.5%	56.5%	36.2%
		EM-seq (std)	1	755M	24	99.3%	29.4%	62.7%
			2	756M	24	99.3%	29.8%	62.4%



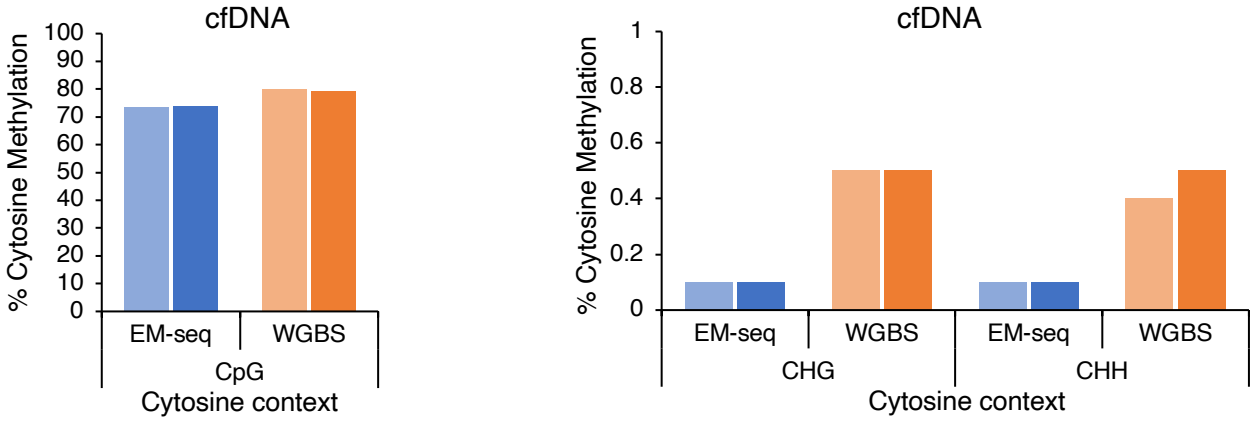
C



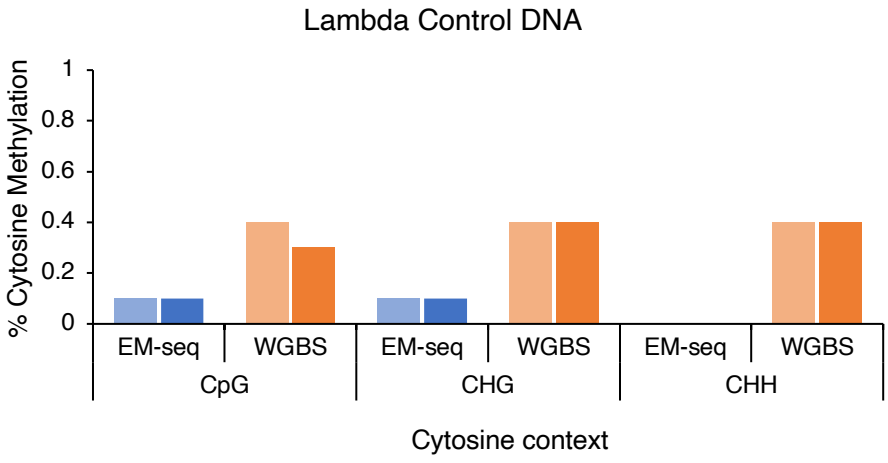
D


cfDNA GC bias plot
EM-seq-1
EM-seq-2
WGBS-1
WGBS-2
Normalized Coverage
GC Content (%)

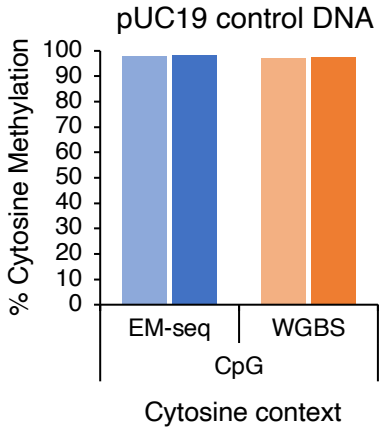
E

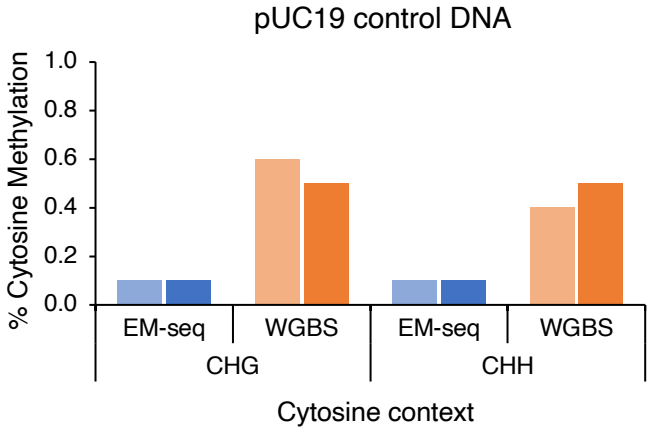

cfDNA
% Cytosine Methylation
EM-seq
WGBS
CpG
Cytosine context
cfDNA
% Cytosine Methylation
EM-seq
WGBS
EM-seq
WGBS
CHG
CHH
Cytosine context

F

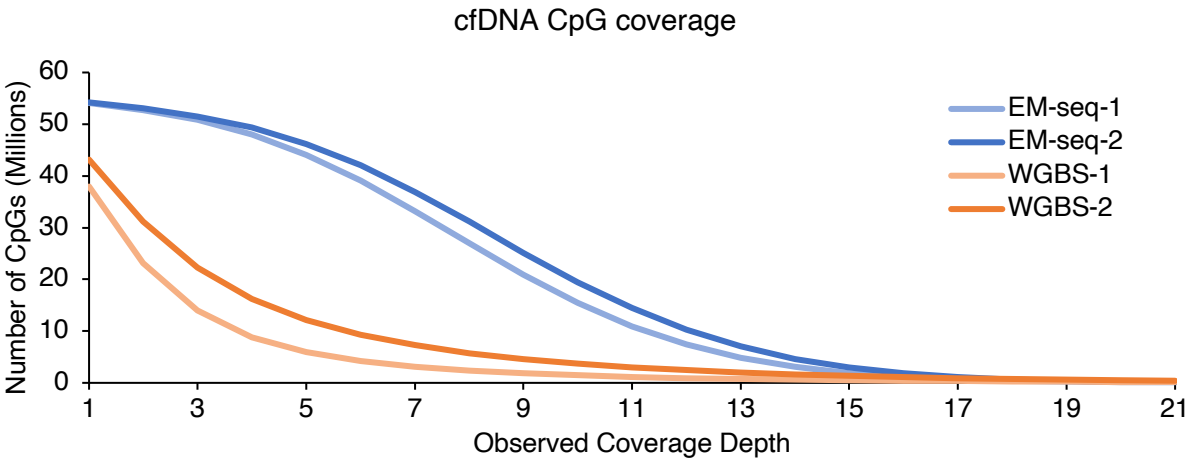

Lambda Control DNA
% Cytosine Methylation
EM-seq
WGBS
EM-seq
WGBS
EM-seq
WGBS
CpG
CHG
CHH
Cytosine context

G

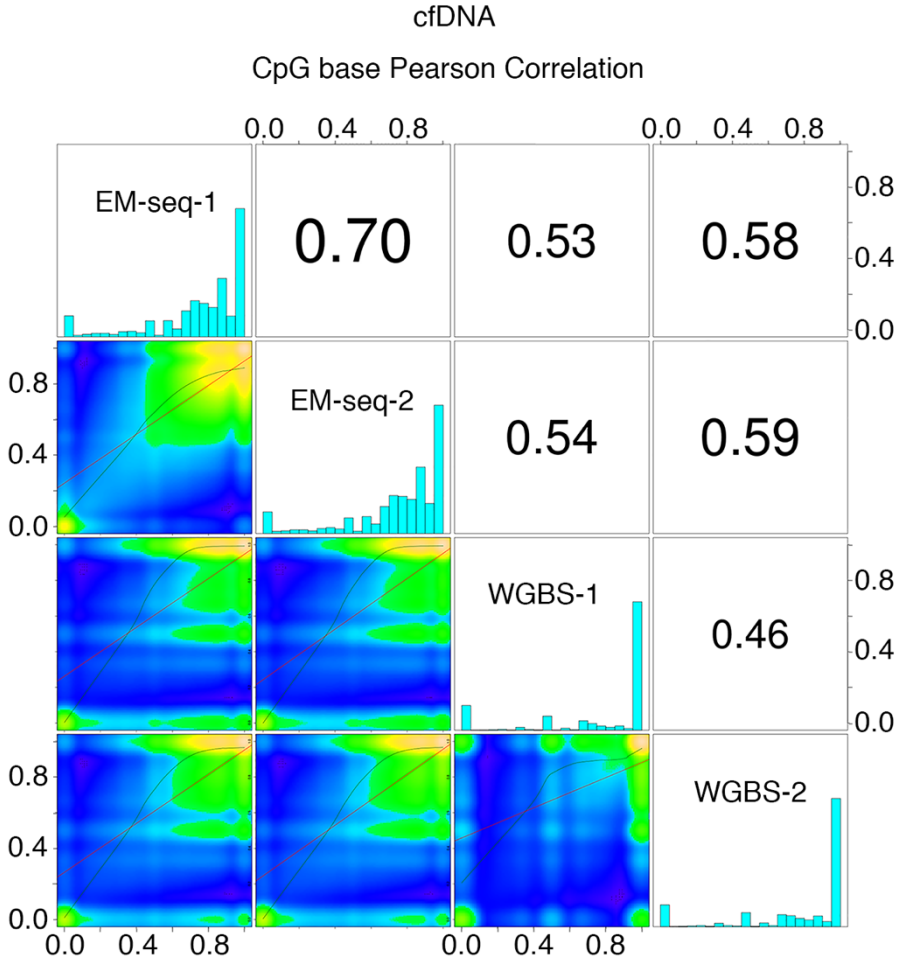

pUC19 control DNA
% Cytosine Methylation
100
90
80
70
60
50
40
30
20
10
0
EM-seq
WGBS
CpG
Cytosine context


pUC19 control DNA
% Cytosine Methylation
1.0
0.8
0.6
0.4
0.2
0.0
EM-seq
WGBS
EM-seq
WGBS
CHG
CHH
Cytosine context

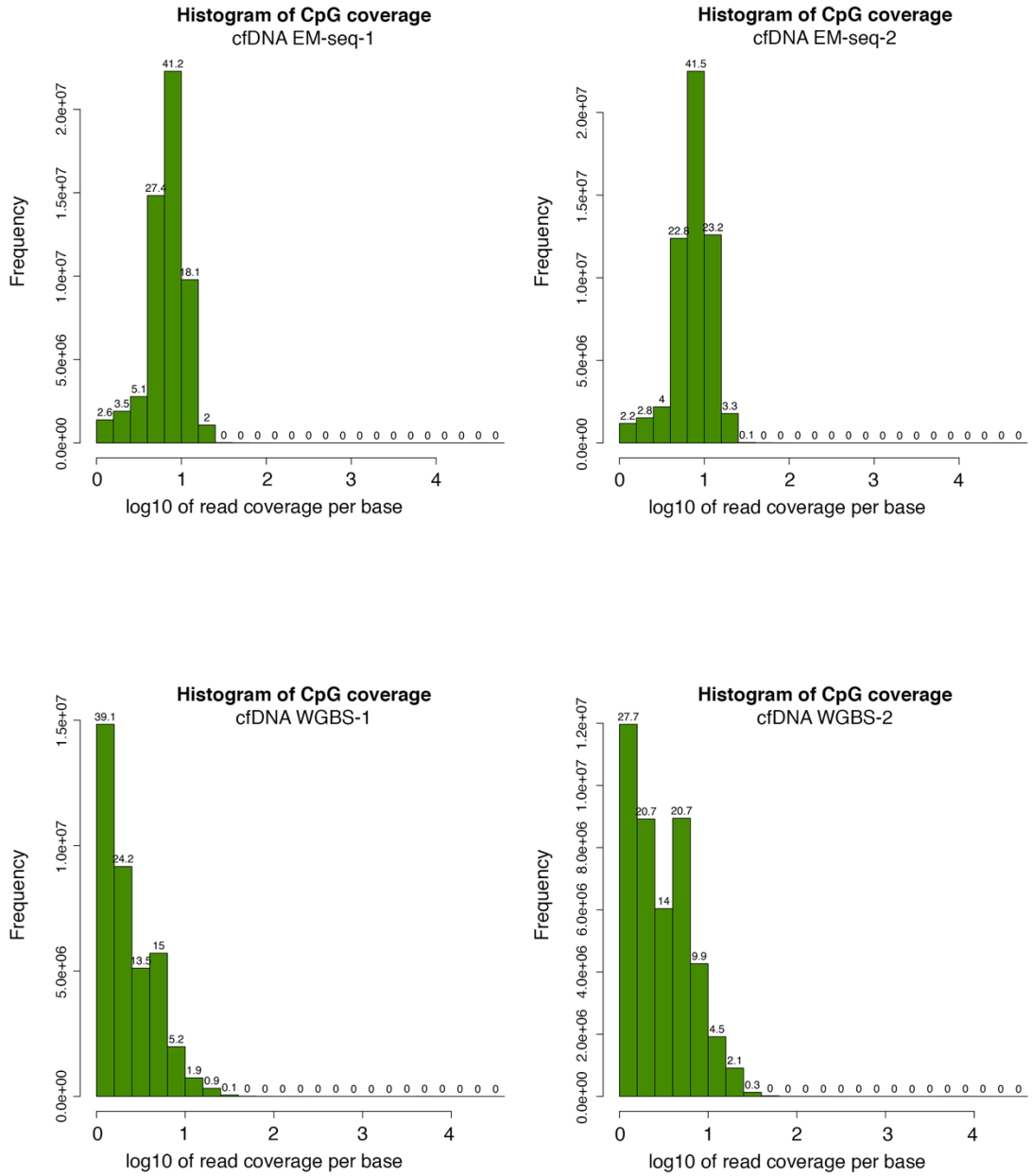
H


cfDNA CpG coverage
Number of CpGs (Millions)
60
50
40
30
20
10
0
1
3
5
7
9
11
13
15
17
19
21
Observed Coverage Depth
EM-seq-1
EM-seq-2
WGBS-1
WGBS-2

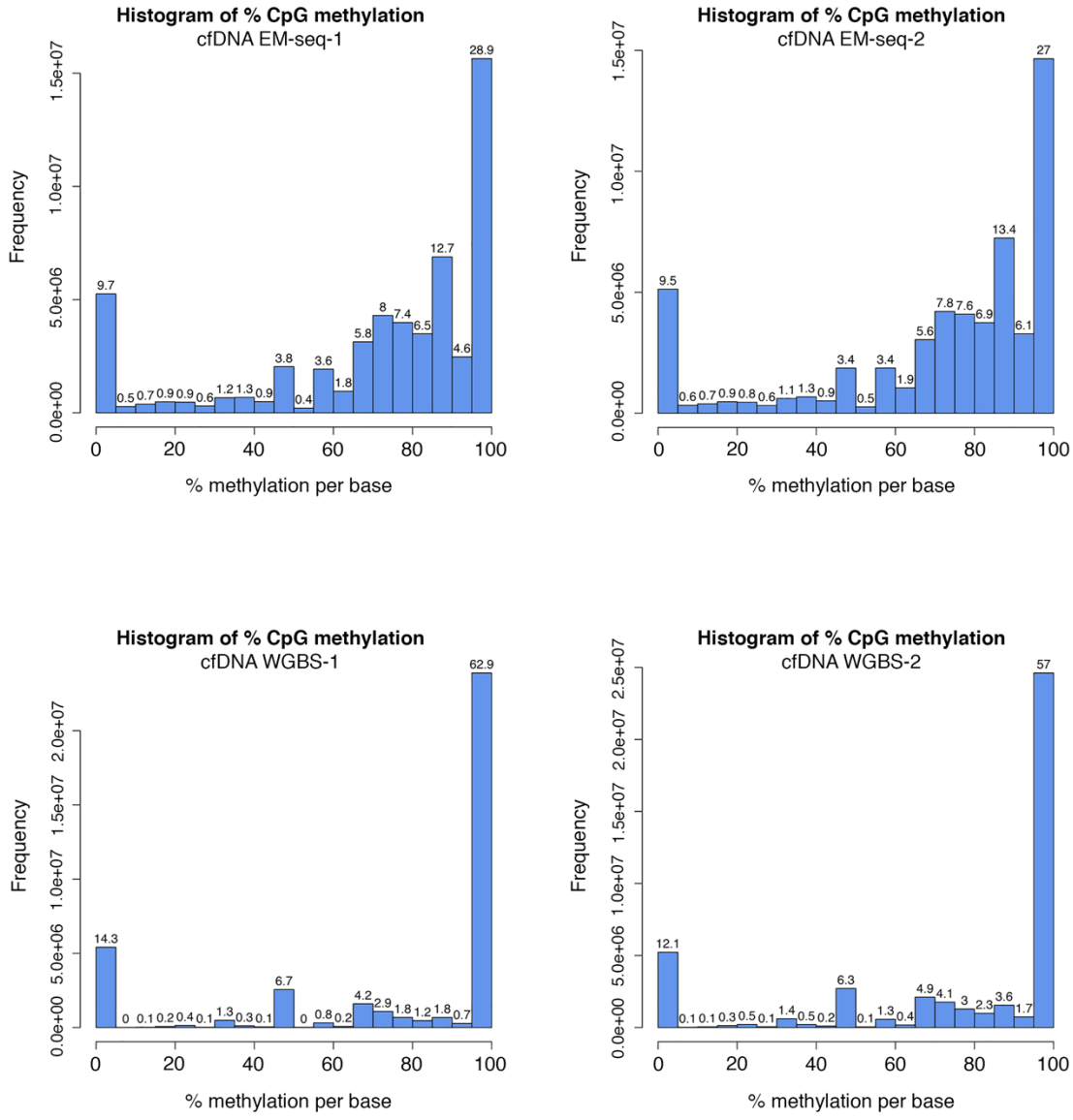
I


cfDNA
CpG base Pearson Correlation
EM-seq-1
EM-seq-2
WGBS-1
WGBS-2
0.70
0.53
0.58
0.54
0.59
0.46
0.0
0.4
0.8

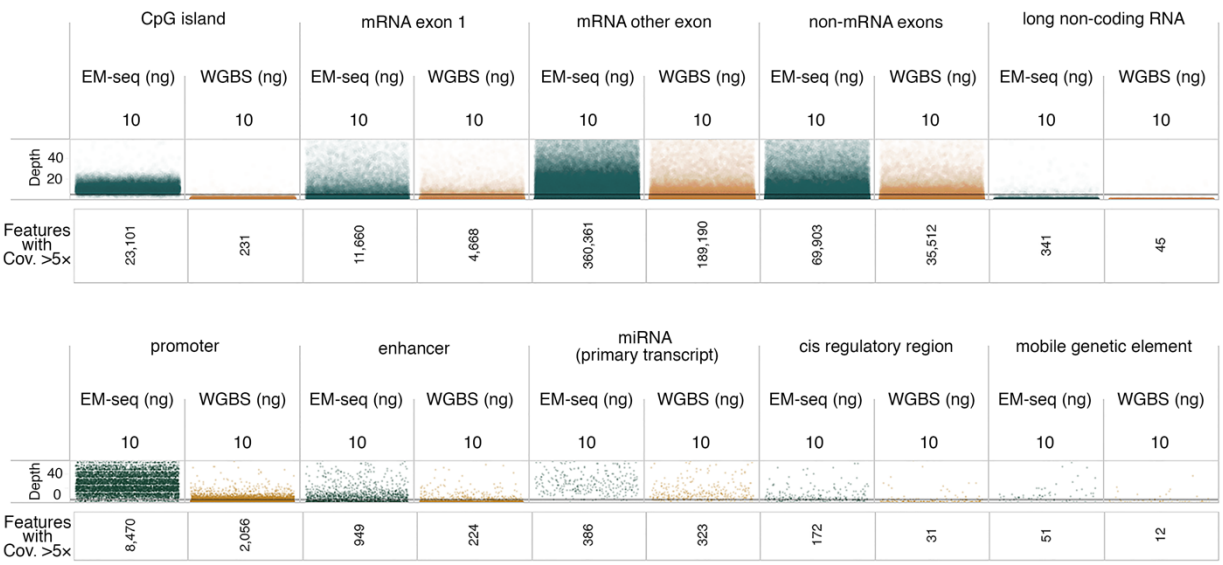
J


Histogram of CpG coverage
cfDNA EM-seq-1
Frequency
log10 of read coverage per base
Histogram of CpG coverage
cfDNA EM-seq-2
Frequency
log10 of read coverage per base
Histogram of CpG coverage
cfDNA WGBS-1
Frequency
log10 of read coverage per base
Histogram of CpG coverage
cfDNA WGBS-2
Frequency
log10 of read coverage per base

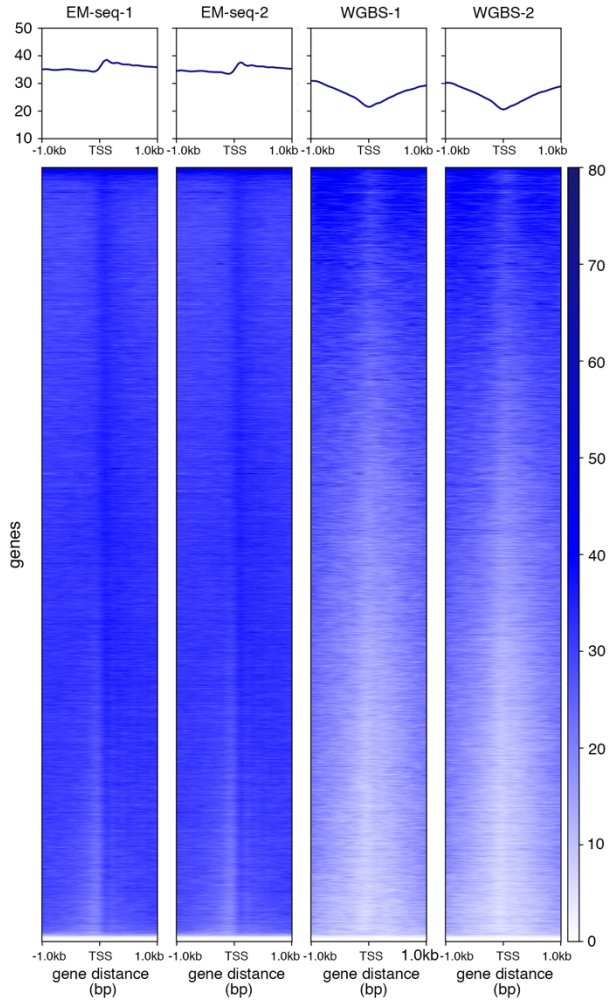
K


Histogram of % CpG methylation
cfDNA EM-seq-1
Frequency
% methylation per base
Histogram of % CpG methylation
cfDNA EM-seq-2
Frequency
% methylation per base
Histogram of % CpG methylation
cfDNA WGBS-1
Frequency
% methylation per base
Histogram of % CpG methylation
cfDNA WGBS-2
Frequency
% methylation per base

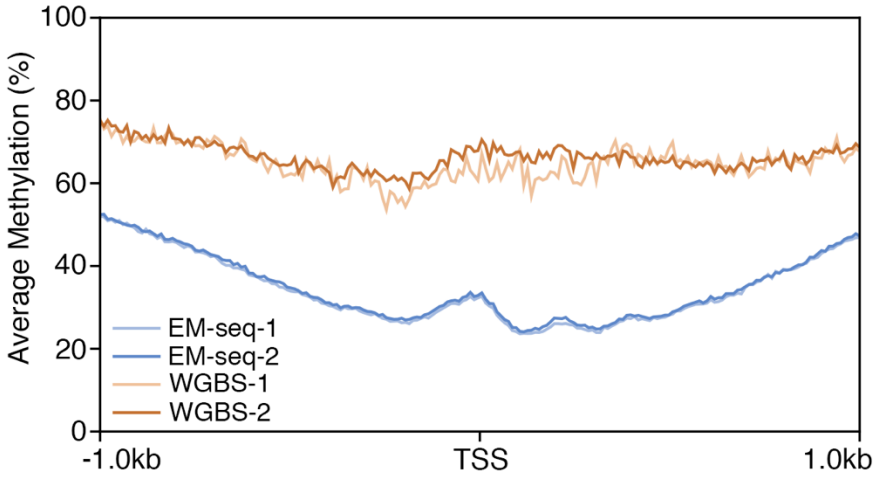
L


CpG island
mRNA exon 1
mRNA other exon
non-mRNA exons
long non-coding RNA
EM-seq (ng)
WGBS (ng)
10
Depth
40
20
Features with Cov. >5×
23,101
231
11,660
4,668
360,361
189,190
69,903
35,512
341
45
promoter
enhancer
miRNA (primary transcript)
cis regulatory region
mobile genetic element
0
8,470
2,056
949
224
386
323
172
31
51
12

M


EM-seq-1
EM-seq-2
WGBS-1
WGBS-2
50
40
30
20
10
-1.0kb
TSS
1.0kb
genes
80
70
60
50
40
30
20
10
0
gene distance (bp)

N



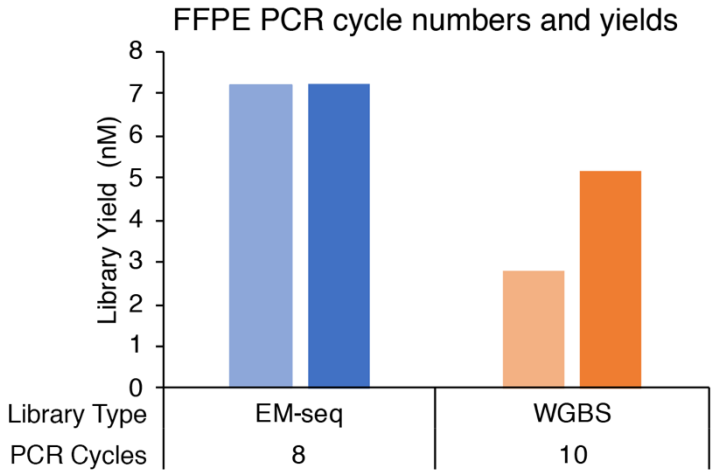
EM-seq and bisulfite cfDNA libraries

10 ng cfDNA was spiked with control DNA (2 ng unmethylated lambda DNA and 0.1 ng CpG methylated pUC19) and libraries were made using either EM-seq or the Zymo Gold Bisulfite Kit prepped for Illumina sequencing using NEBNext Ultra II DNA library kit reagents. Libraries were sequenced on an Illumina NovaSeq 6000 and 750 million reads per library were used for analysis. (A) EM-seq uses less PCR cycles but results in more PCR product than WGBS libraries. (B) Table of sequencing and alignment metrics for EM-seq and WGBS libraries using 750 million Illumina reads. Metrics were calculated using bwa-meth, SAMtools, and Picard. Theoretical coverage is calculated using the number of bases sequenced/total bases in the GRCh38 reference. Percent mapped refers to reads aligned to the reference genome (grch38+controls), Percent Dups refers to reads marked as duplicate by Picard MarkDuplicates, Percent Usable refers to the set of Proper-pair, MapQ 10+, Primary, non-Duplicate reads used in methylation calling (SAMtools view -F 0xF00 -q 10). Effective Coverage is the Percent Usable × Theoretical coverage. (C) Insert sizes of EM-seq and WGBS libraries. (D) GC-bias plot for EM-seq and WGBS libraries. EM-seq libraries display an even GC distribution while WGBS libraries have an AT- rich and GC- poor profile. (E) cfDNA methylation in CpG, CHG and CHH contexts are similarly represented for EM-seq and WGBS

libraries. (F, G) Methylation state for unmethylated lambda control and CpG methylated pUC19 control DNA. (H) The number of CpGs covered for EM-seq and bisulfite libraries were calculated and graphed at minimum coverage depths of 1× through 21×. (I) MethylKit analysis at a minimum of 1× coverage shows good CpG methylation correlation between the 10 ng EM-seq and WGBS libraries. (J, K) CpG coverage and percent methylation per base data were obtained using methylKit. (L) EM-seq and WGBS cover a diverse range of genomic features but EM-seq libraries exhibit greater coverage of all features examined. Coverage of various genomic feature types are represented with one point per region. The vertical position is defined by the average coverage of the feature. Points are staggered horizontally to avoid too much overlapping. Features from NCBI's RefSeq, the Eukaryotic Promoter, UCSC Table Browser and Dfam are shown and the numbers covered at 5× or greater depth are indicated. (M) deepTool Heat map of CpG coverage for +/- 1 kb of the TSS for the 10 ng inputs for EM-seq and bisulfite libraries at a minimum coverage of 1×. Plots were created using random sampling of the same number of reads raw reads from each library (N) Methylation status for EM-seq and bisulfite libraries using 8× minimum coverage depth. EM-seq libraries show less CpG methylation and more accurately represent the expected CpG methylation pattern.

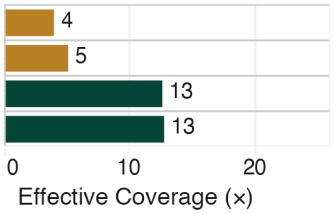
Supplemental Figure 14

A

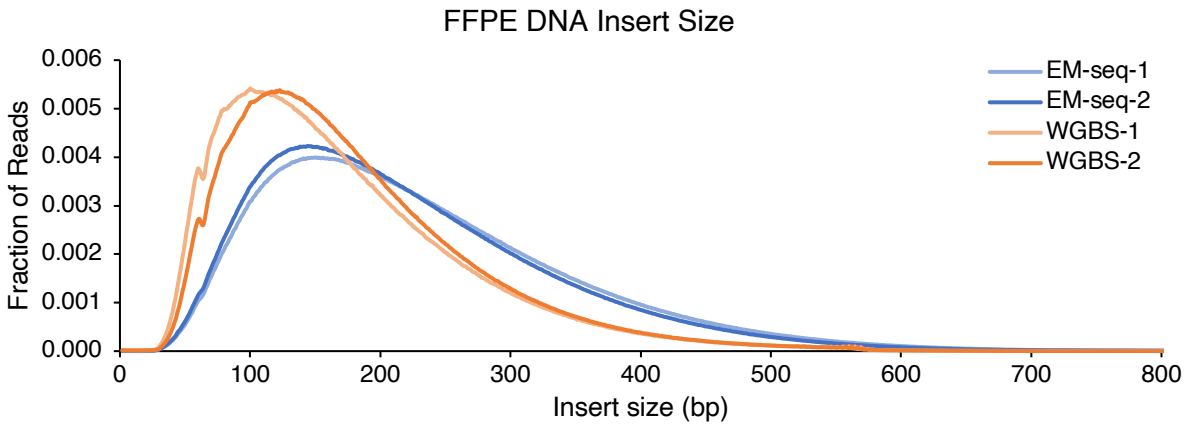


B

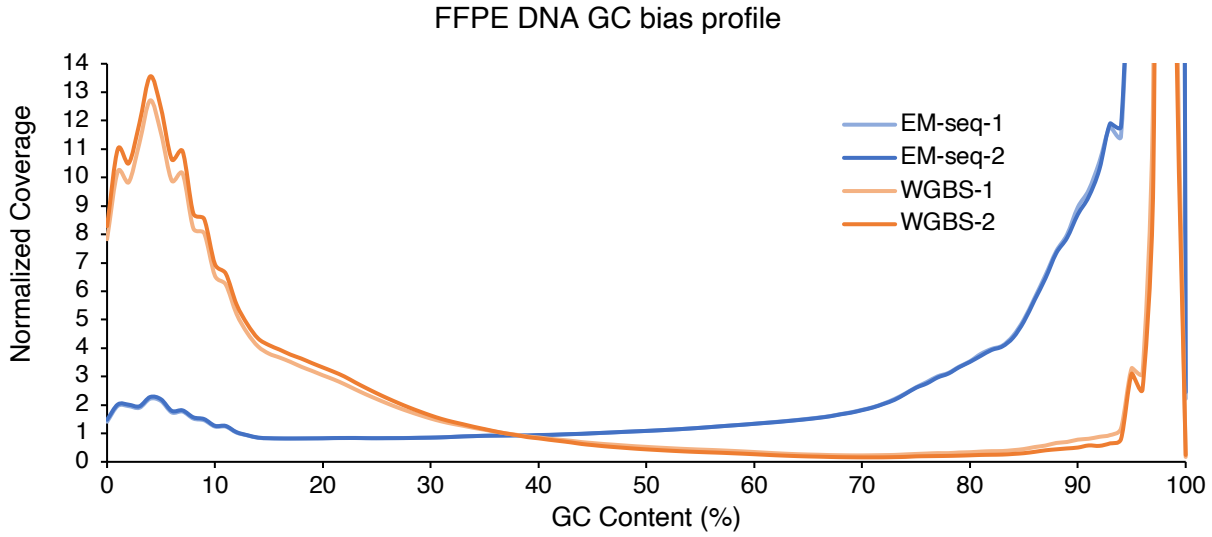
Input (ng)	Read set	Protocol	Rep.	Reads	Expected Coverage	% Mapped	% Dups	% Usable
10	FFPE	Bisulfite (std)	1	755M	24	99.3%	73.8%	16.4%
			2	754M	24	99.5%	68.4%	20.8%
		EM-seq (std)	1	816M	26	91.9%	29.7%	48.0%
			2	815M	26	92.0%	29.3%	48.3%



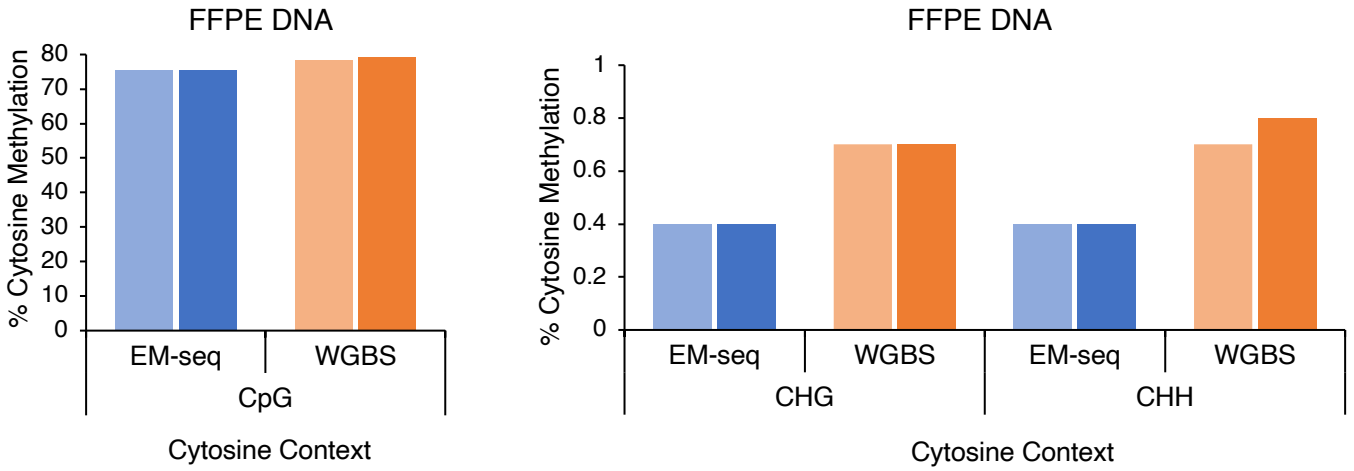
C



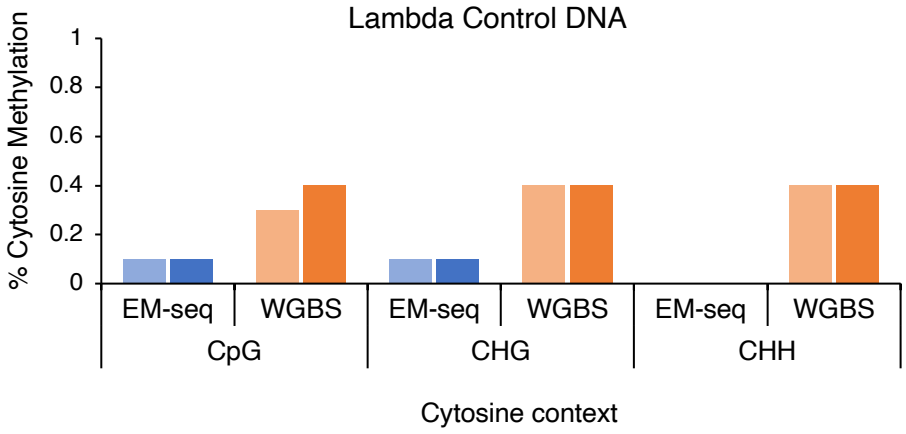
D


FFPE DNA GC bias profile
Normalized Coverage
GC Content (%)
EM-seq-1
EM-seq-2
WGBS-1
WGBS-2

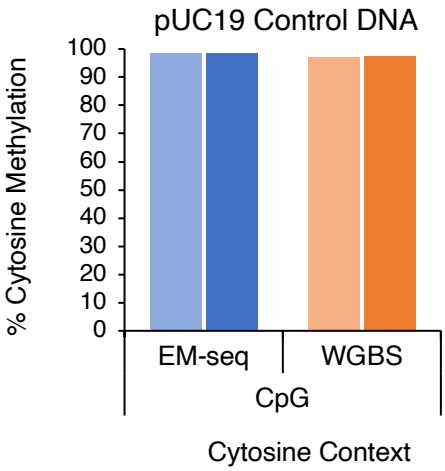
E

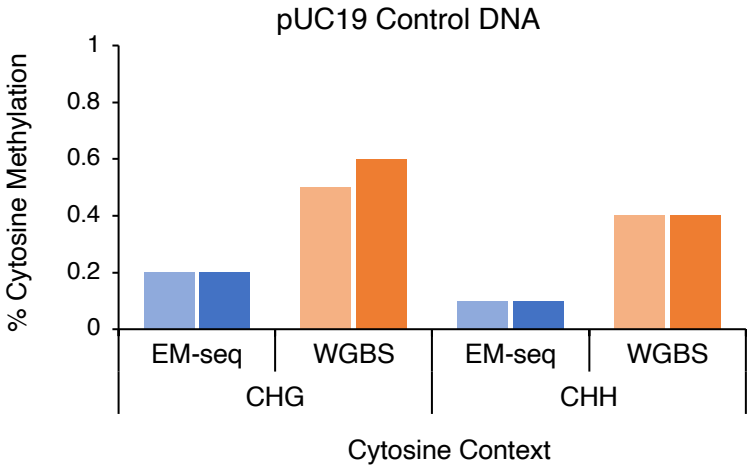

FFPE DNA
% Cytosine Methylation
EM-seq
WGBS
CpG
Cytosine Context
FFPE DNA
% Cytosine Methylation
EM-seq
WGBS
EM-seq
WGBS
CHG
CHH
Cytosine Context

F

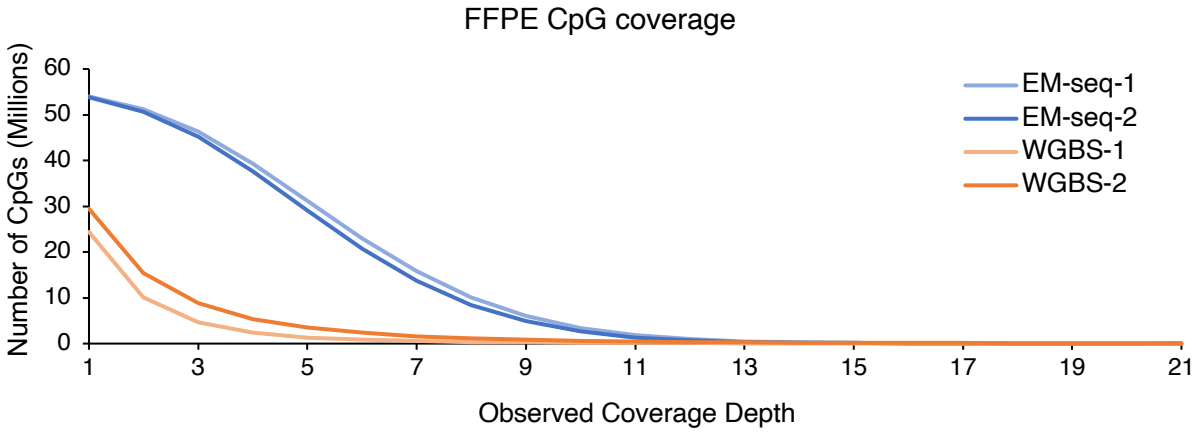

Lambda Control DNA
% Cytosine Methylation
EM-seq
WGBS
EM-seq
WGBS
EM-seq
WGBS
CpG
CHG
CHH
Cytosine context

G

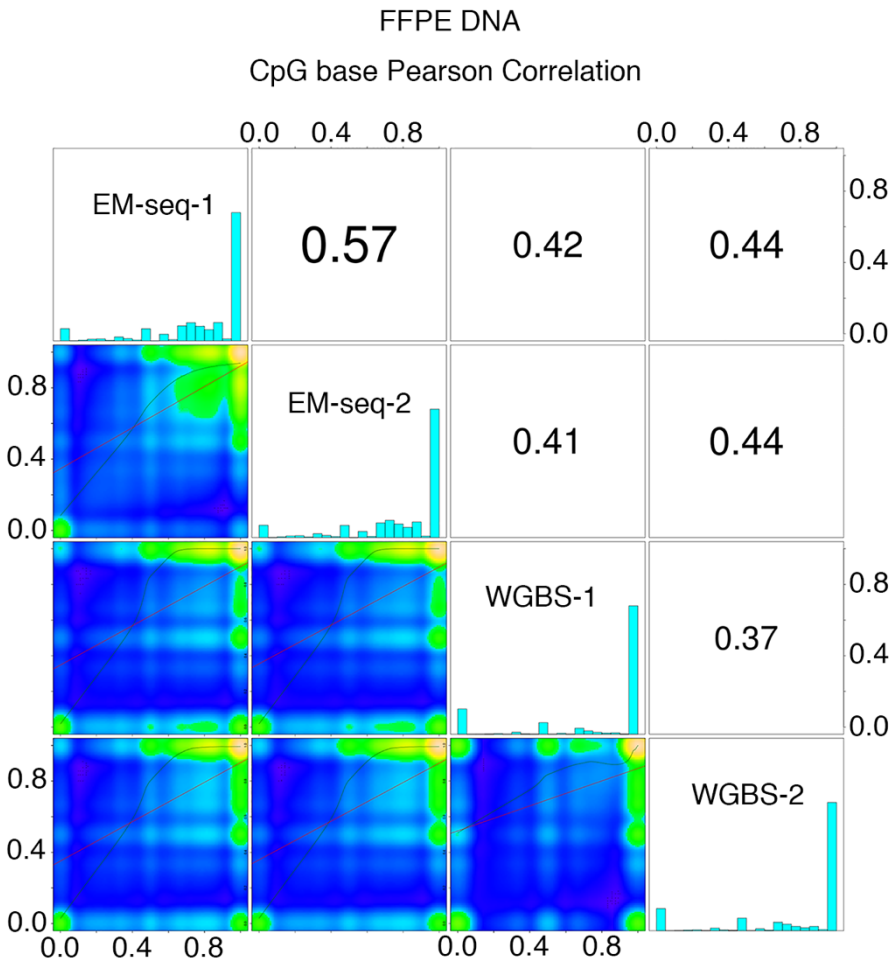

pUC19 Control DNA
% Cytosine Methylation
100
90
80
70
60
50
40
30
20
10
0
EM-seq
WGBS
CpG
Cytosine Context


pUC19 Control DNA
% Cytosine Methylation
1
0.8
0.6
0.4
0.2
0
EM-seq
WGBS
EM-seq
WGBS
CHG
CHH
Cytosine Context

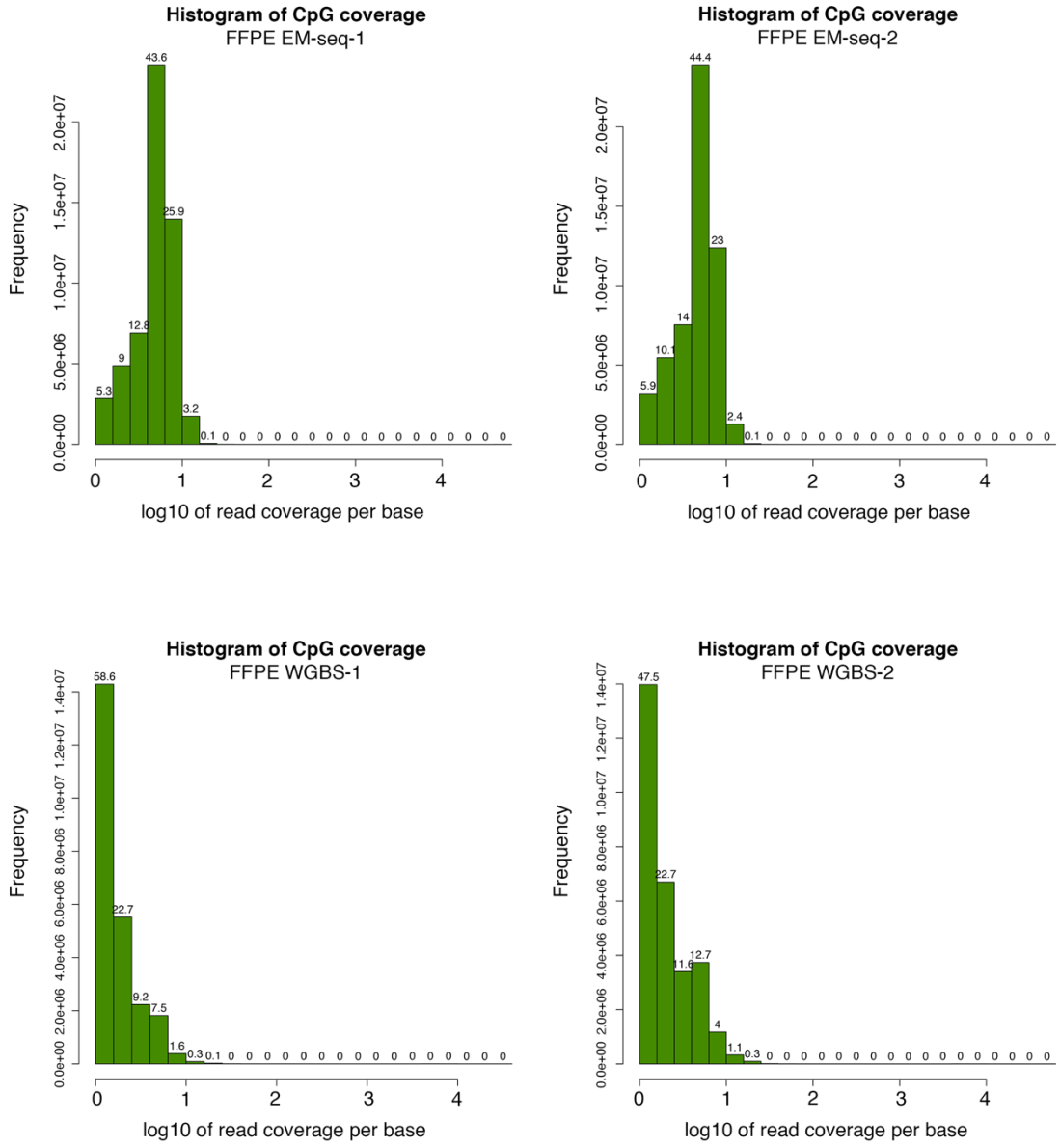
H


FFPE CpG coverage
Number of CpGs (Millions)
60
50
40
30
20
10
0
1
3
5
7
9
11
13
15
17
19
21
Observed Coverage Depth
EM-seq-1
EM-seq-2
WGBS-1
WGBS-2

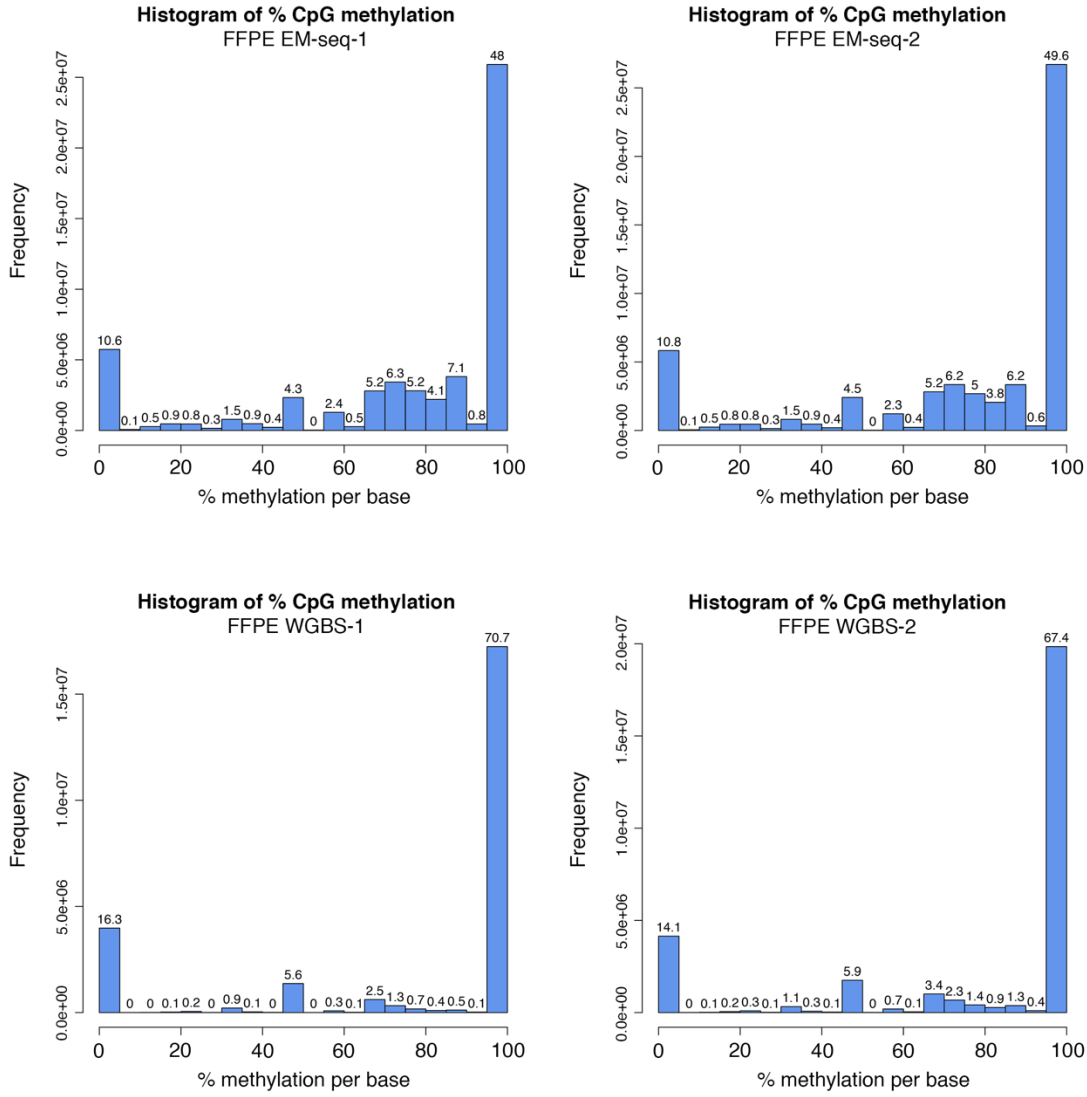
I


FFPE DNA
CpG base Pearson Correlation
EM-seq-1
EM-seq-2
WGBS-1
WGBS-2
0.57
0.42
0.44
0.41
0.44
0.37
0.0
0.4
0.8

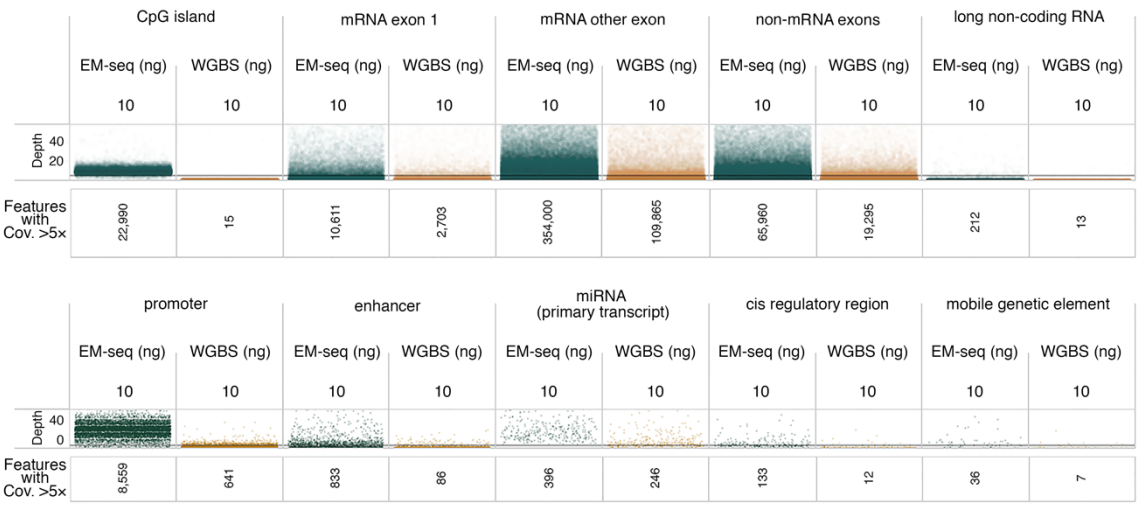
J


Histogram of CpG coverage
FFPE EM-seq-1
Frequency
log10 of read coverage per base
43.6
25.9
12.8
9
5.3
3.2
0.1
Histogram of CpG coverage
FFPE EM-seq-2
Frequency
log10 of read coverage per base
44.4
23
14
10.1
5.9
2.4
0.1
Histogram of CpG coverage
FFPE WGBS-1
Frequency
log10 of read coverage per base
58.6
22.7
9.2
7.5
1.6
0.3
0.1
Histogram of CpG coverage
FFPE WGBS-2
Frequency
log10 of read coverage per base
47.5
22.7
12.7
11.6
4
1.1
0.3

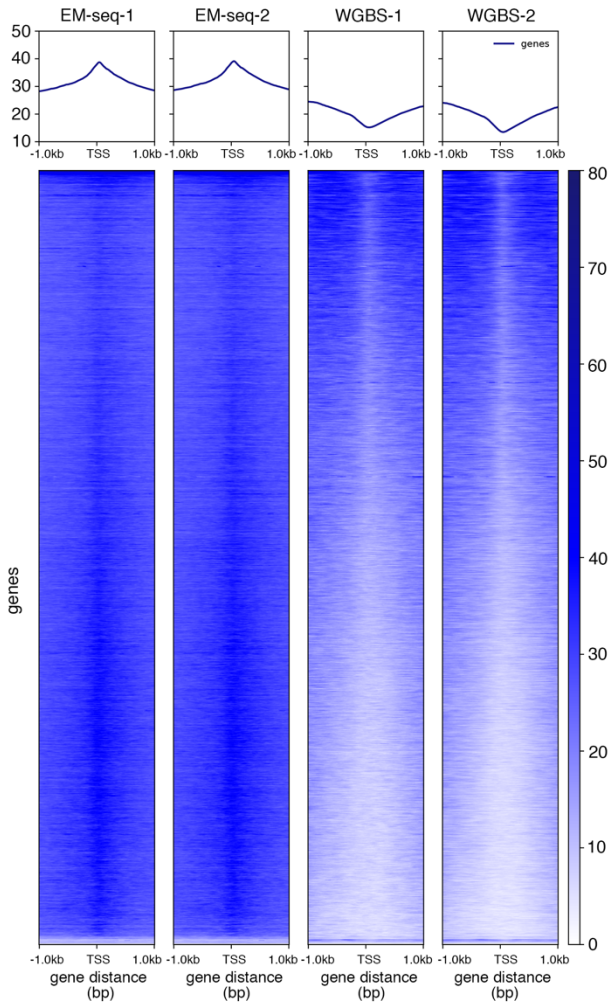
K


Histogram of % CpG methylation
FFPE EM-seq-1
Frequency
% methylation per base
Histogram of % CpG methylation
FFPE EM-seq-2
Frequency
% methylation per base
Histogram of % CpG methylation
FFPE WGBS-1
Frequency
% methylation per base
Histogram of % CpG methylation
FFPE WGBS-2
Frequency
% methylation per base

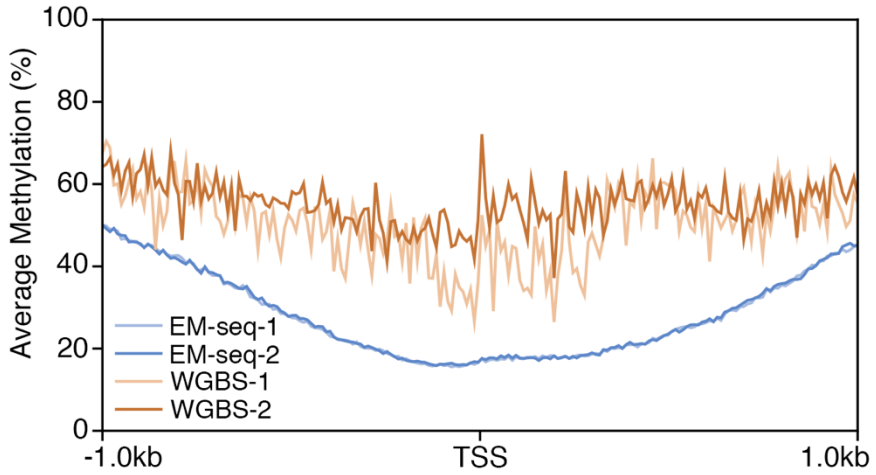
L


CpG island
mRNA exon 1
mRNA other exon
non-mRNA exons
long non-coding RNA
EM-seq (ng)
WGBS (ng)
10
Depth
40
20
Features with Cov. >5×
22,990
15
10,611
2,703
354,000
109,865
65,960
19,295
212
13
promoter
enhancer
miRNA (primary transcript)
cis regulatory region
mobile genetic element
0
8,559
641
833
86
396
246
133
12
36
7

M


EM-seq-1
EM-seq-2
WGBS-1
WGBS-2
genes
50
40
30
20
10
-1.0kb
TSS
1.0kb
80
70
60
50
40
30
20
10
0
gene distance (bp)

N



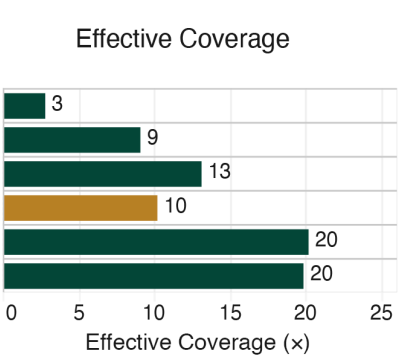
EM-seq and bisulfite FFPE libraries

10 ng FFPE DNA were spiked with control DNA (2 ng unmethylated lambda DNA and 0.1 ng CpG methylated pUC19) and libraries were made using either EM-seq or the Zymo Gold Bisulfite Kit prepped for Illumina sequencing using NEBNext Ultra II DNA library kit reagents. Libraries were sequenced on an Illumina NovaSeq 6000 and 750 - 810 million reads per library were used for analysis. (A) EM-seq uses less PCR cycles but results in more PCR product than whole genome bisulfite libraries (WGBS). (B) Table of sequencing and alignment metrics for EM-seq and WGBS libraries using 750 million read. Metrics were calculated using bwa-meth, SAMtools, and Picard. Theoretical coverage is calculated using the number of bases sequenced/total bases in the GRCh38 reference. Percent mapped refers to reads aligned to the reference genome (grch38+controls), Percent Dups refers to reads marked as duplicate by Picard MarkDuplicates, Percent Usable refers to the set of Proper-pair, MapQ 10+, Primary, non-Duplicate reads used in methylation calling (SAMtools view -F 0xF00 -q 10). Effective Coverage is the Percent Usable × Theoretical coverage. (C) Insert sizes of EM-seq and WGBS libraries. (D) GC-bias plot for EM-seq and WGBS libraries. EM-seq libraries display an even GC distribution while WGBS libraries have an AT- rich and GC- poor profile. (E) cfDNA methylation in CpG, CHG and CHH contexts are similarly represented for EM-seq and WGBS libraries. (F, G) Methylation state for unmethylated lambda control

and CpG methylated pUC19 control DNA. (H) The number of CpGs covered for EM-seq and bisulfite libraries were calculated and graphed at minimum coverage depths of 1× through 21×. (I) MethylKit analysis at a minimum of 1× coverage shows good CpG methylation correlation between the 10 ng EM-seq and WGBS libraries. (J, K) CpG coverage and percent methylation per base data were obtained using methylKit. (L) EM-seq and WGBS cover a diverse range of genomic features but EM-seq libraries exhibit greater coverage of all features examined. Coverage of various genomic feature types are represented with one point per region. The vertical position is defined by the average coverage of the feature. Points are staggered horizontally to avoid too much overlapping. Features from NCBI's RefSeq, the Eukaryotic Promoter, UCSC Table Browser and Dfam are shown and the numbers covered at 5× or greater depth are indicated. (M) deepTool Heat map of CpG coverage for +/- 1 kb of the TSS for the 10 ng inputs for EM-seq and bisulfite libraries at a minimum coverage of 1×. Plots were created using random sampling of the same number of reads raw reads from each library (N) Methylation status for EM-seq and bisulfite libraries using 8× minimum coverage depth. EM-seq libraries show less CpG methylation and more accurately represent the expected CpG methylation pattern.

Supplemental Figure 15

Input (ng)	Protocol	Reads	Expected Coverage	% Mapped	% Dups	% Usable
0.1	EM-seq (low)	811M	26	99.8%	84.5%	10.8%
0.5	EM-seq (low)	811M	26	99.9%	58.7%	34.4%
1	EM-seq (low)	811M	26	99.9%	42.6%	49.8%
10	Bisulfite (std)	809M	26	98.2%	49.9%	38.9%
	EM-seq (low)	811M	26	100.0%	13.7%	77.0%
	EM-seq (std)	811M	26	99.8%	13.7%	75.3%

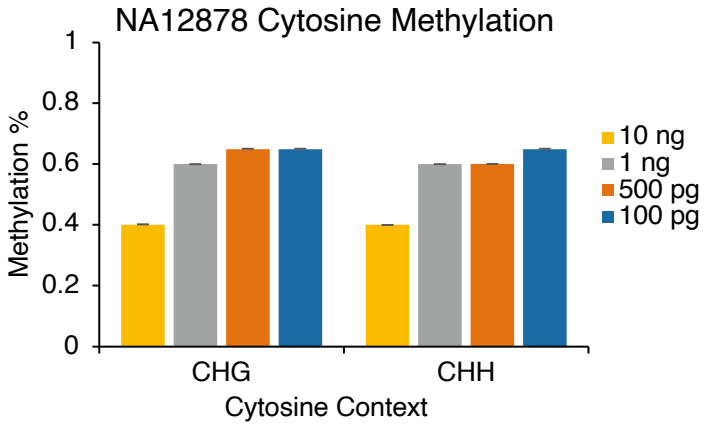


Sequencing Metrics for Low Input EM-seq Libraries

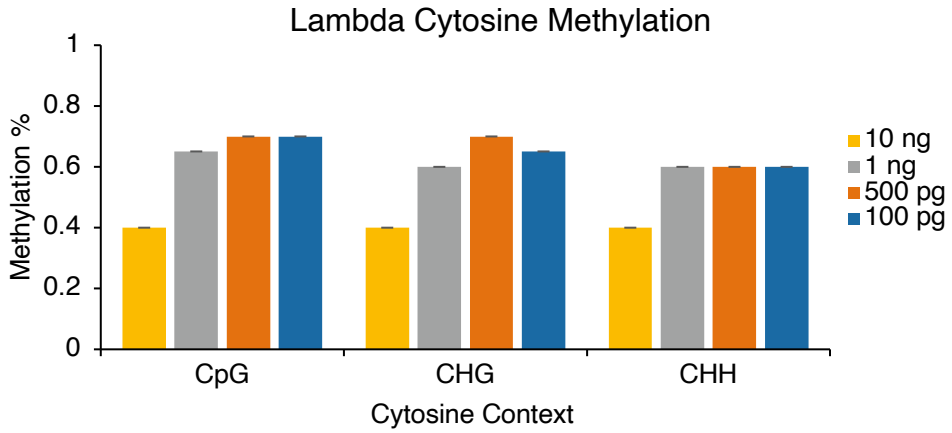
EM-seq libraries were made using 100 pg, 500 pg, 1 ng and 10 ng of NA12878 genomic DNA. Libraries were sequenced on an Illumina NovaSeq 6000, replicates combined, and 810 million total reads per library were used for analysis. Sequencing and alignment metrics for EM-seq libraries are shown. For the 10 ng inputs, the low input EM-seq method metrics were compared to libraries made using the standard EM-seq method and sodium bisulfite conversion. Metrics were calculated using bwa-meth, SAMtools, and Picard. Theoretical coverage is calculated using the number of bases sequenced/total bases in the GRCh38 reference. Percent mapped refers to reads aligned to the reference genome (grch38+controls), Percent Dups refers to reads marked as duplicate by Picard MarkDuplicates, Percent Usable refers to the set of Proper-pair, MapQ 10+, Primary, non-Duplicate reads used in methylation calling (SAMtools view -F 0xF00 -q 10). Effective Coverage is the Percent Usable × Theoretical coverage.

Supplemental Figure 16

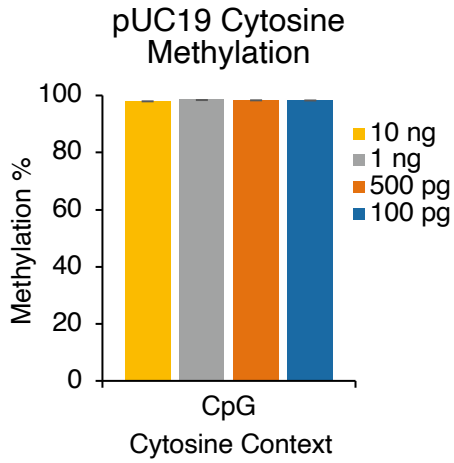
A

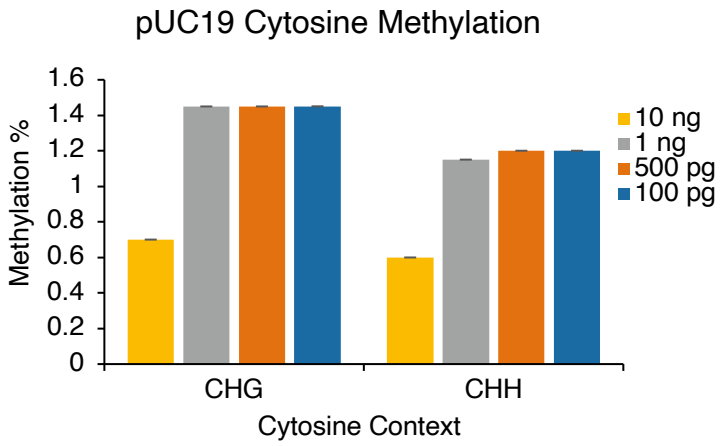


B



C

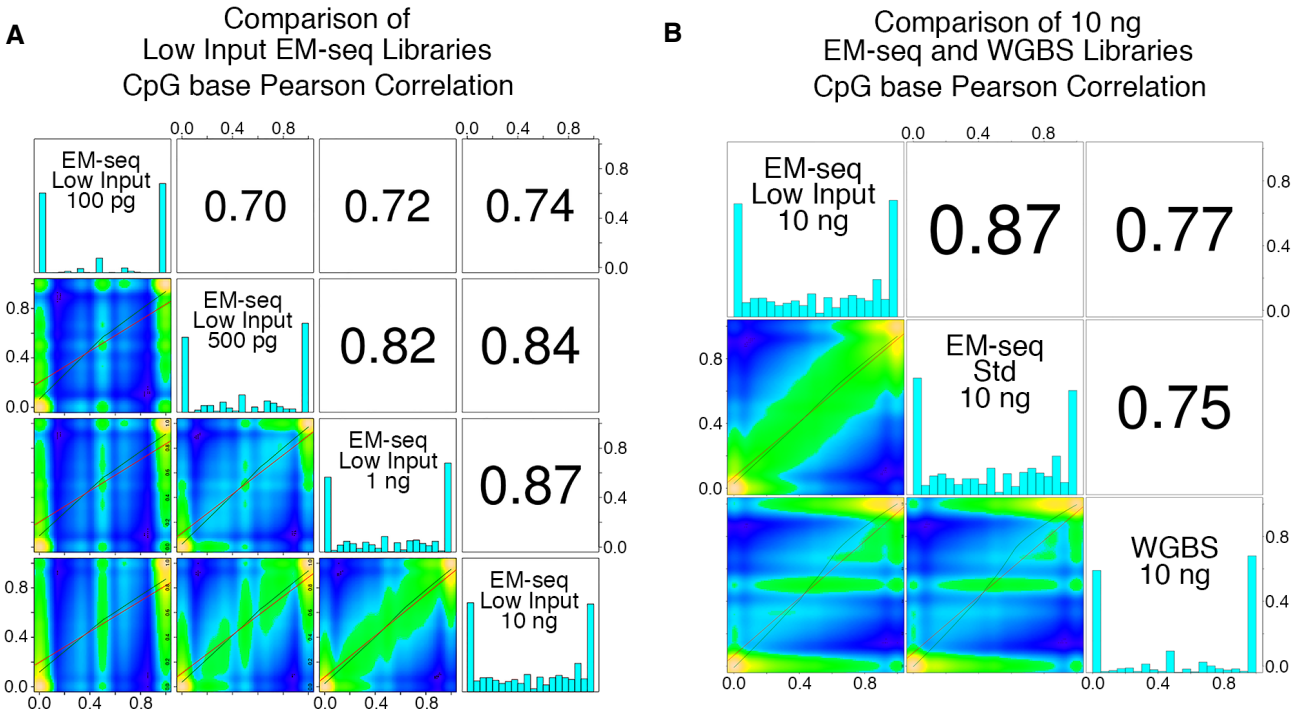




Low Input EM-seq Libraries, Cytosine Methylation of Lambda and pUC19 internal controls

(A) NA12878 DNA cytosine methylation in CHG and CHH contexts. (B) Methylation of cytosines in the unmethylated lambda controls. EM-seq has <1% methylation in CpG, CHG and CHH contexts. (C) Methylation of cytosines in the CpG methylated pUC19 control. EM-seq and WGBS have approximately 97% methylation in the CpG context and around 1% in the CHG and CHH contexts.

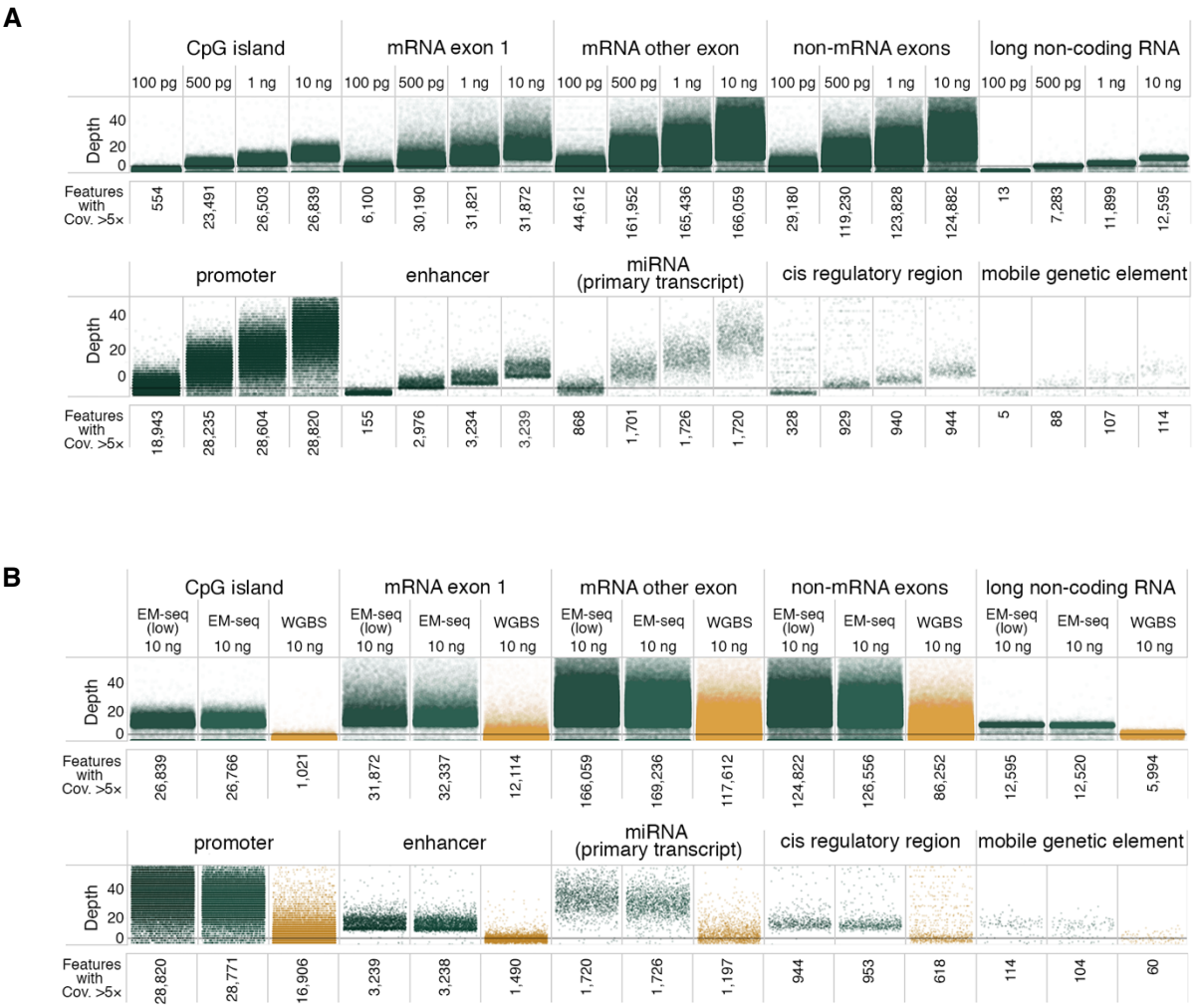
Supplemental Figure 17



Correlations between lower DNA input EM-seq Libraries

Correlations were plotted using methylKit for EM-seq libraries made using NA12878 genomic DNA. (A) MethylKit correlations for 100 pg, 500 pg, 1 ng, and 10 ng NA12878 EM-seq libraries at 1× minimum coverage. Correlations remain strong between libraries but are reduced when considering the 100 pg DNA input. (B) Correlations for 10 ng NA12878 DNA libraries. Data was from low input EM-seq, standard EM-seq and WGBS libraries. The correlation between the two EM-seq methods is higher than for those observed using WGBS.

Supplemental Figure 18

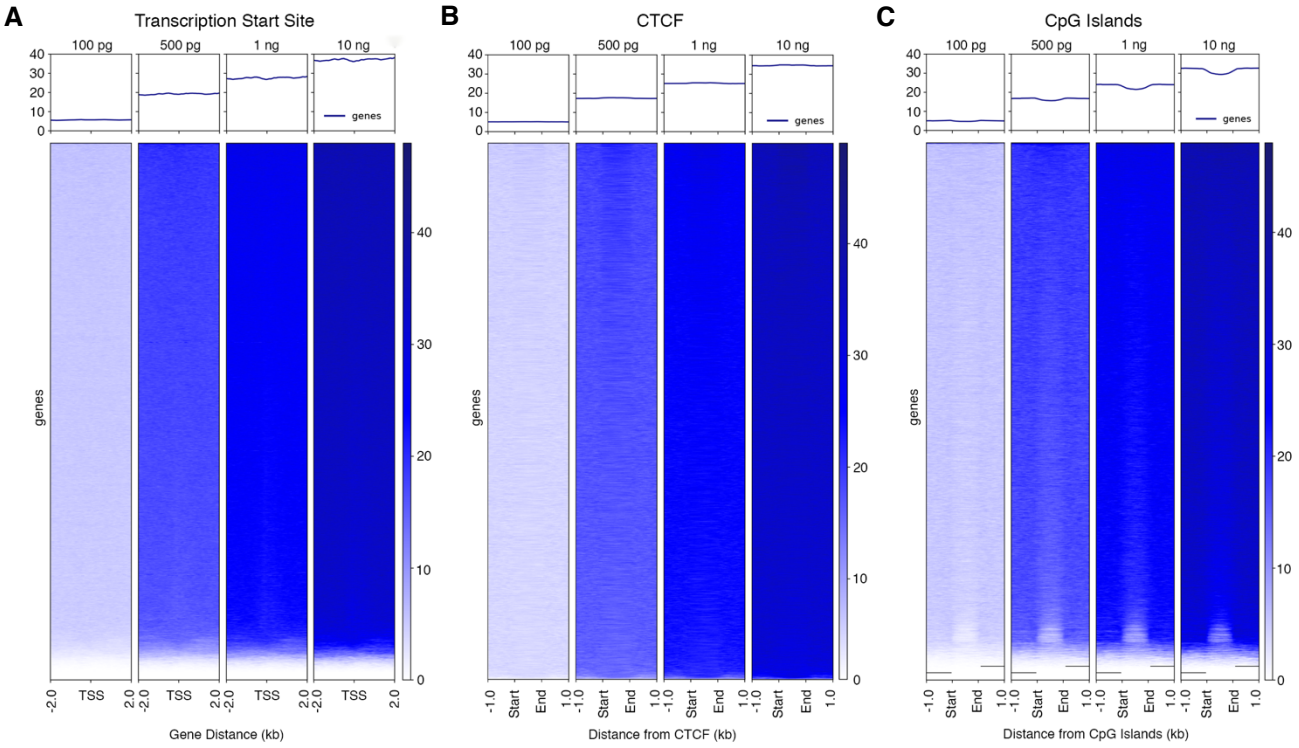


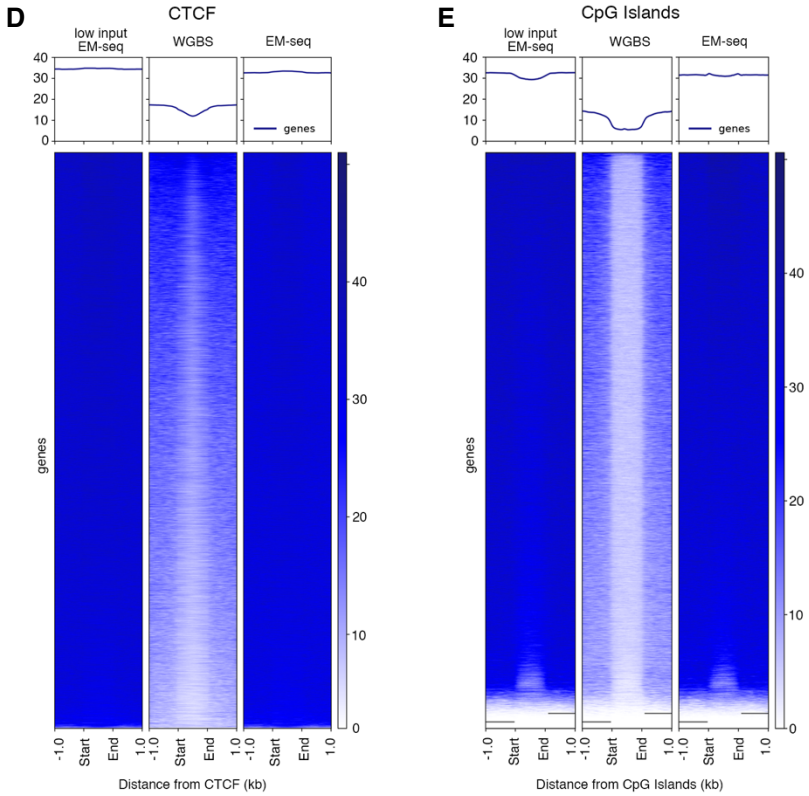
Low Input EM-seq determines methylation status of cytosines at key genomic features

Two technical replicates were combined before randomly selecting reads from each library to produce comparable sets of 810 million total human reads. These were analyzed for each library and annotated using featureCounts. (A) The number of genomic features identified at a coverage depth of >5× is similar for 500 pg, 1 ng, and 10 ng DNA inputs. The 100 pg input library has fewer genomic features identified (likely due to lower library complexity at this extreme input amount) (B) Comparison of genomic

features identified from 10 ng NA12878 DNA libraries. Data was from low input EM-seq, standard EM-seq and WGBS libraries. A wide range of genomic features at >5× coverage were identified. For EM-seq, the same number of genomic features were identified regardless of low input or standard library preparation method. The number of genomic features in each of the categories was less for WGBS libraries compared to EM-seq libraries.

Supplemental Figure 19

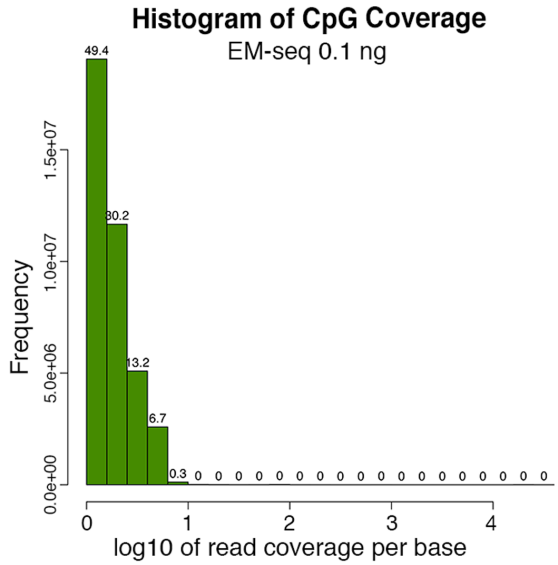


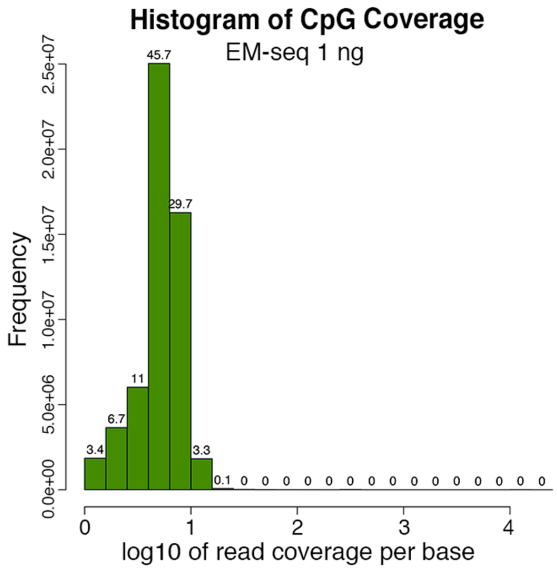


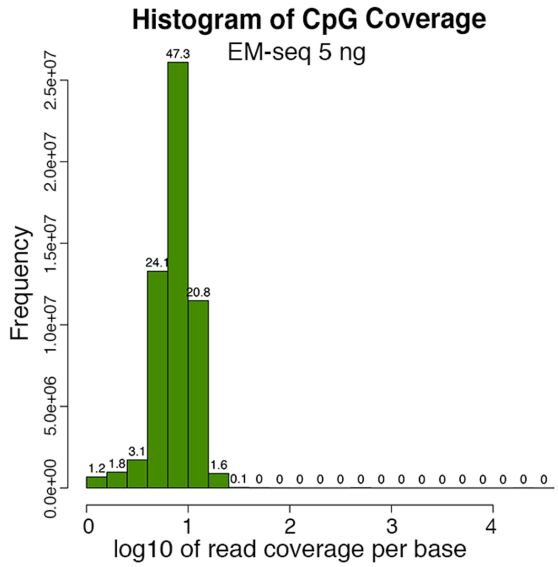
EM-seq provides greater coverage of CpGs over genomic features

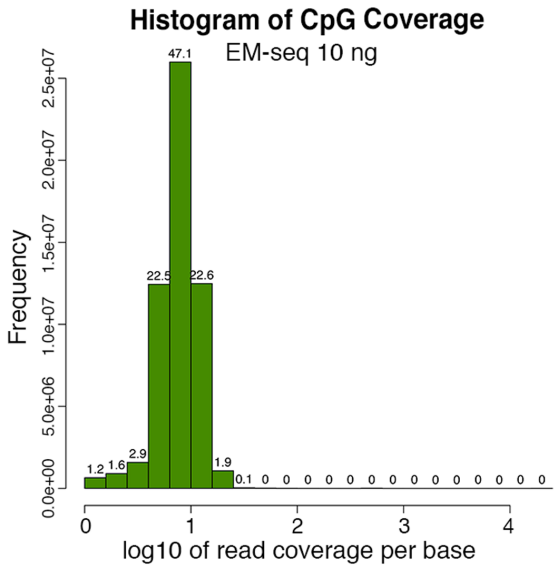
deepTools heat maps representing CpG coverage across genomic features including transcription start sites (TSS), CTCF sites, and CpG islands. (A-C) Heat maps for 100 pg, 500 pg, 1 ng and 10 ng DNA inputs showing coverage around TSS, CTCF and CpG islands. Coverage is even across the features at all inputs. (D, E) Heat maps for 10 ng DNA inputs comparing coverage of genomic features using libraries prepared using low input EM-seq, standard EM-seq and WGBS. Data from low input and standard EM-seq prepped libraries show similar even coverage. WGBS does not have the same evenness of coverage.

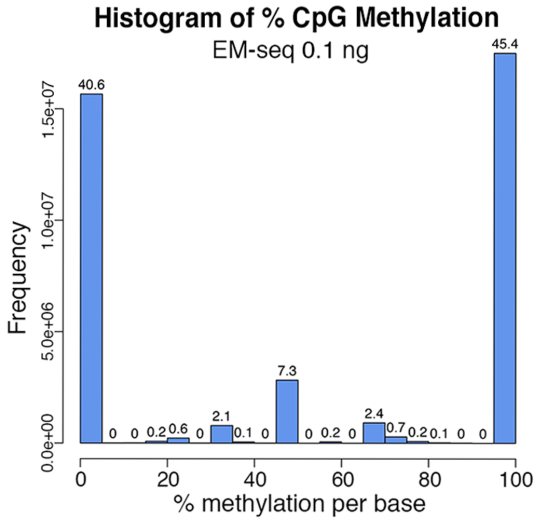
Supplemental Figure 20

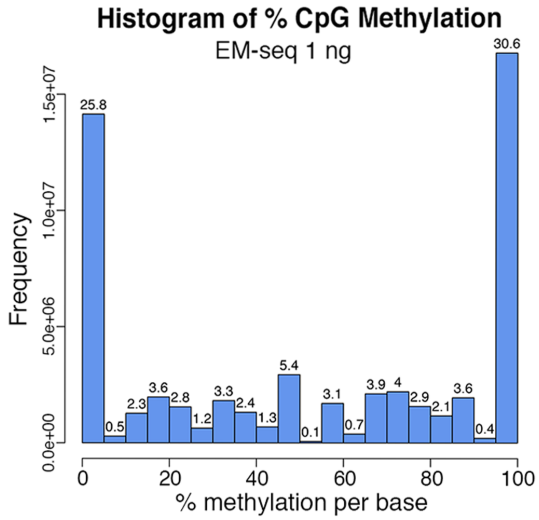


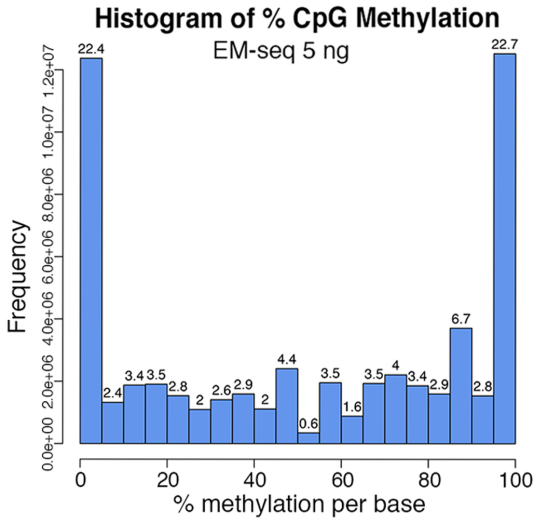


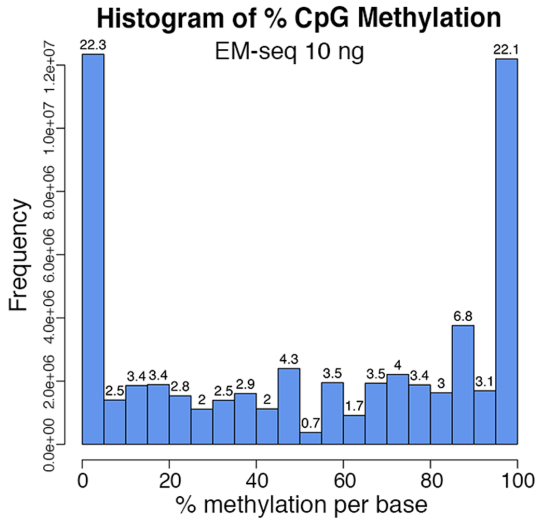








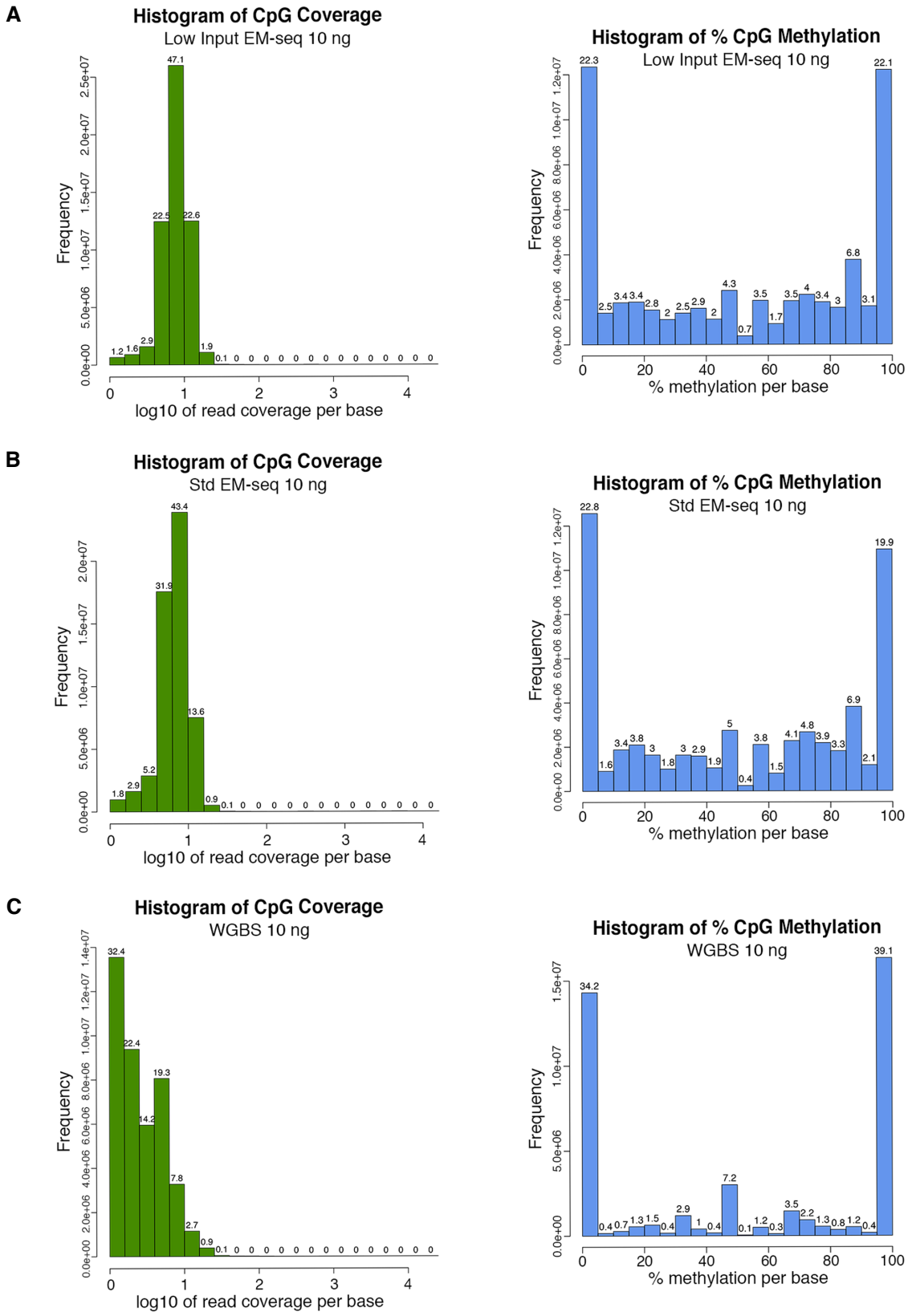




CpG coverage and methylation for low input libraries

The coverage and methylation status of CpGs were extracted using MethylDackel and plotting using methylKit for 100 pg, 500 pg, 1 ng and 10 ng NA12878 DNA inputs. There is deeper coverage for higher DNA inputs. Methylation profiles are as expected, with highest representation of CpG methylation at 0% and 100%.

Supplemental Figure 21



Comparison of WGBS, standard EM-seq and low input EM-seq 10 ng DNA input library types

The coverage and methylation status of CpGs were extracted from MethylDackel output files using methylKit for 10 ng NA12878 genomic DNA inputs. Libraries were made using (A) Low input EM-seq methods (B) Standard EM-seq method and (C) WGBS. Methylation profiles and coverage were similar between the two EM-seq methods with highest representation of CpG methylation at 0% and 100%.

Supplemental Tables

Supplemental Table 1

Pre-nucleotide substrate	Variable Nucleotide	Oligonucleotide Sequence
A	C	ATAAGAATAGAATGAAACGTGAAATGAATATGAAATGAATAGTA
G	C	ATAAGAATAGAATGAAGCGTGAAATGAATATGAAATGAATAGTA
T	C	ATAAGAATAGAATGAATCGTGAAATGAATATGAAATGAATAGTA
U	C	ATAAGAATAGAATGAAUCGTGAAATGAATATGAAATGAATAGTA
5mC	C	ATAAGAATAGAATGAA/5mC/CGTGAAATGAATATGAAATGAATAGTA
A	5mC	ATAAGAATAGAATGAAA/5mC/GTGAAATGAATATGAAATGAATAGTA
G	5mC	ATAAGAATAGAATGAAG/5mC/GTGAAATGAATATGAAATGAATAGTA
T	5mC	ATAAGAATAGAATGAAT/5mC/GTGAAATGAATATGAAATGAATAGTA
U	5mC	ATAAGAATAGAATGAAU/5mC/GTGAAATGAATATGAAATGAATAGTA
C	5mC	ATAAGAATAGAATGAAC/5mC/GTGAAATGAATATGAAATGAATAGTA

Quantitative analysis of APOBEC3A: site preference specificity oligonucleotides

Oligonucleotides used in Supplemental Figure 3A and 3B. For quantitative analysis of APOBEC3A site preference, the same core 44 bp oligonucleotide sequence for investigating the APOBEC3A substrate specificity was used. The base preceding cytosine and 5mC is underlined in the table.

Supplemental Table 2

Variable Nucleotide	Oligonucleotide sequence
C	ATAAGAATAGAATGAATCGTGAAATGAATATGAAATGAATAGTA
5mC	ATAAGAATAGAATGAAT/5mC/GTGAAATGAATATGAAATGAATAGTA
5hmC	ATAAGAATAGAATGAAT/5hmC/GTGAAATGAATATGAAATGAATAGTA
5gmC	ATAAGAATAGAATGAAT/5gmC/GTGAAATGAATATGAAATGAATAGTA
5fC	ATAAGAATAGAATGAAT/5fC/GTGAAATGAATATGAAATGAATAGTA
5caC	ATAAGAATAGAATGAAT/5caC/GTGAAATGAATATGAAATGAATAGTA

Quantitative analysis of APOBEC3A: substrate specificity oligonucleotides

Oligonucleotides used in Supplemental Figure 3C. For quantitative analysis of APOBEC3A substrate specificity, 44 bp oligonucleotides were used. These contained the same core sequence, but varied at one single position with either C, 5mC, 5hmC, 5gmC, 5fC or 5caC.

Supplemental Table 3

Oligonucleotide Modification	APOBEC3A survival in a reaction (half life (h))
No substrate	6
5mC	6
5hmC	6
5gmC	6
5fC	1
5caC	2

APOBEC3A activity half life

Activity of APOBEC3A on different single stranded oligo substrates with a single cytosine, 5mC, 5hmC, 5gmC, 5fC or 5caC was measured over time to determine the half-life of APOBEC3A.

Supplemental Table 4

Forward primer (5' to 3')	Chromosome location (Fwd)	Reverse primer (5' to 3')	Chromosome location (Rev)	Amplicon size, bp
GTAAAATGTTTTGGGTAAAAGTTATATGAATG	Chr1-36100629	TAAAATAATCCTAACTTCTAACCTCTCTTTTC	Chr1-36097684	2945
TGATGAGCGGTAGTTAGGTGAAAG	Chr12-110944577	AATAAATACAATCCCAAAACTATCAAATCTTC	Chr12-110942370	2207
AAAACGGAAATAGGGGTAATGATAG	Chr12-110944120	AATAAATACAATCCCAAAACTATCAAATCTTC	Chr12-110942370	1750
TTTTTTATGATTTGTTGTGATTTGGTTATTAAG	Chr12-110943551	AATAAATACAATCCCAAAACTATCAAATCTTC	Chr12-110942370	1181
GAGTATGGTAAAGTTGAAAATATGTAGATAAG	Chr12-110942913	AATAAATACAATCCCAAAACTATCAAATCTTC	Chr12-110942370	543

DNA integrity assessment

Sequences of primers used for amplicon test designed using GRCh37 as the reference genome.

Supplemental Table 5

Sample Description	From [bp]	To [bp]	Average Size [bp]	Region Molarity [nmol/l]	PCR cycles
0.1 ng-1	100	1000	397	18.5	14
0.1 ng-2	100	1000	382	11.3	14
0.5 ng-1	100	1000	389	13	12
0.5 ng-2	100	1000	393	14.1	12
1 ng-1	100	1000	397	12.5	11
1 ng-2	100	1000	385	11.9	11
10 ng-1	100	1000	384	21.8	8
10 ng-2	100	1000	404	24.2	8

Low input library yields

NA12878 low input libraries. The average size of the libraries, PCR yields and PCR cycles are indicated**.**

Supplemental Table 6

A

Enzymes Used	% 5mC/Total C	STDEV
TET2 + T4-BGT	0.023%	0.00002
TET2	0.018%	0.00001
T4-BGT	2.660%	0.00009
APOBEC3A	0.088%	0.00003
TET2 + T4-BGT + APOBEC3A	0.000%	0.00000
Unconverted DNA	2.675%	0.00019

B

Enzymes Used	% dU/(dC+dU)	STDEV
TET2 + T4-BGT	0.000%	0.00000
TET2	0.000%	0.00000
T4-BGT	0.000%	0.00000
APOBEC3A	97.822%	0.00134
TET2 + T4-BGT + APOBEC3A	97.897%	0.00038
Unconverted DNA	0.000%	0.00000

C

Reaction	Name	dcaC	dC	dhmC	dU	dhmU	dmC	dfC	dT	Total C Mod
TET2+T4-BGT	CP_01	32.18	16879.16	1.14	0.00	160.42	3.46	0.00	20870.13	36.79
TET2+T4-BGT	CP_01	28.74	15667.93	0.99	0.00	148.87	3.25	0.00	19403.74	32.99
TET2+T4-BGT	CP_02	13.10	8978.19	0.42	0.00	70.03	2.28	0.00	10453.93	15.80
TET2+T4-BGT	CP_02	14.96	8454.20	0.44	0.00	70.29	2.00	0.00	10149.15	17.40
TET2	CP_03	357.03	12724.17	90.39	0.00	140.66	2.48	56.20	16280.33	506.11
TET2	CP_03	344.88	12808.22	89.17	0.00	145.25	2.35	52.46	16007.79	488.86
TET2	CP_04	324.89	11916.66	85.04	0.00	137.32	2.12	53.02	14592.71	465.06
TET2	CP_04	314.85	11666.93	81.04	0.00	131.10	2.13	50.91	14427.04	448.93
T4-BGT	CP_05	0.00	11316.63	0.00	0.00	0.00	302.55	0.00	14609.16	302.55
T4-BGT	CP_05	0.00	11213.19	0.00	0.00	0.00	298.00	0.00	13994.27	298.00
T4-BGT	CP_06	0.00	10630.01	0.00	0.00	0.00	282.62	0.00	13686.18	282.62
T4-BGT	CP_06	0.00	10546.38	0.00	0.00	0.00	279.67	0.00	13725.83	279.67
APOBEC	CP_07	0.00	93.12	0.00	4516.74	0.00	4.00	0.00	6314.92	4.00
APOBEC	CP_07	0.00	97.61	0.00	4510.71	0.00	3.92	0.00	6257.50	3.92
APOBEC	CP_08	0.00	95.81	0.00	4140.91	0.00	3.81	0.00	5472.26	3.81
APOBEC	CP_08	0.00	91.56	0.00	3868.68	0.00	3.62	0.00	5220.60	3.62
TET2+T4-BGT+APOBEC	CP_09	5.81	98.59	0.00	4519.24	32.06	0.00	0.00	6243.56	5.81
TET2+T4-BGT+APOBEC	CP_09	4.22	98.60	0.00	4513.06	27.41	0.00	0.00	5923.75	4.22
TET2+T4-BGT+APOBEC	CP_10	5.38	88.21	0.00	4175.77	27.54	0.00	0.00	5530.32	5.38
TET2+T4-BGT+APOBEC	CP_10	4.56	87.56	0.00	4139.12	28.41	0.00	0.00	5567.22	4.56
Unconverted	CP_11	0.00	8851.73	0.00	0.00	0.00	238.75	0.00	11743.82	238.75
Unconverted	CP_11	0.00	8889.26	0.00	0.00	0.00	236.55	0.00	11556.56	236.55
Unconverted	CP_12	0.00	8253.42	0.00	0.00	0.00	219.28	0.00	10850.20	219.28
Unconverted	CP_12	0.00	8217.99	0.00	0.00	0.00	220.59	0.00	11047.84	220.59

LCMS after Enzymatic Treatments

The amount of 5mC and uracil after different enzymatic treatments was investigated. The activity of individual TET2, T4-BGT and APOBEC3A enzymes or the combinations indicated on 200 ng of NA12878 DNA were analyzed using LCMS. Summaries indicating the percentage 5mC (Supplemental Table 6A) and the percentage uracil (Supplemental Table 6B) are shown. The amount of 5mC is reduced by addition of

TET2 and by the combined addition of TET2 and T4-BGT. APOBEC3A is very active as observed by the formation of uracils. There was 0.9% residual 5mC available for deamination by APOBEC3A. This residual 5mC is based upon the percentage of 5mC remaining after the TET2 + APOBEC3A reaction divided by the percentage of unconveted 5mC. Each reaction was in duplicate, and each duplicate was assayed twice using LC-MS. Supplemental Table 6C shows the complete data from this experiment.

Supplemental Table 7

Reference genome sequences

>chr1 AC:CM000663.2 gi:568336023 LN:248956422 rl:Chromosome M5:6aef897c3d6ff0c78aff06ac189178dd AS:GRCh38
>chr2 AC:CM000664.2 gi:568336022 LN:242193529 rl:Chromosome M5:f98db672eb0993dcfdabafe2a882905c AS:GRCh38
>chr3 AC:CM000665.2 gi:568336021 LN:198295559 rl:Chromosome M5:76635a41ea913a405ded820447d067b0 AS:GRCh38
>chr4 AC:CM000666.2 gi:568336020 LN:190214555 rl:Chromosome M5:3210fecf1eb92d5489da4346b3fddc6e AS:GRCh38
>chr5 AC:CM000667.2 gi:568336019 LN:181538259 rl:Chromosome M5:a811b3dc9fe66af729dc0dddf7fa4f13 AS:GRCh38 hm:47309185-49591369
>chr6 AC:CM000668.2 gi:568336018 LN:170805979 rl:Chromosome M5:5691468a67c7e7a7b5f2a3a683792c29 AS:GRCh38
>chr7 AC:CM000669.2 gi:568336017 LN:159345973 rl:Chromosome M5:cc044cc2256a1141212660fb07b6171e AS:GRCh38
>chr8 AC:CM000670.2 gi:568336016 LN:145138636 rl:Chromosome M5:c67955b5f7815a9a1edfaa15893d3616 AS:GRCh38
>chr9 AC:CM000671.2 gi:568336015 LN:138394717 rl:Chromosome M5:6c198acf68b5af7b9d676dfdd531b5de AS:GRCh38
>chr10 AC:CM000672.2 gi:568336014 LN:133797422 rl:Chromosome M5:c0eeee7acfdaf31b770a509bdaa6e51a AS:GRCh38
>chr11 AC:CM000673.2 gi:568336013 LN:135086622 rl:Chromosome M5:1511375dc2dd1b633af8cf439ae90cec AS:GRCh38
>chr12 AC:CM000674.2 gi:568336012 LN:133275309 rl:Chromosome M5:96e414eace405d8c27a6d35ba19df56f AS:GRCh38
>chr13 AC:CM000675.2 gi:568336011 LN:114364328 rl:Chromosome M5:a5437debe2ef9c9ef8f3ea2874ae1d82 AS:GRCh38
>chr14 AC:CM000676.2 gi:568336010 LN:107043718 rl:Chromosome M5:e0f0eecc3bcab6178c62b6211565c807 AS:GRCh38 hm:multiple
>chr15 AC:CM000677.2 gi:568336009 LN:101991189 rl:Chromosome M5:f036bd11158407596ca6bf3581454706 AS:GRCh38
>chr16 AC:CM000678.2 gi:568336008 LN:90338345 rl:Chromosome M5:db2d37c8b7d019caaf2dd64ba3a6f33a AS:GRCh38
>chr17 AC:CM000679.2 gi:568336007 LN:83257441 rl:Chromosome M5:f9a0fb01553adb183568e3eb9d8626db AS:GRCh38
>chr18 AC:CM000680.2 gi:568336006 LN:80373285 rl:Chromosome M5:11eeaa801f6b0e2e36a1138616b8ee9a AS:GRCh38
>chr19 AC:CM000681.2 gi:568336005 LN:58617616 rl:Chromosome M5:85f9f4fc152c58cb7913c06d6b98573a AS:GRCh38 hm:multiple
>chr20 AC:CM000682.2 gi:568336004 LN:64444167 rl:Chromosome M5:b18e6c531b0bd70e949a7fc20859cb01 AS:GRCh38
>chr21 AC:CM000683.2 gi:568336003 LN:46709983 rl:Chromosome M5:974dc7aec0b755b19f031418fdedf293 AS:GRCh38 hm:multiple
>chr22 AC:CM000684.2 gi:568336002 LN:50818468 rl:Chromosome M5:ac37ec46683600f808cdd41eac1d55cd AS:GRCh38 hm:multiple
>chrX AC:CM000685.2 gi:568336001 LN:156040895 rl:Chromosome M5:2b3a55ff7f58eb308420c8a9b11cac50 AS:GRCh38
>chrY AC:CM000686.2 gi:568336000 LN:57227415 rl:Chromosome M5:ce3e31103314a704255f3cd90369ecce AS:GRCh38 hm:10001-2781479,56887903-57217415
>chrM AC:J01415.2 gi:113200490 LN:16569 rl:Mitochondrion M5:c68f52674c9fb33aef52dcf399755519 AS:GRCh38 tp:circular
>chr1_KI270706v1_random AC:KI270706.1 gi:568335410 LN:175055 rg:chr1 rl:unlocalized M5:62def1a794b3e18192863d187af956e6 AS:GRCh38
>chr1_KI270707v1_random AC:KI270707.1 gi:568335409 LN:32032 rg:chr1 rl:unlocalized M5:78135804eb15220565483b7cdd02f3be AS:GRCh38
>chr1_KI270708v1_random AC:KI270708.1 gi:568335408 LN:127682 rg:chr1 rl:unlocalized M5:1e95e047b98ed92148dd84d6c037158c AS:GRCh38

>chr1_KI270709v1_random AC:KI270709.1 gi:568335407 LN:66860 rg:chr1 rl:unlocalized M5:4e2db2933ea96aee8dab54af60ecb37d AS:GRCh38
>chr1_KI270710v1_random AC:KI270710.1 gi:568335406 LN:40176 rg:chr1 rl:unlocalized M5:9949f776680c6214512ee738ac5da289 AS:GRCh38
>chr1_KI270711v1_random AC:KI270711.1 gi:568335405 LN:42210 rg:chr1 rl:unlocalized M5:af383f98cf4492c1f1c4e750c26cbb40 AS:GRCh38
>chr1_KI270712v1_random AC:KI270712.1 gi:568335404 LN:176043 rg:chr1 rl:unlocalized M5:c38a0fecae6a1838a405406f724d6838 AS:GRCh38
>chr1_KI270713v1_random AC:KI270713.1 gi:568335403 LN:40745 rg:chr1 rl:unlocalized M5:cb78d48cc0adbc58822a1c6fe89e3569 AS:GRCh38
>chr1_KI270714v1_random AC:KI270714.1 gi:568335402 LN:41717 rg:chr1 rl:unlocalized M5:42f7a452b8b769d051ad738ee9f00631 AS:GRCh38
>chr2_KI270715v1_random AC:KI270715.1 gi:568335401 LN:161471 rg:chr2 rl:unlocalized M5:b65a8af1d7bbb7f3c77eea85423452bb AS:GRCh38
>chr2_KI270716v1_random AC:KI270716.1 gi:568335400 LN:153799 rg:chr2 rl:unlocalized M5:2828e63b8edc5e845bf48e75fbad2926 AS:GRCh38
>chr3_GL000221v1_random AC:GL000221.1 gi:224183270 LN:155397 rg:chr3 rl:unlocalized M5:3238fb74ea87ae857f9c7508d315babb AS:GRCh38
>chr4_GL000008v2_random AC:GL000008.2 gi:568335399 LN:209709 rg:chr4 rl:unlocalized M5:a999388c587908f80406444cebe80ba3 AS:GRCh38
>chr5_GL000208v1_random AC:GL000208.1 gi:224183050 LN:92689 rg:chr5 rl:unlocalized M5:aa81be49bf3fe63a79bdc6a6f279abf6 AS:GRCh38
>chr9_KI270717v1_random AC:KI270717.1 gi:568335398 LN:40062 rg:chr9 rl:unlocalized M5:796773a1ee67c988b4de887addbed9e7 AS:GRCh38
>chr9_KI270718v1_random AC:KI270718.1 gi:568335397 LN:38054 rg:chr9 rl:unlocalized M5:b0c463c8efa8d64442b48e936368dad5 AS:GRCh38
>chr9_KI270719v1_random AC:KI270719.1 gi:568335396 LN:176845 rg:chr9 rl:unlocalized M5:cd5e932cfc4c74d05bb64e2126873a3a AS:GRCh38
>chr9_KI270720v1_random AC:KI270720.1 gi:568335395 LN:39050 rg:chr9 rl:unlocalized M5:8c2683400a4aeeb40abff96652b9b127 AS:GRCh38
>chr11_KI270721v1_random AC:KI270721.1 gi:568335394 LN:100316 rg:chr11 rl:unlocalized M5:9654b5d3f36845bb9d19a6dbd15d2f22 AS:GRCh38
>chr14_GL000009v2_random AC:GL000009.2 gi:568335393 LN:201709 rg:chr14 rl:unlocalized M5:862f555045546733591ff7ab15bcecbe AS:GRCh38
>chr14_GL000225v1_random AC:GL000225.1 gi:224183274 LN:211173 rg:chr14 rl:unlocalized M5:63945c3e6962f28ffd469719a747e73c AS:GRCh38
>chr14_KI270722v1_random AC:KI270722.1 gi:568335392 LN:194050 rg:chr14 rl:unlocalized M5:51f46c9093929e6edc3b4dfd50d803fc AS:GRCh38
>chr14_GL000194v1_random AC:GL000194.1 gi:224183213 LN:191469 rg:chr14 rl:unlocalized M5:6ac8f815bf8e845bb3031b73f812c012 AS:GRCh38
>chr14_KI270723v1_random AC:KI270723.1 gi:568335391 LN:38115 rg:chr14 rl:unlocalized M5:74a4b480675592095fb0c577c515b5df AS:GRCh38

>chr14_KI270724v1_random AC:KI270724.1 gi:568335390 LN:39555 rg:chr14 rl:unlocalized M5:c3fcb15dddf45f91ef7d94e2623ce13b AS:GRCh38
>chr14_KI270725v1_random AC:KI270725.1 gi:568335389 LN:172810 rg:chr14 rl:unlocalized M5:edc6402e58396b90b8738a5e37bf773d AS:GRCh38
>chr14_KI270726v1_random AC:KI270726.1 gi:568335388 LN:43739 rg:chr14 rl:unlocalized M5:fbe54a3197e2b469ccb2f4b161cfbe86 AS:GRCh38
>chr15_KI270727v1_random AC:KI270727.1 gi:568335387 LN:448248 rg:chr15 rl:unlocalized M5:84fe18a7bf03f3b7fc76cbac8eb583f1 AS:GRCh38
>chr16_KI270728v1_random AC:KI270728.1 gi:568335386 LN:1872759 rg:chr16 rl:unlocalized M5:369ff74cf36683b3066a2ca929d9c40d AS:GRCh38
>chr17_GL000205v2_random AC:GL000205.2 gi:568335385 LN:185591 rg:chr17 rl:unlocalized M5:458e71cd53dd1df4083dc7983a6c82c4 AS:GRCh38
>chr17_KI270729v1_random AC:KI270729.1 gi:568335384 LN:280839 rg:chr17 rl:unlocalized M5:2756f6ee4f5780acce31e995443508b6 AS:GRCh38
>chr17_KI270730v1_random AC:KI270730.1 gi:568335383 LN:112551 rg:chr17 rl:unlocalized M5:48f98ede8e28a06d241ab2e946c15e07 AS:GRCh38
>chr22_KI270731v1_random AC:KI270731.1 gi:568335382 LN:150754 rg:chr22 rl:unlocalized M5:8176d9a20401e8d9f01b7ca8b51d9c08 AS:GRCh38
>chr22_KI270732v1_random AC:KI270732.1 gi:568335381 LN:41543 rg:chr22 rl:unlocalized M5:d837bab5e416450df6e1038ae6cd0817 AS:GRCh38
>chr22_KI270733v1_random AC:KI270733.1 gi:568335380 LN:179772 rg:chr22 rl:unlocalized M5:f1fa05d48bb0c1f87237a28b66f0be0b AS:GRCh38
>chr22_KI270734v1_random AC:KI270734.1 gi:568335379 LN:165050 rg:chr22 rl:unlocalized M5:1d17410ae2569c758e6dd51616412d32 AS:GRCh38
>chr22_KI270735v1_random AC:KI270735.1 gi:568335378 LN:42811 rg:chr22 rl:unlocalized M5:eb6b07b73dd9a47252098ed3d9fb78b8 AS:GRCh38
>chr22_KI270736v1_random AC:KI270736.1 gi:568335377 LN:181920 rg:chr22 rl:unlocalized M5:2ff189f33cfa52f321accddf648c5616 AS:GRCh38
>chr22_KI270737v1_random AC:KI270737.1 gi:568335376 LN:103838 rg:chr22 rl:unlocalized M5:2ea8bc113a8193d1d700b584b2c5f42a AS:GRCh38
>chr22_KI270738v1_random AC:KI270738.1 gi:568335375 LN:99375 rg:chr22 rl:unlocalized M5:854ec525c7b6a79e7268f515b6a9877c AS:GRCh38
>chr22_KI270739v1_random AC:KI270739.1 gi:568335374 LN:73985 rg:chr22 rl:unlocalized M5:760fbd73515fedcc9f37737c4a722d6a AS:GRCh38
>chrY_KI270740v1_random AC:KI270740.1 gi:568335373 LN:37240 rg:chrY rl:unlocalized M5:69e42252aead509bf56f1ea6fda91405 AS:GRCh38
>chrUn_KI270302v1 AC:KI270302.1 gi:568335372 LN:2274 rl:unplaced M5:ee6dff38036f7d03478c707176431 96e AS:GRCh38
>chrUn_KI270304v1 AC:KI270304.1 gi:568335371 LN:2165 rl:unplaced M5:9423c1b46a48aa6331a77ab5c702ac9d AS:GRCh38
>chrUn_KI270303v1 AC:KI270303.1 gi:568335370 LN:1942 rl:unplaced M5:2cb746c78e0faa11e628603a4bc9bd58 AS:GRCh38
>chrUn_KI270305v1 AC:KI270305.1 gi:568335369 LN:1472 rl:unplaced M5:f9acea3395b6992cf3418b6689b98cf9 AS:GRCh38
>chrUn_KI270322v1 AC:KI270322.1 gi:568335368 LN:21476 rl:unplaced M5:7d459255d1c54369e3b64e719061a5a5 AS:GRCh38
>chrUn_KI270320v1 AC:KI270320.1 gi:568335367 LN:4416 rl:unplaced M5:d898b9c5a0118e76481bf5695272959e AS:GRCh38

>chrUn_KI270310v1 AC:KI270310.1 gi:568335366 LN:1201 rl:unplaced M5:af6cb123af7007793bac06485a2a20e9 AS:GRCh38
>chrUn_KI270316v1 AC:KI270316.1 gi:568335365 LN:1444 rl:unplaced M5:6adde7a9fe7bd6918f12d0f0924aa8ba AS:GRCh38
>chrUn_KI270315v1 AC:KI270315.1 gi:568335364 LN:2276 rl:unplaced M5:ecc43e822dc011fae1fcfd9981c46e9c AS:GRCh38
>chrUn_KI270312v1 AC:KI270312.1 gi:568335363 LN:998 rl:unplaced M5:26499f2fe4c65621fd8f4ecafbad31d7 AS:GRCh38
>chrUn_KI270311v1 AC:KI270311.1 gi:568335362 LN:12399 rl:unplaced M5:59594f9012d8ce21ed5d1119c051a2ba AS:GRCh38
>chrUn_KI270317v1 AC:KI270317.1 gi:568335361 LN:37690 rl:unplaced M5:cd4b1fda800f6ec9ea8001994dbf6499 AS:GRCh38
>chrUn_KI270412v1 AC:KI270412.1 gi:568335360 LN:1179 rl:unplaced M5:7bb9612f733fb7f098be79499d46350c AS:GRCh38
>chrUn_KI270411v1 AC:KI270411.1 gi:568335359 LN:2646 rl:unplaced M5:fc240322d91d43c04f349cc58fda3eca AS:GRCh38
>chrUn_KI270414v1 AC:KI270414.1 gi:568335358 LN:2489 rl:unplaced M5:753e02ef3b1c590e0e3376ad2ebb5836 AS:GRCh38
>chrUn_KI270419v1 AC:KI270419.1 gi:568335357 LN:1029 rl:unplaced M5:58455e7a788f0dc82034d1fb109f6f5c AS:GRCh38
>chrUn_KI270418v1 AC:KI270418.1 gi:568335356 LN:2145 rl:unplaced M5:1537ec12b9c58d137a2d4cb9db896bbc AS:GRCh38
>chrUn_KI270420v1 AC:KI270420.1 gi:568335355 LN:2321 rl:unplaced M5:bac954a897539c91982a7e3985a49910 AS:GRCh38
>chrUn_KI270424v1 AC:KI270424.1 gi:568335354 LN:2140 rl:unplaced M5:747c8f94f34d5de4ad289bc604708210 AS:GRCh38
>chrUn_KI270417v1 AC:KI270417.1 gi:568335353 LN:2043 rl:unplaced M5:cd26758fda713f9c96e51d541f50c2d0 AS:GRCh38
>chrUn_KI270422v1 AC:KI270422.1 gi:568335352 LN:1445 rl:unplaced M5:3fce80eb4c0554376b591699031feb56 AS:GRCh38
>chrUn_KI270423v1 AC:KI270423.1 gi:568335351 LN:981 rl:unplaced M5:bdf5a85c001731dccfb150db2bfe58ac AS:GRCh38
>chrUn_KI270425v1 AC:KI270425.1 gi:568335350 LN:1884 rl:unplaced M5:665a46879bbb48294b0cfa87b61e71f6 AS:GRCh38
>chrUn_KI270429v1 AC:KI270429.1 gi:568335349 LN:1361 rl:unplaced M5:ee8962dbef9396884f649e566b78bf06 AS:GRCh38
>chrUn_KI270442v1 AC:KI270442.1 gi:568335348 LN:392061 rl:unplaced M5:796289c4cda40e358991f9e672490015 AS:GRCh38
>chrUn_KI270466v1 AC:KI270466.1 gi:568335347 LN:1233 rl:unplaced M5:530b7033716a5d72dd544213c513fd12 AS:GRCh38
>chrUn_KI270465v1 AC:KI270465.1 gi:568335346 LN:1774 rl:unplaced M5:bb1b2323414425c46531b3c3d22ae00d AS:GRCh38
>chrUn_KI270467v1 AC:KI270467.1 gi:568335345 LN:3920 rl:unplaced M5:db34e0dc109a4afd499b5ec6aaae9754 AS:GRCh38
>chrUn_KI270435v1 AC:KI270435.1 gi:568335344 LN:92983 rl:unplaced M5:1655c75415b9c29e143a815f44286d05 AS:GRCh38
>chrUn_KI270438v1 AC:KI270438.1 gi:568335343 LN:112505 rl:unplaced M5:e765271939b854dd6826aa764e930c87 AS:GRCh38
>chrUn_KI270468v1 AC:KI270468.1 gi:568335342 LN:4055 rl:unplaced M5:0a603090f06108ed7aff75df0767b822 AS:GRCh38
>chrUn_KI270510v1 AC:KI270510.1 gi:568335341 LN:2415 rl:unplaced M5:cd7348b3b5d9d0dfef6aed2af75ce920 AS:GRCh38
>chrUn_KI270509v1 AC:KI270509.1 gi:568335340 LN:2318 rl:unplaced M5:1cdeb8c823d839e1d1735b5bc9a14856 AS:GRCh38
>chrUn_KI270518v1 AC:KI270518.1 gi:568335339 LN:2186 rl:unplaced M5:3fd898b62ca859f50fb8b83e7706352b AS:GRCh38
>chrUn_KI270508v1 AC:KI270508.1 gi:568335338 LN:1951 rl:unplaced M5:7d42a358d472b9cbdfdf30c8742473d0 AS:GRCh38
>chrUn_KI270516v1 AC:KI270516.1 gi:568335337 LN:1300 rl:unplaced M5:1cbaaafbbf016906a5bf5886f5a0ecb7 AS:GRCh38
>chrUn_KI270512v1 AC:KI270512.1 gi:568335336 LN:22689 rl:unplaced M5:ba1021c82d1230af856f59079e2f71b4 AS:GRCh38
>chrUn_KI270519v1 AC:KI270519.1 gi:568335335 LN:138126 rl:unplaced M5:8d754e9c9afd904fba0a2cd577fcc9a1 AS:GRCh38
>chrUn_KI270522v1 AC:KI270522.1 gi:568335334 LN:5674 rl:unplaced M5:070b4678e22501029c2e3297115216bc AS:GRCh38
>chrUn_KI270511v1 AC:KI270511.1 gi:568335333 LN:8127 rl:unplaced M5:907ca34a4a2a6673632ebaf513a4c1a4 AS:GRCh38
>chrUn_KI270515v1 AC:KI270515.1 gi:568335332 LN:6361 rl:unplaced M5:dd7527ee8e0bdb0a43ca0b2a5456c8c3 AS:GRCh38

>chrUn_KI270507v1 AC:KI270507.1 gi:568335331 LN:5353 rl:unplaced M5:311894d0a815eb07c5cac49da851cb4a AS:GRCh38
>chrUn_KI270517v1 AC:KI270517.1 gi:568335330 LN:3253 rl:unplaced M5:913440c38d95c618617ca69bb9296170 AS:GRCh38
>chrUn_KI270529v1 AC:KI270529.1 gi:568335329 LN:1899 rl:unplaced M5:4caf890f2586daab8e4b2e2db904f05f AS:GRCh38
>chrUn_KI270528v1 AC:KI270528.1 gi:568335328 LN:2983 rl:unplaced M5:d75c9235f0b8c449fc4352997c56b086 AS:GRCh38
>chrUn_KI270530v1 AC:KI270530.1 gi:568335327 LN:2168 rl:unplaced M5:04549369e1197c626669a10164613635 AS:GRCh38
>chrUn_KI270539v1 AC:KI270539.1 gi:568335326 LN:993 rl:unplaced M5:19e3a982e67eafef39c5a3e4163f1e17 AS:GRCh38
>chrUn_KI270538v1 AC:KI270538.1 gi:568335325 LN:91309 rl:unplaced M5:d60b72221cc7af871f2c757577e4c92a AS:GRCh38
>chrUn_KI270544v1 AC:KI270544.1 gi:568335324 LN:1202 rl:unplaced M5:e62a14b14467cdf48b5a236c66918f0f AS:GRCh38
>chrUn_KI270548v1 AC:KI270548.1 gi:568335323 LN:1599 rl:unplaced M5:866b0db8e9cf66208c2c064bd09ce0a2 AS:GRCh38
>chrUn_KI270583v1 AC:KI270583.1 gi:568335322 LN:1400 rl:unplaced M5:b127e2e6dbe358ff192b271b8c6ee690 AS:GRCh38
>chrUn_KI270587v1 AC:KI270587.1 gi:568335321 LN:2969 rl:unplaced M5:36be47659719f47b95caa51744aa8f70 AS:GRCh38
>chrUn_KI270580v1 AC:KI270580.1 gi:568335320 LN:1553 rl:unplaced M5:1df30dae0f605811d927dcea58e729fc AS:GRCh38
>chrUn_KI270581v1 AC:KI270581.1 gi:568335319 LN:7046 rl:unplaced M5:9f26945f9f9b3f865c9ebe953cbbc1a9 AS:GRCh38
>chrUn_KI270579v1 AC:KI270579.1 gi:568335318 LN:31033 rl:unplaced M5:fe62fb1964002717cc1b034630e89b1f AS:GRCh38
>chrUn_KI270589v1 AC:KI270589.1 gi:568335317 LN:44474 rl:unplaced M5:211c215414693fe0a2399cf82e707e03 AS:GRCh38
>chrUn_KI270590v1 AC:KI270590.1 gi:568335316 LN:4685 rl:unplaced M5:e8a57f147561b361091791b9010cd28b AS:GRCh38
>chrUn_KI270584v1 AC:KI270584.1 gi:568335315 LN:4513 rl:unplaced M5:d93636c9d54abd013cfc0d4c01334032 AS:GRCh38
>chrUn_KI270582v1 AC:KI270582.1 gi:568335314 LN:6504 rl:unplaced M5:6fd9804a7478d2e28160fe9f017689cb AS:GRCh38
>chrUn_KI270588v1 AC:KI270588.1 gi:568335313 LN:6158 rl:unplaced M5:37ffa850e69b342a8f8979bd3ffc77d4 AS:GRCh38
>chrUn_KI270593v1 AC:KI270593.1 gi:568335312 LN:3041 rl:unplaced M5:f4a5bfa203e9e81acb640b18fb11e78e AS:GRCh38
>chrUn_KI270591v1 AC:KI270591.1 gi:568335311 LN:5796 rl:unplaced M5:d6af509d69835c9ac25a30086e5a4051 AS:GRCh38
>chrUn_KI270330v1 AC:KI270330.1 gi:568335310 LN:1652 rl:unplaced M5:c2c590706a339007b00c59e0b8937e78 AS:GRCh38
>chrUn_KI270329v1 AC:KI270329.1 gi:568335309 LN:1040 rl:unplaced M5:f023f927ae84c5cc48dc4dce11ba90f2 AS:GRCh38
>chrUn_KI270334v1 AC:KI270334.1 gi:568335308 LN:1368 rl:unplaced M5:53afe12d1371f250a3d1de655345d374 AS:GRCh38
>chrUn_KI270333v1 AC:KI270333.1 gi:568335307 LN:2699 rl:unplaced M5:57baf650c47bba9b3a8b7c6d0fb55ad6 AS:GRCh38
>chrUn_KI270335v1 AC:KI270335.1 gi:568335306 LN:1048 rl:unplaced M5:eb27188639503b524d2659a23b8262ea AS:GRCh38
>chrUn_KI270338v1 AC:KI270338.1 gi:568335305 LN:1428 rl:unplaced M5:301ef75a6b2996d745eb3464bd352b57 AS:GRCh38
>chrUn_KI270340v1 AC:KI270340.1 gi:568335304 LN:1428 rl:unplaced M5:56b462bac20d385cdfcde0155fe4c3a1 AS:GRCh38
>chrUn_KI270336v1 AC:KI270336.1 gi:568335303 LN:1026 rl:unplaced M5:69ad2d85d870c8b0269434581e86e30e AS:GRCh38
>chrUn_KI270337v1 AC:KI270337.1 gi:568335302 LN:1121 rl:unplaced M5:16fc8d71a2662a6cfec7bdeec3d810c6 AS:GRCh38
>chrUn_KI270363v1 AC:KI270363.1 gi:568335301 LN:1803 rl:unplaced M5:6edd17a912f391022edbc192d49f2489 AS:GRCh38
>chrUn_KI270364v1 AC:KI270364.1 gi:568335300 LN:2855 rl:unplaced M5:6ff66a8e589ca27d93b5bac0e5b13a87 AS:GRCh38
>chrUn_KI270362v1 AC:KI270362.1 gi:568335299 LN:3530 rl:unplaced M5:bc82401ffd9a5ae711fa0ea34da8d2f0 AS:GRCh38
>chrUn_KI270366v1 AC:KI270366.1 gi:568335298 LN:8320 rl:unplaced M5:44a0b65b7ba6bcff37eca202e7d966ea AS:GRCh38
>chrUn_KI270378v1 AC:KI270378.1 gi:568335297 LN:1048 rl:unplaced M5:fc13bda7dbd914c92fb7e49489d1350f AS:GRCh38
>chrUn_KI270379v1 AC:KI270379.1 gi:568335296 LN:1045 rl:unplaced M5:3218bef25946cd95de585dfc7750f63b AS:GRCh38
>chrUn_KI270389v1 AC:KI270389.1 gi:568335295 LN:1298 rl:unplaced M5:2c9b08c57c27e714d4d5259fd91b6983 AS:GRCh38
>chrUn_KI270390v1 AC:KI270390.1 gi:568335294 LN:2387 rl:unplaced M5:7a64d89ea14990c16d20f4d6e7283e10 AS:GRCh38
>chrUn_KI270387v1 AC:KI270387.1 gi:568335293 LN:1537 rl:unplaced M5:22a12462264340c25e912b8485cdfa91 AS:GRCh38

>chrUn_KI270395v1 AC:KI270395.1 gi:568335292 LN:1143 rl:unplaced M5:7c03ca4756c1620f318fb189214388d8 AS:GRCh38
>chrUn_KI270396v1 AC:KI270396.1 gi:568335291 LN:1880 rl:unplaced M5:9069bed3c2efe7cc87227d619ad5816f AS:GRCh38
>chrUn_KI270388v1 AC:KI270388.1 gi:568335290 LN:1216 rl:unplaced M5:76f9f3315fa4b831e93c36cd88196480 AS:GRCh38
>chrUn_KI270394v1 AC:KI270394.1 gi:568335289 LN:970 rl:unplaced M5:d5171e863a3d8f832f0559235987b1e5 AS:GRCh38
>chrUn_KI270386v1 AC:KI270386.1 gi:568335288 LN:1788 rl:unplaced M5:b9b1baaa7abf206f6b70cf31654172db AS:GRCh38
>chrUn_KI270391v1 AC:KI270391.1 gi:568335287 LN:1484 rl:unplaced M5:1fa5cf03b3eac0f1b4a64948fd09de53 AS:GRCh38
>chrUn_KI270383v1 AC:KI270383.1 gi:568335286 LN:1750 rl:unplaced M5:694d75683e4a9554bcc1291edbcaee43 AS:GRCh38
>chrUn_KI270393v1 AC:KI270393.1 gi:568335285 LN:1308 rl:unplaced M5:3724e1d70677d6b5c4bcf17fd40da111 AS:GRCh38
>chrUn_KI270384v1 AC:KI270384.1 gi:568335284 LN:1658 rl:unplaced M5:b06e44ea15d0a57618d6ca7d2e6ac5d2 AS:GRCh38
>chrUn_KI270392v1 AC:KI270392.1 gi:568335283 LN:971 rl:unplaced M5:59b3ca8de65fb171683f8a06d3b4bf0d AS:GRCh38
>chrUn_KI270381v1 AC:KI270381.1 gi:568335282 LN:1930 rl:unplaced M5:2a9297cfd3b3807195ab9ad07e775d99 AS:GRCh38
>chrUn_KI270385v1 AC:KI270385.1 gi:568335281 LN:990 rl:unplaced M5:112a8b1df94ef0498a0bfe2d2ea5cc23 AS:GRCh38
>chrUn_KI270382v1 AC:KI270382.1 gi:568335280 LN:4215 rl:unplaced M5:e7085cdcee6ad62f359744e13d3209fc AS:GRCh38
>chrUn_KI270376v1 AC:KI270376.1 gi:568335279 LN:1136 rl:unplaced M5:59e8fc80b78d62325082334b43dffdba AS:GRCh38
>chrUn_KI270374v1 AC:KI270374.1 gi:568335278 LN:2656 rl:unplaced M5:dbc92c9a92e558946e58b4909ec95dd5 AS:GRCh38
>chrUn_KI270372v1 AC:KI270372.1 gi:568335277 LN:1650 rl:unplaced M5:53a9d5e8fd28bce5da5efcfd9114dbf2 AS:GRCh38
>chrUn_KI270373v1 AC:KI270373.1 gi:568335276 LN:1451 rl:unplaced M5:b174fe53be245a840cd6324e39b88ced AS:GRCh38
>chrUn_KI270375v1 AC:KI270375.1 gi:568335275 LN:2378 rl:unplaced M5:d678250c97e9b94aa390fa46e70a6d83 AS:GRCh38
>chrUn_KI270371v1 AC:KI270371.1 gi:568335274 LN:2805 rl:unplaced M5:a0af3d778dfeb7963e8e6d84c0c54fba AS:GRCh38
>chrUn_KI270448v1 AC:KI270448.1 gi:568335273 LN:7992 rl:unplaced M5:0f40827c265cb813b6e723da6c9b926b AS:GRCh38
>chrUn_KI270521v1 AC:KI270521.1 gi:568335272 LN:7642 rl:unplaced M5:af5bef7cefec7bd7efa729ac6c5be088 AS:GRCh38
>chrUn_GL000195v1 AC:GL000195.1 gi:224183229 LN:182896 rl:unplaced M5:5d9ec007868d517e73543b005ba48535 AS:GRCh38
>chrUn_GL000219v1 AC:GL000219.1 gi:224183268 LN:179198 rl:unplaced M5:f977edd13bac459cb2ed4a5457dba1b3 AS:GRCh38
>chrUn_GL000220v1 AC:GL000220.1 gi:224183269 LN:161802 rl:unplaced M5:fc35de963c57bf7648429e6454f1c9db AS:GRCh38
>chrUn_GL000224v1 AC:GL000224.1 gi:224183273 LN:179693 rl:unplaced M5:d5b2fc04f6b41b212a4198a07f450e20 AS:GRCh38
>chrUn_KI270741v1 AC:KI270741.1 gi:568335271 LN:157432 rl:unplaced M5:86eaea8a15a3950e37442eaaa5c9dc92 AS:GRCh38
>chrUn_GL000226v1 AC:GL000226.1 gi:224183275 LN:15008 rl:unplaced M5:1c1b2cd1fccbc0a99b6a447fa24d1504 AS:GRCh38
>chrUn_GL000213v1 AC:GL000213.1 gi:224183287 LN:164239 rl:unplaced M5:9d424fdcc98866650b58f004080a992a AS:GRCh38
>chrUn_KI270743v1 AC:KI270743.1 gi:568335270 LN:210658 rl:unplaced M5:3b62d9d3100f530d509e4efebd98502c AS:GRCh38
>chrUn_KI270744v1 AC:KI270744.1 gi:568335269 LN:168472 rl:unplaced M5:e90aee46b947ff8c32291a6843fde3f9 AS:GRCh38
>chrUn_KI270745v1 AC:KI270745.1 gi:568335268 LN:41891 rl:unplaced M5:1386fe3de6f82956f2124e19353ff9c1 AS:GRCh38

>chrUn_KI270746v1 AC:KI270746.1 gi:568335267 LN:66486 rl:unplaced M5:c470486a0a858e14aa21d7866f83cc17 AS:GRCh38
>chrUn_KI270747v1 AC:KI270747.1 gi:568335266 LN:198735 rl:unplaced M5:62375d812ece679c9fd2f3d08d4e22a4 AS:GRCh38
>chrUn_KI270748v1 AC:KI270748.1 gi:568335265 LN:93321 rl:unplaced M5:4f6c6ab005c852a4352aa33e7cc88ded AS:GRCh38
>chrUn_KI270749v1 AC:KI270749.1 gi:568335264 LN:158759 rl:unplaced M5:c899a7b4e911d371283f3f4058ca08b7 AS:GRCh38
>chrUn_KI270750v1 AC:KI270750.1 gi:568335263 LN:148850 rl:unplaced M5:c022ba92f244b7dc54ea90c4eef4d554 AS:GRCh38
>chrUn_KI270751v1 AC:KI270751.1 gi:568335262 LN:150742 rl:unplaced M5:1b758bbdee0e9ca882058d916cba9d29 AS:GRCh38
>chrUn_KI270752v1 AC:KI270752.1 gi:568335261 LN:27745 rl:unplaced M5:e0880631848337bd58559d9b1519da63 AS:GRCh38
>chrUn_KI270753v1 AC:KI270753.1 gi:568335260 LN:62944 rl:unplaced M5:25075fb2a1ecada67c0eb2f1fe0c7ec9 AS:GRCh38
>chrUn_KI270754v1 AC:KI270754.1 gi:568335259 LN:40191 rl:unplaced M5:fe9e16233cecbc244f06f3acff3d03b8 AS:GRCh38
>chrUn_KI270755v1 AC:KI270755.1 gi:568335258 LN:36723 rl:unplaced M5:4a7da6a658955bd787af8add3ccb5751 AS:GRCh38
>chrUn_KI270756v1 AC:KI270756.1 gi:568335257 LN:79590 rl:unplaced M5:2996b120a5a5e15dab6555f0bf92e374 AS:GRCh38
>chrUn_KI270757v1 AC:KI270757.1 gi:568335256 LN:71251 rl:unplaced M5:174c73b60b41d8a1ef0fbaa4b3bdf0d3 AS:GRCh38
>chrUn_GL000214v1 AC:GL000214.1 gi:224183298 LN:137718 rl:unplaced M5:46c2032c37f2ed899eb41c0473319a69 AS:GRCh38
>chrUn_KI270742v1 AC:KI270742.1 gi:568335255 LN:186739 rl:unplaced M5:2f31c013a4a8301deb8ab7ed1ca1cd99 AS:GRCh38
>chrUn_GL000216v2 AC:GL000216.2 gi:568335254 LN:176608 rl:unplaced M5:725009a7e3f5b78752b68afa922c090c AS:GRCh38
>chrUn_GL000218v1 AC:GL000218.1 gi:224183305 LN:161147 rl:unplaced M5:1d708b54644c26c7e01c2dad5426d38c AS:GRCh38
>chrEBV AC:AJ507799.2 gi:86261677 LN:171823 rl:decoy M5:6743bd63b3ff2b5b8985d8933c53290a SP:Human_herpesvirus_4 tp:circular
>phage_lambda J02459.1_48502
>plasmid_puc19c L09137.2 Cloning vector pUC19c, complete sequence
>phage_T4 168903
>phage_Xp12 65076