Supplementary Materials for

Mapping the nascent DNA methylome reveals inheritance of hemi-methylation at CTCF/cohesin sites

Chenhuan Xu and Victor G. Corces

correspondence to: vgcorces@gmail.com

This PDF file includes:

Materials and Methods

Cell culture

H9 and HUES64 human embryonic stem cells were cultured in STEMPRO hESC SFM (Thermo Fisher) on cultureware coated with Geltrex Matrix (Thermo Fisher) at 37 °C under 5% CO_2. Medium was changed every day. CTCF-AID E14 mESCs were cultured and induced as described in (*29*). *Drosophila* Kc167 cells were cultured in SFX-Insect medium (HyClone) at 25 °C. H9 cells were grown to 60-70% confluence before EdU (5-ethynyl-2'-deoxyuridine) (Thermo Fisher) labeling. Kc167 cells were grown to 50% confluence and treated with 2 mM hydroxyurea (Sigma) for 20 hr before EdU labeling.

EdU labeling and chromatin extraction

For pulse, H9 cells were replenished with medium containing 50 μM EdU and cultured at 37 °C for 20 min before formaldehyde crosslinking. For chase, after labeling with EdU for 20 min, cells were washed with DPBS twice, replenished with medium containing 50 μM dT (Sigma), and cultured at 37 °C for 8 hr before formaldehyde crosslinking. Kc167 cells were replenished with medium containing 40 μM EdU and cultured at 25 °C for 20 min before formaldehyde crosslinking. Crosslinking with 1% formaldehyde was done in DPBS at room temperature for 10 min and was quenched with 0.125 M glycine (Sigma). Cells were collected and washed with DPBS, and lysed in Lysis Buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40, 10% glycerol). Pellets were collected and re-suspended in Sonication Buffer (10 mM Tris-HCl pH 8.0, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.1% sodium deoxycholate, 0.5% sodium lauroylsarcosine). Sonication was done on ice using a Bioruptor (Diagenode). Triton X-100 was added to a final concentration of 1% before centrifugation at 15,000 x *g*. Supernatant was collected as the chromatin fraction. This part of the protocol is common to nasBS-seq, nasChIP-seq, and nasChIP-BS-seq.

nasBS-seq

Approximately 10^8 H9 hESCs were used to make a pair of parent and daughter strand nasBS-seq libraries under each condition. After chromatin extraction, Elution Buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0, 1% SDS) was added to chromatin in a 1:1 ratio. Crosslinking reversal was done by incubation at 65 °C for 8 hr. Proteinase K was added to a final concentration of 0.2 mg/ml before incubation at 55 °C for 30 min. Phenol:chloroform:isoamyl alcohol extraction and ethanol precipitation was done sequentially to purify genomic DNA. Click chemistry was performed in a 200 μl volume (174 μl DNA in DPBS, 2 μl 1 mM biotin-azide (Thermo Fisher), 4 μl 100 mM $CuSO_4$ (Sigma), 20 μl 100 mM sodium ascorbate (Sigma)) at 37 °C for 1 hr. After ethanol precipitation, DNA was incubated with Dynabeads Streptavidin C1 beads (Thermo Fisher) on a rotor for 30 min at room temperature, and washed according to the manufacturer's instructions. Beads were incubated with End Repair Enzyme Mix (NEB) on a rotator for 30 min at room temperature. After washing, beads were incubated with Klenow Fragment (3'→5' exo^-) reaction mix (NEB) on a rotator for 30 min at 37 °C. Beads were washed and then incubated with T4 DNA ligase reaction mix (NEB) and methylated adaptors (Illumina) on a rotator for 4 hr at room temperature. After washing, beads were incubated in 150 mM NaOH for 3 min at room temperature. Supernatant was

collected as the fraction containing the parental strand DNA (biotin$^-$). Beads were washed 3 times with NaOH, re-suspended in 95% formamide and 10 mM EDTA pH 8.2, and incubated at 90 °C for 3 min. Supernatant was collected as the fraction containing the daughter strand DNA (biotin$^+$). Both fractions were ethanol precipitated, mixed with 1 pg of *lambda* DNA, and bisulfite converted with the EZ DNA Methylation-Lightning Kit (Zymo). PCR amplification was done with HiFi Uracil+ polymerase (KAPA). For each library, the minimal number of PCR cycles were used to yield at least 20 ng of DNA measured after purification with AMpure XP beads (Beckman Coulter).

nasChIP-seq

Approximately 1x10^7 H9 hESCs were used to make each nasChIP-seq library. Anti-CTCF (Millipore) or -SMC1A (Millipore) antibodies were pre-incubated with Dynabeads Protein A beads (Thermo Fisher). Extracted chromatin was incubated with antibody-beads complex on a rotator at 4 °C for 5 hr. Beads were washed once with Wash Buffer (20 mM Tris-HCl pH 8.1, 500 mM NaCl, 2 mM EDTA pH 8.0, 1% Triton X-100, 0.1% SDS), twice with LiCl Buffer (10 mM Tris-HCl pH 8.1, 250 mM LiCl, 1 mM EDTA pH 8.0, 1% NP-40, 1% sodium deoxycholate), and once with 1x TE pH 8.0 with 50 mM NaCl. DNA was eluted by incubating beads in Elution Buffer at 65 °C for 30 min. Crosslinking-reversal, proteinase K digestion, phenol:chloroform:isoamyl alcohol extraction, ethanol precipitation, click chemistry, ethanol precipitation, streptavidin beads pull-down, and Illumina library preparation steps were the same as described above for nasBS-seq.

nasChIP-BS-seq

Approximately 3-5x10^7 H9 hESCs were used to make a pair of parent and daughter strand nasChIP-BS-seq libraries under each condition. Anti-DNMT1 (Abcam), -DNMT3A (Abcam) or -DNMT3B (Abcam) antibodies were used. All steps were the same as described above for nasChIP-seq except that ligation with methylated adaptors, strand separation, and bisulfite conversion steps were done as described above for nasBS-seq to generate a pair of parental and daughter strand libraries for each condition.

ChIP-hairpinBS-seq

Approximately 3x10^7 H9 hESCs were used to make each ChIP-hairpinBS-seq library. Anti-CTCF (Millipore) antibodies were pre-incubated with Dynabeads Protein A beads (Thermo Fisher). Extracted chromatin was incubated with antibody-beads complex on a rotator at 4 °C for 5 hr. Beads were washed once with Wash Buffer (20 mM Tris-HCl pH 8.1, 500 mM NaCl, 2 mM EDTA pH 8.0, 1% Triton X-100, 0.1% SDS), twice with LiCl Buffer (10 mM Tris-HCl pH 8.1, 250 mM LiCl, 1 mM EDTA pH 8.0, 1% NP-40, 1% sodium deoxycholate), and once with 1x TE pH 8.0 with 50 mM NaCl. DNA was eluted by incubating beads in Elution Buffer at 65 °C for 30 min. Crosslinking-reversal, proteinase K digestion, phenol:chloroform:isoamyl alcohol extraction, ethanol precipitation, end repair reaction, and A-tailing reaction were performed. The rest of the protocol followed the method reported in (*31*).

Hi-ChIP

See (*32*).

Processing of nasBS-seq and nasChIP-BS-seq sequence data

50 bp paired-end reads were first trimmed with Trimmomatic (0.36) (*33*) to remove any remnant adaptor sequence and bases with Phred quality score <20. Only pairs with both reads ≥20 bp were retained for subsequent analysis. Reads were aligned to the human genome (hg38, XX) using bismark (0.19.0) (*34*) with bowtie2 (2.2.9) (*35*). In rare cases of mapping rate <50%, two adjustments were made sequentially to improve the yield: 1) The 5' end of the mate 2 reads were trimmed 5-10 bases off, which generally increased the mapping rate by 10-20%. 2) The mate 1 reads from the unmapped paired-end reads were retrieved for single-end mapping, given the fact that mate 1 reads alone could achieve >70% mapping rate under such circumstances. Alignment score was controlled by applying bowtie2 option --score-min L,0,-0.4. Mapping uniqueness was controlled by selecting alignments with mapping quality Q>10 using Samtools (1.6) (*36*). For each library, reads were specifically mapped to either the Watson or Crick strand, and duplicate reads were removed by bismark. Methylation of single cytosines was called from strand-specific alignments. For each strand, methylation calls from all biological replicates were pooled together unless otherwise noted. Methylation of CpGs was extracted from certain pairs of Watson/Crick strand methylation calls using bedtools (2.27.0) (*37*). The unmapped reads from bismark were re-aligned to the *lambda* genome to determine the bisulfite conversion rate of each library.

nasChIP-seq sequencing data processing

50 bp paired-end reads were aligned to the human genome (hg38, XX) using bowtie2. Alignment score was controlled by applying bowtie2 option --score-min L,0,-0.4. Mapping uniqueness was controlled by selecting alignments with mapping quality Q>10 using Samtools. Peak calling and read density calculation were done using MACS2 (2.1.0) with BAMPE option (*38*). Matrix underlying heatmaps was generated with deepTools (2.5.4) (*39*) and was visualized using Java Treeview (*40*).

iSA method

For certain biological replicates of pairs of parental and daughter strand libraries, alignments with exactly the same two ends were searched and counted between $parent^{Watson}$ and $daughter^{Crick}$ strands, and between $daughter^{Watson}$ and $parent^{Crick}$ strands, using bedtools. The pairs of alignments with the same step shift at both ends (1-5 bases upstream or downstream) was also counted to calculate the fold enrichment of genuinely paired alignments over random pairing. Methylation of cytosines and intraCpGs was called from each pair of alignments. Methylation of intraCpGs was recorded in one of the four possible states: methylation, unmethylation, hemi-methylation with respect to the parent-daughter axis. For WGBS datasets, searching for pairs was done between alignments from Watson and Crick strands. Methylation of intraCpGs was recorded in one of the four possible states: methylation, unmethylation, hemi-methylation with respect to the Watson-Crick axis.

Expanded discovery of concordant hemiCpGs

Based on the observation that methylation on all four strands for concordant hemiCpGs changed very little from pulse to chase, and that it is the only type of

hemiCpGs showing this pattern, we applied two criteria to discover additional CpGs of this type: 1) Any CpG with four Cs on all four strands either from pulse or chase mapped at least four times, and with ΔmC≥67% or ≤-67% on both DNA duplexes in a concordant way. 2) Any CpG with both Cs on the same DNA duplex mapped at least four times in both pulse and chase, and with ΔmC≥67% or ≤-67% in both conditions.

Average methylation profiling around genomic features

Relative position of Cs to the anchor point was determined using bedtools. A sliding window of 5-20 bp with 1 bp step was applied and the mean methylation value was calculated for each step. The mean methylation values were plotted according to the relative position of the window center base to the anchor point base.

Annotation of the occupancy frequency of CTCF motifs

Two lists of 90,521 and 75,993 CTCF motifs ever detected as occupied in at least one cell type in human or mouse were compiled from all available CTCF ChIP-seq experiments in human or mouse cell lines deposited in ENCODE. For calculating the reads per million (RPM) value of ChIP-seq/nasChIP-seq data, a global division of the read density with the total number of alignments (in million) was done. For annotating the CTCF RPM value of CTCF motifs, the mean RPM value from a 40 bp window surrounding the CTCF motif center base was calculated. For annotating the RAD21/SMC1A RPM value of CTCF motifs, the mean RPM value from a 100 bp window surrounding the CTCF motif center base was calculated. CTCF/cohesin co-occupied CTCF motifs refer to CTCF motifs with CTCF RPM>1 and SMC1A/RAD21 RPM>0.4 unless otherwise noted.

Calculation of the Hemi Index of CTCF motifs

CTCF motifs were oriented using the G-rich strand (*41*). All CpGs with two Cs mapped at least four times and within the upstream (-165 to -20 bp from motif center base) or downstream (20 to 165 bp from motif center base) regions of oriented CTCF motifs were retrieved. For each CpG, the ΔmC value (%) was calculated as:

$\Delta mC = mC^{motif} - mC^{oppo}$

where mC^{motif} is the methylation frequency (%) of the C on the same strand as the CTCF motif sequence, and mC^{oppo} is the methylation frequency (%) of the C on the complementary strand as the CTCF motif sequence. For each CTCF motif, the Hemi Index (HI) was calculated as:

$HI = \sum_{i=1}^{M} \Delta mC_i - \sum_{j=1}^{N} \Delta mC_j$

where ΔmC_i is the ΔmC value of the *i*th downstream CpG, and ΔmC_j is the ΔmC value of the *j*th upstream CpG. To obtain the Hemi Index of the pulse condition in H9 cells, all CpGs from the two nascent DNA duplexes were taken into account.

Analysis of Hi-ChIP data

Juicer (*42*) was used to analyze RAD21 Hi-ChIP data. Only interaction contacts with mapping quality score q≥5 were retained. The WT library was down-sampled to match the number of interaction contacts in the DNMT3B-KO library. Juicebox (*43*) was used to obtain interaction matrices at 25 kb resolution. The total interaction contacts of CTCF motif-containing 25 kb bins with each consecutive downstream or upstream 25 kb

bin were retrieved from the matrix. The ratio of total interaction contacts (WT over DNMT3B-KO) was calculated for each 25 kb bin within a ±1 Mb window surrounding CTCF motif-containing bins.

Published datasets

The WGBS datasets were obtained from GEO under accession numbers: H9 (study 1: GSM491349; study 2: SRX026814, SRX056693, SRX056694, SRX056695; study 3: GSM1493983, GSM1493984, GSM1493985), HUES64 (GSM1112840, GSM1112841), HUES64-DNMT3-KO (GSM1545005, GSM1545006, GSM1545007), mESC (GSM748786, GSM748787), mouse early embryonic stages (GSE56697), MEF (GSE58610), and from ENCODE under accession numbers: H1 (ENCSR000AJJ, ENCSR617FKV), GM12878 (ENCSR890UQO), HepG2 (ENCSR881XOU), IMR-90 (ENCSR888FON), K562 (ENCSR765JPC). The RRBS datasets were obtained from GEO under accession numbers: HUES64-DNMT1-KO (GSM1545009, GSM1545010). The TAB-seq dataset was obtained from GEO under accession number: H1 (GSE36173). The hairpinBS-seq dataset was obtained from GEO under accession number: E14 (GSM1173118). The ChIP-seq datasets were obtained from GEO under accession numbers: CTCF in HUES64 (GSM1505621, GSM1505622), CTCF in mESC (GSM747534, GSM747535, GSM747536), CTCF-AID in mESC (GSE98671), MBD proteins in mESC (GSE39610), and from ENCODE under accession numbers: CTCF in H1 (ENCSR000BNH), RAD21 in H1 (ENCSR000BLD), ZNF143 in H1 (ENCSR000EBW). The RNA-seq datasets were obtained from GEO under accession numbers: mouse ICM (GSM2229972, GSM2229973). The MNase-seq datasets were obtained from GEO under accession numbers: H1 (GSM1194220), H9 (GSM1194221). The Repli-seq datasets were obtained from GEO under accession numbers: Kc167 (GSM948519, GSM948521, GSM948522, GSM948524).

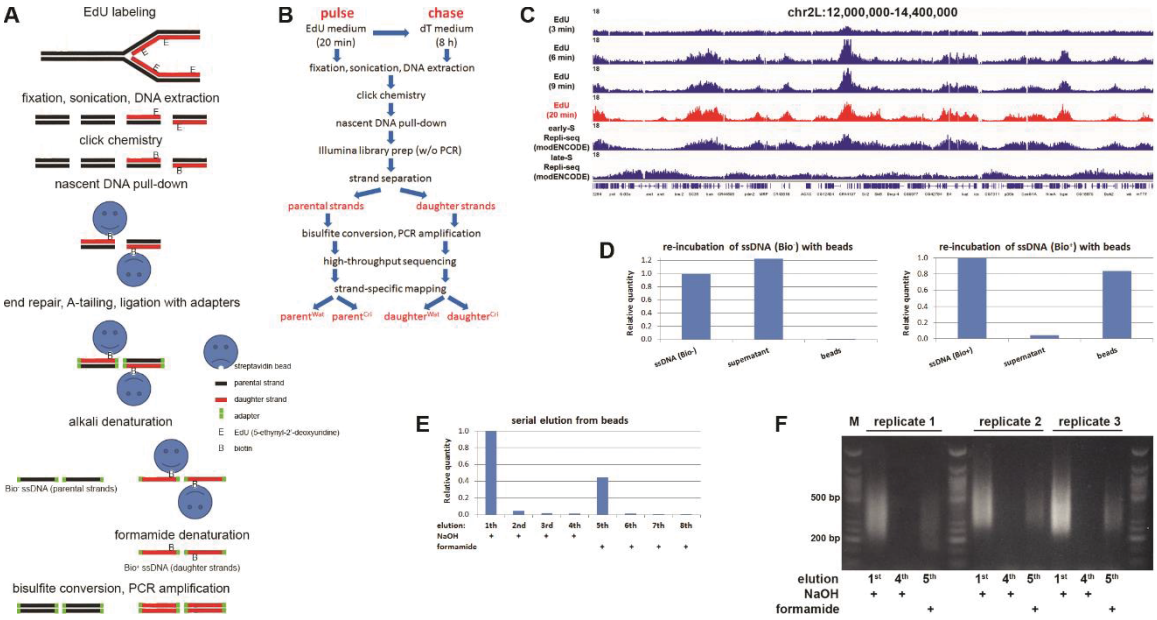
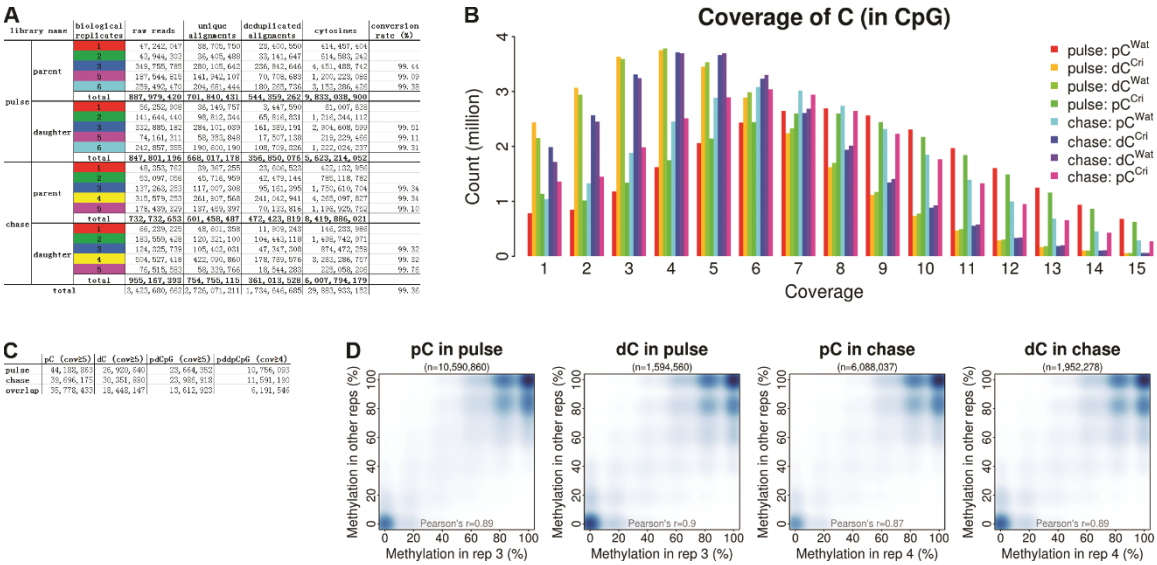


Fig. S1. Proof of concept and quality measurements of nasBS-seq.

(**A**) Schematic overview of nasBS-seq. (**B**) Workflow for nasBS-seq. (**C**) Different labeling times of EdU were tested in *Drosophila* Kc167 cells synchronized at the G1/S boundary, and compared with published Repli-seq datasets in Kc167. (**D**) Bio^- and bio^+ fractions of ssDNA in Kc167 cells after strand separation were re-incubated with streptavidin beads. DNA in the pre-incubation fraction, the supernatant, and bead fractions after re-incubation, were quantified by QPCR using primers targeting the TSS region of the *cbt* gene. (**E**) Serial elution was done with bead-captured Kc167 nascent DNA. NaOH was used for the 1st to 4th elution, and formamide was used for the 5th to 8th elution. DNA in the eluate was quantified by QPCR using the *cbt* primers. (**F**) Three replicates of nasBS-seq libraries in H9 hESC. Mock libraries from the 4th elution with NaOH were made in parallel and yielded DNA in undetectable levels.



library name		biological replicates	raw reads	unique alignments	deduplicated alignments	cytosines	conversion rate (%)
pulse	parent	1	47,242,047	38,705,750	23,400,550	514,457,484	
		2	45,944,303	36,405,488	33,141,647	514,583,242	
		3	349,755,735	280,105,642	236,842,646	4,451,488,742	99.44
		5	187,544,815	141,942,107	70,708,683	1,270,225,982	99.09
		6	255,492,470	204,681,444	180,265,736	3,152,286,426	99.38
		total	**887,979,420**	**701,840,431**	**544,359,262**	**9,833,038,909**	
	daughter	1	56,262,308	38,149,797	3,447,590	61,007,838	
		2	141,644,440	98,812,344	65,816,851	1,216,544,112	
		3	332,985,182	284,101,089	161,389,191	2,904,608,599	99.51
		5	74,161,311	58,353,048	17,507,138	315,229,466	99.11
		6	242,867,335	190,600,190	108,709,226	1,922,024,237	99.31
		total	**847,891,196**	**668,017,178**	**356,850,076**	**5,823,214,052**	
chase	parent	1	45,353,762	39,367,255	23,606,523	422,132,956	
		2	53,097,096	45,716,959	42,479,144	785,118,762	
		3	137,265,253	117,097,308	95,161,395	1,750,619,704	99.34
		4	315,579,283	261,307,568	241,042,941	4,265,097,827	99.34
		5	175,439,329	137,459,397	70,133,816	1,185,929,752	99.10
		total	**732,732,653**	**601,488,487**	**472,423,819**	**8,415,886,021**	
	daughter	1	86,239,225	48,601,358	11,809,243	148,233,986	
		2	183,559,428	120,321,100	104,443,118	1,498,742,971	
		3	124,325,779	105,402,631	47,347,508	874,472,253	99.33
		4	504,527,418	422,090,860	178,769,976	3,283,286,757	99.32
		5	76,515,583	58,339,766	18,544,283	228,058,206	99.76
		total	**955,167,393**	**754,755,115**	**361,013,528**	**6,007,794,179**	
total			3,423,669,662	2,726,071,211	1,734,646,685	29,883,933,152	99.36

C	pC (cov25)	dC (cov25)	pdCpG (cov25)	pdgpCpG (cov24)
pulse	44,188,863	26,920,640	23,664,352	10,756,983
chase	39,696,175	30,351,880	23,985,918	11,591,180
overlap	35,778,433	18,448,147	13,612,923	6,191,546

Fig. S2. Statistics and reproducibility of nasBS-seq.

(**A**) The number of paired-end reads, uniquely mapped and deduplicated alignments, Cs mapped, and the conversion rate determined by methylation frequency of spiked-in *lambda* DNA for each biological replicate of nasBS-seq are shown. (**B**) Coverage of all mapped Cs in the context of CpG from all eight strands in pulse and chase. (**C**) The number of parental Cs, daughter Cs, parent-daughter CpGs, and CpGs with all four Cs mapped in pulse or chase are shown. (**D**) For each library, methylation frequency of Cs from the biggest contributing biological replicate and the rest of the replicates were compared. Only Cs mapped at least 5 times in both pools were compared. Pearson's correlation coefficient is shown for each comparison.

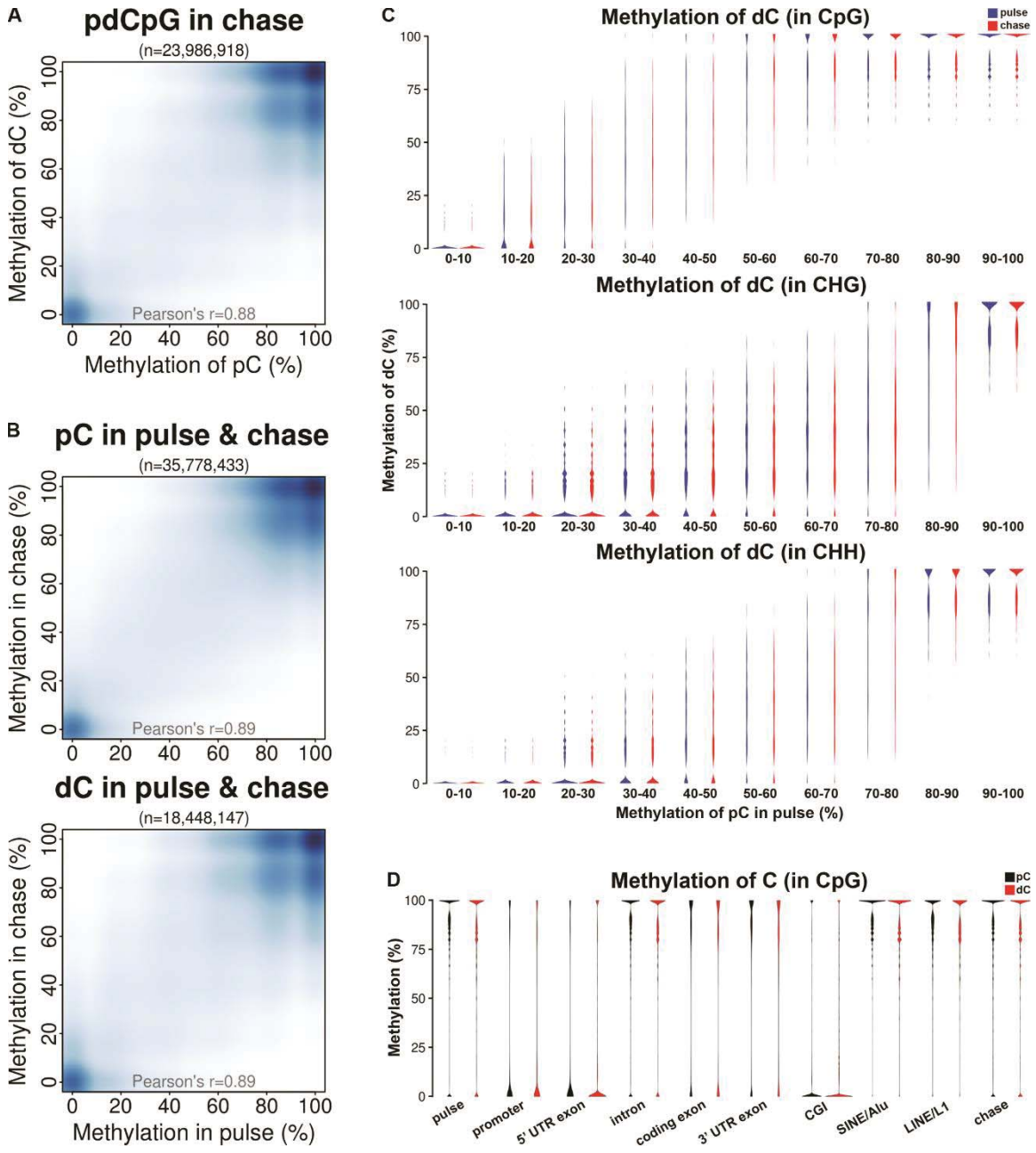


Fig. S3. Methylation of Cs in all context is maintained after 20 min of replication.

(**A**) Correlation of methylation frequency between pCs and dCs from pdCpGs in chase. (**B**) Correlation of methylation frequency of pCs or dCs between pulse and chase. (**C**) pCs in the context of CpG, CHG or CHH in pulse were separated into 10 intervals according to methylation frequency. The distribution of methylation frequency of dCs in pulse and chase associated with each pC group is shown. (**D**) Distribution of methylation frequency of pCs and dCs from pulse, different genomic features (pulse), and chase are shown.

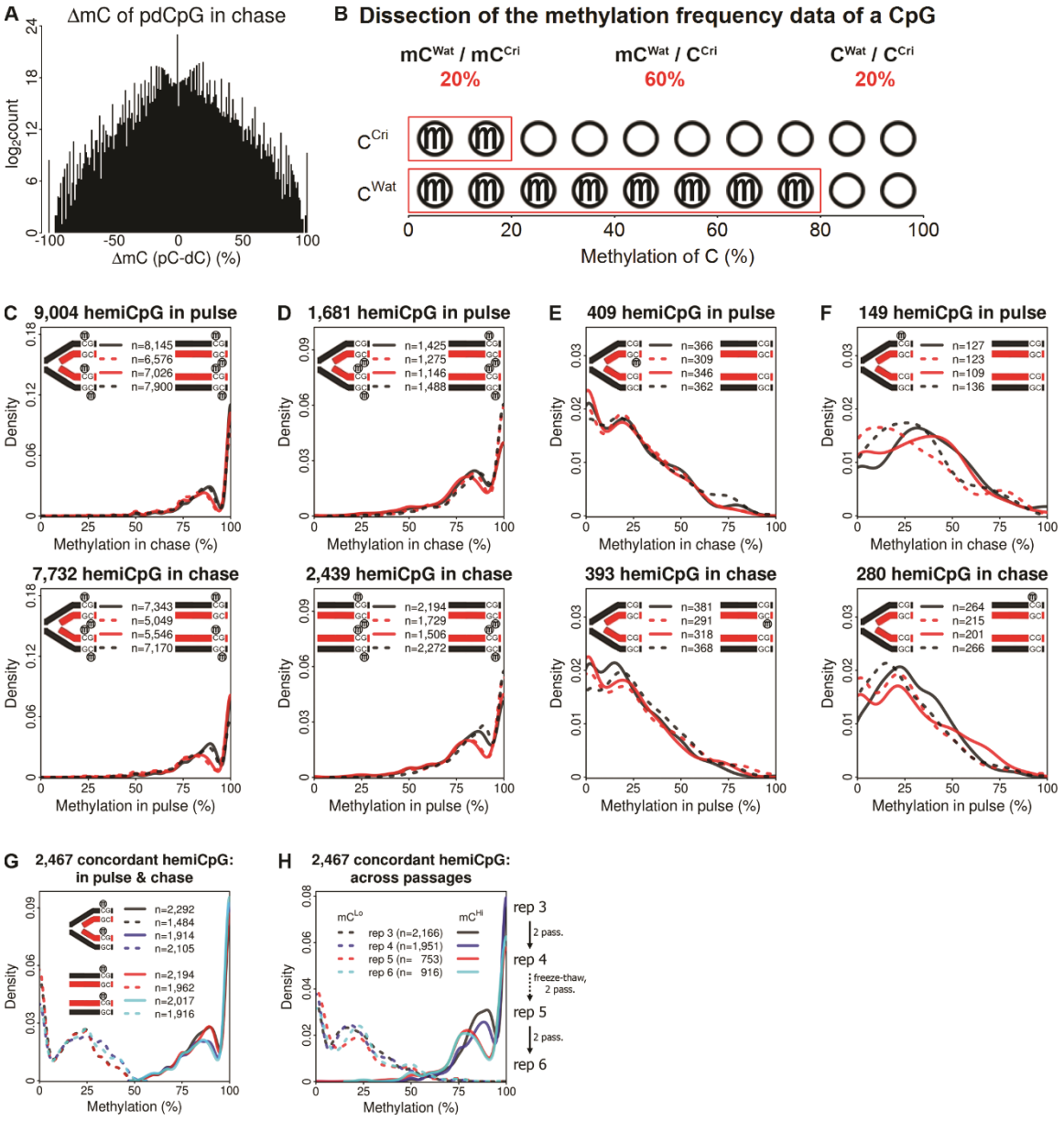


Fig. S4. Transient and stably inherited hemiCpGs during DNA replication.

(**A**) A histogram showing the count of pdCpGs in chase with differential ΔmC values. (**B**) A schematic example showing the simplification of BS-seq data representing methylation frequency of two Cs in the same CpG into methylation frequency of CpG *per se*. (**C-F**) Different types of hemiCpGs in pulse were intersected with chase, and distribution of methylation frequency for each of the 4 Cs in chase is shown (upper), and *vice versa* (lower). (**G**) All concordant hemiCpGs were intersected with each of the 8 strands in pulse and chase, and the distribution of methylation frequency is shown for each C. (**H**) The chronological relationship between biological replicates is shown. All concordant hemiCpGs were intersected with Watson or Crick strands in each biological replicate, and the distribution of methylation frequency is shown for each strand.

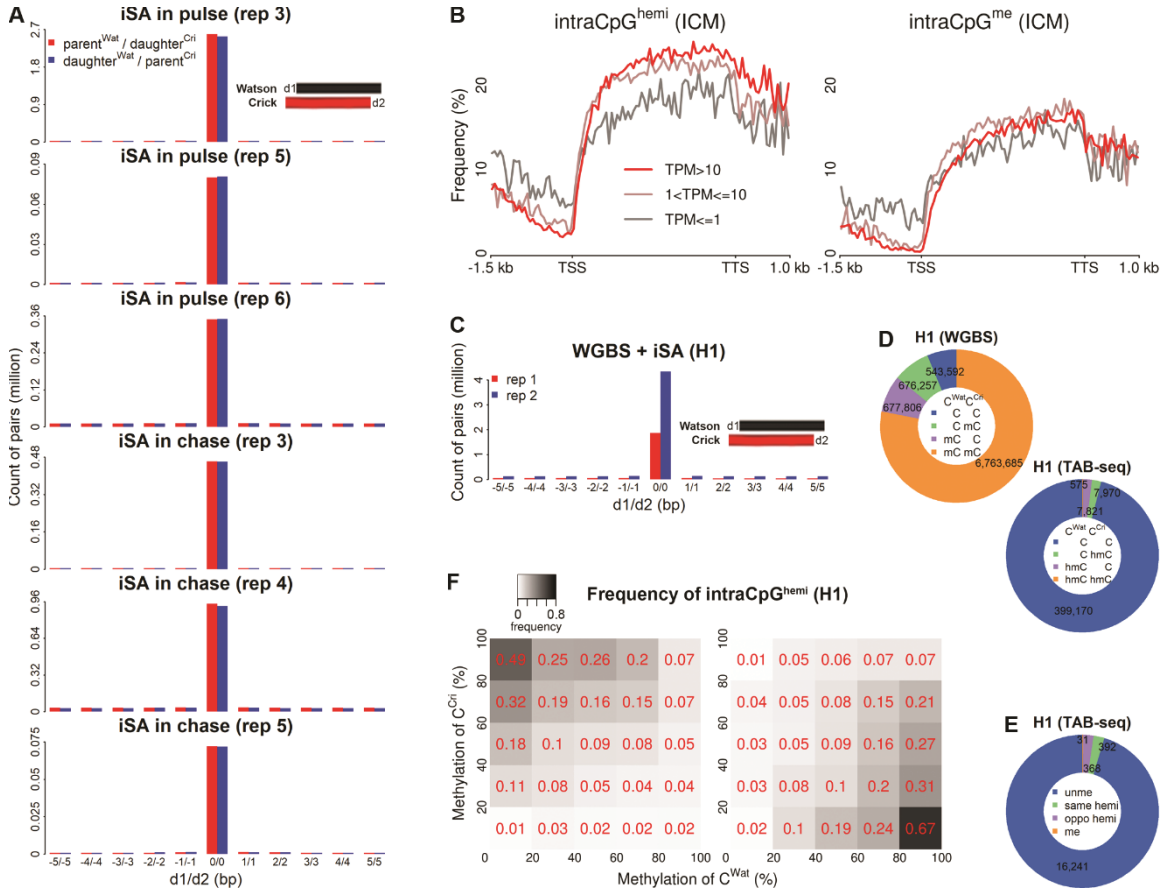


Fig. S5. The iSA method unveils methylation status of CpGs from single dsDNA fragments.

(**A**) Pairs of alignments with differential end-matching were searched and counted between parent^{Watson} and daughter^{Crick} strands, and between daughter^{Watson} and parent^{Crick} strands in all applicable replicates in pulse or chase. (**B**) The frequency of intraCpG^{hemi} and intraCpG^{me} at genic regions in mouse inner cell mass. TPM: transcripts per million. (**C**) Pairs of alignments with differential end-matching were searched and counted between Watson and Crick strands in WGBS datasets from H1 hESC. (**D**) The fraction of all four types of intraCpGs called from WGBS or TAB-seq in H1 hESC using the iSA method. (**E**) The intraCpG^{hemi} from WGBS in H1 hESC were intersected with intraCpGs from TAB-seq in H1 hESC, and the fraction of different hydroxymethylation states is shown. (**F**) All intraCpGs in H1 hESC were allocated to the appropriate cells according to their methylation frequency in WGBS. The frequency of Watson- (right) or Crick- (left) methylated intraCpG^{hemi} out of all intraCpGs within each cell is shown.

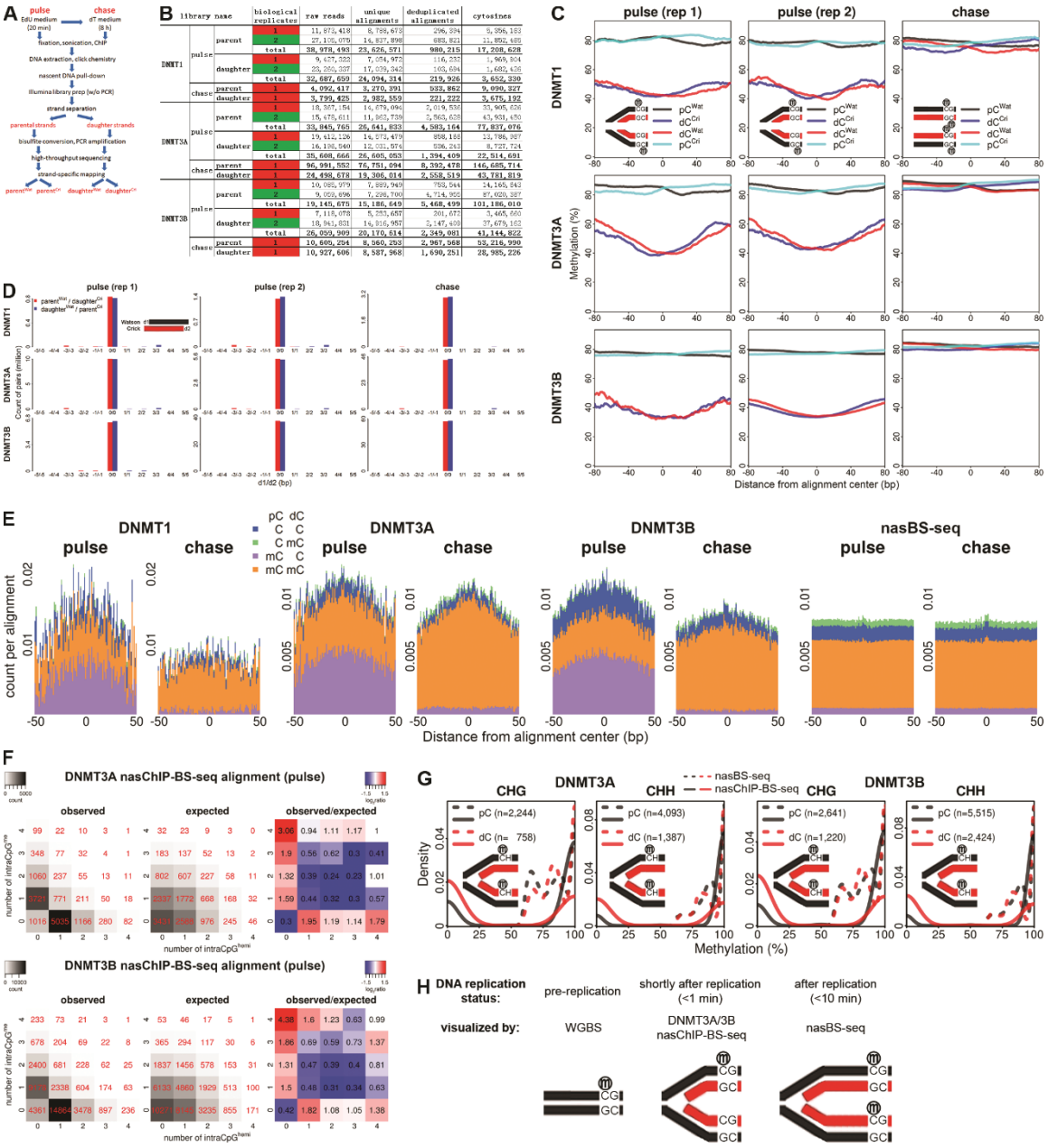


Fig. S6. nasChIP-BS-seq visualizes transient interactions between DNMTs and substrate dCs for both maintenance and de novo methylation.

(**A**) Workflow for nasChIP-BS-seq. (**B**) Statistics of DNMT nasChIP-BS-seq in H9 hESC. (**C**) Average methylation profiles of all four strands around alignment centers from DNMT nasChIP-BS-seq in pulse or chase. (**D**) Pairs of alignments with differential end-matching were searched and counted between strands of parent^{Watson} and daughter^{Crick}, and between daughter^{Watson} and parent^{Crick} in all DNMT nasChIP-BS-seq replicates. (**E**) Stacked frequency of all four types of DNMT-targeted or nasBS-seq intraCpGs around alignment centers in pulse or chase. (**F**) 2-D histograms showing the count of DNMT3A/3B nasChIP-BS-seq alignments with different numbers of intraCpG^{hemi} and intraCpG^{me}. The expected distribution was simulated by randomly

allocating the same numbers of intraCpG^{hemi} and intraCpG^{me} to the same number of alignments 100 times and the mean counts were taken. (**G**) Distribution of methylation frequency of the 2 Cs in maintained CHG/CHH methylation viewed through nasBS-seq and DNMT3A/3B nasChIP-BS-seq. (**H**) The essential steps during inheritance of concordant hemiCpGs.

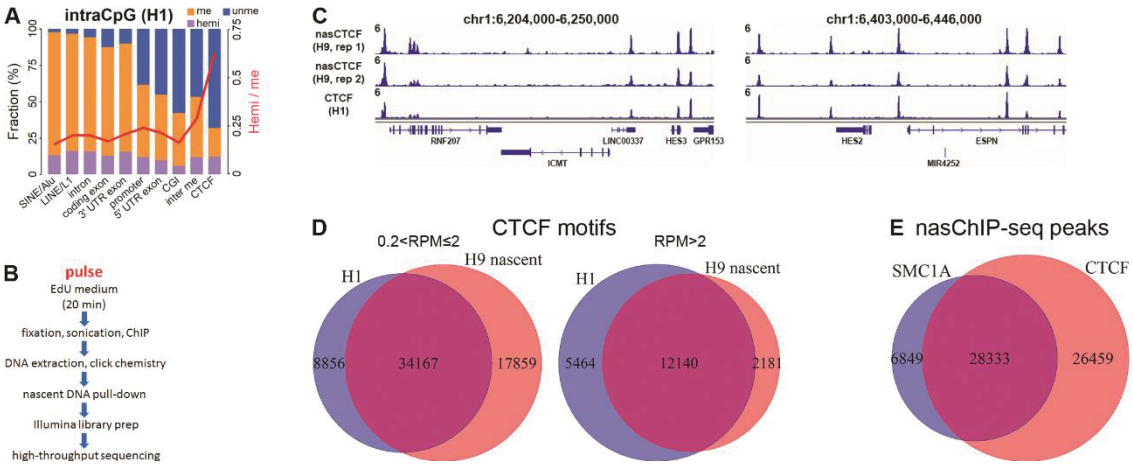


Fig. S7. nasChIP-seq maps the protein-DNA interactome on nascent chromatin.

(**A**) The fraction of 3 types of intraCpGs (with 2 types of intraCpG^{hemi} combined) at different genomic features in H1 hESC. The ratio of intraCpG^{hemi} over intraCpG^{me} is also shown. inter me: regions with intermediate DNA methylation (*44*). CTCF: 400 bp windows surrounding occupied CTCF motifs. (**B**) Workflow for nasChIP-seq. (**C**) Genome browser screenshots for two CTCF nasChIP-seq replicates in H9 hESC and the published CTCF ChIP-seq in H1 hESC. (**D**) Venn diagrams showing the overlap of CTCF motifs with medium (left) or high (right) occupancy between CTCF nasChIP-seq in H9 hESC and CTCF ChIP-seq in H1 hESC. (**E**) Venn diagram showing the overlap of MACS2 peaks between CTCF and SMC1A nasChIP-seq in H9 hESC.

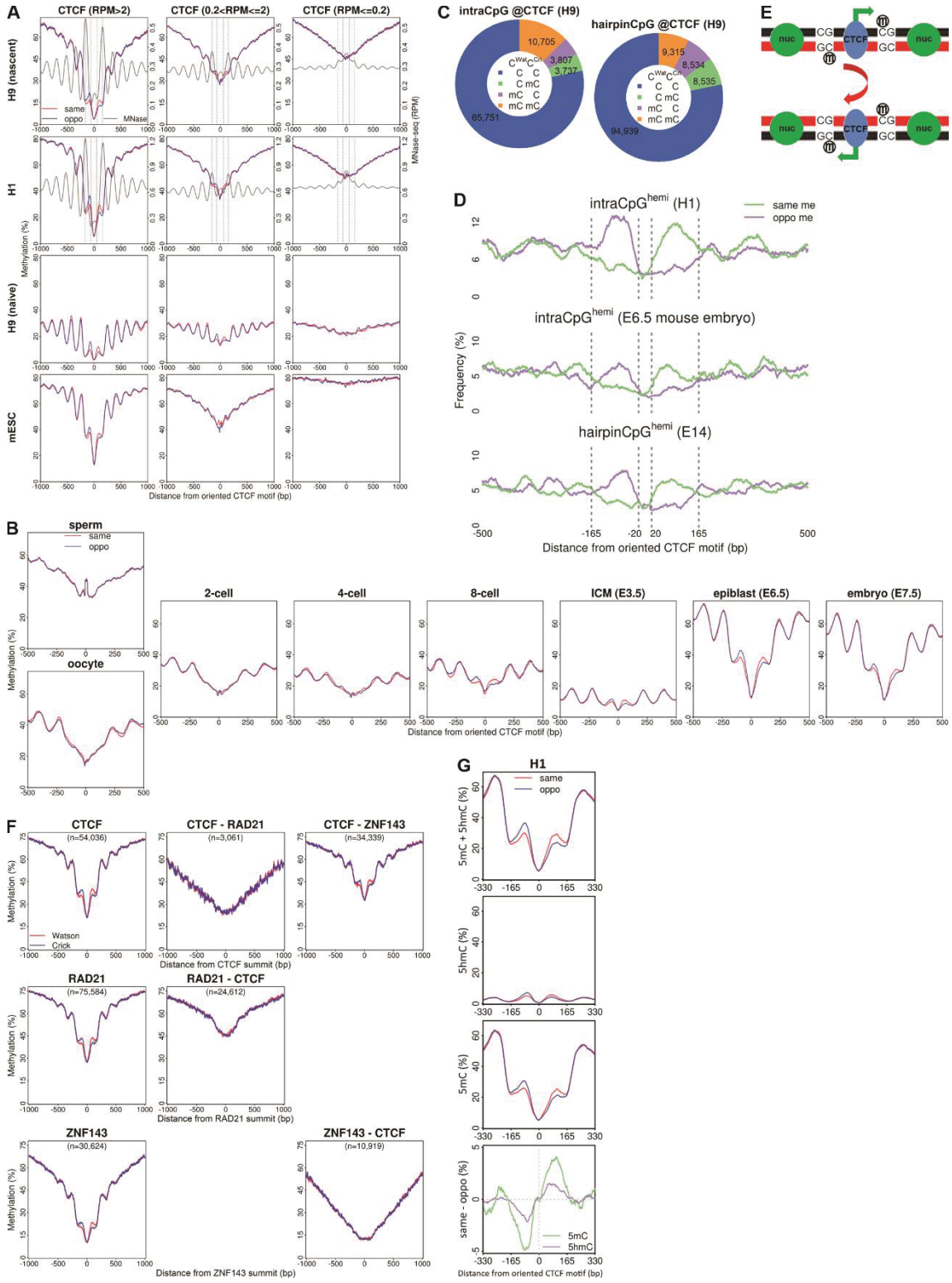


Fig. S8. Stable inheritance of enriched hemiCpGs flanking CTCF/cohesin binding sites.

(**A**) Average profiles of the motif (same) or opposite (oppo) strand methylation, and MNase-seq around oriented CTCF motifs with high, medium or low CTCF occupancy in

H9 hESC nascent chromatin, H1 hESC, naïve H9 hESC, and 159-2 mESC. (**B**) Average methylation profile of the motif or opposite strand at mouse early embryonic stages around oriented CTCF motifs occupied in mESC. (**C**) The fraction of all four types of intraCpGs from nasBS-seq and hairpinCpGs from CTCF ChIP-hairpinBS-seq in H9 hESC around occupied CTCF motifs. (**D**) The frequency of motif or opposite strand-methylated (same me or oppo me) intraCpG^{hemi} around occupied CTCF motifs in H1 hESC and E6.5 mouse embryos. The frequency of hairpinCpG^{hemi} from hairpinBS-seq in E14 mESC is also shown. (**E**) Schematic diagram showing that conformation of hemiCpGs around CTCF motifs is the same after rotating the DNA duplex by 180°. Hence, the conformation of hemiCpGs is irrespective of CTCF motif orientation marked by the green arrows. (**F**) Average methylation profiles of Watson and Crick strands around CTCF, RAD21 or ZNF143 peak summits with or without the occupancy of other factors in H1 hESC. The lists of peak summits were obtained from ENCODE. (**G**) Average profile of 5hmC (from TAB-seq) and 5mC (methylation frequency of WGBS subtracted by that of TAB-seq) from the motif or opposite strand around occupied CTCF motifs in H1 hESC. Methylation profiles of motif strand subtracted by opposite strand is also shown.

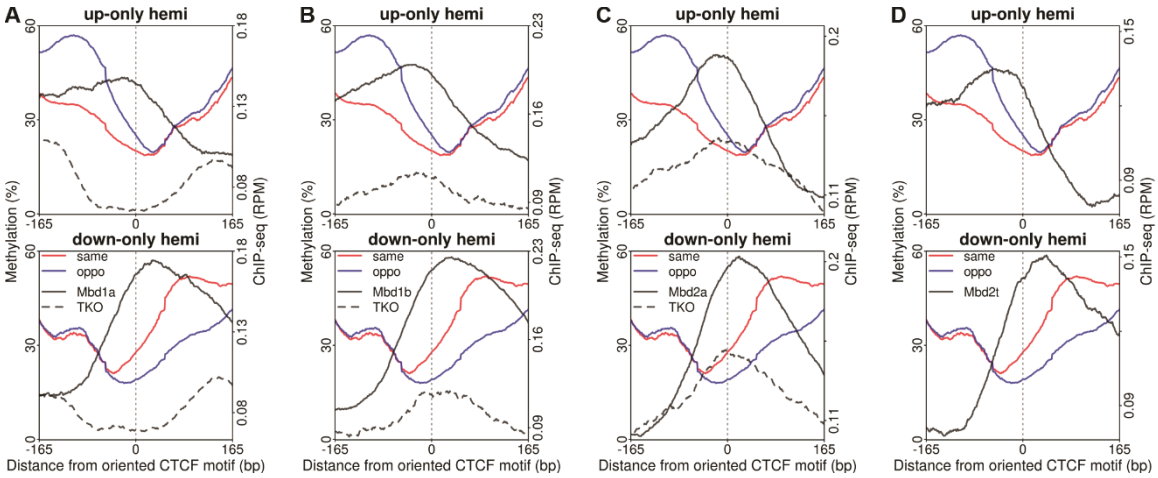


Fig. S9. Co-localization of MBD proteins with hemi-methylated sites flanking CTCF/cohesin binding sites.

(**A-D**) Occupancy of Mbd1a (**A**), Mbd1b (**B**), Mbd2a (**C**), and Mbd2t (**D**) in WT mESC, and DNMT1/3A/3B triple-knockout (TKO) mESC was profiled around CTCF motifs showing hemi-methylation upstream- or downstream-only in WT mESC. See legends in the bottom panels.

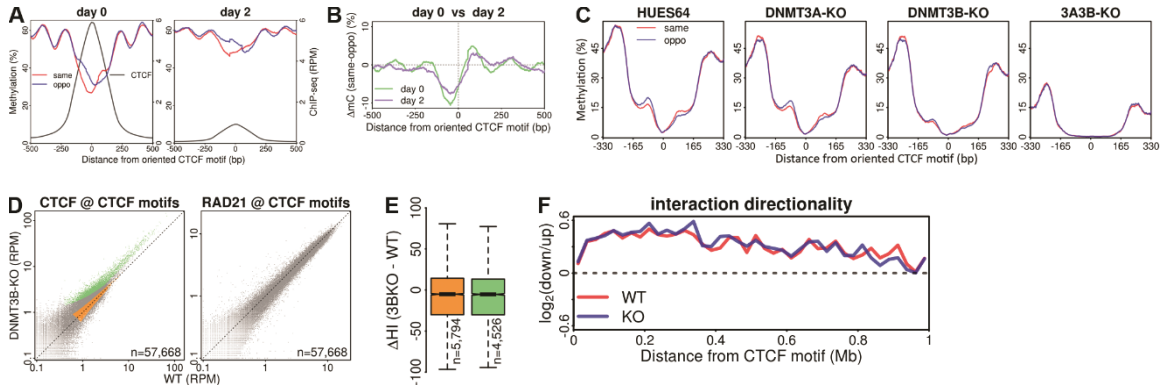


Fig. S10. Loss of DNMT3B has very subtle effect on CTCF/cohesin occupancy.

(**A**) Average profiles of the motif (same) or opposite (oppo) strand methylation, and CTCF ChIP-seq around occupied CTCF motifs in E14 mESC without (day 0) or with induced CTCF degradation (day 2). (**B**) Methylation profiles of motif strand subtracted by opposite strand in **A** is shown. (**C**) Average methylation profiles of the motif or opposite strand around occupied CTCF motifs in HUES64 hESC and its DNMT3A-KO, DNMT3B-KO, DNMT3A/3B-double KO counterpart cells. (**D**) Reads per million (RPM) values of CTCF motifs from CTCF or RAD21 ChIP-seq were compared between WT and DNMT3B-KO HUES64 hESC. (**E**) The changes of Hemi Index under DNMT3B-KO were compared between CTCF motifs with subtle increase (green in **D**) and with no changes (orange in **D**), and show no difference. (**F**) The ratio of downstream over upstream interaction contacts emanating from occupied CTCF motifs within a ±1 Mb window from RAD21 Hi-ChIP in WT and DNMT3B-KO HUES64 hESC is shown.

References

1. R. Holliday, J. E. Pugh, DNA modification mechanisms and gene activity during development. Science 187, 226-232 (1975).
2. J. A. Law, S. E. Jacobsen, Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet 11, 204-220 (2010).
3. L. S. Chuang et al., Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. Science 277, 1996-2000 (1997).
4. M. Bostick et al., UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science 317, 1760-1764 (2007).
5. J. Sharif et al., The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. Nature 450, 908-912 (2007).
6. G. Liang et al., Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. Mol Cell Biol 22, 480-491 (2002).
7. H. Gowher, A. Jeltsch, Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. J Mol Biol 309, 1201-1208 (2001).
8. A. Aoki et al., Enzymatic properties of de novo-type mouse DNA (cytosine-5) methyltransferases. Nucleic Acids Res 29, 3506-3512 (2001).
9. T. Chen, Y. Ueda, J. E. Dodge, Z. Wang, E. Li, Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. Mol Cell Biol 23, 5594-5605 (2003).
10. A. Tsumura et al., Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes Cells 11, 805-814 (2006).
11. J. Arand et al., In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. PLoS Genet 8, e1002750 (2012).
12. J. Liao et al., Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. Nat Genet 47, 469-478 (2015).
13. Y. He, J. R. Ecker, Non-CG Methylation in the Human Genome. Annu Rev Genomics Hum Genet 16, 55-77 (2015).
14. Z. Shipony et al., Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. Nature 513, 115-119 (2014).
15. L. Wang et al., Programming and inheritance of parental DNA methylomes in mammals. Cell 157, 979-991 (2014).
16. M. Yu et al., Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell 149, 1368-1380 (2012).
17. S. Jeong et al., Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. Mol Cell Biol 29, 5366-5376 (2009).
18. S. Sharma, D. D. De Carvalho, S. Jeong, P. A. Jones, G. Liang, Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. PLoS Genet 7, e1001286 (2011).
19. C. T. Ong, V. G. Corces, CTCF: an architectural protein bridging genome topology and function. Nat Rev Genet 15, 234-246 (2014).
20. E. P. Consortium, An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74 (2012).

21. R. R. Meehan, J. D. Lewis, S. McKay, E. L. Kleiner, A. P. Bird, Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell 58, 499-507 (1989).
22. H. W. Gabel et al., Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. Nature 522, 89-93 (2015).
23. T. Baubec, R. Ivanek, F. Lienert, D. Schubeler, Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell 153, 480-492 (2013).
24. K. D. Kernohan et al., ATRX partners with cohesin and MeCP2 and contributes to developmental silencing of imprinted genes in the brain. Dev Cell 18, 191-202 (2010).
25. K. D. Kernohan, D. Vernimmen, G. B. Gloor, N. G. Berube, Analysis of neonatal brain lacking ATRX or MeCP2 reveals changes in nucleosome density, CTCF binding and chromatin looping. Nucleic Acids Res 42, 8356-8368 (2014).
26. S. Horike, S. Cai, M. Miyano, J. F. Cheng, T. Kohwi-Shigematsu, Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. Nat Genet 37, 31-40 (2005).
27. E. P. Nora et al., Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell 169, 930-944 e922 (2017).
28. C. D. Laird et al., Hairpin-bisulfite PCR: assessing epigenetic methylation patterns on complementary strands of individual DNA molecules. Proc Natl Acad Sci U S A 101, 204-209 (2004).
29. L. Zhao et al., The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. Genome Res 24, 1296-1307 (2014).
30. E. P. Nora et al., Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell 169, 930-944 e922 (2017).
31. L. Zhao et al., The dynamics of DNA methylation fidelity during mouse embryonic stem cell self-renewal and differentiation. Genome Res 24, 1296-1307 (2014).
32. M. J. Rowley et al., Evolutionarily Conserved Principles Predict 3D Chromatin Organization. Mol Cell 67, 837-852 e837 (2017).
33. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120 (2014).
34. F. Krueger, S. R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics 27, 1571-1572 (2011).
35. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-359 (2012).
36. H. Li et al., The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079 (2009).
37. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841-842 (2010).
38. Y. Zhang et al., Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137 (2008).
39. F. Ramirez, F. Dundar, S. Diehl, B. A. Gruning, T. Manke, deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 42, W187-191 (2014).

40. A. J. Saldanha, Java Treeview--extensible visualization of microarray data. Bioinformatics 20, 3246-3248 (2004).
41. H. S. Rhee, B. F. Pugh, Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147, 1408-1419 (2011).
42. N. C. Durand et al., Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst 3, 95-98 (2016).
43. N. C. Durand et al., Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst 3, 99-101 (2016).
44. G. Elliott et al., Intermediate DNA methylation is a conserved signature of genome regulation. Nat Commun 6, 6363 (2015).