Supplementary Materials

Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development

Vladimir B. Teif[1,*], Daria A. Beshnova[1], Yevhen Vainshtein[2], Caroline Marth[1], Jan-Philipp Mallm[1], Thomas Höfer[2] and Karsten Rippe[1,*]

[1] Research Group Genome Organization and Function, Deutsches Krebsforschungszentrum (DKFZ) and BioQuant, Heidelberg, Germany. [2] Division Theoretical Systems Biology, Deutsches Krebsforschungszentrum (DKFZ) and BioQuant, Heidelberg, Germany.

*Correspondence should be addressed to Vladimir Teif (V.Teif@dkfz.de) or Karsten Rippe (Karsten.Rippe@dkfz.de).

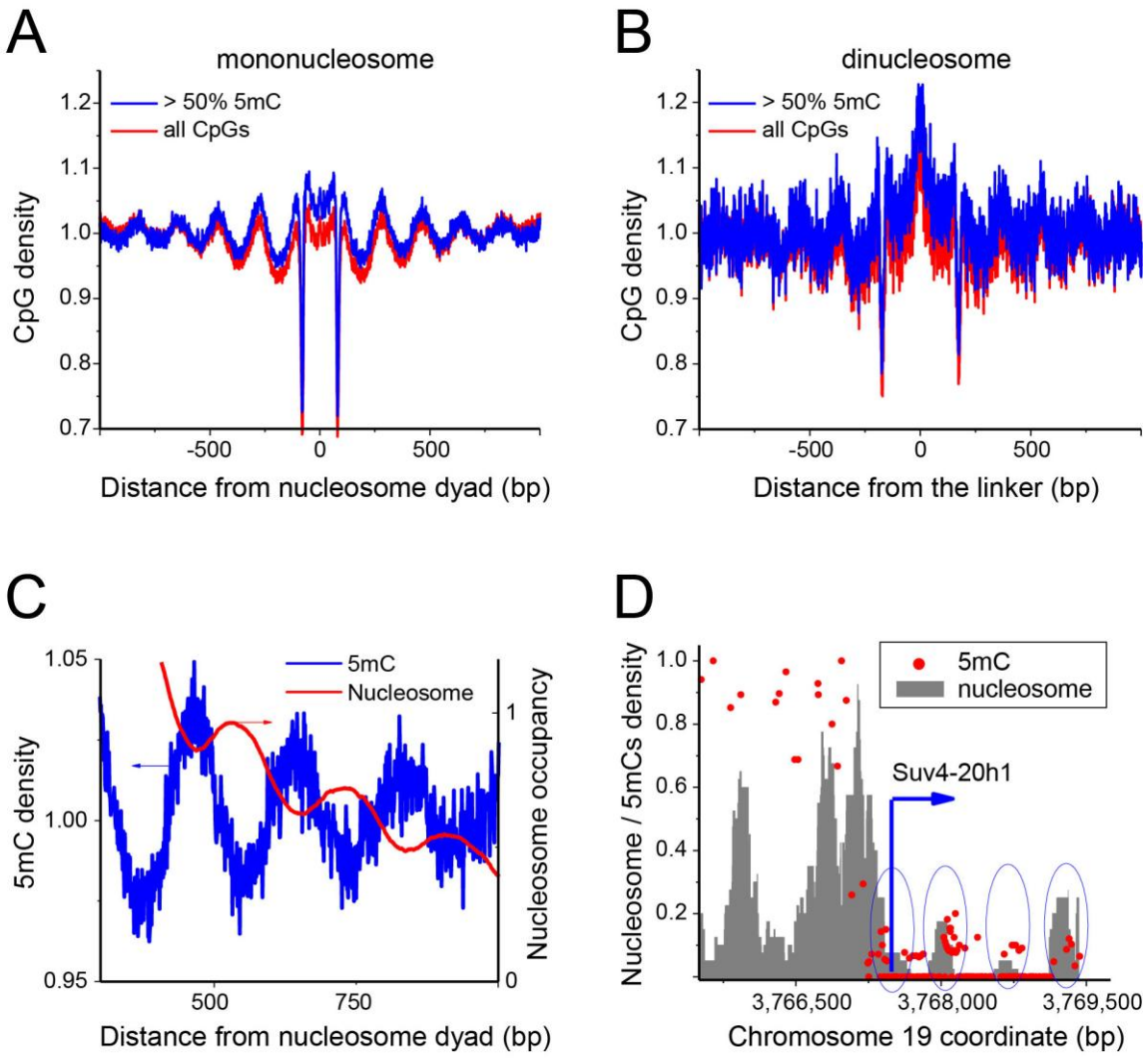

Figure S1. DNA methylation patterns along the nucleosome.

(A) Densities of all CpGs (red line) and methylated CpGs (blue line) calculated as average of all individual mononucleosome reads obtained from MNase-seq analysis. Methylated CpGs were defined as CpGs where the methylation level was >50%. (B) Same as in panel A, but for dinucleosomes. The linker between two nucleosomes in the dinucleosome has ~20% enrichment of DNA methylation. (C) Enlarged view of the DNA methylation density plot shown in panel A together with the nucleosome occupancy for the same region. The nucleosome occupancy is scaled as 100 x log (nucleosome occupancy). The original average nucleosome density oscillated in the interval (1, 1.04). (D) Exemplary region inside a CpG island, which shows the opposite trend to that found in panels A-C.

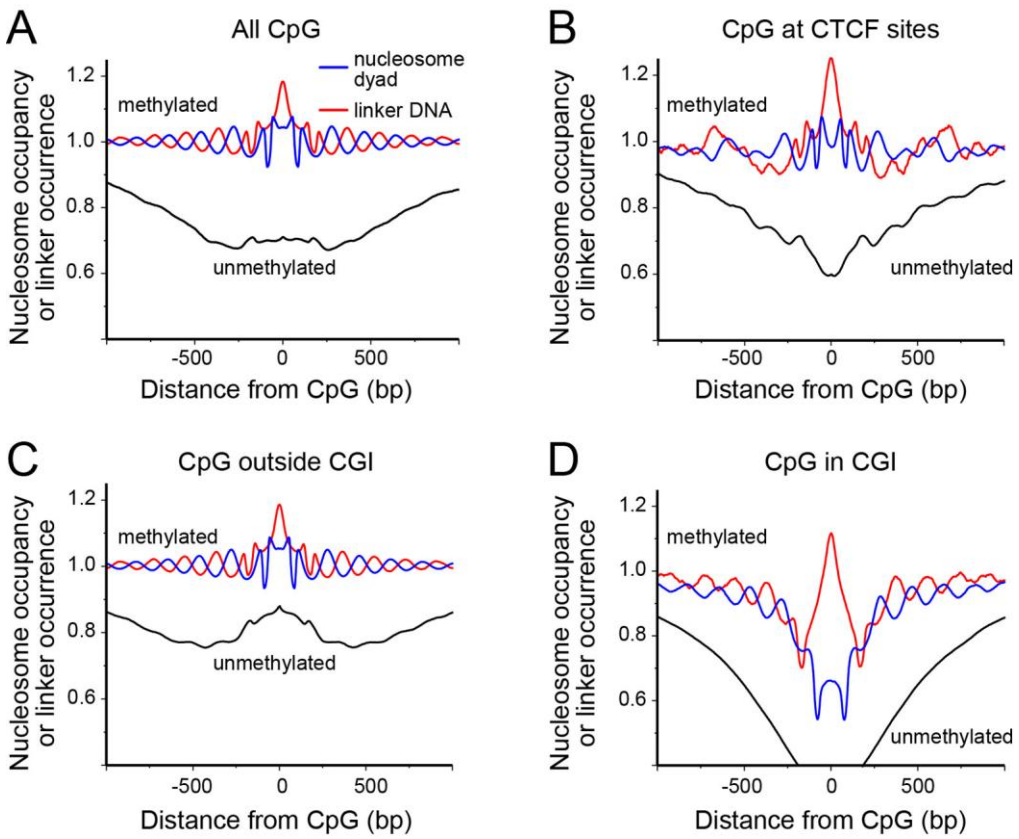

Figure S2. Nucleosome occupancy patterns at different classes of CpGs.

Black lines correspond to the average nucleosome occupancy level of unmethylated CpGs (methylation <10%) calculated for the mononucleosome preparation. Red lines refer to methylated CpGs (methylation >50%) and represent the probability to find the DNA linker as defined by a 40 bp region in the middle of mapped dinucleosomes at a given position relative to the 5mC site. Blue lines show the corresponding probability to position the nucleosome dyad as determined by the central 40 bp of a mononucleosome fragment. (A) All CpGs. (B) CpGs within a [-500 bp, 500 bp] interval around CTCF binding sites. (C) CpGs outside of CpG islands (CGI). (D) CpGs within CGIs. The minimum of the line of average nucleosome occupancy of unmethylated CpGs at CGIs decreases to 27% of the average genome-wide level.

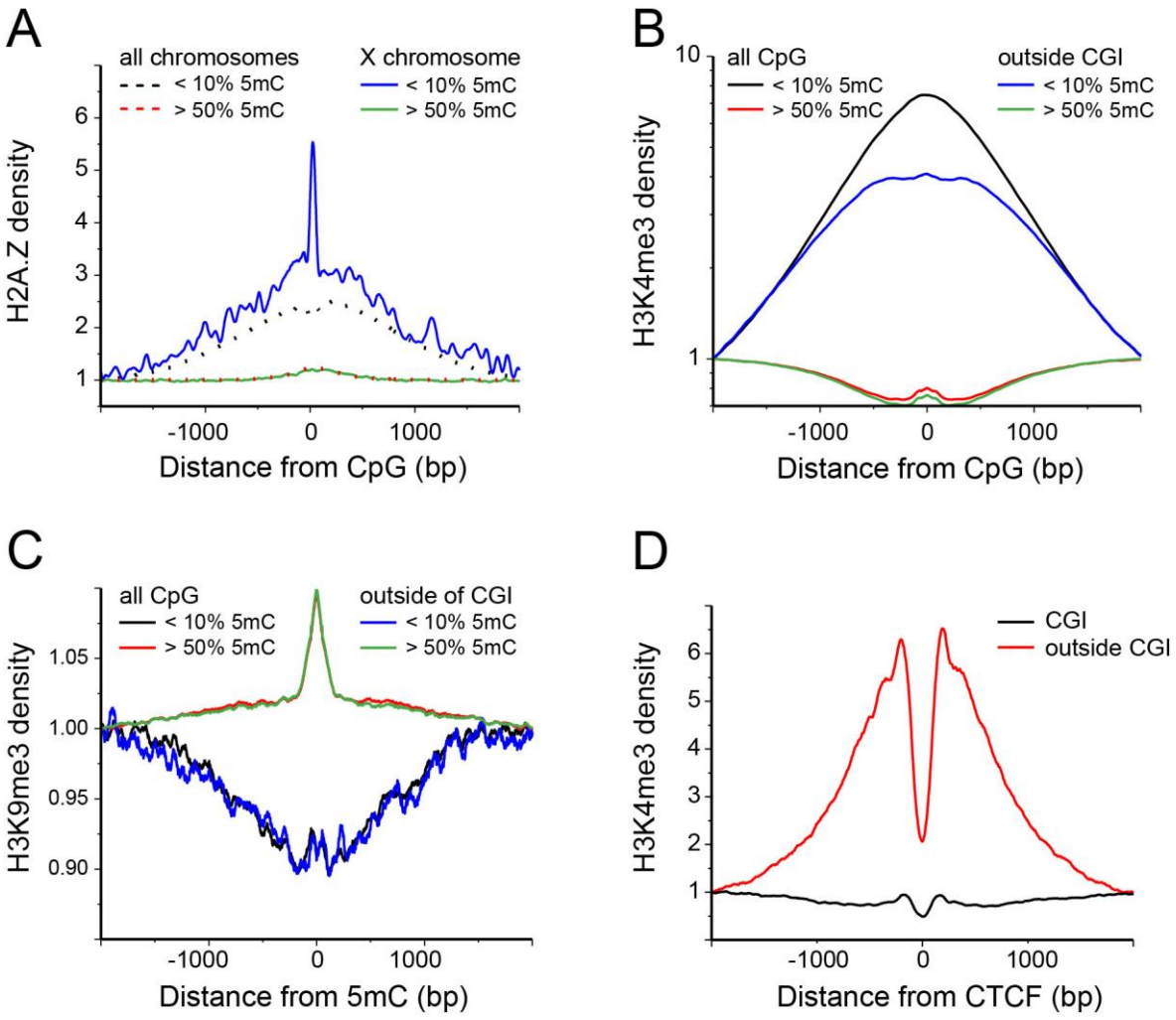

Figure S3. Densities of histone variant H2A.Z and histone H3K4me3 and H3K9me3 modifications around CpGs.

(A) Enrichment of histone variant H2A.Z as mapped by Ku et al. (Ku et al. 2012) at unmethylated (< 10% 5mC) and methylated (> 50% 5mC) CpGs. Dashed lines: all chromosomes, unmethylated CpGs (black) and methylated CpGs (red). Solid lines: chromosome X, unmethylated (blue) and methylated (green). (B) The density of the active chromatin mark H3K4me3 (determined in this work) around all unmethylated (black) and methylated CpGs (red) as well as unmethylated CpGs (blue) and methylated (green) CpGs outside of CGIs. (C) Density of the H3K9me3 modification (determined in this work) around CpGs. Same color code as in panel B. The modification was depleted around unmethylated CpGs. (D) H3K4me3 enrichment around CTCF binding sites inside and outside of CGIs.

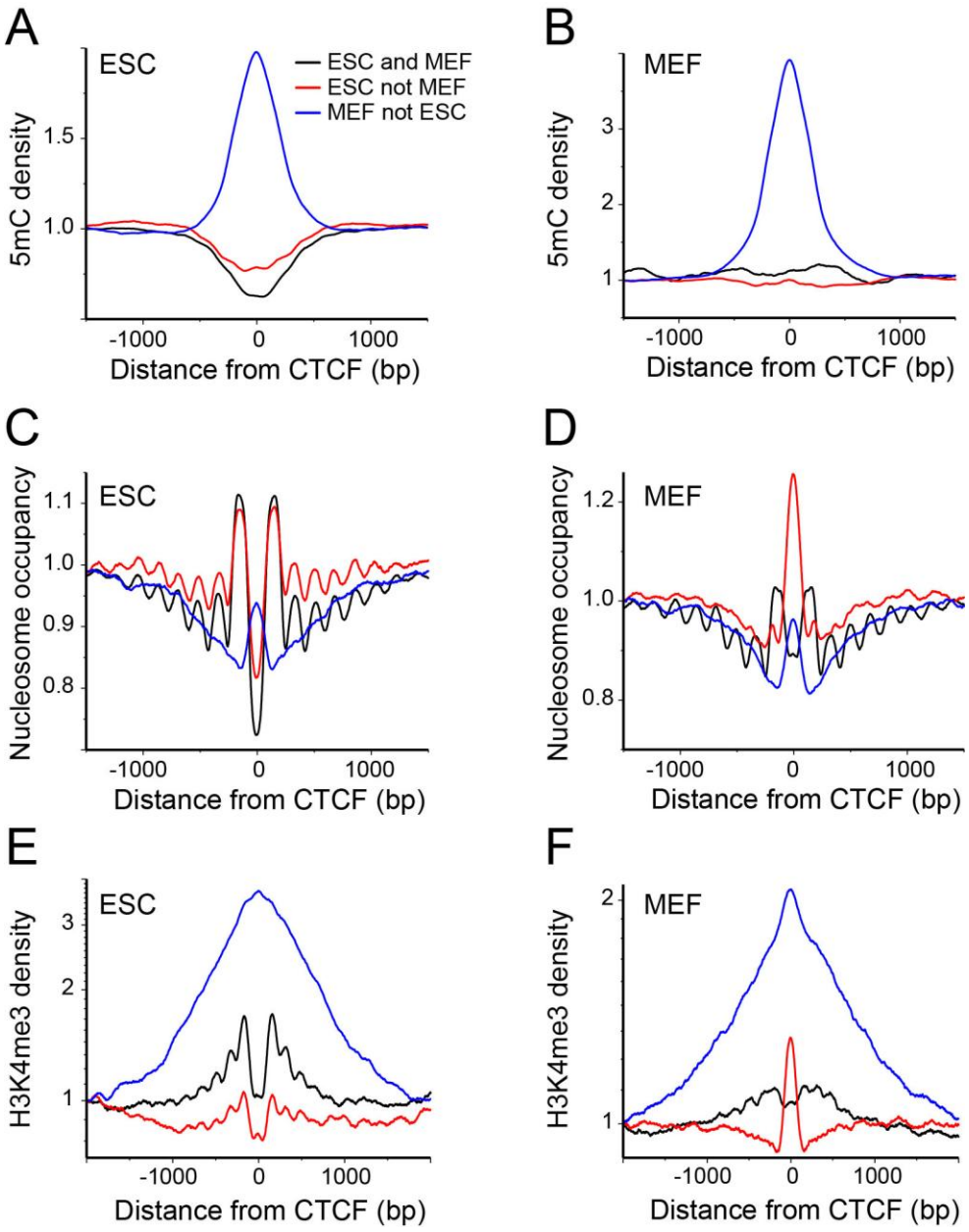

Figure S4. Nucleosome occupancy and 5mC or H3K4me3 density at CTCF binding sites. Constitutive CTCF sites occupied both in ESCs and in MEFs ("ESC and MEF", black line), variable sites ("ESC not MEF", red line) where CTCF occupancy decreases by more than 50% in MEFs, and a class of very weak sites, where CTCF occupancy increases from ~10% in ESCs to ~15% in MEFs ("MEF not ESC", blue line). (A) 5mC enrichment in ESCs from MeDIP experiments (Wilson et al., 2012) around different classes of CTCF sites. (B) Same as panel A but for MEFs. (C) Nucleosome occupancy around CTCF sites in ESCs according to the data set of Teif et al. (Teif et al, 2012). (D) Same as panel C but for MEFs. (E) H3K4me3 patterns around CTCF binding sites in ESCs. (F) Same as panel E but for MEFs.

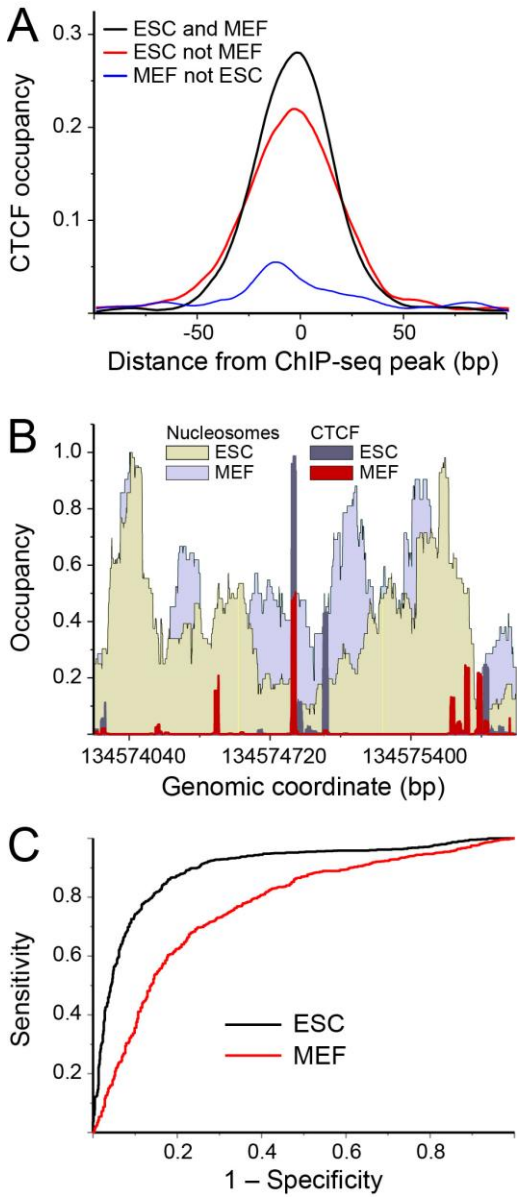

Figure S5. The interplay of CTCF binding and nucleosome positioning.

(A) CTCF binding profiles predicted from the DNA sequence using TFnuc algorithm without taking into account nucleosomes. The resulting profiles resemble experimental CTCF occupancies in ESCs but not in MEFs, which are plotted in Figure 6A. (B) Exemplary genomic region on chromosome 1 with experimentally determined nucleosome and CTCF occupancies. (C) Receiver-operator curves and the corresponding area under the curve (AUC) values for CTCF binding site prediction in ESCs and in MEFs using TFnuc algorithm without taking into account nucleosomes. Values of AUC = 0.90 (ESCs) and AUC = 0.77 (NPC) were calculated.

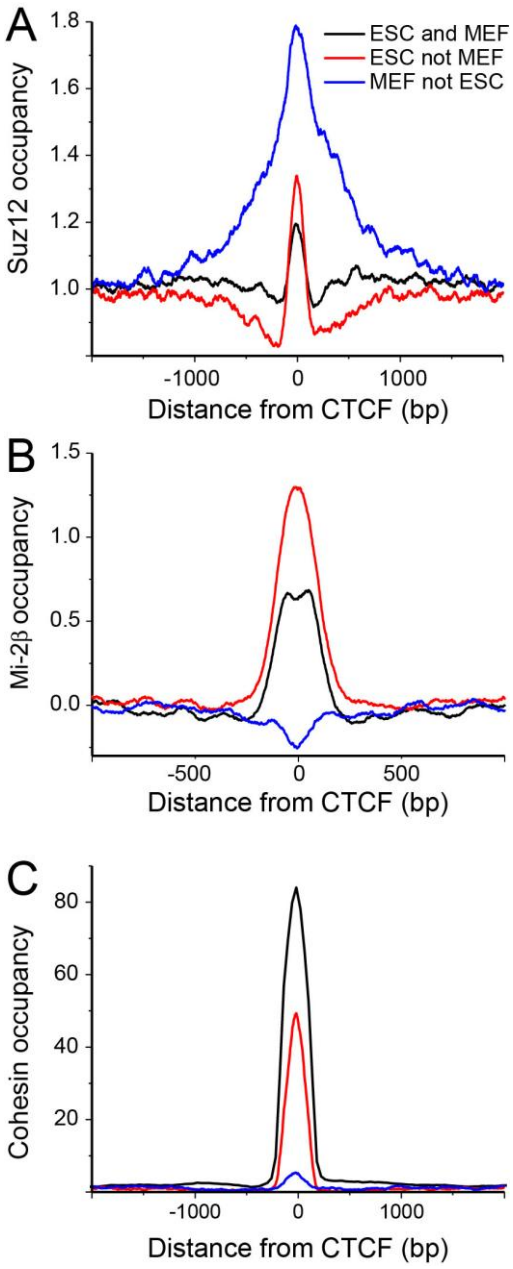

Figure S6. Suz12, Mi-2β and Cohesin occupancy around CTCF binding sites in ESCs.

(A) Enrichment of PRC2 subunit SUZ12 at CTCF sites calculated with SUZ12 ChIP-seq data from Hu et al. (Hu et al. 2013). SUZ12 enrichment was highest at weak "MEF not ESC" sites, suggesting that these regions undergo chromatin compaction by PRC2 complexes. (B) Enrichment of Mi-2β (also known as Chd4) around CTCF binding sites calculated with Mi-2β ChIP-seq data from Whyte et al. (Whyte et al. 2012). Mi-2β enrichment was highest at variable "ESC not MEF" sites. (C) Cohesin enrichment around CTCF sites calculated with cohesin ChIP-seq data from Kagey et al. (Kagey et al. 2010). Cohesin binding affinity correlated with that of CTCF.

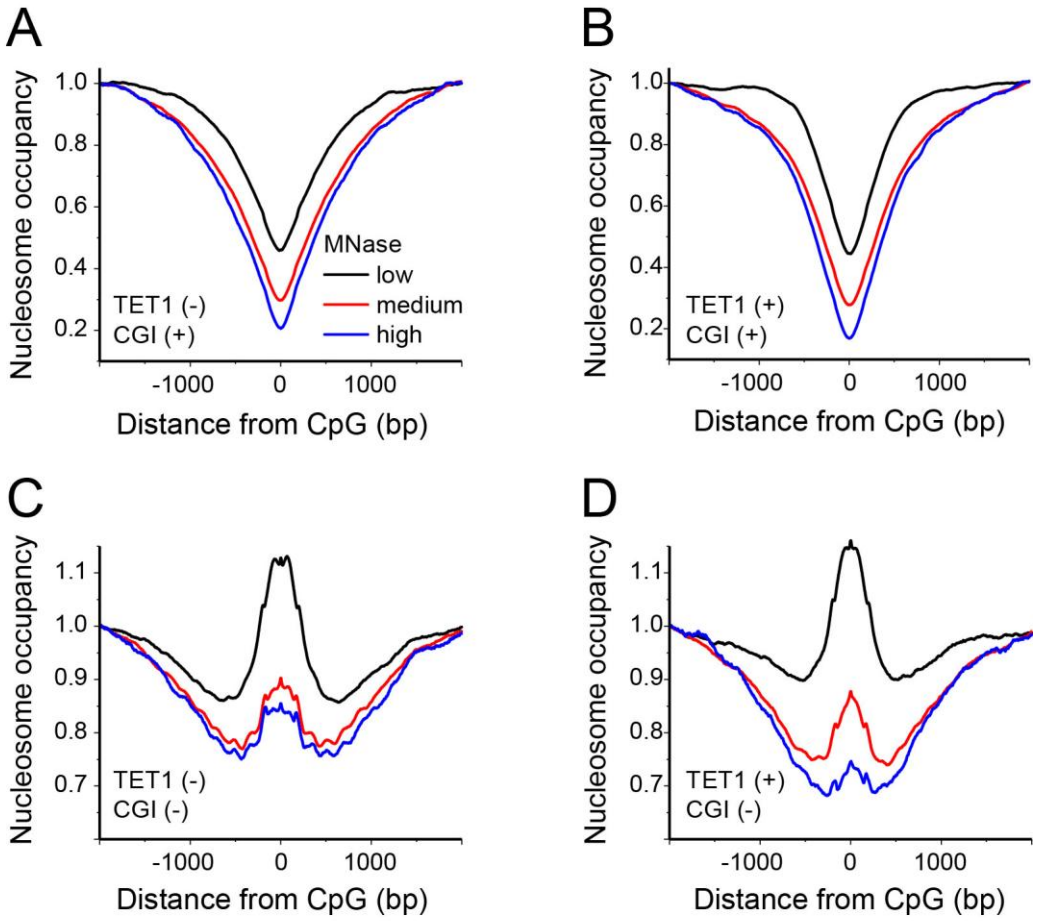

Figure S7. Nucleosome occupancy around different classes of CpGs that lack cytosine (hydroxy)methylation.

Nucleosome occupancy was determined at low (black), medium (red) and high (blue) degrees of MNase digestion. The presence/absence of TET1 was derived ChIP-seq data (Yu et al. 2012). (A) CGIs without TET1. (B) CGIs with TET1. (C) CpGs without TET1 outside of CGIs. (D) CpGs with TET1 outside of CGIs.

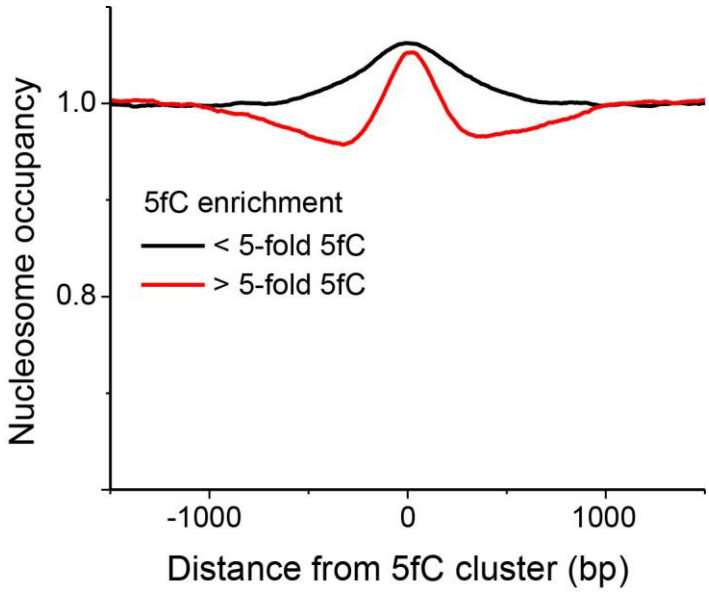

Figure S8. **Average nucleosome occupancy around 5fC sites in ESCs.**

Nucleosome occupancy in regions with low (black line) and high (red line) 5fC density. The 5fC data were from Raiber et al. (Raiber et al. 2012). The coordinates of the regions and their relative 5fC enrichment values were quantified using the RSEG peak calling software (Song and Smith 2011) as provided by the authors of this publication.

Table S1. Data sources

What is measured	Experiment type	Cell type	GEO accession #
Mononucleosomes	MNase-seq, high and medium MNase	ESC, NPC, MEF	GSE40951
Mononucleosomes	MNase-seq, Low MNase	ESC	GSE56938
Dinucleosomes	MNase-seq	ESC	
H3K4me3	ChIP-seq	ESC	
H3K9me3	ChIP-seq	ESC	
5mC	Bis-seq	ESC, NPC	GSE30206
5mC	MeDIP	ESC, MEF	GSE27468
5hmC	TAB-seq	ESC	GSM882244
5hmC	hMeDIP	ESC, NPC	GSM1002268
H2A.Z	ChIP-seq	ESC	GSE39237
Mi-2β	ChIP-seq	ESC	GSM687289
TET1	ChIP-seq	ESC	GSE24843
CTCF	ChIP-seq	NPC	GSE36203
CTCF	ChIP-seq	ESC	GSM918743
CTCF	ChIP-seq	MEF	GSM918748
SUZ12	ChIP-seq	ESC	GSM970527
Cohesin	ChIP-seq	ESC	GSE22557

Supplementary References

Hu G, Cui K, Northrup D, Liu C, Wang C, Tang Q, Ge K, Levens D, Crane-Robinson C, Zhao K. 2013. H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* **12**(2): 180-192.

Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS et al. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**(7314): 430-435.

Ku M, Jaffe JD, Koche RP, Rheinbay E, Endoh M, Koseki H, Carr SA, Bernstein BE. 2012. H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. *Genome Biol* **13**(10): R85.

Raiber EA, Beraldi D, Ficz G, Burgess HE, Branco MR, Murat P, Oxley D, Booth MJ, Reik W, Balasubramanian S. 2012. Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol* **13**(8): R69.

Song Q, Smith AD. 2011. Identifying dispersed epigenomic domains from ChIP-Seq data. *Bioinformatics* **27**(6): 870-871.

Whyte WA, Bilodeau S, Orlando DA, Hoke HA, Frampton GM, Foster CT, Cowley SM, Young RA. 2012. Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**(7384): 221-225.

Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B et al. 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**(6): 1368-1380.