Vol 462|19 November 2009|doi:10.1038/nature08514 nature

ARTICLES

Human DNA methylomes at base resolution show widespread epigenomic differences

Ryan Lister[1]*, Mattia Pelizzola[1]*, Robert H. Dowen[1], R. David Hawkins[2], Gary Hon[2], Julian Tonti-Filippini[4], Joseph R. Nery[1], Leonard Lee[2], Zhen Ye[2], Que-Minh Ngo[2], Lee Edsall[2], Jessica Antosiewicz-Bourget[5,6], Ron Stewart[5,6], Victor Ruotti[5,6], A. Harvey Millar[4], James A. Thomson[5,6,7,8], Bing Ren[2,3] & Joseph R. Ecker[1]

DNA cytosine methylation is a central epigenetic modification that has essential roles in cellular processes including genome regulation, development and disease. Here we present the first genome-wide, single-base-resolution maps of methylated cytosines in a mammalian genome, from both human embryonic stem cells and fetal fibroblasts, along with comparative analysis of messenger RNA and small RNA components of the transcriptome, several histone modifications, and sites of DNA–protein interaction for several key regulatory factors. Widespread differences were identified in the composition and patterning of cytosine methylation between the two genomes. Nearly one-quarter of all methylation identified in embryonic stem cells was in a non-CG context, suggesting that embryonic stem cells may use different methylation mechanisms to affect gene regulation. Methylation in non-CG contexts showed enrichment in gene bodies and depletion in protein binding sites and enhancers. Non-CG methylation disappeared upon induced differentiation of the embryonic stem cells, and was restored in induced pluripotent stem cells. We identified hundreds of differentially methylated regions proximal to genes involved in pluripotency and differentiation, and widespread reduced methylation levels in fibroblasts associated with lower transcriptional activity. These reference epigenomes provide a foundation for future studies exploring this key epigenetic modification in human disease and development.

Introduction

- DNA cythosine methylation is a central epigenetic modification
- it affects gene regulation ad cellular differentiation
- it alters chromatine density and accessibility to the DNA
- here we present the first genome wide, single-base resolution study of cytosine methylation in human embryonic stem cells and fetal fibroblasts
- this will show a difference in composition and patterning of mC between the two cell lines

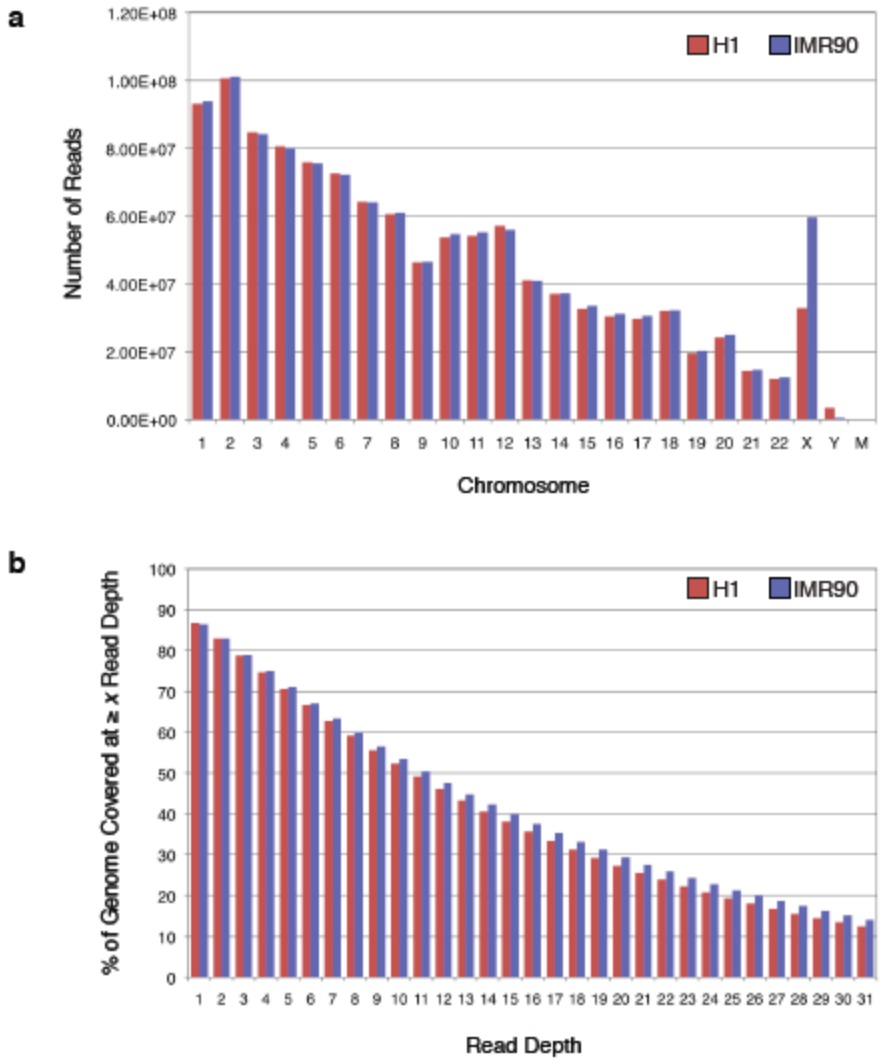
Methods

- Cellular cultures of human H1, H9, BMP-4 induced, IMR90 and iPS (IMR90) cells were used.
- Two biological replicates were prepared for each cell line.

- MethylC-Seq was performed: genomic DNA was fragmented by sonication (50-500 bp), then ends were repaired with a nucleotide triphosphate mix free of dCTP. Cytosine-methylated adapters were ligated for genomic DNA library construction. DNA was isolated by 2% agarose gel electrophoresis and sodium bisulphite (BS) conversion was performed, followed by four cycles of PCR amplification.
- mRNA-Seq, ChIP-Seq and smRNA-Seq were performed.
- MethylC-Seq libraries were sequenced using the Illumina Genome Analyzer II (GA II) and reads were aligned to the human reference genome hg18 using the Bowtie alignment algorithm. Methylated cytosines were identified at 1% false discovery rate (FDR).
- The reads from the two biological replicates were pooled to provide greater coverage for identification of the methylcytosines.

Results

Single-base-resolution maps for DNA methylation for two human cell lines



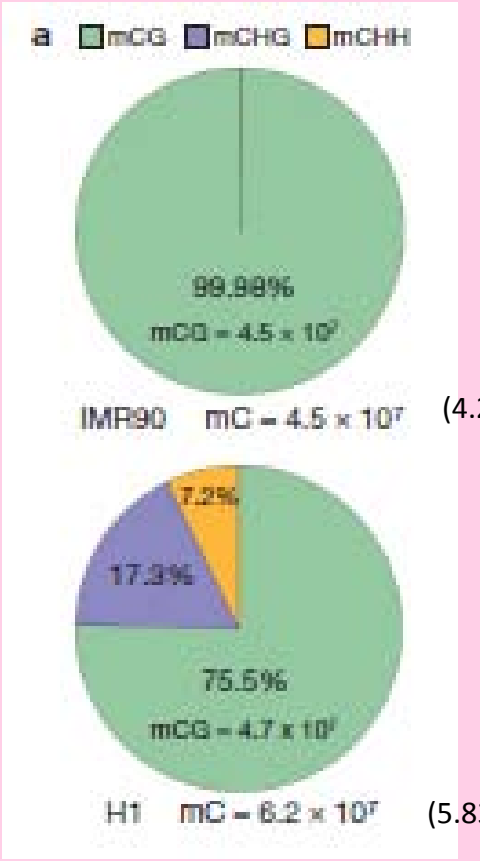
Supplementary Figure 1 | Uniquely Mapped Reads and Coverage for H1 MethylC-Seq. **a,** The number of uniquely mapped MethylC-seq reads for each chromosome of H1 and IMR90. **b,** The percent of the H1 and IMR90 genomes that is covered by differing minimum number of MethylC-seq reads.

- H1 human embryonic stem cells
- IMR90 fetal lung fibroblasts

- DNA is treated with BS then MethylC-Seq is performed
- This generates 1.16 and 1.18 billion reads
- The sequence yield is 87.5 and 91.0 Gb
- Average read depth is of 14.2X and 14.8X

- Over 86% of the genome is covered by at least one read, accounting for 94% of the total cytosines

Non-CG methylation is probably a general feature of human embryonic stem cells



(4.25% of C)

(5.83 % of C)

Figure 1 | Global trends of human DNA methylomes. a, The percentage of methylcytosines identified for H1 and IMR90 cells in each sequence context.

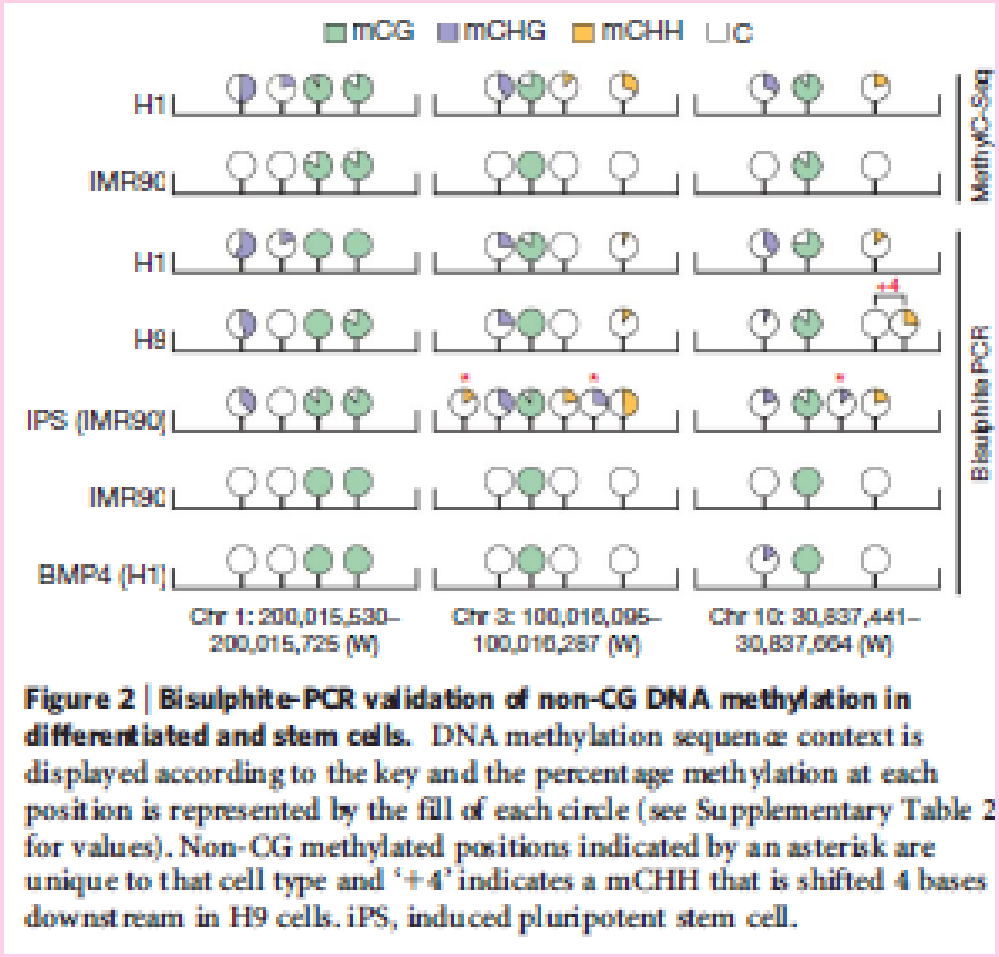
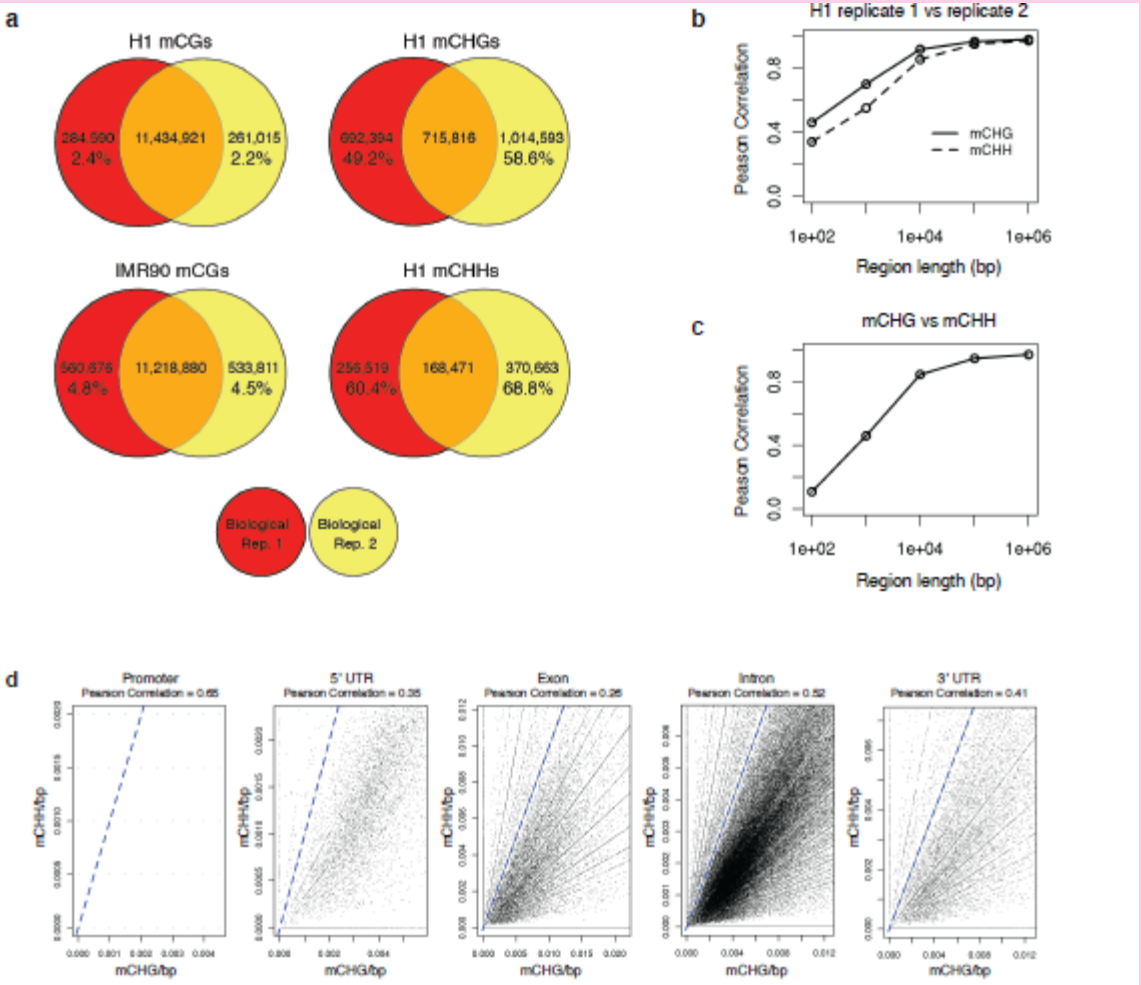


Figure 2 | Bisulphite-PCR validation of non-CG DNA methylation in differentiated and stem cells. DNA methylation sequence context is displayed according to the key and the percentage methylation at each position is represented by the fill of each circle (see Supplementary Table 2 for values). Non-CG methylated positions indicated by an asterisk are unique to that cell type and '+4' indicates a mCHH that is shifted 4 bases downstream in H9 cells. iPS, induced pluripotent stem cell.

- Like IMR90 cells, BMP4-induced H1 cells lost non-CG methylation
- induced pluripotent stem (iPS) cells restore non-CG methylation



Supplementary Figure 2 | Direct Overlap in Methylcytosines Between the H1 and IMR90 Cell Types, and Regional Correlation of non-CG Methylation Between Biological Replicates and mCHG/mCHH. a, Methylcytosines with similar sequencing depth were compared and classified as unique to biological replicate 1 (red), unique to replicate 2 (yellow) or common to both replicates (orange). The number of methylcytosines in each category is listed, as well as the percent methylcytosines unique within each biological replicate. **b,** Pearson correlation of the density of non-CG methylation sites within adjacent regions of chromosome 1 of varying length between the two H1 biological replicates. The correlation was determined independently for mCHG and mCHH. **c,** Pearson correlation was computed as in panel **b**, comparing mCHG to mCHH density from methylcytosine sites identified in the composite of the two biological replicates. **d,** Scatter plot of mCHG and mCHH density for each promoter, 5' UTR, exon, intron and 3'UTR occurrence. A blue dashed line with slope 1 along regions with equal mCHG and CHH density is displayed. Pearson correlation is reported in the plot title.

Two biological replicates with different passage number are used for each cell line

This reveales a high concordance of cytosine methylation status between replicates

OCT4 gene shows different methylation between H1 and IMR90

There is a ~50-fold reduction of transcription in IMR90 cells

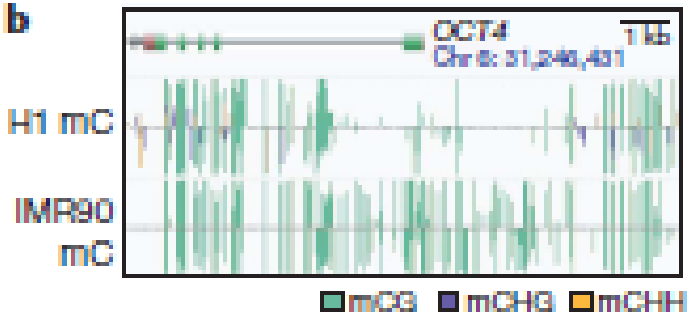
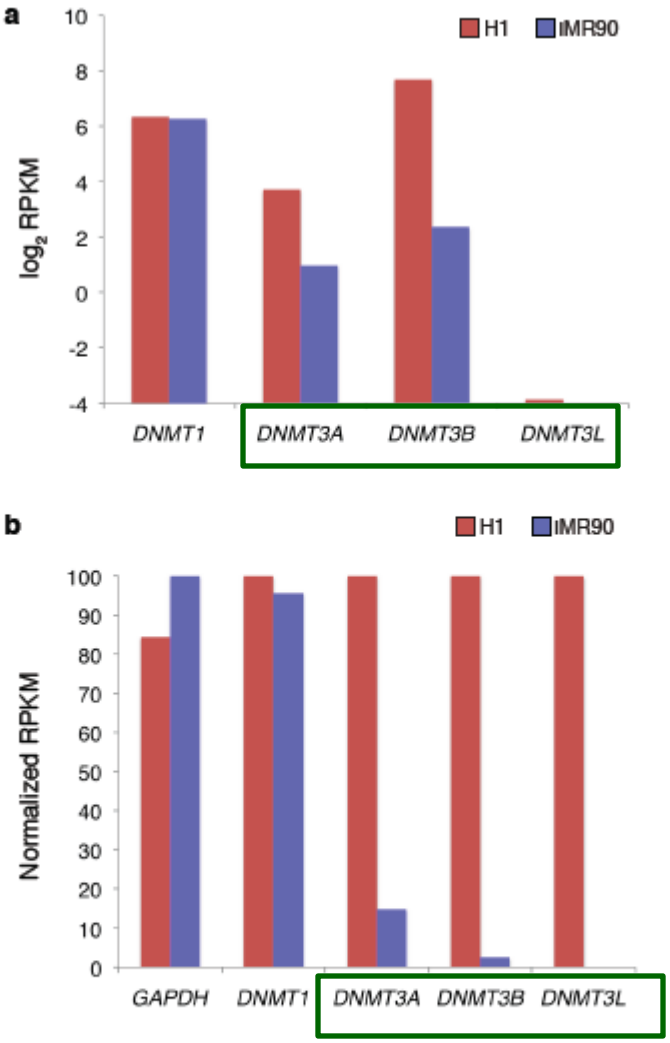


Figure 1 | Global trends of human DNA methylomes. b, Anno] browser representation of *OCT4*.

DNMT3A, 3B and 3L show lower expression in IMR90 cells, due to absence of mCHG and mCHH methylations



Supplementary Figure 3 | Differentually Expression of *DNMT* Genes In H1 and IMR90. **a,** log2RPKM and **b,** Maximum normalized RPKM measurements of transcript abundance for *DNMT1, DNMT3a, DNMT3b, DNMT3L* and *GAPDH* from RNA-seq. Abbreviations: RPKM, reads per kilobase of exon model per million mapped reads.

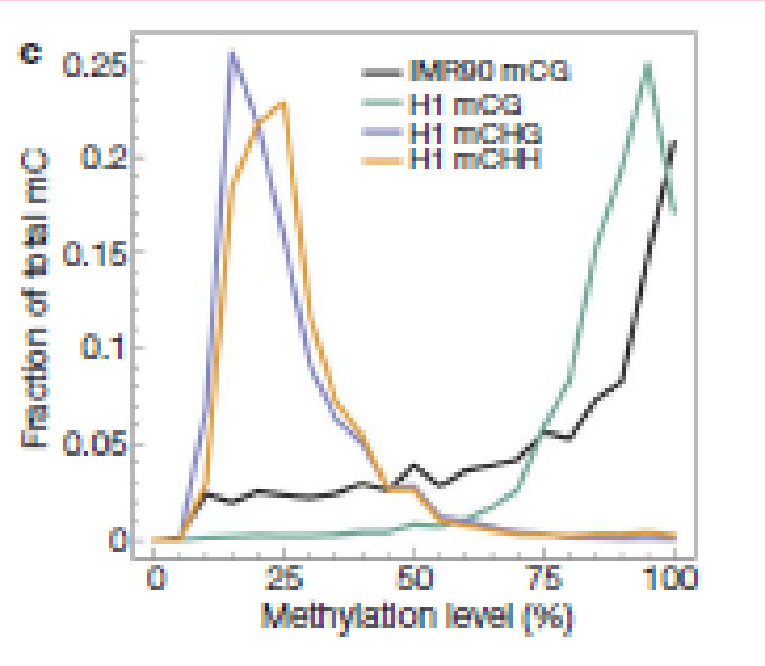


Figure 1 | Global trends of human DNA methylomes.

b, AnnoJ browser representation of *OCT4*. **c**, Distribution of the methylation level in each sequence context. The *y* axis indicates the fraction of all methylcytosines that display each methylation level (*x* axis), where methylation level is the mC/C ratio at each reference cytosine (at least 10 reads required). **d**, Blue dots indicate methylcytosine density in H1 cells in

H1 77% mCG 80-100% methylated
H1 85% mCHG and mCHH 10-40% methylated
IMR90 56% mCG 80-100% methylated

Although the total number of mCG sites in H1 and IMR90 cells is similar, IMR90 mCG sites are tipically less frequently methylated

82.7% and 67.7% of all CG sites are methylated in H1 and IMR90 cells, respectively

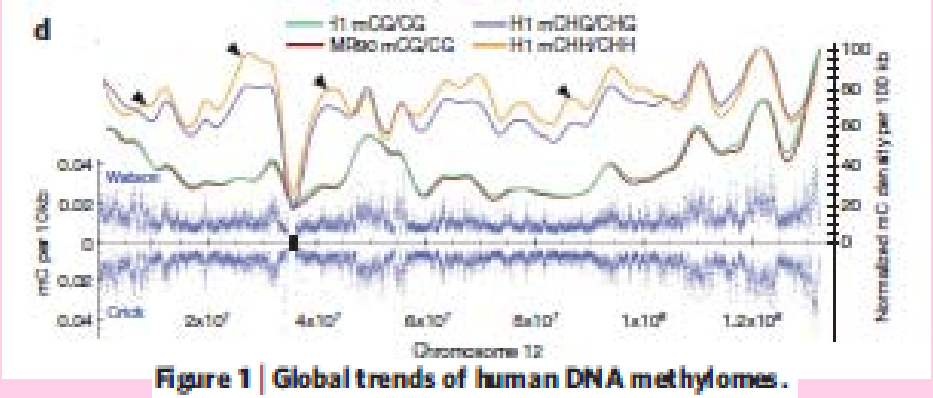
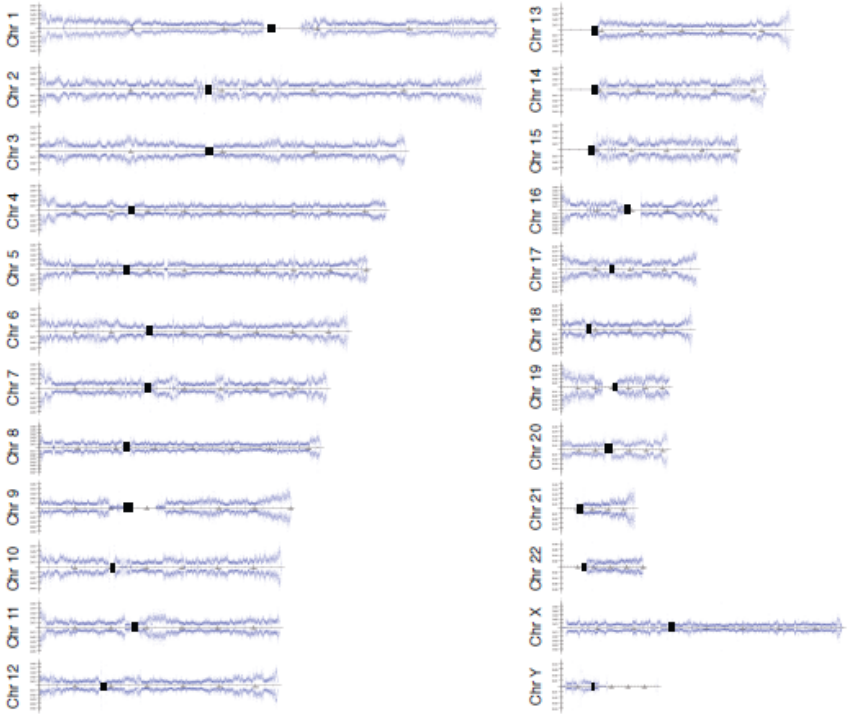


Figure 1 | Global trends of human DNA methylomes.

reads required). **d**, Blue dots indicate methylcytosine density in H1 cells in 10-kb windows throughout chromosome 12 (black rectangle, centromere). Smoothed lines represent the methylcytosine density in each context in H1 and IMR90 cells. Black triangles indicate various regions of contrasting trends in CG and non-CG methylation. mC, methylcytosine.



Supplementary Figure 4 | The Density of Methylcytosines Identified in All Chromosomes in H1 Cells. Blue dots indicate the density of all methylcytosines in 10 kb windows. Black rectangles indicate approximate centromere positions.

Density of methylcytosines identified in all chromosomes in H1 cells

- sub-telomeric regions frequently show higher DNA methylation density
- methylation is lower at the centromere
- mCG density profile of H1 and IMR90 is similar
- mCHG and mCHH densities show a moderate correlation (Pearson correlation = 0.5, See Supplementary figure 2d)

Pervasive non-CG DNA methylation in embryonic stem cells

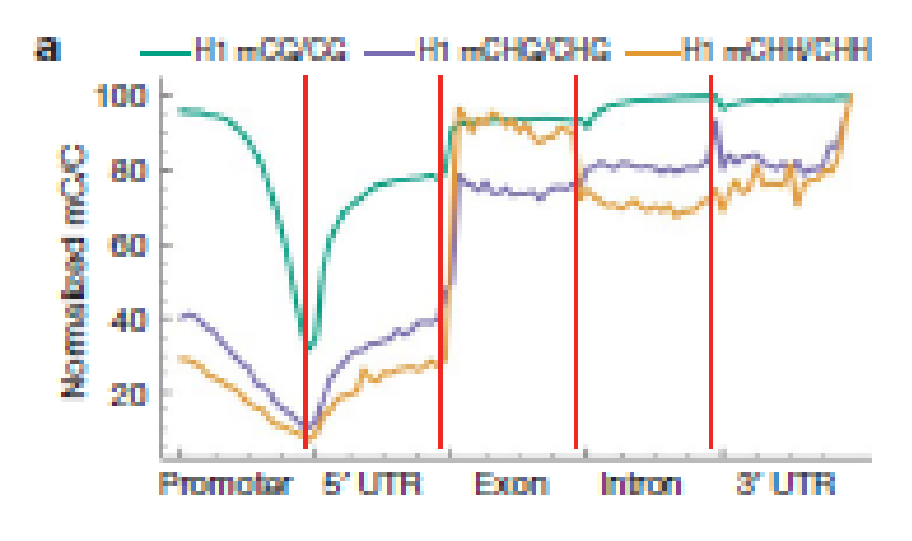
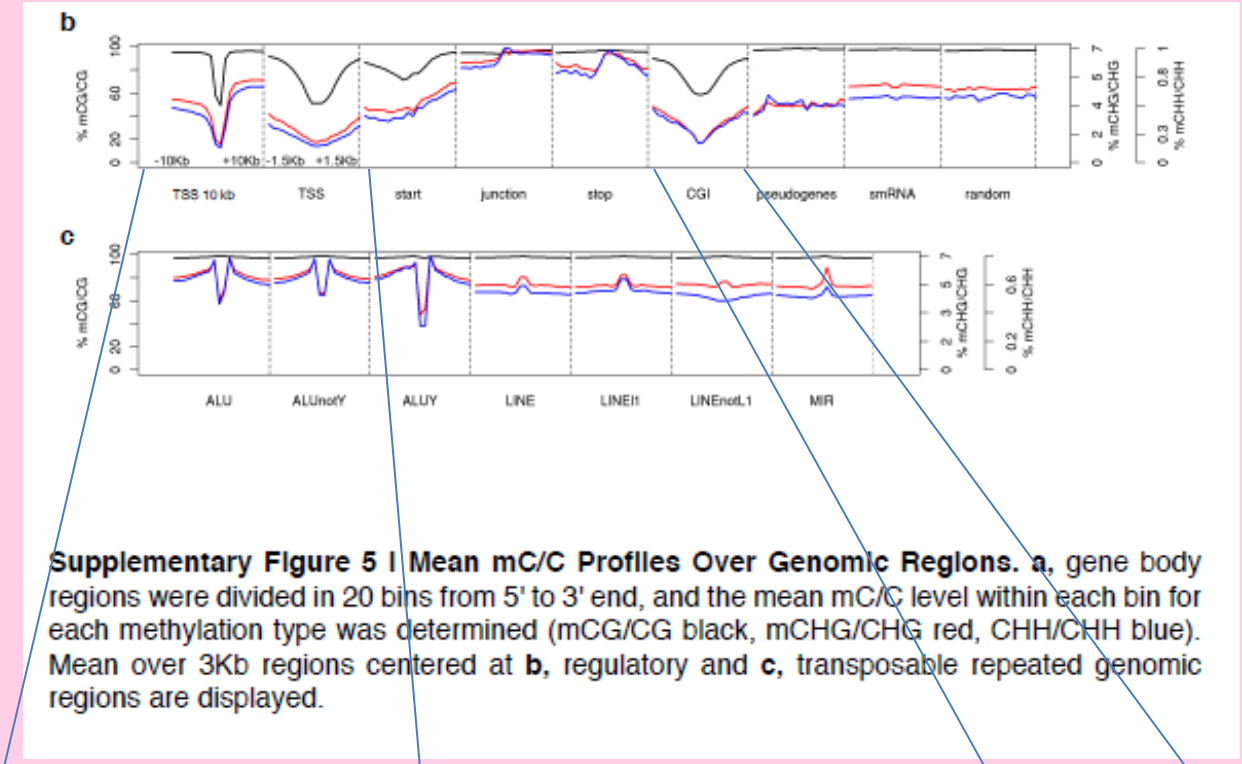


Figure 3 | Non-CG DNA methylation in H1 embryonic stem cells. a, Relative methylation density (the ratio of methylcytosines to reference cytosines) in H1 throughout different gene-associated regions (promoters encompass 2 kb upstream of the transcriptional start site). The mean mC/C profile was normalized to the maximum value. **b**, Relative methylation density within

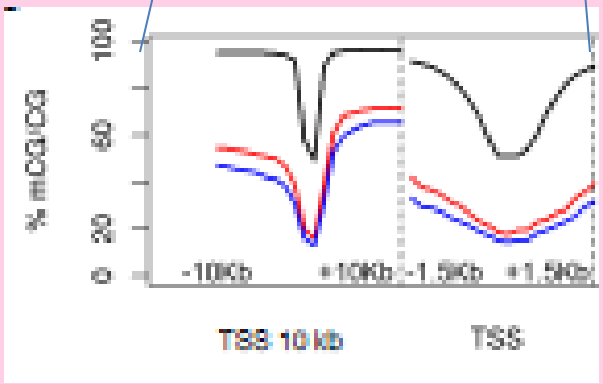
We observe a correlation in the density of mCG and the distance from the TSS:

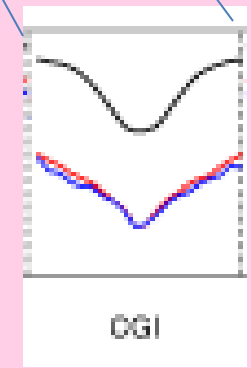
- it decreases from 2 kb upstream of the TSS
- it increases in 5'-UTR
- it reaches similar levels in exons, introns and 3'-UTR (as to 2 kb upstream of the TSS)

mCHG and mCHH densities within exons, introns and 3'-UTR are twice as high as in the promoter
mCHH density is 15-20% higher in exons than within introns and the 3'-UTR

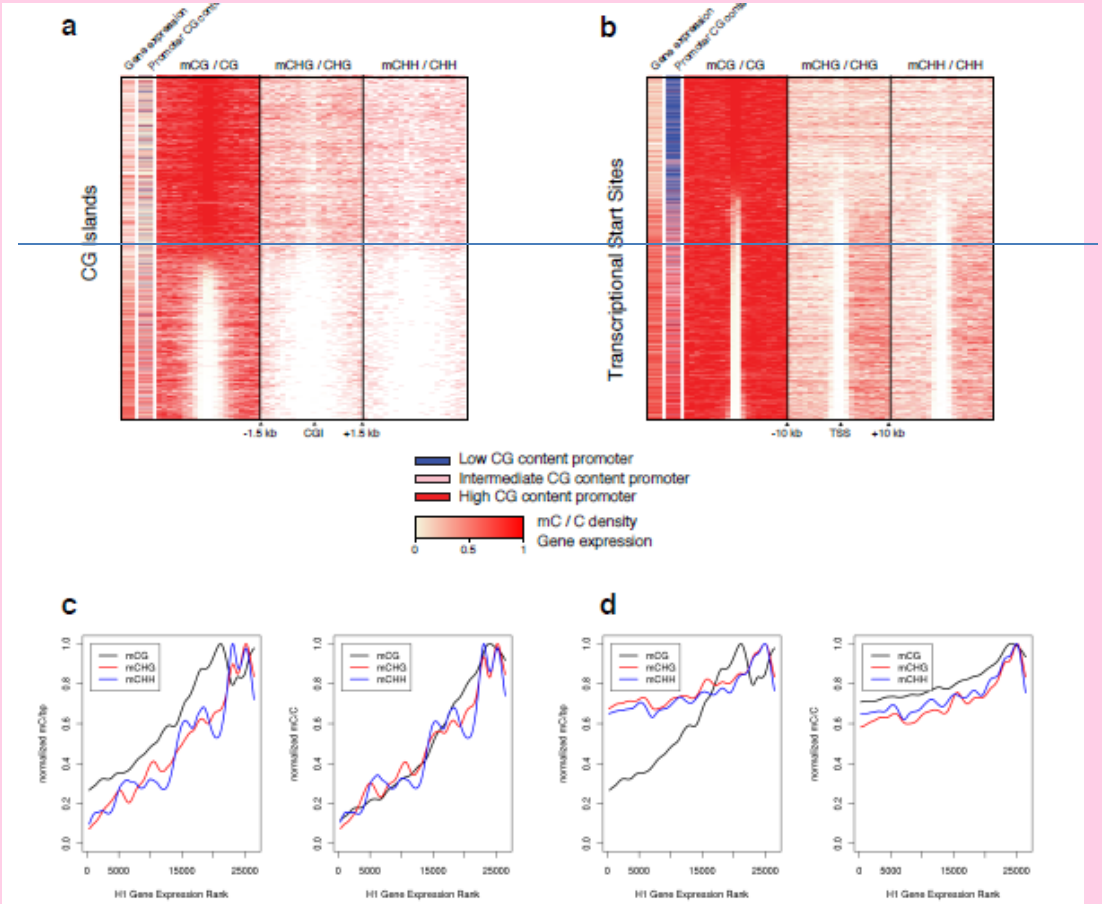


Supplementary Figure 5 | Mean mC/C Profiles Over Genomic Regions. a, gene body regions were divided in 20 bins from 5' to 3' end, and the mean mC/C level within each bin for each methylation type was determined (mCG/CG black, mCHG/CHG red, CHH/CHH blue). Mean over 3Kb regions centered at **b,** regulatory and **c,** transposable repeated genomic regions are displayed.





There are lower relative densities of methylation at CG islands and TSSs



Supplementary Figure 6 I DNA Methylation at CG Islands, Transcriptional Start Sites and Promoters. Relative DNA methylation density at **a,** CG islands (1.5 kb upstream/downstream) and **b,** transcriptional start sites (10 kb upstream/downstream) is displayed with downstream gene expression and promoter CG content. Each CG island was assigned to the closest gene whose TSS is within 10Kb. As expected, low CG content promoters are highly methylated, or close to highly methyalted CG islands, and close to low expressed genes. High CG content promoters are poorly methylated and usually close to highly expressed genes. CG and non-CG methylation density was profiled upstream of the transcriptional start site (TSS) and have compared this to the expression of the downstream gene, for all genes. For both proximal TSS (**c,** defined as -150 bp to +150 bp across TSS) and promoter (**d,** defined as the region 1.5 kb upstream of the TSS) there is a clear anti-correlation of gene expression in respect to both the absolute and relative mC content (mC/bp and mC/C, respectively). This trend is more evident for the region proximal the TSS. Abbrevitations: CGI, CG island. mC, methylcytosine. TSS, transcriptional start site.

…with exceptions

In regions 1.5 kb upstream and downstream of CG islands and TSSs of low expressed genes

Methylation is mainteined high (this corresponds to low CG content promoters)

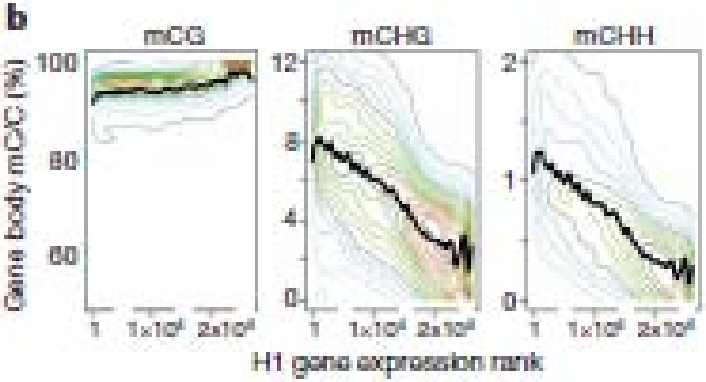
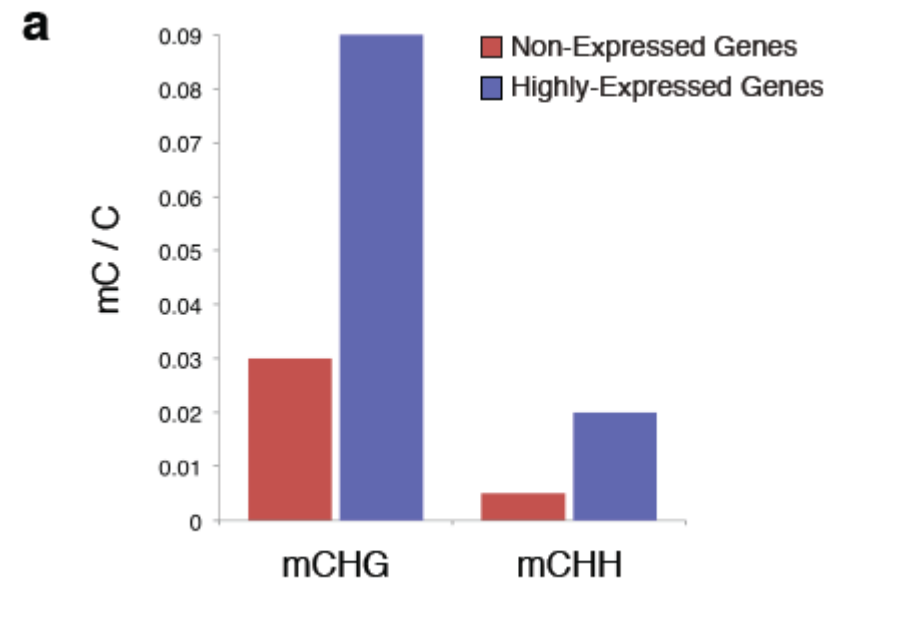


Figure 3 | Non-CG DNA methylation in H1 embryonic stem cells.

normalized to the maximum value. **b**, Relative methylation density within gene bodies (*y* axis) as a function of gene expression (*x* axis), with transcript abundance increasing from right to left. Coloured lines represent data point density and smoothing with cubic splines is displayed in black. **c**, Graphical



Supplementary Figure 7 | a, Enrichment of non-CG methylation in non-expressed and highly-expressed genes in H1. **b,** Over-representation of GO terms of genes within 20 kb of

Identification of links between gene activity and non-CG methylation level within the gene body

- strand-specific RNA-Seq shows a positive correlation between gene expression and mCHG (r=0.60) or mCHH (r=0.58) density
- no correlaton between CG methylation density and gene expression is evident in H1 cells

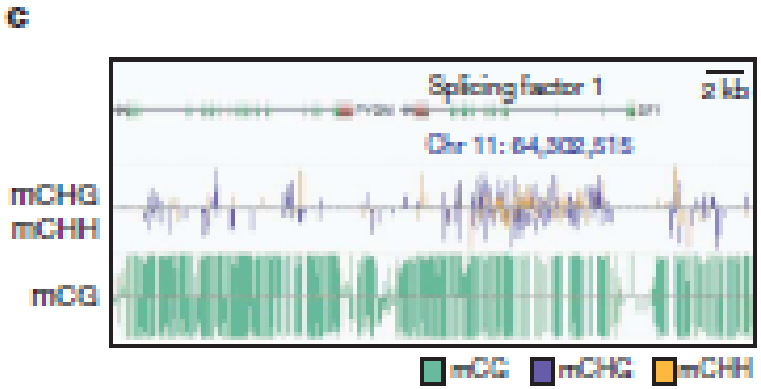


Figure 3 | Non-CG DNA methylation in H1 embryonic stem cells.

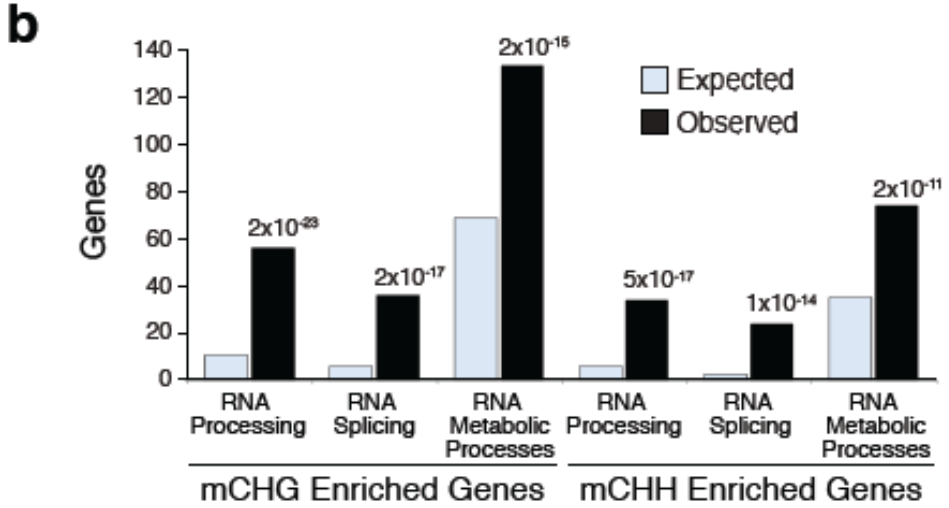
density and smoothing with cubic splines is displayed in black. **c**, Graphical representation of methylation at a non-CG methylation enriched gene, splicing factor 1. **d**, Average relative methylation densities in each sequence

447 and 226 genes are proximal to genomic regions highly enriched for mCHG and mCHH, respectively, with 180 genes in common

e.g. Splicing Factor 1

Analysis of gene ontology terms reveals genes involved in

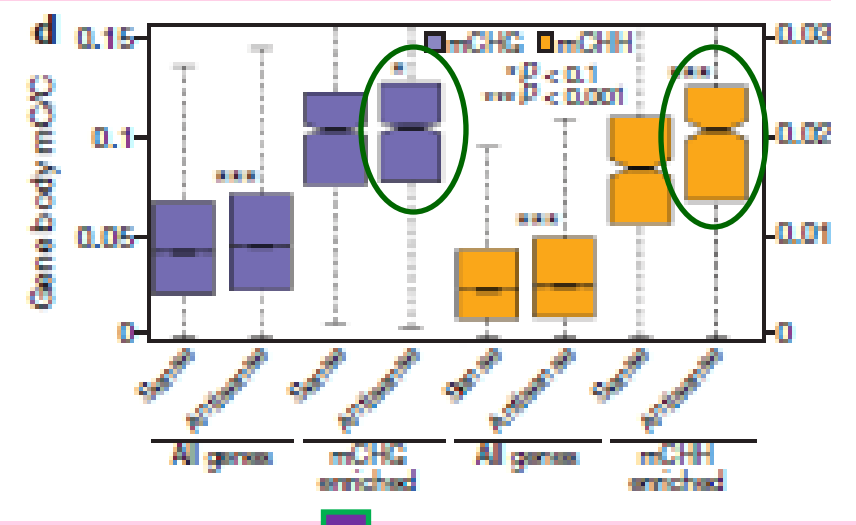
- RNA processing
- RNA splicing
- RNA metabolic processes

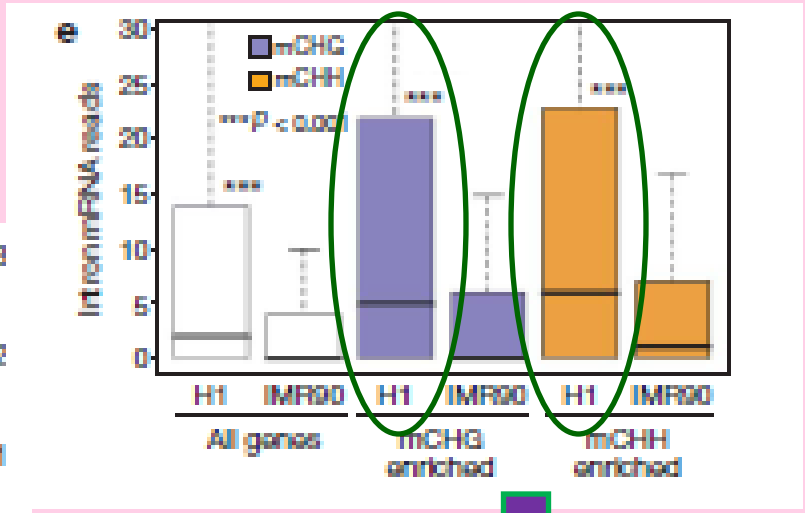


Supplementary Figure 7

highly-expressed genes in H1. **b,** Over-representation of GO terms of genes within 20 kb of genomic regions displaying the highest enrichment of CHG and CHH methylation. The enrichment P-value is shown for each GO term.

Figure 3 | Non-CG DNA methylation in H1 embryonic stem cells.
splicing factor 1. **d**, Average relative methylation densities in each sequence context within gene bodies on the sense or antisense strand relative to gene directionality. *P*-values for differences between sense and antisense densities are indicated. Boxes in **d** and **e** represent the quartiles and whiskers mark the minimum and maximum values. **e**, Number of mRNA intronic reads in all genes or genes associated with non-CG enriched regions, in H1 and IMR90 cells. *P*-values for differences between H1 and IMR90 reads are indicated.

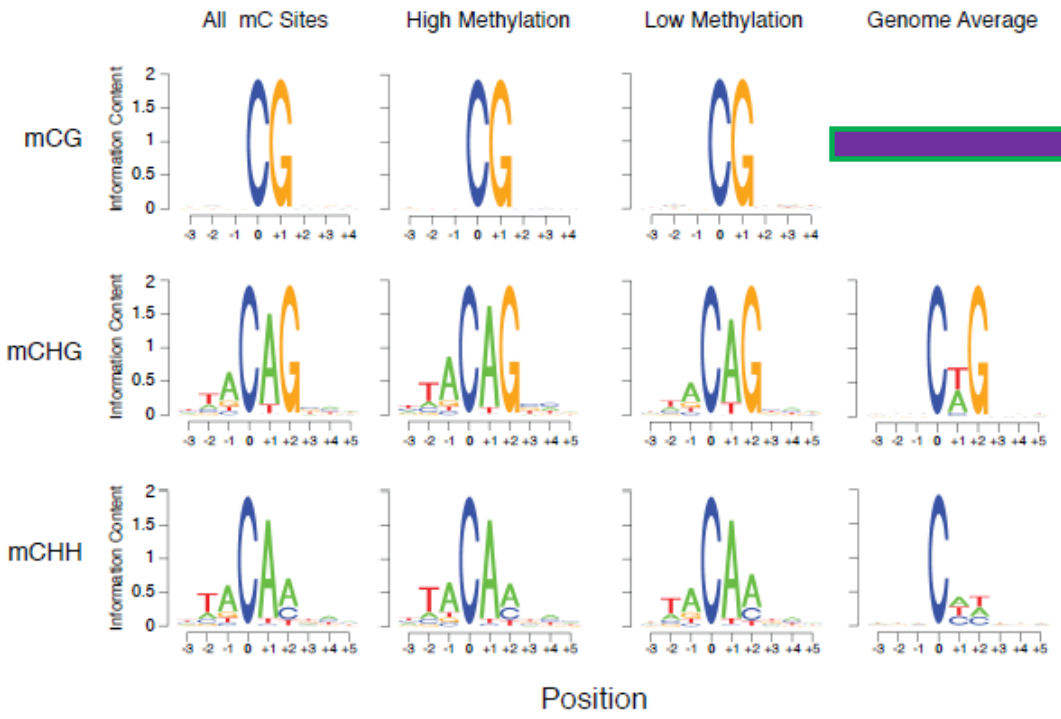




Enrichment of non-CG methylation of the antisense strand of genes enriched for both mCHG and mCHH

H1 cells show more RNA originating from introns than IMR90
This discrepancy is significantly enhanced in the mCHG and mCHH enriched genes

The higher abundance of intronic reads is associated with higher non-CG methylation within gene bodies, rather than differential non-CG methylation of exons versus introns



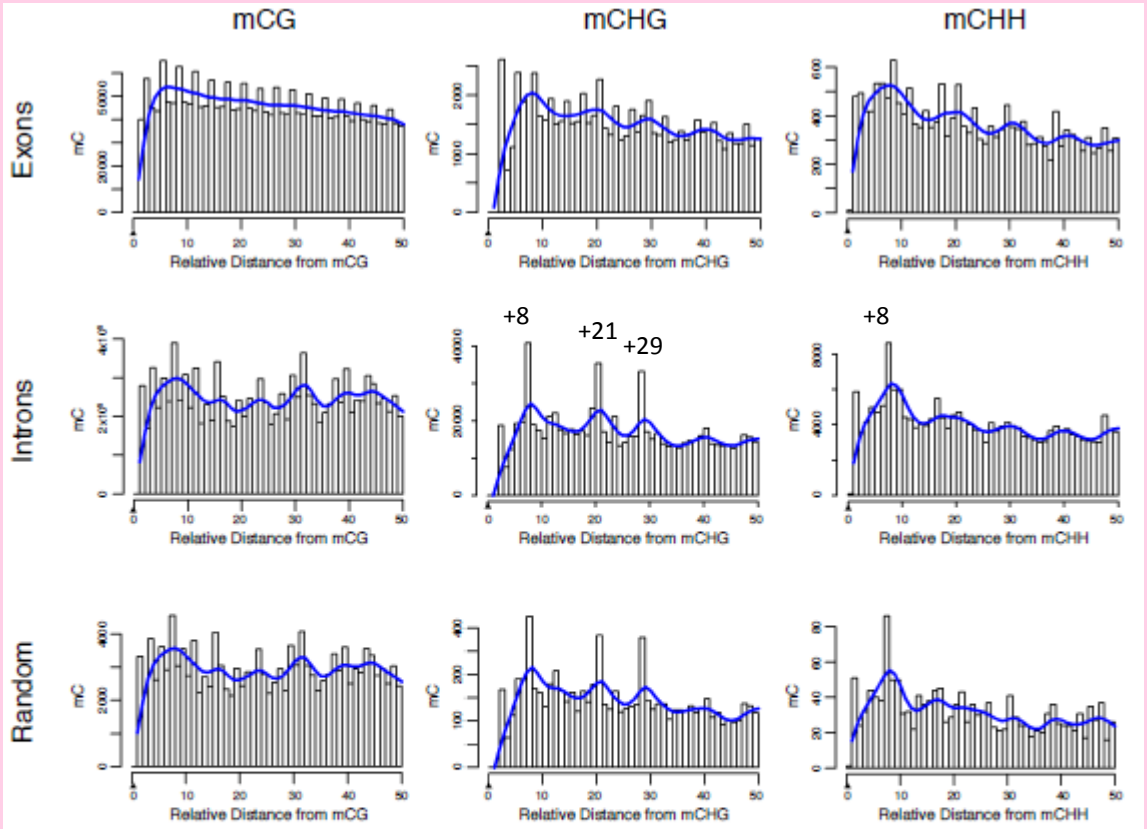
Supplementary Figure 8 | Logo Plots of the Sequences Proximal to Sites of DNA Methylation in Each Sequence Context in H1 Cells. Logo plots are presented for all methylcytosines, and methylcytosines that display a high methylation level (CG ≥75% methylated, non-CG ≥25% methylated), and low methylation level (CG <75% methylated, non-CG <25% methylated). Three bases flanking every site of methylation were analysed to identify local sequence preferences. The information content of each base represents the level of sequence enrichment. Local sequence enrichments were not evident when all cytosines were analyzed, regardless of their methylation status, and the level of methylation at a non-CG methylation site did not appear to influence the local sequence enrichment.

No local sequence enrichment

- preference for TA upstream
- A (or T) following C

This sequence is preferred by DNMT3 methyltransferases

Analysis of the distance between methylcytosines in 50 nucleotides regions



Supplementary Figure 9 | Spacing of Adjacent Methylcytosines in Different Contexts. Prevalence of mCHG/mCHH sites (y-axis) as a function of the number of bases between adjacent mCHG/mCHH sites (x-axis) based on all non redundant pair-wise distances up to 50 nt in exons, introns and random sequences. The blue line represents smoothing by cubic splines.

Exons:
No periodicity

Introns and random:
- no periodicity for mCG
- two pairs of 8-base separated mCHG sites spaced with 13 bases between them
- periodicity of 8-10 bases for mCHH

8-10 base periodicity corresponding to a single turn of the DNA helix; this is important for DNMT3A - 3L heterotetramer

Depleted DNA methylation at DNA-protein interaction sites

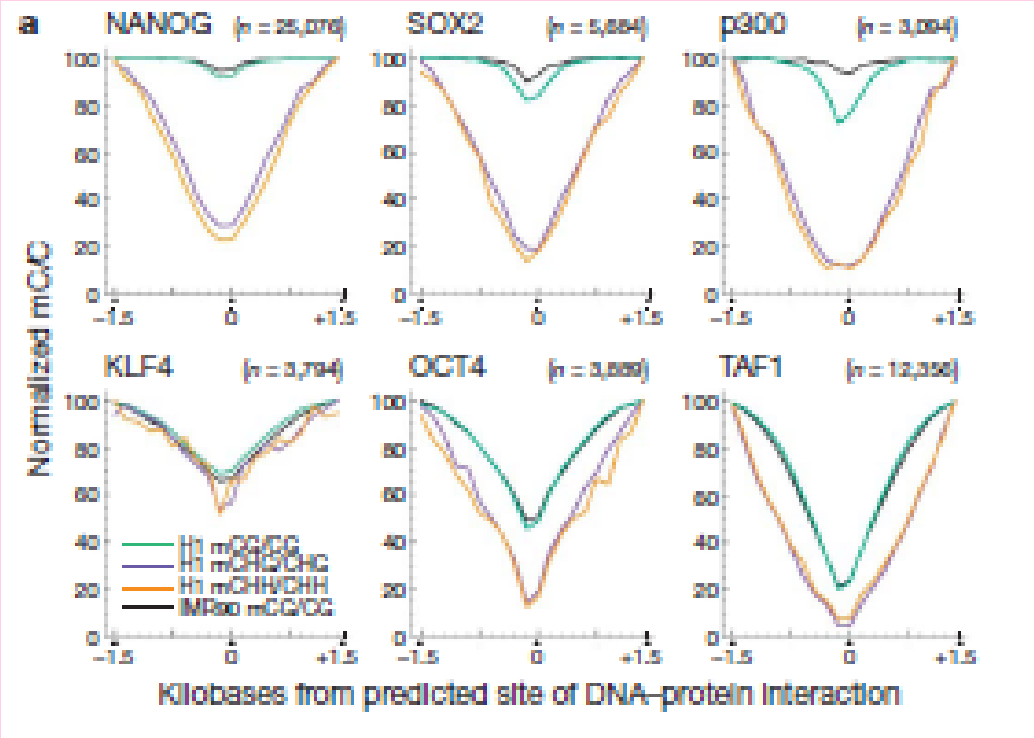


Figure 4 | Density of DNA methylation at sites of DNA–protein interaction. **a**, Average relative DNA methylation densities 1.5 kb upstream and downstream of predicted sites of DNA–protein interaction. **b**. Co-

- ChIP-Seq is performed in H1 cells

General decrease in the profile of relative methylation density towards the site of interaction, particularly in the non-CG context

NANOG, KLF4, OCT4 and TAF1 interaction sites are hypomethylated but proteins are not expressed in IMR90

ChIP-Seq detects location of enhancers (regions of simultaneous enrichment of H3K4me1 and H3K27ac) throughout the H1 and IMR90 genomes

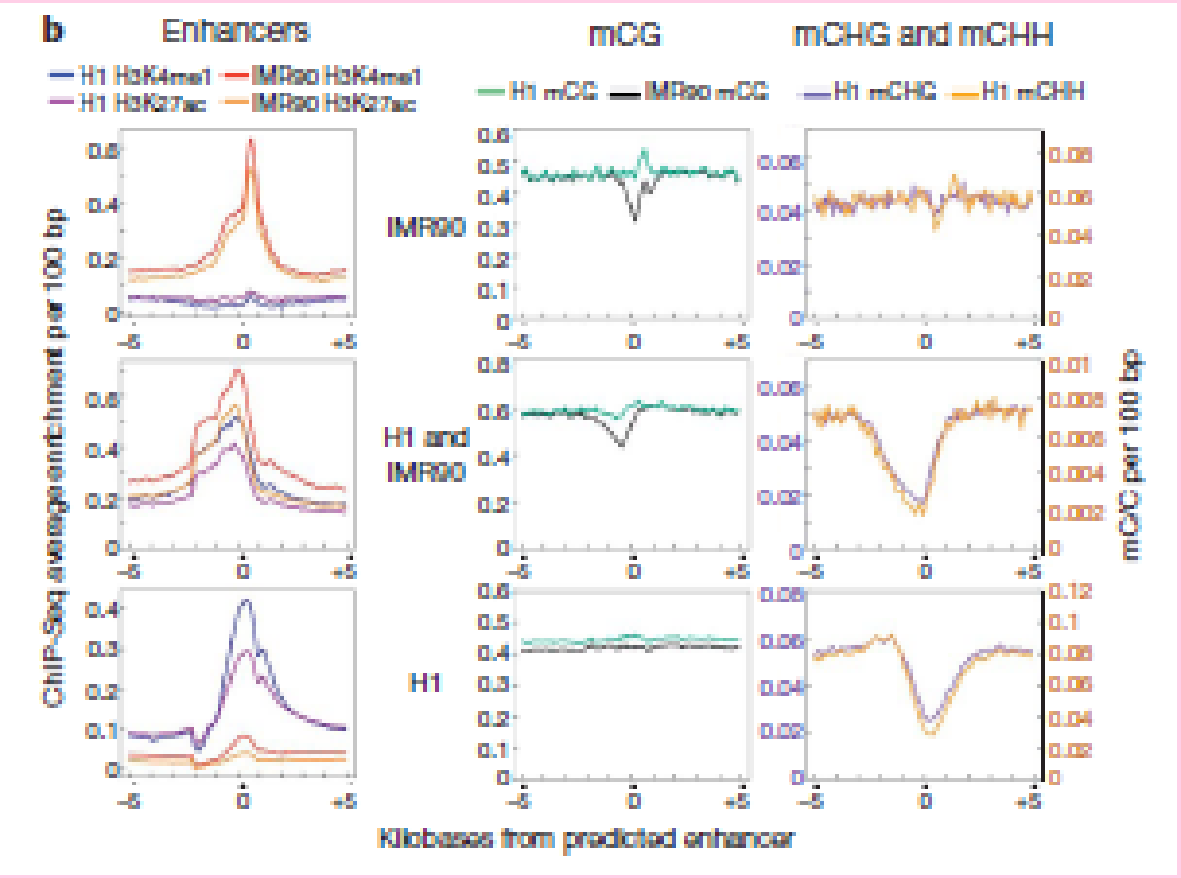


Figure 4 | Density of DNA methylation at sites of DNA–protein interaction. **a**, Average relative DNA methylation densities 1.5 kb upstream and downstream of predicted sites of DNA–protein interaction. **b**, Co-localization of H3K4me1 and H3K27ac ChIP-Seq tag enrichment indicative of enhancer sites that have been grouped into three sets: specific to IMR90 cells (top), H1 cells (bottom), or common to both H1 and IMR90 cells (middle). Average relative DNA methylation densities in each sequence context in 100-bp windows are displayed throughout 5 kb upstream and downstream of the enhancers in each of the sets.

- ↓ IMR90 mCG
- ↑ H1 mCG

- ↓ IMR90 mCG
- ↓ H1 mCHG and mCHH

- ↓ H1 mCHG and mCHH

Cell-type-specific utilization of different categories of DNA methylation

Widespread cell-specific patterns of DNA methylation

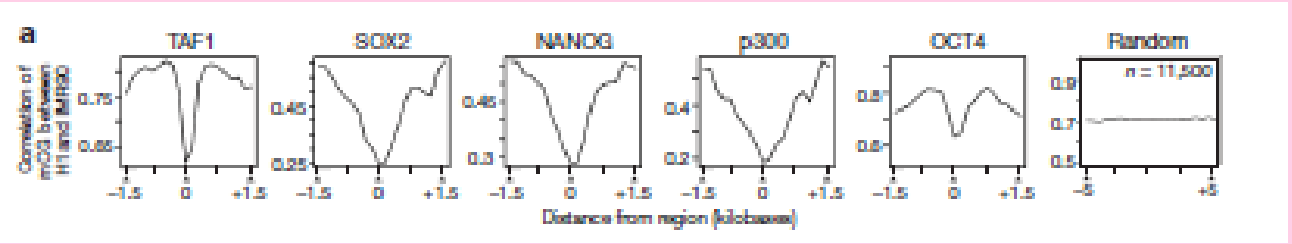
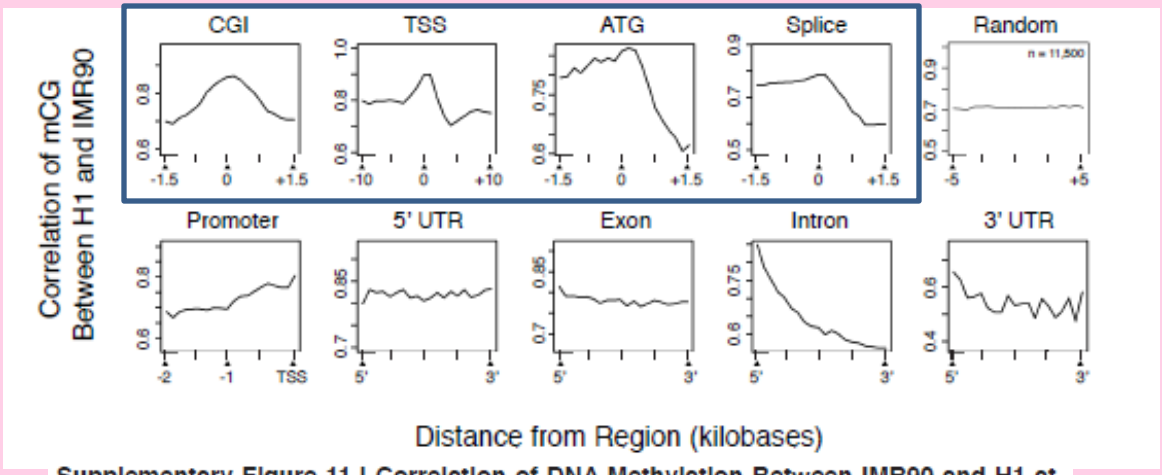


Figure 5 | Cell-type variation in DNA methylation. **a**, Pearson correlation coefficient of mCG methylation density (*y* axis) between H1 and IMR90 at various genomic features. Regions were divided into 20 equally sized bins from 5′ to 3′. Pearson correlation was determined in each bin considering all the H1 and IMR90 occurrences of the given genomic region. **b**, DNA



Supplementary Figure 11 | Correlation of DNA Methylation Between IMR90 and H1 at Different Genomic Features. The Pearson correlation coefficient of mCG methylation density (y-axis) between H1 and IMR90 at various genomic features. Regions were divided in 20 equally sized bins from 5' to 3' or based on the distance from the localization of the genomic feature as indicated. Pearson correlation was determined in each bin considering all the H1 and IMR90 occurrences of the given genomic region. An increased and high mCG correlation level was observed in correspondence to genomic regions expected to display a more constitutive epigenetic state, such as CG islands and TSS. We also observed a greater correlation at translational start sites and splice junctions. Gene promoters displayed an increase in correlation as the distance from the TSS decreased. We observed that the correlation in introns is highest toward the 5' exon-intron junction and decreased throughout the length of the introns. Abbreviations: CGI, CG islands. mC, methylcytosine. TSS, transcriptional start site.

Patterns of methylation vary in different cell types

They control differentiation

- pairwise comparison of methylation calculated for 20 equally sized windows for various genomic features

Even though both genomes show mCG depletion there is a significant decrease in the conservation of methylation

- decrease of correlation at the sites of protein-DNA interaction
- high correlation for regions expected to be conserved

PMDs: partially methylated domains

- contiguous regions with average methylation <70%
- in IMR90 cells
- average length of 153 kb
- 38.4% of autosomes and 80% of X chr.

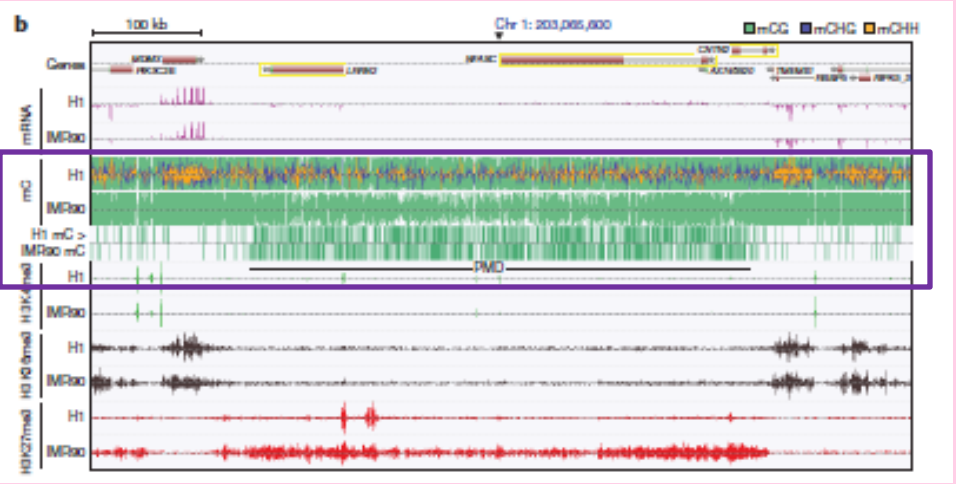
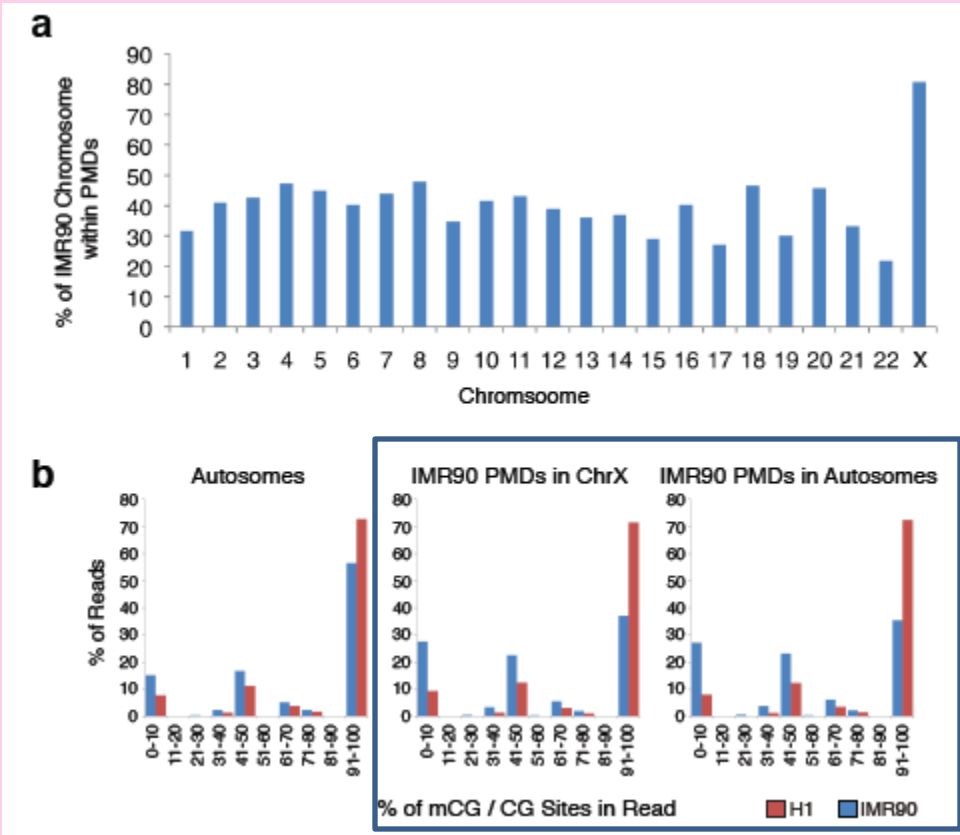


Figure 5 | Cell-type variation in DNA methylation.

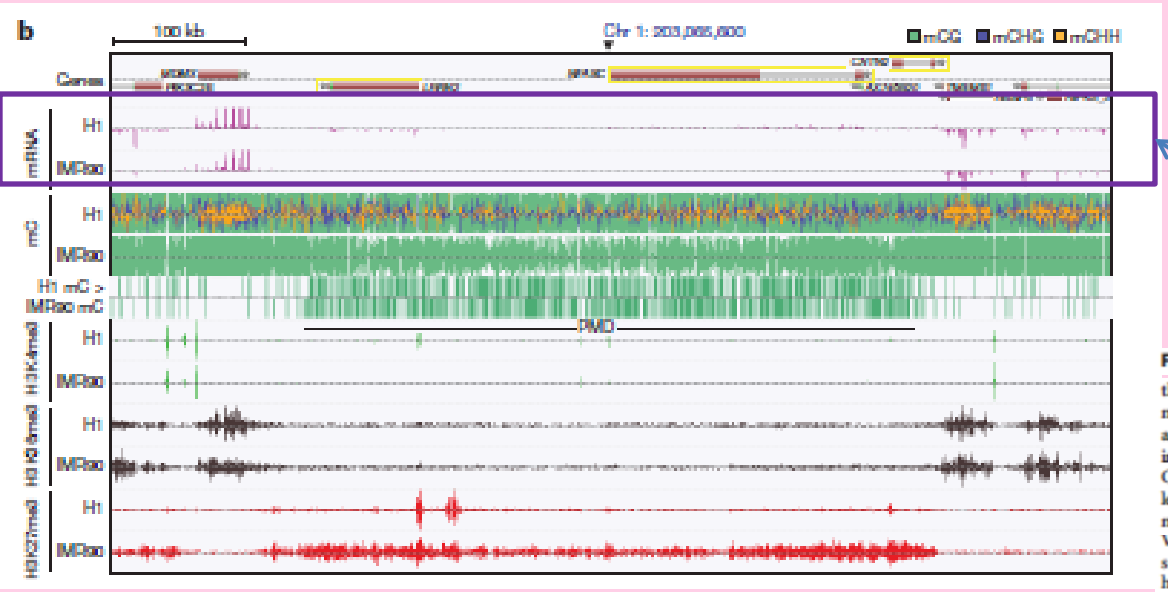
the H1 and IMR90 occurrences of the given genomic region. **b**, DNA methylation, mRNA and histone modifications in H1 and IMR90 cells associated with a PMD. Vertical lines above and below the dotted central line in DNA methylation tracks indicate methylcytosines on the Watson and Crick strands, respectively. Line vertical height indicates the methylation level. The H1 mC > IMR90 mC track indicates methylcytosines significantly more methylated in H1 than IMR90 at a 5% FDR (Fisher's exact test).

Vertical bars in the mRNA and histone modification tracks represent sequence tag enrichment. A yellow box indicates any gene with ≥30-fold higher mRNA abundance in H1 than IMR90. **c**, Comparison of transcript



Supplementary Figure 12 | PMDs and the Distribution of Unmethylated, Partial, and Completely Methylated Reads. a, The percent of each IMR90 chromosome that is within PMDs. **b,** For MethylC-seq reads located within a set of genomic regions, the percentage of CG sites within each read that were methylated was calculated, and the percent of all reads within the regions (y-axis) that were methylated at given percentages (x-axis) is displayed. This is presented for H1 and IMR90 MethylC-seq reads in autosomes, in IMR90 partially methylated domains on chromosome X, and IMR90 partially methylated domains in autosomes. Abbreviations: mC, methylcytosine. PMD, partially methylated domain.

They are frequently partially methylated or unmethylated compared to all IMR90 reads aligned to the autosomes and to H1 reads



Inspection of 5,644 genes with a TSS located in or within 10 kb of a PMD

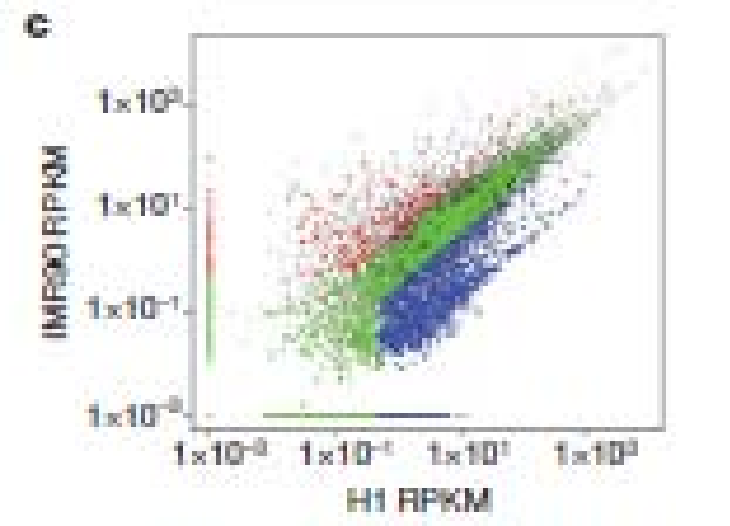
- they are less expressed in IMR90

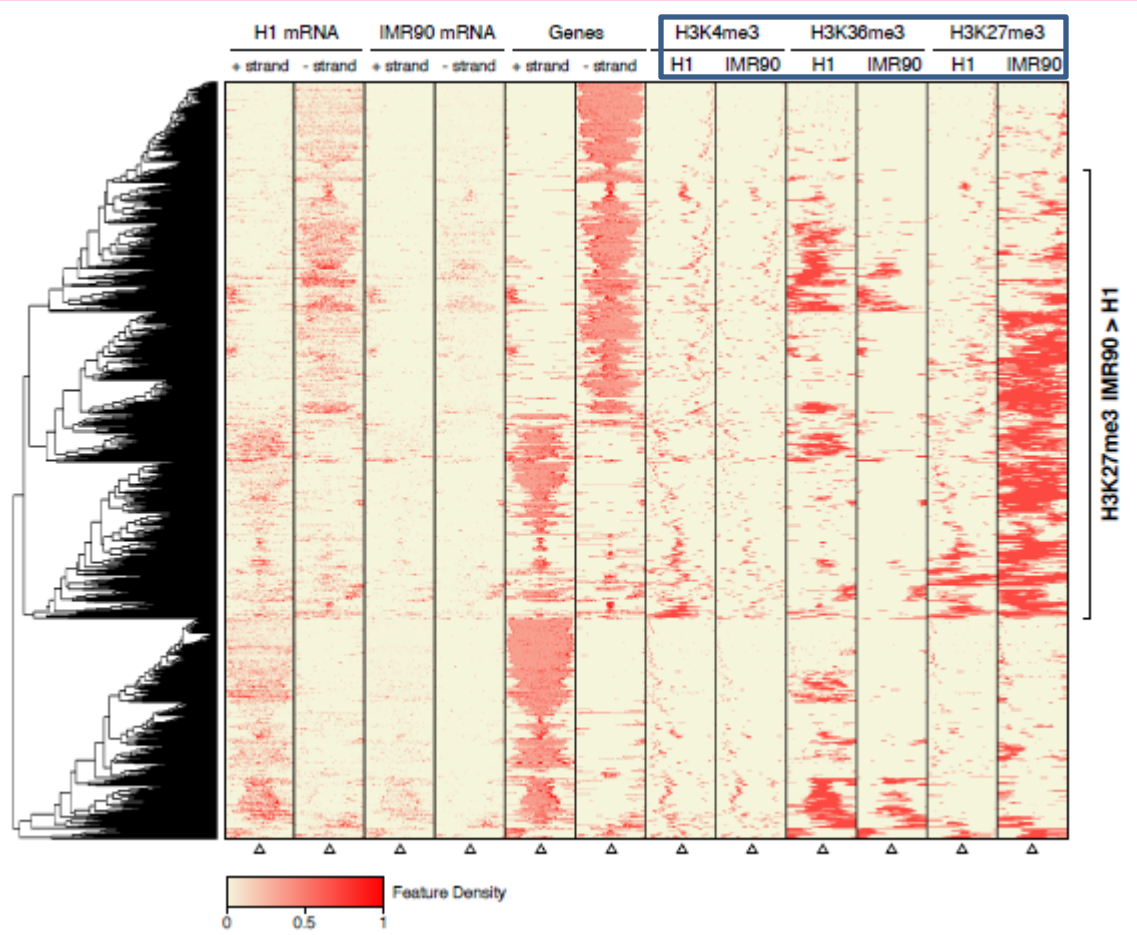
Figure 5 | Cell-type variation in DNA methylation.

the H1 and IMR90 occurrences of the given genomic region. **b**, DNA methylation, mRNA and histone modifications in H1 and IMR90 cells associated with a PMD. Vertical lines above and below the dotted central line in DNA methylation tracks indicate methylcytosines on the Watson and Crick strands, respectively. Line vertical height indicates the methylation level. The H1 mC > IMR90 mC track indicates methylcytosines significantly more methylated in H1 than IMR90 at a 5% FDR (Fisher's exact test). Vertical bars in the mRNA and histone modification tracks represent sequence tag enrichment. A yellow box indicates any gene with ≥30-fold higher mRNA abundance in H1 than IMR90. **c**, Comparison of transcript abundance between H1 and IMR90 cells of genes with a transcriptional start site located in or within 10 kb of a PMD. Black dots indicate all genes in the genome; blue, red and green indicate PMD genes expressed ≥3-fold higher in H1, IMR90 or not differentially expressed, respectively. **d**, Mean gene

Comparison of transcripts between H1 and IMR90

- 42% of the genes more expressed in H1 (3X IMR90) are located within PMDs
- 13% of the genes more expressed in IMR90 are located in PMDs





Supplementary Figure 13 | Transcriptional Activity and Epigenetic Modifications at Partially Methylated Domains. The density of strand-specific mRNA reads, and the presence of domains of H3K4me3, H3k36me3 and H3K27me3 in H1 and IMR90 was profiled 20 kb upstream to 20 kb downstream of each gene located in an IMR90 PMD. Open triangles indicate the central point in each 40 kb window. Also displayed is the presence within the Human reference sequence of genes on each strand, where pink coloring indicates the gene body and dark red boxes represent exons. The complete linkage hierarchical clustering of the regions based on these data is presented. Abbreviations: mC, methylcytosine. PMD, partially methylated domain.

IMR90 PMD genes show lower mRNA abundance, lower levels of H3K4me3 and H3K36me3 modifications; they show higher proximal H3K27me3 levels

(see slide 22)

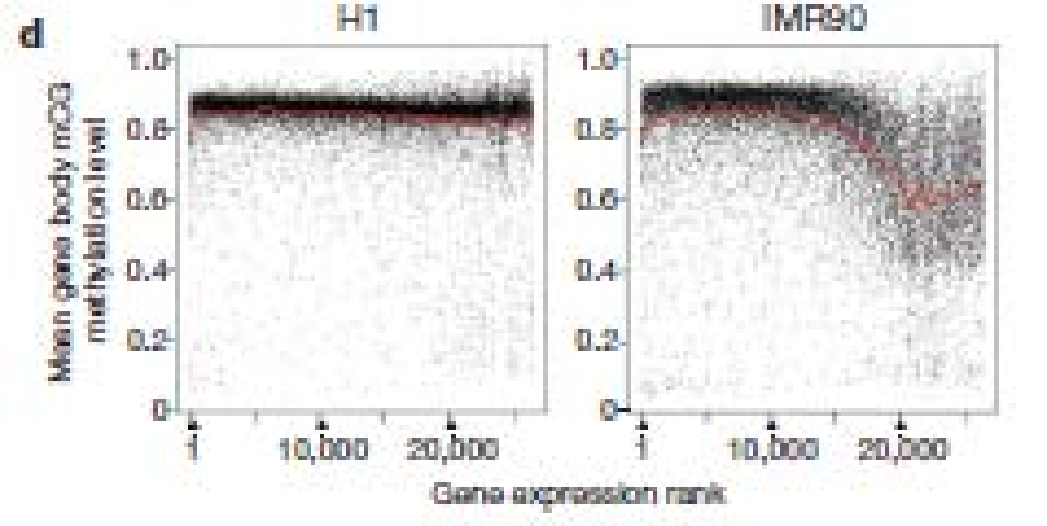


Figure 5 | Cell-type variation in DNA methylation.

in H1, IMR90 or not differentially expressed, respectively. **d**, Mean gene body mCG methylation (at least 10 reads required) as a function of the gene expression rank, 1 being the most expressed. mC, methylcytosine; PMD, partially methylated domain; RPKM, reads per kilobase of transcript per million reads.

In IMR90 cells there is a positive correlation between mCG methylation level and gene expression;
in H1 no correlation is discernible

The positive correlation between gene expression and and gene body methylation could be re-interpreted as a depletion of methylation in repressed genes in differentiated cells

Stem cell hypomethylated regions

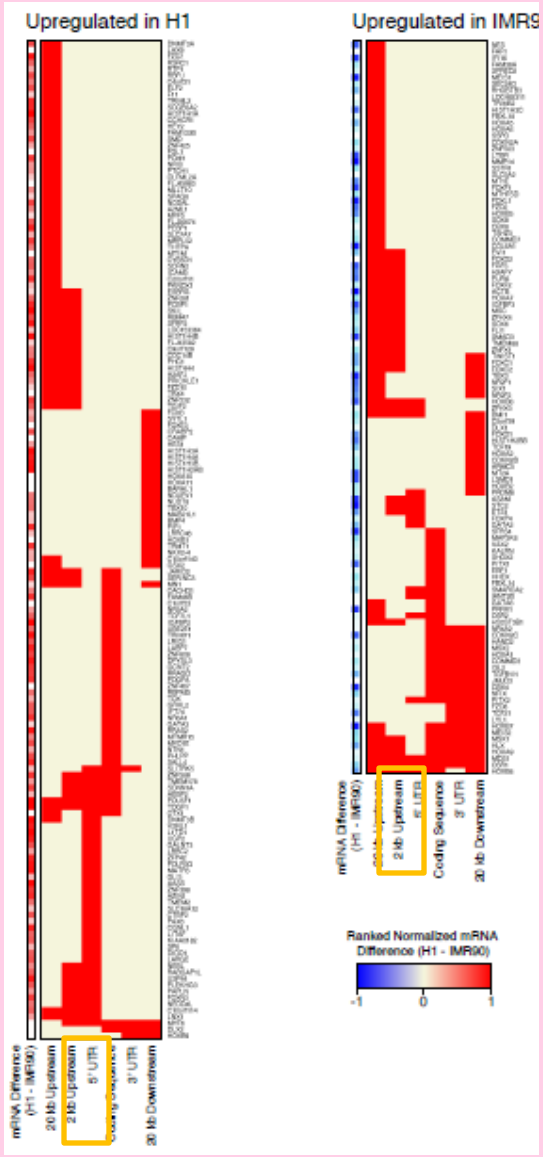
Supplementary Figure 14 | Differentially Methylated Regions proximal to *DNMT3B*. AnnoJ genome browser display of DNA methylation and mRNA at two DMRs upstream of *DNMT3B*. For DNA methylation tracks, vertical lines above and below the dotted central line indicate the presence of methylcytosines on the Watson and Crick strands, respectively. The color represents the context of DNA methylation, as indicated, and the vertical height of the line indicates the methylation level of each methylcytosine. The IMR90 > H1 mC track indicates methylcytosines that are significantly more methylated in IMR90 relative to H1 at a 5% FDR (Fisher's Exact Test), and the color represents the context of DNA methylation. Abbreviations: mC, methylcytosine. DMR, differentially methylated region.

DMRs: differentially methylated regions

- IMR90 are highly methylated than H1

491 DMRs are associated with 139 and 113 genes more highly expressed in H1 and IMR90

>50% of these genes with DMRs located within 2 kb upstream of the TSS or the 5'-UTR



Supplementary Figure 15 | Genes Within 20 kb of IMR90 Hypermethylated Regions Being More Expressed in H1 or IMR90. Gene regions bearing differential methylation are indicated in red. Side colorbar displays normalized differential expression (red and blue for genes upregulated in H1 and IMR90, respectively).

Genomic features density in a 20 kb up- and downstream window for each DMR

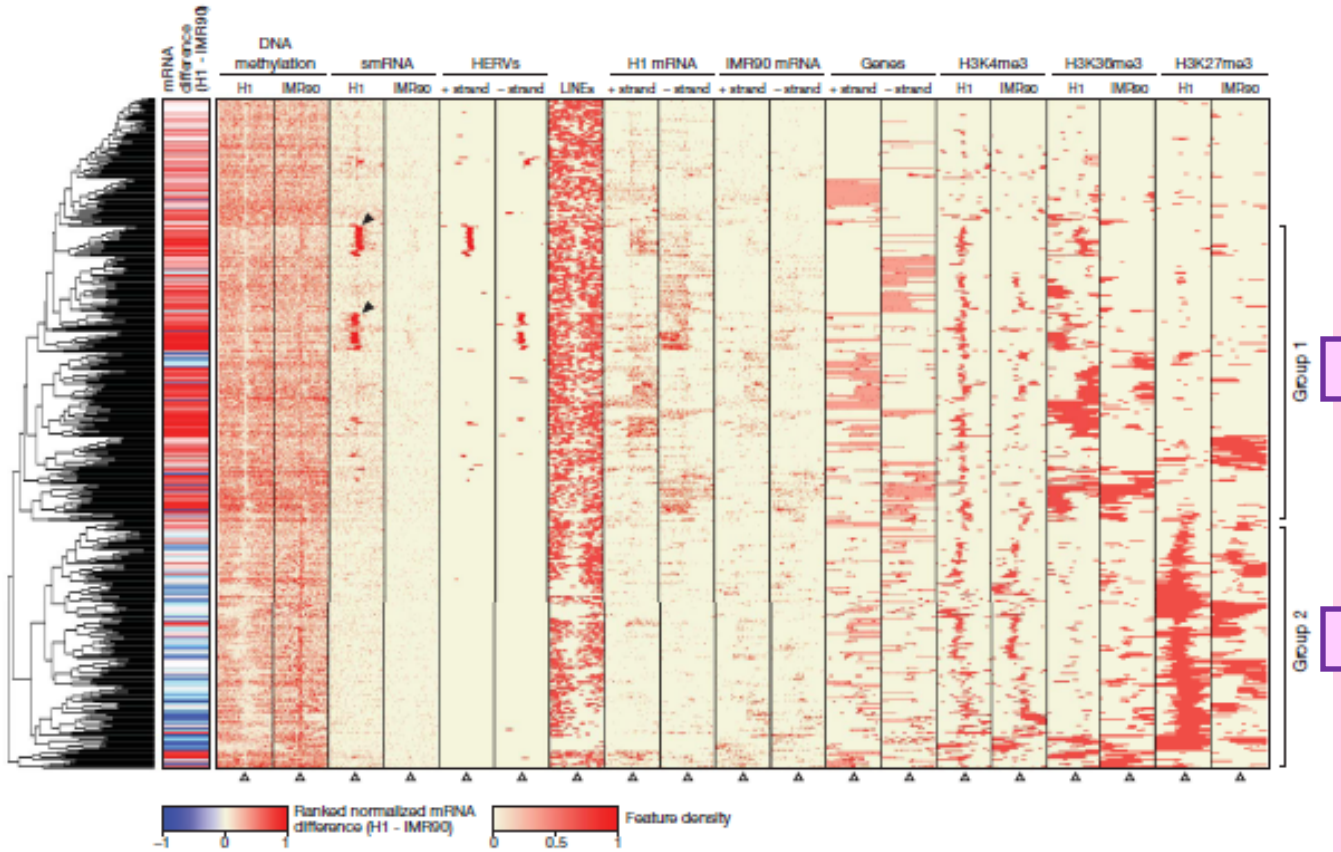


Figure 6 | Clustering of genomic, epigenetic and transcriptional features at differentially methylated regions. The density of DNA methylation, smRNA reads, strand-specific mRNA reads and the presence of domains of H3K4me3, H3k36me3 and H3K27me3 in H1 and IMR90 was profiled through 20 kb upstream and downstream of each of the 491 DMRs where DNA methylation was more prevalent in IMR90 than H1. Open triangles indicate the central point in each window. The side colour bar indicates the difference between H1 and IMR90 mRNA levels. The location of HERVs, LINEs and genes is displayed on each strand, where pink colouring indicates the gene body and dark red boxes represent exons. Black triangles indicate regions enriched for smRNAs that are coincident with HERVs. Group 1 and 2 are discussed in the text. DMRs, differentially methylated regions; HERVs, human endogenous retroviruses.

Group 1 DMRs:
high proximal
- H3K4me3
- H3K36me3
- transcriptional activity (in H1 relative to IMR90)

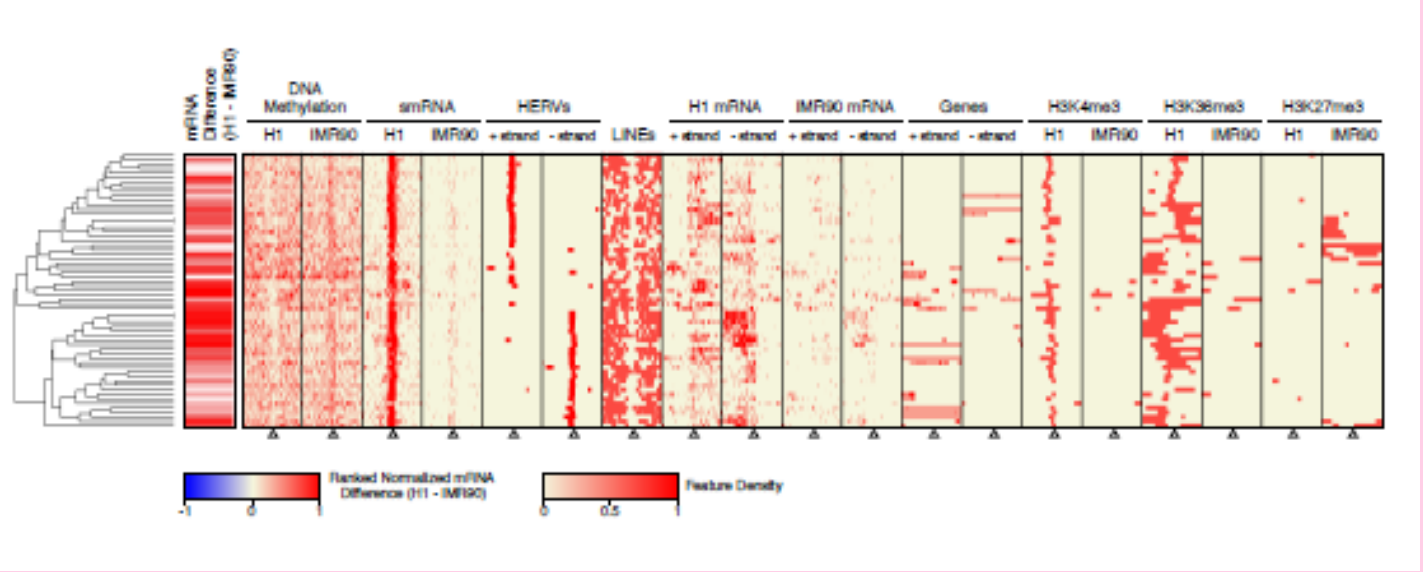
low proximal
- H3K27me3

Group 2 DMRs:
high proximal
- expression in IMR90
- H3K27me3

low proximal
- LINEs in the flanking sequence
- DNA methylation

Group 2 regions in H1 show H3K4me3 and H3K27me3 → suppressed but poised transcriptional status

Genomic features density in a 20 kb up- and downstream window for each HERV



Supplementary Figure 16 | Clustering of Genomic, Epigenetic and Transcriptional Features at Differentially Methylated HERVs. The density of DNA methylation, smRNA reads, strand-specific mRNA reads, and the presence of domains of H3K4me3, H3k36me3 and H3K27me3 in H1 and IMR90 was profiled 20 kb upstream to 20 kb downstream of each of the 61 smRNA clusters that co-localize with DMRs. Abbreviations: DMRs, Differentially Methylated Regions. HERVs, Human Endogenous Retroviruses.

HERVs: human endogenous retroviruses
A subset of Group 1 DMRs co-localizes with small RNAs that map to HERVs

- less methylated in H1
- high transcriptional activity in H1

These small RNAs, like in plants, could be involved in accurate targeting of DNA methylation

Concluding remarks

- There are extensive differences between the DNA methylomes of the two human cell types
- There is abundant methylation in the non-CG context
- Different roles can be suggested by profiling of enhancers and different patterning of CG and non-CG methylation
- Non-CG may have a key role in the origin and maintenance of pluripotency
- Future studies will explore methylation variations throughout differentiation and re-establishment in induced pluripotent states