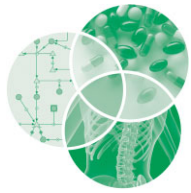
Advanced Review



Mediators and dynamics of DNA methylation

Robert Shoemaker,[1] Wei Wang[1] and Kun Zhang[2]*

As an inherited epigenetic marker occurring mainly on cytosines at CpG dinucleotides, DNA methylation occurs across many higher eukaryotic organisms. Looking at methylation patterns genome-wide classifies cell types uniquely and in several cases discriminates between healthy and cancerous cell types. DNA methylation can occur allele-specifically, which allows the cellular regulatory machinery to recognize each allele separately. Although only a small number of allele specifically methylated (ASM) regions are known, genome-wide experiments show that ASM is prevalent throughout the human genome. These DNA methylation patterns can be modified via DNA demethylation, which is important for induced pluripotent stem reprogramming and primordial germ cells. Recent evidence shows that the protein activation-induced cytidine deaminase plays a critical role in these demethylation events. Many transcription factors mediate DNA methylation patterns. Some transcription factors bind specifically to methylated or unmethylated sequences and other transcription factors protect genomic regions (e.g., promoter regions) from nearby DNA methylation encroachment. Possibly acting as another epigenetic regulatory layer, methylated cytosines are also converted to 5-hydroxyethylcyotines, which is a new modification type whose biological significance has yet been defined. *WIREs Syst Biol Med* 2011 3 281–298 DOI: 10.1002/wsbm.124

INTRODUCTION

DNA methylation is an inherited epigenetic chemical modification that occurs primarily on the cytosines in CpG dinucleotides. However, examples of non-CpG methylation are found in plants and human embryonic stem cells.[1–3] The proteins DNA Methyltransferase 1 (DNMT1), DNMT3a, and DNMT3b are known to methylate cytosines, while protein families, including Methyl-CpG Binding Domain (MBD) containing proteins, SET and RING finger-associated (SRA) domain containing proteins, and other zinc finger proteins, recognize the presence of methylated cytosines.[4–13] DNMT1 maintains methylation patterns and this maintenance is required for normal cell division.[14–16] Originally thought to bind DNA due to its similar structure and sequence to other members within the DNMT family, DNMT2 was later found to specifically target RNA.[14,17,18] DNMT3a and DNMT3b have *de novo* methylation activity that is required for embryonic development.[19] Another protein, DNMT3L, colocalizes with DNMT3a and DNMT3b, and it enhances *de novo* methylation *in vitro* and *in vivo*.[20–26] DNMT3L itself lacks methyltransferase activity and thus enhances methylation via its interaction with DNMT3a. DNA methylation is widespread across many organisms, including bacteria, insects, and mammals, and it is most commonly associated with transcriptional silencing.[27–30] As many repetitive regions and transposons are methylated in various genomes, DNA methylation is thought to act as a genomic defense mechanism that prevents the activation of these sequences.[31–34] DNA methylation is also present in genes and regulatory regions, where it is correlated with transcriptional repression or activation depending on its context. Given the regulatory effects associated with DNA methylation, characterizing the methylation states of cytosines genome-wide reveals the biological state of the host cell on a global scale.[1–3,35–37] This article (1) discusses the biological meaning of DNA methylation patterns

*Correspondence to: kzhang@bioeng.ucsd.edu

[1]Department of Chemistry and Biochemistry, University of California at San Diego, La Jolla, CA, USA

[2]Department of Bioengineering, University of California at San Diego, La Jolla, CA, USA

DOI: 10.1002/wsbm.124

on intercellular and intracellular levels, (2) explores the importance of the mechanisms that erase these patterns, (3) discusses the role of transcription factors in DNA methylation, and (4) reviews the presence of a new cytosine modification, hydroxymethylcytosine, which is a product of DNA methylation.

INTERCELLULAR DNA METHYLATION SIGNATURES: CONSERVATION AND APPLICATIONS

DNA Methylation in Higher Eukaryotes

A comparison of DNA methylation patterns across eight species (Arabidopsis, green algae, rice, poplar, honey bee, mouse, sea squirt, and zebrafish) revealed not only many common features but also striking differences.[27] Repetitive regions and transposons were enriched for CpG methylation relative to nearby regions for all eight organisms, but these methylation patterns in sea squirt were much weaker. Regions of high CpG density, which are known as CpG islands, are mostly unmethylated in vertebrates (i.e., zebrafish and mouse) but overall vertebrate methylation levels are high (70–80% global CpG methylation). Non-CpG (CHG and CHH) methylation was found in all three flowering plants; CHG and CHH methylation was enriched at transposons and repetitive regions. CHH methylation describes a methylated cytosine followed by two nucleotides that may not be guanine. CHG methylation entails a methylated cytosine that precedes an adenine, thymine, or cytosine, followed by guanine. Unlike CG or CHG methylation, CHH sites are strand specific, which means, for example, a CHH site on the Watson strand is not found on the reverse complementary Crick strand. Although exhibiting non-CpG methylation, green algae's CHH and CHG patterns showed no enrichment in any particular genomic region. The remaining organisms displayed a much lower percentage of non-CpG methylation. Vertebrates did not show a significant enrichment of CpG methylation in exons relative to introns as was seen in all three flowering plants. The honeybee, although displaying low levels of global CpG methylation, displayed significant CpG methylation in exon regions. Green algae showed very low global CpG methylation, but exons were more methylated. Sea squirt also showed preferential methylation of exon regions. This conservation study shows that DNA methylation is present in higher eukaryotes and that DNA methylation often targets certain genomic regions. However, the targeted regions are not entirely consistent across all of the eight studied organisms.

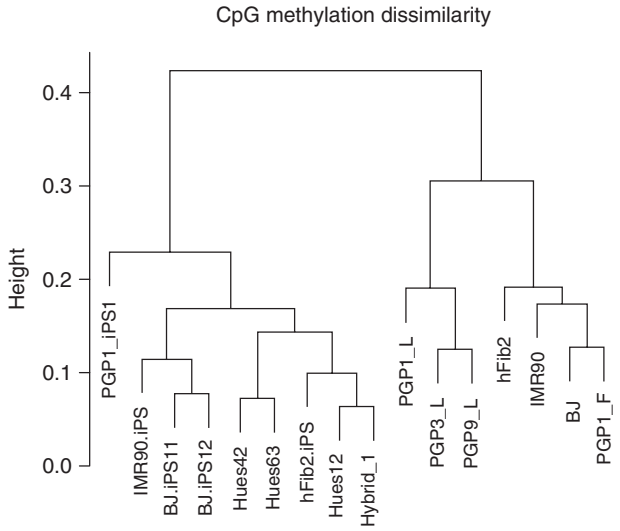


FIGURE 1 | Cell type specificity of DNA methylation. This figure shows the dissimilarity of cell types based on their methylation frequencies from targeted bisulfite data that covered CpG islands on chr12 and chr20. Cell types that are closer together share a more similar methylation signature. The fibroblasts, PGP1F (Personal Genome Foundation 1 Fibroblast), BJ, IMR90, and hFib2 (human Fibroblast), cluster closely together, while the lymphoblasts PGP9L (Personal Genome Project 9 Lymphoblast), PGP3L, and PGP1L also cluster together closely. Regarding the pluripotent cell lines, the ES cells (Hues12, Hues42, and Hues63) cluster very closely together. The hybrid cell line, which consists of fused nuclei from Hues6 and BJ, also clusters closely with the ES cell group. The induced pluripotent stem (iPS) lines appear to be much more similar to the ES cell group than the differentiated fibroblasts and lymphoblasts but the iPS group exhibits a wider spectrum of methylation signatures than the ES cell group or the two differentiate cell groups.

Since CpG methylation sites are known in a sequenced genome, one can easily compare methylation signatures across intraspecies cell lines. Genome-wide DNA methylation patterns are cell type specific. Our group targeted bisulfite sequencing of CpG islands across chr12 and chr20, and found that the methylation frequency data clearly distinguished between human embryonic stem cell lines, fibroblasts, and lymphoblasts[38,39] (Figure 1). An additional study reported a conserved set of regions that are uniquely methylated between human liver, spleen, and brain tissues.[40] These differentially methylated regions tended to be located just outside of CpG islands, and were thus called CpG island shores. Methylation of these CpG island shores had a strong inverse relationship with the expression of associated genes. The authors labeled 16,379 regions as differentially methylated across examined tissues (T-DMRs), and the median length of these regions was 255 bp. Extending this analysis to 13 colorectal cancer cell samples with matched normal mucosal samples,

the authors found a separate set of differentially methylated regions (C-DMRs) that showed significant differences in methylation between the normal mucosal and the matched cancer samples. Forty-five percent of the 2707 identified C-DMRs overlapped with T-DMRs (P-value $< 10^{-14}$), which showed that these many of these DMRs could discriminate between tissues and colorectal cancer. A continuation of this study found 4401 regions were differentially methylated in iPS cells relative to the untransformed fibroblast cells (R-DMRs).[41] These R-DMRs, similar to the C-DMRs and T-DMRs, overlapped significantly with CpG island shore regions (over 70%). Many R-DMRs also overlapped with T-DMRs (56%). These studies showed that DNA methylation signatures are unique across many different types of cell lines.

DNA Methylation as a Clinical Biomarker

DNA methylation can classify cell types into subpopulations, which can exhibit unique phenotypes. One cancer study examined the methylation profiles of blast cells taken from 344 diagnosed acute myeloid leukemia (AML) patients.[42] Clustering these methylation profiles created 16 unique AML subtype clusters. Three of those patient clusters were defined by the WHO classification,[43] eight were enriched for specific genetic or epigenetic lesions, and the remaining five could not be explained by current knowledge; all of these subtypes were distinct when compared to normal bone marrow cells. The authors used the methylation signatures of 18 methylation probe sets that covered 15 genes and developed a classifier that predicted the overall survival and event-free survival of an AML patient (P-value < 0.001, multivariate Cox proportional hazards model). These authors showed that DNA methylation signatures can act as biomarkers that foretell patients' clinical outcomes. Another study focused on breast cancer and used methylation signatures to distinguish between the different breast cancer mutation types (BRCA1, BRCA2, BRCAx).[44] The authors report that DNA methylation profiling predicted BRCA1, BRCA2, and BRCAx tumors with error rates of 11, 31, and 36%, respectively. Classification based on DNA methylation signatures was significantly more accurate than using gene expression data, which resulted in error rates of 11, 44, and 71%, respectively. However, the gene expression data were able to cluster the breast cancer samples into intrinsic subtypes (i.e., basal, luminal A, luminal B, HER2-amplified, and normal-like) while the DNA methylation data could not.[45] Another group used the DNA methylation signatures from over 300 peripheral blood samples as an indicator for ovarian cancer.[46] Methylation studies involving colorectal cancer and breast cancer have shown that peripheral blood samples can indicate the presence of cancer when compared to healthy controls.[47–51] Starting with 25,642 CpG sites, the authors found that a 100 CpG sites in peripheral blood cells that accurately discriminated between healthy and pretreatment ovarian cancer patients. Using these selected 100 CpG sites, the authors were able to correctly identify ovarian cancer samples (58 healthy samples and 43 pretreatment cases) with an area under the receiver operating characteristic (ROC) curve (AUC) of 0.82. AUC describes the ability of a test to detect true positives versus false positives. Tests with AUC values close to 1 represent high true positive and low false positive detection rates, and AUC values close to 0.5 represent discriminatory power that is similar to random selection. Tests with AUC less than 0.5 perform worse than random selection. Comparing a sample post-treatment population that showed signs of active disease with one that served as a healthy control, the classifier correctly identified those active disease patients with an AUC of 0.76. Finally the classifier was able to discriminate between post-treatment samples showing active disease and post-treatment samples without active disease (0.74 AUC). Additional tests showed the CpGs used in the classifier were not dependent on age. DNA methylation can be used to predict cancer-related phenotypes as seen in these three studies.

INTRACELLULAR DNA METHYLATION: ALLELE SPECIFIC INSTANCES

A mammalian cell contains a paternal and a maternal chromosome set. Most biological activity is thought to be symmetrical across the paternal and maternal chromosomes, but there are regions where each parental allele is regulated uniquely. Allele-specific methylation, ASM, (Figure 2a) is a common feature of these regions and ASM can be coupled with allele-specific expression, ASE (Figure 2b). Imprinted genes consist of genes exhibiting ASM and ASE in a parent-of-original specific manner. ASM near the loci of these ASE genes reveals significant methylation differences, where one allele is significantly more methylated than the other. There are between 50 and 100 known imprinted genes in mouse and human.[52–54] However, over 1000 genes have been found to exhibit parent-of-origin allelic gene expression effects in the mouse brain.[55] Besides imprinting, ASM can also occur in a manner that is independent of parental origin. This occurs in chromosome X inactivation randomly, but it has also been unexpectedly observed in numerous autosomal regions.[39,56] This section explores ASM

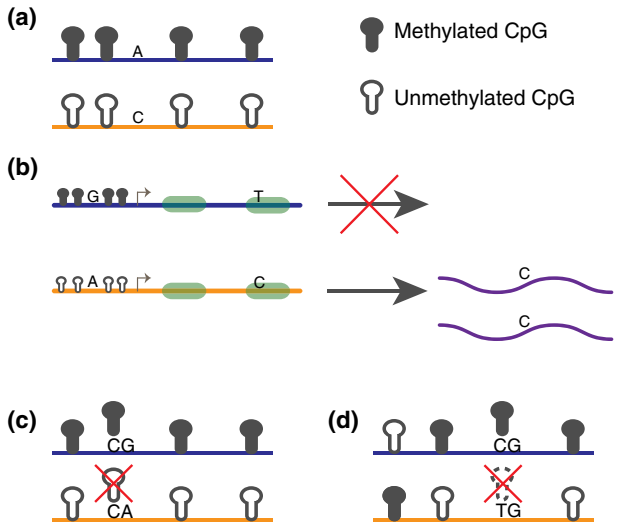


FIGURE 2 | ASM examples. (a) A schematic of typical ASM. The two alleles are represented separately as blue and orange lines. A SNP distinguishes the alleles and the blue allele is methylated while the orange allele is not. This is considered an ASM region. (b) This figure shows a gene whose promoter region is methylated in the blue allele and unmethylated in the orange allele. A SNP site at the promoter distinguishes the alleles. Additionally, another SNP site in an exon differentiates the alleles. This figure shows how ASM can be used to predict ASE behavior. Additional haplotyping can associate the SNPs thereby allele specifically linking the methylation status of a regulatory region to its expression. (c) An instance of a SNP overlapping with a CpG site in an ASM region. The SNP in this example disrupts a CpG site such that the CpG site no longer exists in the orange allele. The ASM behavior seen in this region may be caused by the SNP itself or the SNP/CpG overlap. (d) An instance of a SNP overlapping with a CpG site in a non-ASM region. Similar to (c), the overlapped CpG site no longer exists in the orange allele.

and places it in the context of other regulatory mechanisms that co-occur at these sites.

Chromosome X Inactivation

The extensive research into chromosome X inactivation (XCI) has illuminated a pathway for allele specific methylation via noncoding RNAs (ncRNAs). To maintain proper dosage specificity in female cells, one chromosome X copy is inactivated upon differentiation. The inactivated chromosome is heavily methylated, compacted into a heterochromatic state, and is largely transcriptionally silent. Experiments have shown that this process is directed by large noncoding RNAs, mainly *Xist* and *TsiX*, which are regulated at the X inactivation center.[52,57–62] Prior to XCI, the two X chromosomes pair at and around the X inactivation center.[63–65] The pairing and XCI initiation rely on the binding of Oct4 and the CCCTC-binding factor (CTCF) at the XCI region.[63,66,67]

Changes in the allele-specific transcription of *TsiX* and *Xist* loci initiate a cascade that results in the methylation-mediated silencing of one copy of chromosome X via a yet poorly understood counting mechanism.[63,68,69] Transcription of *Xist* is up regulated, while *TsiX* is down regulated on the selected X_i chromosome (inactive chromosome X); whereas on X_a (active chromosome X), *Xist* is down regulated, and *TsiX* is up regulated. On X_a, *TsiX* has been shown to associate with Dnmt3a at the *Xist* promoter region. This leads to the *de novo* methylation of the *Xist* promoter and thus the silencing of *Xist* transcription on X_a.[70] *Xist* contains a RepA element to which PRC2 binds and this interaction recruits PRC2 to chromosome X.[71,72] As *Xist* transcription remains high on X_i, *Xist* localizes to various locations across X_i and its recruitment of PRC2 leads to a chromosome-wide H3K27me3 modification of nucleosomes on X_i.[62,73] PRC2 then likely recruits DNA methyltransferases to X_i, where it is globally methylated.[74] The XCI process reveals an RNA directed DNA methylation (RdDM) pathway in human. Experiments have observed active RdDM pathways in Arabidopsis, but there are few such mammalian examples.[2,75–78]

Imprinted Genes

Imprint control elements (ICE) regulate the allele specific expression of imprinted genes via its own methylated state. Imprinting patterns (i.e., methylation states of ICEs) are formed during gametogenesis with paternal ICEs methylated in sperm and maternal ICEs methylated during oocyte growth.[79–81] DNMT3A and DNMT3L are required for the establishment of these imprinting patterns in both cell types and DNMT1 is necessary for maintenance of these imprints.[82,83] Histone modifications may play a role in the establishment of these methylation patterns. DNMT3A was shown to recognize H4R2me2 modifications via its plant homeodomain (PHD) motif; the recruitment of DNMT3A leads to *de novo* methylation of nearby sequences, which was shown in the beta-globulin locus.[84] Another experiment has shown that DNMT3L binds to H3, but it is repelled by the H3K4me3 modification.[85] This repulsion would prevent *de novo* methylation of regions marked for poised or active transcription.[86,87] The methylation of ICEs results in the transcription of protein-coding genes in cis (e.g., *GNAS*, *IGF2*, *PWS/AS*, *KCNQ1*). Since DNA methylation is associated with transcriptional silencing, this is an unexpected consequence. DNA methylation at ICEs serves to repress transcriptionally repressive factors, which explains its appearance as an activator. Unmethylated ICEs can form insulators, which prevent mRNA transcription, or unmethylated

ICEs lead to the transcription of ncRNAs, which then repress transcription of nearby protein coding genes in a cis fashion.[88–90]

Imprinting defects, which are seen as the improper establishment of regional DNA methylation haplotypes (i.e., epi-haplotypes), are associated with severe disorders. For example, imprinting defects in chromosome region 15q11.2–q13 are directly linked to Angelman Syndrome (severe developmental delay) and Prader–Willi Syndrome (severe hypotonia).[91] Imprinting defects within the imprinted *GNAS1* gene lead to pseudohypoparathyroidism, which causes electrolyte imbalances and low-levels of calcium in the blood.[92,93] Loss of imprinting at the *H19|IGF2* region leads to Beckwith–Wiedemann syndrome (large body size, embryonal tumors, and visceromegaly).[94] Allele-specific DNA methylation aberrations also extend beyond imprinted genes. Experiments have shown the expression of the X-linked gene *MeCP2* is tightly regulated *in vivo*. Mutations in this gene cause Rett Syndrome, while a 2X over expression of the gene leads to neurological disorders.[95] Severe phenotypes result from deviations from expected ASM patterns, which demonstrate the importance of a tightly regulated methylation regulatory network.

General Allele-Specific Methylation

Beyond imprinted genes whose methylation patterns are determined during gametogenesis, there are additional ASM patterns in mouse and human. A human methylation-sensitive single nucleotide polymorphism (SNP) analysis involving 12 samples covering 6 tissues revealed 16 ASM candidates.[56] The authors found that 12 of those sites strongly associated with nearby SNPs. Unlike in imprinted genes, where methylation patterns are determined by parent-of-origin, these ASM sites were dependent on sequence. The authors labeled these ASM regions as epi-haplotypes. Further studying ASM in human, a recent targeted genome survey of CpG islands in chromosome 12 and chromosome 20 across 16 cell lines found that many CpGs with methylation frequencies between 0.25 and 0.75 ('fuzzily methylated regions').[38,39] Using SNPs to separate alleles in order to explore areas of fuzzy methylation, the authors found between 23 and 37% of heterozygous SNPs showed significant ASM behavior. Due to the high CpG content of the targeted regions, many SNPs overlapped with CpG sites, and many of the ASM examples were due to sequence differences between the alleles (Figures 2c and d). For example, a G/A heterozygous SNP at the second position of a CpG dinucleotide would appear as a CG sequence on one allele and a CA on another allele. The CA sequence would then be completely unmethylated based solely on its sequence. Additional examples of ASM without SNPs at a CpGs were also found. Interestingly, different ASM scales were detected as some regions contained only a single CpG with ASM, while other regions showed ASM spanning several CpGs. A comparison across cell lines revealed that only 6% of ASM behavior was conserved across cells that contained the same heterozygous SNP. Additional experiments in mouse have found hundreds of candidate ASM regions genome-wide.[96] Examining 18 of those sites in detail, the authors found 15 of them showed the same methylation levels in the male germ line and were only differentially methylated in somatic cells. Unlike imprinting, the methylation states of these 15 sites are not determined until after fertilization, which means a mechanism separate than imprinting is available for establishing ASM. Many of the ASM regions in this study were associated with sequence variants, which demonstrate the likely role of sequence specific factors in regulating ASM in mouse.[97–99] Another genome-wide analysis showed that 2704 SNPs displayed ASM and 90.3% of them were associated with cis-acting ASM.[100] ASM flipping has also been observed. ASM flipping describes a SNP region, where allele A is more methylated than allele B in one cell line and allele B is more methylated than allele A in another cell line. Instances of ASM flipping were found across cell line families in our targeted bisulfite sequencing study.[39] These studies validate the presence of cis-dependent ASM in various cell lines in both human and mouse. In the mammalian genome, ASM occurs outside of imprinted regions, at different scales, and may be regulated by cis-sequences.

Recent experiments have shown that cell line clonality can lead to biological artifacts and diminishes the *in vivo* applicability of results based on clonal cell lines. A study found that about 20% of tested lymphoblastoid cell lines were pauciclonal (consisting of only a few clones) or monoclonal.[101] In cell lines where only a few clones are present, sensitive allele-specific studies may not reflect the whole cell population *in vivo*. Such clonal cell lines are likely to exhibit artifacts like random monoallelic expression, which mask the biological signal found in the *in vivo* cell population. Affected studies include gene expression and methylation studies. Another study found that many traits (e.g., RNA transcript levels, drug responses) are highly variable across lymphoblastoid cell lines and that gene expression is better explained by artifacts (e.g., Epstein-Barr virus (EBV) copy numbers and growth rate) than by genetic variants.[102] However, another group that investigated multiple individual-specific cell types showed that clonality

did not significantly affect their ASE results.[103] Given these findings, experimentalists should account for the clonality of their investigated cell lines when evaluating results, especially in allele-specific studies.

Linked with ASM in many cases, ASE has been observed to occur randomly in clonal cell lines. Experiments comparing the ASE behavior of clonal cell lines derived from B-lymphoblastoid cells revealed the presence of a class of genes whose ASE behavior was random.[104] This class describes genes that are expressed on the paternal, maternal, or both alleles across clones. Such genes are spread throughout the genome and do not cluster together in enriched locations. Many of the genes (80%) that showed monoallelic expression in at least one clonal cell line showed biallelic expression in other clonal lines. This demonstrates that these genes can be expressed from either allele or both simultaneously. Overall 5–10% of human autosomal genes were found to be randomly monoallelically expressed. Additional studies need to be performed to demonstrate the prevalence of random monoallelic expression outside of the B-lymphoblastoid cell lines investigated in this paper. Connecting DNA methylation to ASE remains an interesting challenge but methylation studies have the potential to explain many of the recently observed ASE phenomena, including random monoallelic expression.

METHYLATION PATTERN ERASURE: iPS AND PGC DYNAMICS

Reprogramming involves the activation of a small set of transcription factors within a differentiated cell, which transforms the cell into an ES-like state.[105–108] The ability to reprogram differentiated cells into a pluripotent state raises hopes for improved transplantation and disease therapies. An issue with transforming cells into iPS cell lines is the low transformation efficiency (<0.1%). The activation of certain transcription factors during transformation may bias the iPS cell toward certain differentiated cell types or lead to tumorigenesis.[106,109] In order for iPS-based therapies to become widespread, iPS transformation efficiency and safety must be improved.[108,110–115]

Methylation and iPS Reprogramming

During differentiation, which is reversed during iPS transformation, massive changes in methylation occur genome-wide. Genome-wide methylome constructions of a fibroblast, IMR90, and an ES cell, H1, have revealed that many regions of the IMR90 genome are methylated at much lower frequency than in H1[1] (Figure 3a). However, the authors found 491 regions where IMR90 was more methylated than in H1 (DMRs). The 139 genes associated with these DMRs showed higher expression in H1 relative to IMR90 and 113 genes showed lower expression. The majority of these genes had DMRs within 2 kb upstream from the transcription start site (TSS) or 5′ untranslated region. The H1 cell line also showed non-CpG methylation (CHH and CHG) throughout the genome while only CpG methylation was present in IMR90. Relating the methylation signatures of iPS cells to fibroblasts and ES cells, a targeted bisulfite study showed that the methylation signatures of human reprogrammed cells from two fibroblast cell lines clustered closely to ES cells and were distinct from their respective untransformed cell lines[38] (Figure 3b). However, the iPS cell lines were still distinguishable from ES cell lines, showing that differences between the iPS cell lines and ES cell lines exist. This difference was also found in another study where iPS cells had regions with methylation signatures that were dissimilar to both ES cells and their respective untransformed cell lines.[41] These studies show that there are significant methylation differences between ES and differentiated cells. Successful iPS transformation involves massive methylation changes in order for a differentiated cell to arrive at a pluripotent state.

Reprogramming involves DNA demethylation and inhibiting DNA demethylation thus diminishes reprogramming efficiency. A mammalian protein associated with demethylation is activation-induced cytidine deaminase, AID, which is a 5-meC deaminase. AID was first discovered as a necessary component in Class Switch Recombination and high levels of Somatic Hypermutation processes, which take place within B-cells.[116–118] Initially thought to be expressed only in immune cells, further research showed that AID was also expressed in pluripotent cells.[119] AID's association with demethylation stems from its conversion of methylated cytosines to thymines.[120,121] The converted base is removed via G:T mismatch repair, which, for example, can be performed by Mbd4.[119,122,123] Demonstrating that the protein AID is important in iPS reprogramming, a group used mouse ES cells fused with human fibroblasts cells to show that AID regulates the transcription of Oct4 and Nanog.[124] The fusion of a mouse ES cell and a human fibroblast cell creates a heterokaryon and this process efficiently produces iPS cells. Seventy percent of heterokaryons expressed high levels of human Oct4, Nanog, and GAPDH (negative control protein) transcripts on the third day after fusion. Noticing that AID transcripts were found in the heterokaryons and that AID was bound to the methylated promoter

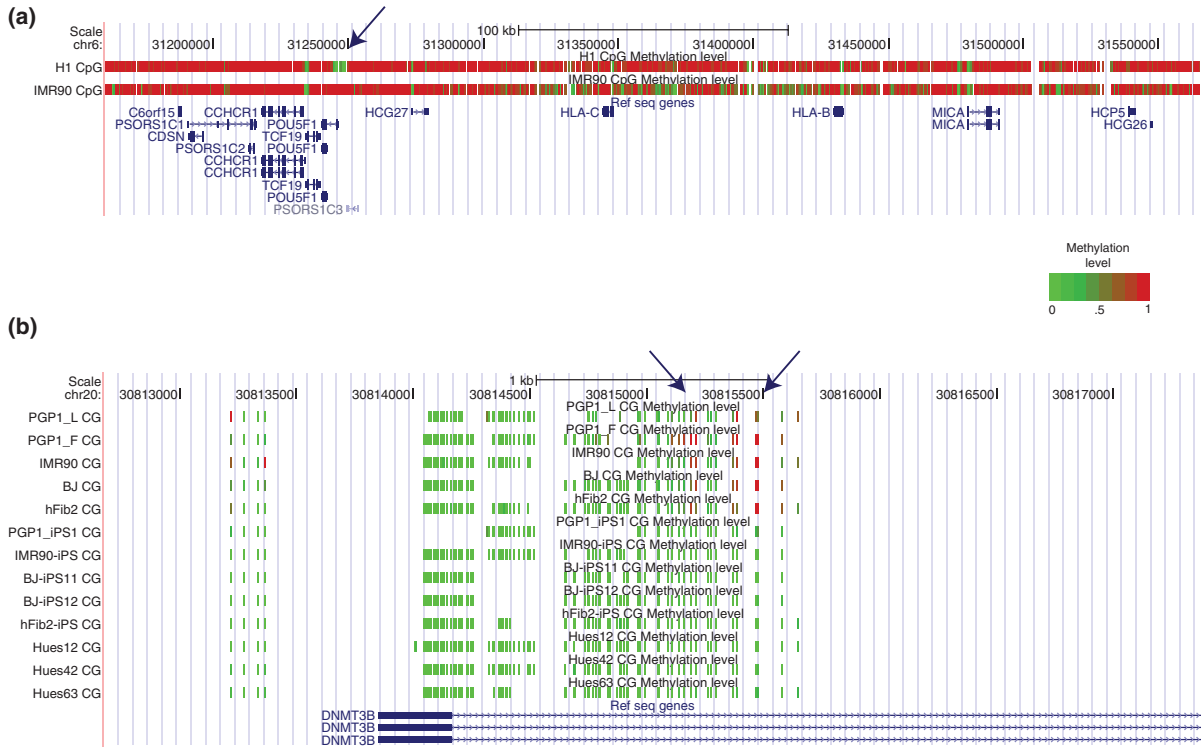


FIGURE 3 | Methylation frequency differences between differentiated and pluripotent cell lines. (a) This figure shows the methylation status of H1 (ES cell) and IMR90 (fibroblast cell) across a 3 kb region of chromosome 6. Each vertical bar represents a CpG site whose color is dependent on its methylation level. The arrow indicates the promoter of Pou5F1, which codes for the Oct4 protein. Oct4 is a master regulatory of pluripotency in ES cells. The promoter is unmethylated in H1, where Oct4 is transcribed, and methylated in IMR90, where Oct4 is repressed. Regions upstream of the Pou5F1 promoter show larges swaths of differential methylation. Successfully reprogramming involving large methylation changes in such areas. (b) A detailed view of the DNMT3b TSS site across 13 cell lines. The arrows indicate CpG sites that show cell type specific methylation. The iPS lines have methylation signatures that match the ES cells instead of their pretransformed cell types. PGP_1 is a lymphoblastoid cell line while PGP1_F, IMR90, BJ, and hFib2 are fibroblast cell lines. PGP1_iPS1 was transformed from the PGP1 fibroblast. Hues12, Hues42, and Hues63 are embryonic stem cell lines.

regions of Oct4 and Nanog in human fibroblasts, the authors used several siRNAs that recognized different regions on the AID transcript to knock down AID in both the mouse ES and human fibroblast cells. These siRNAs were then transfused into both cells lines 24 h before the cells were fused. Oct4 and Nanog expression levels were reduced by at least 80% in the AID-knockout heterokaryons relative to the control and bisulfite sequencing revealed significant methylation at the promoter regions of Oct4 and Nanog relative to the control. The results show that the presence of AID likely leads to demethylation at Oct and Nanog promoters, which results in their transcription. Although the knockdown of AID resulted in a significant inhibition of iPS reprogramming, over expression of AID did not change the efficiency of iPS reprogramming. The discovery of AID's activity during reprogramming does not exclude the possibility of other factors being involved in DNA demethylation. Additionally, the details of AID's demethylation mechanism in pluripotent cells are still unknown. Additional research is needed to eliminate these ambiguities and undoubtedly show that AID is a necessary factor in active DNA demethylation. Given these caveats, this experiment demonstrates the role of AID in demethylation is important for the reprogramming of somatic cells into a pluripotent state.

Primordial Germ Cell Methylation Erasure

In additional to affecting reprogramming efficiency, AID has been shown to play a role in DNA demethylation in primordial germ cells (PGCs). The erasure of DNA methylation patterns, including those in imprinted regions and the inactivated X_i, is important for PGCs as it eliminates epimutations and allows a return to pluripotency.[125–128] Studies have shown that methylation signatures in imprinted regions, single copy genes, and certain repeats (e.g. LINE1) are significantly demethylated in PGCs between 11 and 13.5 days *post coitum*.[129] A recent study expanded on

these past findings and looked at methylation effects caused by AID in various genetic elements within murine PGCs.[130] Demethylation in PGCs was found be global and included gene regions, transposons, and repeats; the final epigenetic state of PGCs at 13.5 days *post coitum* was termed an 'epigenetic ground state', where the genome is mostly demethylated and histone marks are mostly absent.[126,130–133] Comparing $Aid^{-/-}$ knockout to wild-type mice, the authors found that the AID deficient PGCs were significantly more methylated; there was a sex bias as the female AID-knockout cell lines were more methylated than the male knockout cell lines. Interestingly the promoters between the AID knockout and wild-type mice did not show a significant difference; transposons, introns, and exons were more methylated in the knockout line.

RELATIONSHIPS BETWEEN TRANSCRIPTION FACTORS AND DNA METHYLATION

Although genome-wide methylomes have recently become available for various human cell lines, the reasons why genomic regions are unmethylated or methylated are mostly unclear. For example, the observation that CpG islands in the human genome are mostly unmethylated has not been explained on a genome-wide scale. However, studies are beginning to reveal a deeper relationship between specific transcription factors and local DNA methylation patterns (Table 1). These studies serve to build a foundation that will thoroughly explain the nonrandom nature of DNA methylation. A set of three transcription factors that tightly associate with DNA methylation are discussed below.

TABLE 1 | Summary of Mentioned DNA Methylation-Related Transcription Factors

Transcription Factor	Methylated Binding Specificity	DNA Methylation-Related Consequence of Transcription Factor Binding
Cfp1	Unmethylated	(1) Binds to unmethylated CpG islands and (2) recruits Set1a and Set1b complexes, which trimethylates histones at H3K4 independently of RNA PolII binding.[134–136]
CTCF	Unmethylated	(1) Blocks enhancer-promoter interactions by acting as an insulator[137,138] and (2) protects genomic regions from nearby DNA methylation encroachment.[139–141]
MeCP2	Methylated	Associates genome-wide with methylated CpGs in mouse. Leads to H3 deacetylation, which results in global chromatin structure changes.[95] Loss of MeCP2 leads to up-regulation of 2,184 genes and down-regulation of 377 genes in mouse.[142]
Sp1	Unmethylated	Protects genomic regions from nearby DNA methylation encroachment.[143,144]

The CCCTC-Binding Factor

CTCF contains an 11 zinc finger DNA-binding domain that is highly conserved in higher eukaryotes.[145,146] Although initially reported as a silencer, CTCF is a versatile transcription factor that has been seen to enhance and repress transcription at promoters as well as act as an insulator.[137,147–150] It binds to a range of sequences and its zinc fingers allosterically customize according to DNA sequence and nearby cofactors, which confers CTCF's ability to recognize a wide variety of sequences.[145]

CTCF plays a critical role in the establishment and maintenance of the imprinted *H19|IGF2* region. Directly upstream of the *H19* locus is an imprint control element (ICE) that contains four CTCF binding sites. This ICE is necessary for the allele specific expression of both *H19* and *IGF2* genes. Methylation of this ICE on the paternal allele prevents CTCF binding within the ICE and also suppresses *H19* expression via promoter methylation.[151–153] CTCF binding on the maternal allele prevents enhancers downstream of *H19* from interacting with the promoter region of *IGF2*. Without CTCF's insulating function on the paternal allele, enhancers downstream of *H19* are able to interact with the *IGF2* promoter, which leads to the paternal expression of *IGF2*. Hypomethylation of the *H19* promoter on the maternal allele leads to maternal *H19* expression. Recent data has shown that CTCF is an essential factor in the formation of long-range contacts (i.e., loops) between this imprinted locus and distal enhancer regions.[138,154] In addition to mediating long-range regulatory interactions, CTCF can protect nearby regions from encroaching DNA methylation. An experiment looking at the retinoblastoma tumor suppressor (*Rb*) gene found that CTCF binding prevents DNA methylation from spreading into the *Rb*'s CpG island promoter region.[155] The ability of CTCF to protect promoters from repressive DNA methylation was also seen in the

c-MYC promoter[140] and in the *BRCA1* promoter.[141] These examples show that CTCF binding depends on the methylation state of its binding sequence, and its binding can mediate long-range regulatory interactions and affect local methylation patterns.

CXXC Finger Protein 1

CXXC finger protein 1 (Cfp1) contains a conserved CXXC domain that is sufficient for specific binding to unmethylated CpG dinucleotides and two PHD fingers.[156,157] The yeast analog of Cfp1, Spp1, has been found to bind to the H3K4me3 histone modification, which is associated with poised or active transcription, via PHD finger interactions.[158] In addition to recognizing H3K4me3 modifications in yeast, Cfp1 associates with the mammalian Set1a and Set1b methyltransferase complexes,[135,136] which are known to trimethylate H3K4.

Using ChIP-Seq technology on mouse brain nuclei, a recent study found that Cfp1 was localized at unmethylated CpG islands.[134] CpG islands bound by Cfp1 were found to also exhibit H3K4me3 modifications. Unmethylated CpG islands not occupied by Cfp1 or H3K4me3 where found to align with the repressive histone modification mark H3K27me3.[159] To test the direct link between Cfp1 and H3K4me3 modification, the authors inserted promoterless CpG-rich DNA into regions not associated with H3K4me3. The results showed Cfp1 binding peaks at the DNA sequence insertion sites and H3K4me3 was present as these locations. However, there was no sign of RNA Pol II binding. This test showed that the H3K4me3 modification is not a byproduct of transcription since it occurs independently of RNA Pol II binding. Cfp1 is a transcription factor that binds to unmethylated CpG islands and modifies nearby histones in a manner that promotes RNA Pol II recruitment.

MeCP2

While the previous two transcription factors bind to unmethylated sequences, methyl CpG binding protein 2 (MeCP2) recognizes CpG methylated DNA.[160,161] MeCP2 is a part of the MBD protein family (MeCP2, MBD1, MBD2, and MBD4), which consists of proteins that share a homologous methylated DNA recognition domain, but contain different transcriptional repression domains.[162–164] The mammalian form of MBD3 is an exception, since there is a mutation in its MBD that prevents it from binding to methylated CpGs.[165] MeCP2 binding was originally thought to cause transcriptional repression due to its recruitment of histone modifying enzymes, which results in repressive histone marks around regions bound by MeCP2, namely histone deacetylation and trimethylation of H3K9.[166–168] MeCP2 is expressed highly only in neurons and its expression level is tightly regulated. Embryonic *MeCP2*-null mice die at week 12 while mature mice under- or over-expressing *MeCP2* (e.g., heterozygous females) exhibit neurological disorders.[169–172]

The transcriptional role of MeCP2 has recently changed from a transcriptional repressor to a global chromatin remodeler that can lead to the expression and repression of various genes. Since MeCP2 binding leads to traditionally associated transcriptionally repressive histone modification patterns, MeCP2 binding near genes was thought to silence them. However, a recent study showed that mice over-expressing *MeCP2* led to a significant number of up regulated genes relative to *MeCP2*-null mice (2184 genes up regulated while 377 genes down regulated).[142] These authors found that a global transcriptional activator CREB1 and MeCP2 co-occupied promoters of many activated genes, including the *Creb1* promoter, and CREB1 was also found to copurify with MeCP2. These findings are inconsistent with the model of *MeCP2* as a transcriptional repressor. Further changing the view of *MeCP2*'s role as a repressor, another study performed on mouse brain tissue revealed MeCP2 binding correlates with DNA methylation levels genome-wide. MeCP2 is a protein highly expressed in neuronal cells and its concentration in neurons is similar to that of nucleosomes.[95] Using ChIP-Seq to enrich for sequences bound to MeCP2, the authors found that 56% of the mouse genome showed some MeCP2 binding activity. The intensity of the MeCP2 signal increased as DNA methylation density increased. MeCP2 binding was not found at unmethylated CpG islands, suggesting a genome-wide methylated DNA binding preference for MeCP2. To test the relationship between H3 acetylation and MeCP2 abundance, the level of H3 acetylation was measured between wild-type and *MeCP2*-null neurons. There was a 2.6-fold increase in H3 acetylation in the *MeCP2*-null neurons relative to the wild-type. As a control, wild-type and *MeCP2*-null glial cells were also examined, but no significant differences in H3 acetylation were found. The *MeCP2*-null brain cell lines also showed a 1.6-fold increase in transcription of repetitive elements, while no expression change was seen in Actb, c-Myc, or tyrosine hydroxylase genes. These authors present MeCP2 as a protein that binds to methylated DNA on a global scale. In accordance with previous experiments, MeCP2 does reduce H3 acetylation levels and inhibits transcription of repetitive elements. Due to the global scale of histone deacetylation in the presence of MeCP2, MeCP2's association with deacetylation, which is traditionally

associated with silencing transcription, and activation of gene transcription are not necessarily mutually exclusive. The consequences of global histone modification changes are unknown and a rigorous study into this subject may bridge these two seemingly opposing characteristics of MeCP2.

Transcription Factors Regulate and Are Regulated by DNA Methylation

Experiments focused on specific transcription factors have shown the dependence of DNA methylation states on transcription factor binding. However, the list of transcription factors affected by methylation is short and the relationships between DNA methylation and most transcription factors are still unclear. A recent motif based study looked at CpG island sequences that are resistant to de novo methylation in colorectal and leukemia cells. Using motif analysis tools, the authors found a set of motifs that are strongly associated with de novo methylation resistance.[173] The most significant motifs are YY1, Sp1, and NRF1. Sp1 has been known to protect CpG islands from methylation.[143,144] YY1 recruits the Polycomb Repressive Complex 2, PRC2, which is known for transcriptional silencing via the H3K27me3 histone modification.[174] PRC2 may also act in a context-specific manner as an allele-specific antagonist to DNA methylation.[175] Elucidating the causes and consequences of DNA methylation will yield a better understanding of cell differentiation, reprogramming to pluripotency, and cancer development since these processes involve massive genome-wide methylation changes.

HYDROXYMETHYLCYTOSINE: A NEW CYTOSINE MODIFICATION

Complementing AID's demethylation activities via deamination and G:T mismatch repair, a new family of mammalian proteins have been reported to convert methylated cytosine (5meC) bases into hydroxymethylcytosine (5hmC) *in vivo*[176,177] (Figure 4). The first study identified an unknown base that started appearing on Thin Layer Chromatography plates.[176] The authors found that 5hmC comigrated with this unknown spot and mass spectrometry later confirmed the presence of this compound as 5hmC. This new cytosine modification was found the nuclei of Purkinje neurons and granule cells. Hmc5 was found to make up 0.59% of all bases in Purkinje DNA and 0.23% in granule cell nuclei. A seemingly insignificant fraction of DNA exhibits this modification but relative to CpG, which makes up about 1% of all bases, the abundance of 5hmC is indeed significant.

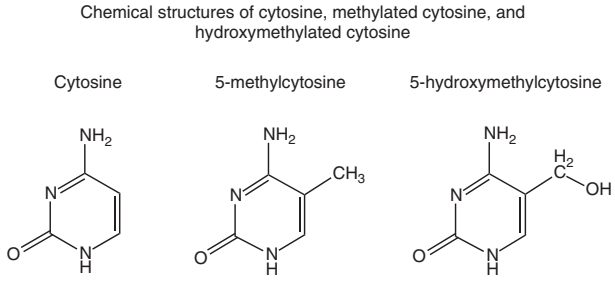


FIGURE 4 | Structures of *in vivo* cytosine modifications found in humans. This figure shows the structure of cytosine and its two modified forms found in the human genome.

Another study searched for proteins responsible for this novel modified base. Trypanosome proteins JPB1 and JPB2 are known to hydroxylate and glucosylate the methyl group in a thymine, which results in b-D-glucosyl hydroxymethyluracil.[178] Using the predicted oxygenase domain in those proteins to find proteins with similar functions in human, this study found that the predicated trypanosome oxygenase domain shared significant homology with the human proteins TET1, TET2, and TET3. To test for TET1's activity, the authors transfected hemagglutinin-tagged TET1 into embryonic kidney cells (HEK 293). Kidney cells that showed an increasingly strong signal for hemagglutinin (HA) also showed a decreasing signal for methylated cytosine (5meC) relative to a mock control. Using methylation-sensitive restriction enzyme techniques, the authors discovered a base of unknown identity. Using mass spectrometry and comparing with the fragmentation pattern of 5hmC, the authors found that the unknown base was hmC. Furthermore, the authors found that 5hmC made up 4–6% of all cytosine species at MspI cleavage sites in mouse ES cells while 5meC made up 55–60%. Knockdown of TET1 via RNAi led to a 40% decrease in 5hmC levels. The continued presence of 5hmC in the absence of TET1 was attributed to the presence of additional proteins, such as TET2 and TET3. Finally, the induced differentiation of mouse ES cells by removal of LIF for 5 days led to an 80% TET1 transcript decline and about a 40% drop in 5hmC levels. Protocols for large-scale identification of 5hmC have yet been presented, but a recent study evaluated the ability of current methylation detection techniques to detect 5meC, 5hmC, or both. 5hmC appears to be resistant to bisulfite and bisulfite sequencing results thusly do not distinguish between 5meC and 5hmC. Polymerase chain reaction (PCR) appears to amplify 5hmC and meC sequences with similar efficiency. The monoclonal antibody used in MeDIP experiments,[179–181] however, is 5meC specific. The proteins MBD2b, MBD1, MBD4, and

MeCP2 also bind 5meC specifically.[162,182–184] These studies show the presence of another type of modified cytosine *in vivo*. Bilsulfite sequencing cannot discern between the cytosine types but MeDIP and certain protein complexes can. Recent evidence showed that single molecule real time sequencing technology (SMRT) discriminates between the two types of cytosine modifications, too. The presence of 5meC and 5hmC affects polymerase kinetics in SMRT and this can be exploited to differentiate between the two modifications at a base pair resolution.[185]

CONCLUSION

Methylation is present in various higher eukaryotes but methylation targeting of genomic regions is variable. Vertebrates tend to have highly methylated genomes. In humans, embryonic stem cells tend to be more methylated than differentiated cell lines and human ES cells also exhibit non-CpG methylation (CHH and CHG). Methylation signatures are cell type specific and the methylation frequencies of CpGs can identify human tissues, ES cells, and iPS cells. Methylation signatures can also discriminate between normal and cancer cells and even detect cancer subtypes. DNA methylation can also be allele specific, which differentiates the parental alleles. Chromosome X inactivation and imprinted genes regions are examples of this behavior. Recent research has found that allele specific methylation (ASM) is also found throughout autosomal regions outside of imprinted gene regions. The number of identified allele specifically methylated sites (ASM) is expanding rapidly as genome-wide experiments show that ASM is prevalent throughout the genome.

One reason for the cell type specificity of DNA methylation is due to its interactions with transcription factors. Transcription factors like Cfp1, Sp1, and CTCF bind to unmethylated sequences while other transcription factors like MeCP2 recognize methylated sequences only. Cfp1 binding in CpG islands leads to H3K4me3 modification of nearby histones, which is an RNA PolII positive mark. Sp1 and CTCF can protect certain regions (e.g., promoter regions) from nearby DNA methylation encroachment. MeCP2 binds to methylated sequences genome-wide and regulates chromatin structure globally. Motif analyses have revealed transcription factor motifs, YY1, Sp1, and NRF1, are associated with a genomic region's resistance to methylation. Although the mammalian DNMT protein family has been established as the source of de novo methylation and methylation maintenance, only recently have mammalian proteins been identified to be involved in methylation erasure. AID is one such candidate protein whose presence is important for successful iPS reprogramming and the establishment of an 'epigenetic ground state' in primordial germ cells.

DNA methylation investigations have revealed the presence of another modified cytosine type in human, 5-hydroxymethylcytosine. TET1 protein modulates levels of 5-hydroxymethylcytosine base and TET2 and TET3 are suspected to play a similar role. The currently popular technique of bisulfite sequencing does not discriminate between the two modifications but newer single molecule real time sequencing technology shows promise in identifying both 5-methylcytosine and 5-hydroxymethylcytosine individually.

Researchers are slowly unraveling the mammalian proteins involved in DNA methylation and the complex mammalian regulatory mechanisms of which DNA methylation plays a part, which include histone modifications, noncoding RNAs, and transcription factors. Additional work in this field will elucidate the role of this epigenetic marker in the context of the human transcriptional regulatory network. However, even as DNA methylation is not completely understood, it stills offers clinical applications and guides researches to biologically active regions in the genome.

REFERENCES

1. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-filippini J, Nery JR, Lee L, Ye Z, Ngo QM, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 2009, 462:315–322.

2. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 2008, 133:523–536.

3. Cokus SJ. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 2008, 452:215–219.

4. Sasai N, Nakao M, Defossez PA. Sequence-specific recognition of methylated DNA by human zinc-finger proteins. *Nucleic Acids Res* 2010, 38:5015–5022.

5. Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 2007, 317:1760–1764.

6. Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE. The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr Biol* 2007, 17:379–384.
7. Filion GJ, Zhenilo S, Salozhin S, Yamada D, Prokhortchouk E, Defossez PA. A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol Cell Biol* 2006, 26:169–181.
8. Daniel JM, Spring CM, Crawford HC, Reynolds AB, Baig A. The p120(ctn)-binding partner Kaiso is a bimodal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Res* 2002, 30:2911–2919.
9. Defossez PA, Kelly KF, Filion GJ, Perez-Torrado R, Magdinier F, Menoni H, Nordgaard CL, Daniel JM, Gilson E. The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. *J Biol Chem* 2005, 280:43017–43023.
10. Margot JB, Ehrenhofer-Murray AE, Leonhardt H. Interactions within the mammalian DNA methyltransferase family. *BMC Mol Biol* 2003, 4.7.
11. Li E. Chromatin modification and epigenetic reprogramming in mammalian development. *Nat Rev Genet* 2002, 3:662–673.
12. Lan J, Hua S, He X, Zhang Y. DNA methyltransferases and methyl-binding proteins of mammals. *Acta Biochim Biophys Sin (Shanghai)* 2010, 42:243–252.
13. Bestor TH. The DNA methyltransferases of mammals. *Hum Mol Genet* 2000, 9:2395–2402.
14. Goll MG, Kirpekar F, Maggert KA, Yoder JA, Hsieh CL, Zhang X, Golic KG, Jacobsen SE, Bestor TH. Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* 2006, 311:395–398.
15. Liang G, Chan MF, Tomigahara Y, Tsai YC, Gonzales FA, Li E, Laird PW, Jones PA. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol Cell Biol* 2002, 22:480–491.
16. Chen T, Hevi S, Gay F, Tsujimoto N, He T, Zhang B, Ueda Y, Li E. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nat Genet* 2007, 39:391–396.
17. Dong A, Yoder JA, Zhang X, Zhou L, Bestor TH, Cheng X. Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Res* 2001, 29:439–448.
18. Yoder JA, Bestor TH. A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum Mol Genet* 1998, 7:279–284.
19. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992, 69:915–926.
20. Gowher H, Liebert K, Hermann A, Xu G, Jeltsch A. Mechanism of stimulation of catalytic activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-methyltransferases by Dnmt3L. *J Biol Chem* 2005, 280:13341–13348.
21. Chen ZX, Mann JR, Hsieh CL, Riggs AD, Chedin F. Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. *J Cell Biochem* 2005, 95:902–917.
22. Hata K, Okano M, Lei H, Li E. Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* 2002, 129:1983–1993.
23. Chedin F, Lieber MR, Hsieh CL. The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc Natl Acad Sci U S A* 2002, 99:16916–16921.
24. Holz-Schietinger C, Reich NO. The inherent processivity of the human de novo DNA methyltransferase 3A (DNMT3A) is enhanced by DNMT3L. *J Biol Chem* 2010.
25. Glass JL, Fazzari MJ, Ferguson-Smith AC, Greally JM. CG dinucleotide periodicities recognized by the Dnmt3a-Dnmt3L complex are distinctive at retroelements and imprinted domains. *Mamm Genome* 2009, 20:633–643.
26. Jurkowska RZ, Anspach N, Urbanke C, Jia D, Reinhardt R, Nellen W, Cheng X, Jeltsch A. Formation of nucleoprotein filaments by mammalian DNA methyltransferase Dnmt3a in complex with regulator Dnmt3L. *Nucleic Acids Res* 2008, 36:6656–6663.
27. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, et al. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A* 2010, 107:8689–8694.
28. Zemach A, McDaniel IE, Silva P, Zilberman D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 2010, 328:916–919.
29. Field LM, Lyko F, Mandrioli M, Prantera G. DNA methylation in insects. *Insect Mol Biol* 2004, 13:109–115.
30. Gold M, Gefter M, Hausmann R, Hurwitz J. Methylation of DNA. *J Gen Physiol* 1966, 49:5–28.
31. Doerfler W. Patterns of DNA methylation–evolutionary vestiges of foreign DNA inactivation as a host defense mechanism. A proposal. *Biol Chem Hoppe Seyler* 1991, 372:557–564.
32. Matzke MA, Mette MF, Aufsatz W, Jakowitsch J, Matzke AJ. Host defenses to parasitic sequences and the evolution of epigenetic control mechanisms. *Genetica* 1999, 107:271–287.
33. Wolffe AP, Matzke MA. Epigenetics: regulation through repression. *Science* 1999, 286:481–486.
34. Zilberman D. The evolving functions of DNA methylation. *Curr Opin Plant Biol* 2008, 11:554–559.

35. Hughes T, Webb R, Fei Y, Wren JD, Sawalha AH. DNA methylome in human CD4+ T cells identifies transcriptionally repressive and non-repressive methylation peaks. *Genes Immun* 2010.

36. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Kin Sung KW, Rigoutsos I, Loring J, et al. Dynamic changes in the human methylome during differentiation. *Genome Res* 2010, 20:320–331.

37. Xiang H, Zhu J, Chen Q, Dai F, Li X, Li M, Zhang H, Zhang G, Li D, Dong Y, et al. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat Biotechnol* 2010, 28:516–520.

38. Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, Egli D, Maherali N, Park IH, Yu J, et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotechnol* 2009, 27:353–360.

39. Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res* 2010, 20:883–889.

40. Irizarry Ra, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, Cui H, Gabo K, Rongione M, Webster M, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genetics* 2009, 41:178–186.

41. Doi A, Park I-H, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genetics* 2009, 41:1350–1353.

42. Figueroa ME, Lugthart S, Li Y, Erpelinck-Verschueren C, Deng X, Christos PJ, Schifano E, Booth J, van Putten W, Skrabanek L, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. *Cancer Cell* 2010, 17:13–27.

43. International Agency for Research in Cancer. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissue*. 4th ed. Geneva: World Health Organization, 2008, 441.

44. Flanagan JM, Cocciardi S, Waddell N, Johnstone CN, Marsh A, Henderson S, Simpson P, da Silva L, Khanna K, Lakhani S, et al. DNA methylome of familial breast cancer identifies distinct profiles defined by mutation status. *American Journal of Human Genetics* 2010, 86:420–433.

45. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001, 98:10869–10874.

46. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, Jacobs IJ, et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One* 2009, 4.e8274.

47. Grutzmann R, Molnar B, Pilarsky C, Habermann JK, Schlag PM, Saeger HD, Miehlke S, Stolz T, Model F, Roblick UJ, et al. Sensitive detection of colorectal cancer in peripheral blood by septin 9 DNA methylation assay. *PLoS One* 2008, 3.e3759.

48. deVos T, Tetzner R, Model F, Weiss G, Schuster M, Distler J, Steiger KV, Grutzmann R, Pilarsky C, Habermann JK, et al. Circulating methylated SEPT9 DNA in plasma is a biomarker for colorectal cancer. *Clin Chem* 2009, 55:1337–1346.

49. Muller HM, Fiegl H, Widschwendter A, Widschwendter M. Prognostic DNA methylation marker in serum of cancer patients. *Ann N Y Acad Sci* 2004, 1022:44–49.

50. Muller HM, Widschwendter A, Fiegl H, Ivarsson L, Goebel G, Perkmann E, Marth C, Widschwendter M. DNA methylation in serum of breast cancer patients: an independent prognostic marker. *Cancer Res* 2003, 63:7641–7645.

51. Widschwendter M, Apostolidou S, Raum E, Rothenbacher D, Fiegl H, Menon U, Stegmaier C, Jacobs IJ, Brenner H. Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. *PLoS One* 2008, 3.e2656.

52. Luedi PP, Dietrich FS, Weidman JR, Bosko JM, Jirtle RL, Hartemink AJ. Computational and experimental identification of novel human imprinted genes. *Genome Res* 2007, 17:1723–1730.

53. Frost JM, Moore GE. The importance of imprinting in the human placenta. *Plos Genet* 2010, 6.e1001015.

54. Hudson QJ, Kulinski TM, Huetter SP, Barlow DP. Genomic imprinting mechanisms in embryonic and extraembryonic mouse tissues. *Heredity* 2010, 105:45–56.

55. Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 2010, 329:643–648.

56. Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, Li K, Murty VV, Schupf N, Vilain E. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet* 2008, 40:904–908.

57. Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 1991, 349:38–44.

58. Brown CJ, Lafreniere RG, Powers VE, Sebastio G, Ballabio A, Pettigrew AL, Ledbetter DH, Levy E,

Craig IW, Willard HF. Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* 1991, 349:82–84.

59. Chureau C, Prissette M, Bourdet A, Barbe V, Cattolico L, Jones L, Eggen A, Avner P, Duret L. Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Res* 2002, 12:894–908.
60. Willard HF. X chromosome inactivation, XIST, and pursuit of the X-inactivation center. *Cell* 1996, 86:5–7.
61. Lee JT, Davidow LS, Warshawsky D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet* 1999, 21:400–404.
62. Lee JT. Lessons from X-chromosome inactivation: long ncRNA as guides and tethers to the epigenome. *Genes & Development* 2009, 23:1831–1842.
63. Donohoe ME, Silva SS, Pinter SF, Xu N, Lee JT. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* 2009, 460:128–132.
64. Anguera MC, Sun BK, Xu N, Lee JT. X-chromosome kiss and tell: how the Xs go their separate ways. *Cold Spring Harb Symp Quant Biol* 2006, 71:429–437.
65. Xu N, Tsai CL, Lee JT. Transient homologous chromosome pairing marks the onset of X inactivation. *Science* 2006, 311:1149–1152.
66. Lee JT. Molecular links between X-inactivation and autosomal imprinting: X-inactivation as a driving force for the evolution of imprinting? *Curr Biol* 2003, 13:R242–254.
67. Xu N, Donohoe ME, Silva SS, Lee JT. Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat Genet* 2007, 39: 1390–1396.
68. Barakat TS, Jonkers I, Monkhorst K, Gribnau J. X-changing information on X inactivation. *Exp Cell Res* 2010, 316:679–687.
69. Jonkers I, Barakat TS, Achame EM, Monkhorst K, Kenter A, Rentmeester E, Grosveld F, Grootegoed JA, Gribnau J. RNF12 is an X-Encoded dose-dependent activator of X chromosome inactivation. *Cell* 2009, 139:999–1011.
70. Sun BK, Deaton AM, Lee JT. A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell* 2006, 21: 617–628.
71. Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 2008, 322:750–756.
72. Maenner S, Blaud M, Fouillen L, Savoye A, Marchand V, Dubois A, Sanglier-Cianferani S, Van Dorsselaer A, Clerc P, Avner P, et al. 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol* 2010, 8.e1000276.
73. Wutz A, Rasmussen TP, Jaenisch R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* 2002, 30:167–174.
74. Vire E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM, et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 2006, 439:871–874.
75. Zilberman D, Cao X, Johansen LK, Xie Z, Carrington JC, Jacobsen SE. Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. *Curr Biol* 2004, 14:1214–1220.
76. Qi Y, He X, Wang XJ, Kohany O, Jurka J, Hannon GJ. Distinct catalytic and non-catalytic roles of ARGONAUTE4 in RNA-directed DNA methylation. *Nature* 2006, 443:1008–1012.
77. Gao Z, Liu HL, Daxinger L, Pontes O, He X, Qian W, Lin H, Xie M, Lorkovic ZJ, Zhang S, et al. An RNA polymerase II- and AGO4-associated protein acts in RNA-directed DNA methylation. *Nature* 2010, 465:106–109.
78. He XJ, Hsu YF, Zhu S, Wierzbicki AT, Pontes O, Pikaard CS, Liu HL, Wang CS, Jin H, Zhu JK. An effector of RNA-directed DNA methylation in arabidopsis is an ARGONAUTE 4- and RNA-binding protein. *Cell* 2009, 137:498–508.
79. Lucifero D, Mertineit C, Clarke HJ, Bestor TH, TraslerJM. Methylation dynamics of imprinted genes in mouse germ cells. *Genomics* 2002, 79:530–538.
80. Li JY, Lees-Murdock DJ, Xu GL, Walsh CP. Timing of establishment of paternal methylation imprints in the mouse. *Genomics* 2004, 84:952–960.
81. Davis TL, Trasler JM, Moss SB, Yang GJ, Bartolomei MS. Acquisition of the H19 methylation imprint occurs differentially on the parental alleles during spermatogenesis. *Genomics* 1999, 58:18–28.
82. Bourc'his D, Proudhon C. Sexual dimorphism in parental imprint ontogeny and contribution to embryonic development. *Mol Cell Endocrinol* 2008, 282:87–94.
83. Hirasawa R, Chiba H, Kaneda M, Tajima S, Li E, Jaenisch R, Sasaki H. Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev* 2008, 22:1607–1616.
84. Zhao Q, Rank G, Tan YT, Li H, Moritz RL, Simpson RJ, Cerruti L, Curtis DJ, Patel DJ, Allis CD, et al. PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat Struct Mol Biol* 2009, 16:304–311.
85. Jia D, Jurkowska RZ, Zhang X, Jeltsch A, Cheng X. Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* 2007, 449:248–251.

86. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 2007, 130:77–88.

87. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 2007, 39:311–318.

88. Koerner MV, Pauler FM, Huang R, Barlow DP. The function of non-coding RNAs in genomic imprinting. *Development* 2009, 136:1771–1783.

89. Mancini-Dinardo D, Steele SJ, Levorse JM, Ingram RS, Tilghman SM. Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes Dev* 2006, 20:1268–1282.

90. Shin JY, Fitzpatrick GV, Higgins MJ. Two distinct mechanisms of silencing by the KvDMR1 imprinting control region. *EMBO J* 2008, 27:168–178.

91. Sahoo T, del Gaudio D, German JR, Shinawi M, Peters SU, Person RE, Garnica A, Cheung SW, Beaudet AL. Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat Genet* 2008, 40:719–721.

92. Bastepe M, Pincus JE, Sugimoto T, Tojo K, Kanatani M, Azuma Y, Kruse K, Rosenbloom aL, Koshiyama H, Jüppner H. Positional dissociation between the genetic mutation responsible for pseudohypoparathyroidism type Ib and the associated methylation defect at exon A/B: evidence for a long-range regulatory element within the imprinted GNAS1 locus. *Human Molecular Genetics* 2001, 10:1231–1241.

93. Lecumberri B, Fernández-Rebollo E, Sentchordi L, Saavedra P, Bernal-Chico A, Pallardo LF, Bustos JMJn, Castaño L, de Santiago M, Hiort O, et al. Coexistence of two different pseudohypoparathyroidism subtypes (Ia and Ib) in the same kindred with independent Gs{alpha} coding mutations and GNAS imprinting defects. *Journal of Medical Genetics* 2010, 47:276–280.

94. Cerrato F, Sparago A, Verde G, Crescenzo AD, Citro V, Cubellis MV, Rinaldi MM, Boccuto L, Neri G, Magnani C, et al. Different mechanisms cause imprinting defects at the IGF2/H19 locus in Beckwith-Wiedemann syndrome and Wilms, Äô tumour. *Hum. Mol Genet* 2008, 17:1427–1435.

95. Skene PJ, Illingworth RS, Webb S, Kerr ARW, James KD, Turner DJ, Andrews R, Bird AP. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Molecular Cell* 2010, 37:457–468.

96. Schilling E, El Chartouni C, Rehli M. Allele-specific DNA methylation in mouse strains is mainly determined by cis-acting sequences. *Genome Res* 2009, 19:2028–2035.

97. Murrell A, Heeson S, Cooper WN, Douglas E, Apostolidou S, Moore GE, Maher ER, Reik W. An association between variants in the IGF2 gene and Beckwith-Wiedemann syndrome: interaction between genotype and epigenotype. *Hum Mol Genet* 2004, 13:247–255.

98. Flanagan JM, Popendikyte V, Pozdniakovaite N, Sobolev M, Assadzadeh A, Schumacher A, Zangeneh M, Lau L, Virtanen C, Wang SC, et al. Intra- and interindividual epigenetic variation in human germ cells. *Am J Hum Genet* 2006, 79:67–84.

99. Heijmans BT, Kremer D, Tobi EW, Boomsma DI, Slagboom PE. Heritable rather than age-related environmental and stochastic factors dominate variation in DNA methylation of the human IGF2/H19 locus. *Hum Mol Genet* 2007, 16:547–554.

100. Schalkwyk LC, Meaburn EL, Smith R, Dempster EL, Jeffries AR, Davies MN, Plomin R, Mill J. Allelic skewing of DNA methylation is widespread across the genome. *Am J Hum Genet* 2010, 86:196–212.

101. Plagnol V, Uz E, Wallace C, Stevens H, Clayton D, Ozcelik T, Todd JA. Extreme clonality in lymphoblastoid cell lines with implications for allele specific expression analyses. *PLoS One* 2008, 3.e2966.

102. Choy E, Yelensky R, Bonakdar S, Plenge RM, Saxena R, De Jager PL, Shaw SY, Wolfish CS, Slavik JM, Cotsapas C, et al. Genetic analysis of human traits in vitro: drug response and gene expression in lymphoblastoid cell lines. *Plos Genet* 2008, 4.e1000287.

103. Lee JH, Park IH, Gao Y, Li JB, Li Z, Daley GQ, Zhang K, Church GM. A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. *Plos Genet* 2009, 5.e1000718.

104. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A. Widespread monoallelic expression on human autosomes. *Science* 2007, 318:1136–1140.

105. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006, 126:663–676.

106. Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, et al. Induced pluripotent stem cell lines derived from human somatic cells. *Science* 2007, 318:1917–1920.

107. Yu J, Vodyanik MA, He P, Slukvin, II, Thomson JA. Human embryonic stem cells reprogram myeloid precursors following cell-cell fusion. *Stem Cells* 2006, 24:168–176.

108. Yu J, Hu K, Smuga-Otto K, Tian S, Stewart R, Slukvin, II, Thomson JA. Human induced pluripotent stem cells free of vector and transgene sequences. *Science* 2009, 324:797–801.

109. Okita K, Ichisaka T, Yamanaka S. Generation of germline-competent induced pluripotent stem cells. *Nature* 2007, 448:313–317.

110. Banito A, Gil J. Induced pluripotent stem cells and senescence: learning the biology to improve the technology. *EMBO Rep* 2010, 11:353–359.

111. Kaji K, Norrby K, Paca A, Mileikovsky M, Mohseni P, Woltjen K. Virus-free induction of pluripotency and subsequent excision of reprogramming factors. *Nature* 2009, 458:771–775.

112. Kawamura T, Suzuki J, Wang YV, Menendez S, Morera LB, Raya A, Wahl GM, Belmonte JC. Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* 2009, 460:1140–1144.

113. Woltjen K, Michael IP, Mohseni P, Desai R, Mileikovsky M, Hamalainen R, Cowling R, Wang W, Liu P, Gertsenstein M, et al. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* 2009, 458:766–770.

114. Okita K, Nakagawa M, Hyenjong H, Ichisaka T, Yamanaka S. Generation of mouse induced pluripotent stem cells without viral vectors. *Science* 2008, 322:949–953.

115. Jia F, Wilson KD, Sun N, Gupta DM, Huang M, Li Z, Panetta NJ, Chen ZY, Robbins RC, Kay MA, et al. A nonviral minicircle vector for deriving human iPS cells. *Nat Methods* 2010, 7:197–199.

116. Longacre A, Storb U. A novel cytidine deaminase affects antibody diversity. *Cell* 2000, 102:541–544.

117. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 2000, 102:553–563.

118. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, Catalan N, Forveille M, Dufourcq-Labelouse R, Gennery A, et al. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 2000, 102:565–575.

119. Morgan HD, Dean W, Coker HA, Reik W, Petersen-Mahrt SK. Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J Biol Chem* 2004, 279:52353–52360.

120. Conticello SG, Langlois MA, Yang Z, Neuberger MS. DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv Immunol* 2007, 94:37–73.

121. Macduff DA, Harris RS. Directed DNA deamination by AID/APOBEC3 in immunity. *Curr Biol* 2006, 16:R186–189.

122. Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* 2008, 135:1201–1212.

123. Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* 1999, 401:301–304.

124. Bhutani N, Brady JJ, Damian M, Sacco A, Corbel SpY, Blau HM. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* 2010, 463:1042–1047.

125. Sasaki H, Matsui Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat Rev Genet* 2008, 9:129–140.

126. Hajkova P, Ancelin K, Waldmann T, Lacoste N, Lange UC, Cesari F, Lee C, Almouzni G, Schneider R, Surani MA. Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* 2008, 452:877–881.

127. Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. *Science* 2001, 293:1089–1093.

128. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. *Nature* 2000, 403:501–502.

129. Hajkova P, Erhardt S, Lane N, Haaf T, El-Maarri O, Reik W, Walter J, Surani MA. Epigenetic reprogramming in mouse primordial germ cells. *Mech Dev* 2002, 117:15–23.

130. Popp C, Dean W, Feng S, Cokus SJ, Andrews S, Pellegrini M, Jacobsen SE, Reik W. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 2010, 463:1101–1105.

131. Seki Y, Yamaji M, Yabuta Y, Sano M, Shigeta M, Matsui Y, Saga Y, Tachibana M, Shinkai Y, Saitou M. Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* 2007, 134:2627–2638.

132. Daujat S, Weiss T, Mohn F, Lange UC, Ziegler-Birling C, Zeissler U, Lappe M, Schubeler D, Torres-Padilla ME, Schneider R. H3K64 trimethylation marks heterochromatin and is dynamically remodeled during developmental reprogramming. *Nat Struct Mol Biol* 2009, 16:777–781.

133. Chuva de Sousa Lopes SM, Hayashi K, Shovlin TC, Mifsud W, Surani MA, McLaren A. X chromosome activity in mouse XX primordial germ cells. *Plos Genet* 2008, 4.e30.

134. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr ARW, Deaton Ae, Andrews R, James KD, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 2010, 464:1082–1086.

135. Lee JH, Tate CM, You JS, Skalnik DG. Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. *J Biol Chem* 2007, 282:13419–13428.

136. Lee JH, Skalnik DG. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1

histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J Biol Chem* 2005, 280:41725–41731.

137. Xie X, Mikkelsen TS, Gnirke A, Lindblad-Toh K, Kellis M, Lander ES. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. *Proc Natl Acad Sci U S A* 2007, 104:7145–7150.
138. Li T, Hu JF, Qiu X, Ling J, Chen H, Wang S, Hou A, Vu TH, Hoffman AR. CTCF regulates allelic expression of Igf2 by orchestrating a promoter-polycomb repressive complex 2 intrachromosomal loop. *Mol Cell Biol* 2008, 28:6473–6482.
139. Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* 2009, 19:24–32.
140. Gombert WM, Krumm A. Targeted deletion of multiple CTCF-binding elements in the human C-MYC gene reveals a requirement for CTCF in C-MYC expression. *PloS One* 2009, 4.e6109.
141. Ishihara K, Oshimura M, Nakao M. CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol Cell* 2006, 23:733–742.
142. Chahrour M, Jung SY, Shaw C, Zhou X, Wong ST, Qin J, Zoghbi HY. MeCP2, a key contributor to neurological disease, activates and represses transcription. *Science* 2008, 320:1224–1229.
143. Boumber Ya, Kondo Y, Chen X, Shen L, Guo Y, Tellez C, Estécio MRH, Ahmed S, Issa J-PJ. An Sp1/Sp3 binding polymorphism confers methylation protection. *Plos Genet* 2008, 4:e1000162–e1000162.
144. Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, Temper V, Razin A, Cedar H. Sp1 elements protect a CpG island from de novo methylation. *Nature* 1994, 371:435–438.
145. Ohlsson R, Renkawitz R, Lobanenkov V. CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet* 2001, 17:520–527.
146. Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell* 2009, 137:1194–1211.
147. Baniahmad A, Steiner C, Kohne AC, Renkawitz R. Modular structure of a chicken lysozyme silencer: involvement of an unusual thyroid hormone receptor binding site. *Cell* 1990, 61:505–514.
148. Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, Neiman PE, Lobanenkov VV. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol Cell Biol* 1993, 13:7612–7624.
149. Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, Goodwin GH. A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. *Oncogene* 1990, 5:1743–1753.
150. Vostrov AA, Quitschke WW. The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. *J Biol Chem* 1997, 272:33353–33359.
151. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 2000, 405:482–485.
152. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 2000, 405:486–489.
153. Szabo P, Tang SH, Rentsendorj A, Pfeifer GP, Mann JR. Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. *Curr Biol* 2000, 10:607–610.
154. Yoon YS, Jeong S, Rong Q, Park KY, Chung JH, Pfeifer K. Analysis of the H19ICR insulator. *Mol Cell Biol* 2007, 27:3499–3510.
155. De La Rosa-Velazquez IA, Rincon-Arano H, Benitez-Bribiesca L, Recillas-Targa F. Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. *Cancer Res* 2007, 67:2577–2585.
156. Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol Cell Biol* 2000, 20:2108–2121.
157. Lee JH, Voo KS, Skalnik DG. Identification and characterization of the DNA binding domain of CpG-binding protein. *J Biol Chem* 2001, 276:44669–44676.
158. Shi X, Kachirskaia I, Walter KL, Kuo JH, Lake A, Davrazou F, Chan SM, Martin DG, Fingerman IM, Briggs SD, et al. Proteome-wide analysis in Saccharomyces cerevisiae identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. *J Biol Chem* 2007, 282:2450–2455.
159. Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A. Dissecting direct reprogramming through integrative genomic analysis. *Nature* 2008, 454: 49–55.
160. Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 1989, 58:499–507.

161. Nan X, Campoy FJ, Bird A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* 1997, 88:471–481.

162. Bogdanovic O, Veenstra GJ. DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma* 2009, 118: 549–565.

163. Clouaire T, Stancheva I. Methyl-CpG binding proteins: specialized transcriptional repressors or structural components of chromatin? *Cell Mol Life Sci* 2008, 65:1509–1522.

164. Hendrich B, Tweedie S. The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *Trends Genet* 2003, 19:269–277.

165. Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* 1999, 13:1924–1935.

166. Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T. The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem* 2003, 278:4035–4040.

167. Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 1998, 393:386–389.

168. Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 1998, 19:187–191.

169. Balmer D, Goldstine J, Rao YM, LaSalle JM. Elevated methyl-CpG-binding protein 2 expression is acquired during postnatal human brain development and is correlated with alternative polyadenylation. *J Mol Med* 2003, 81:61–68.

170. Gonzales ML, LaSalle JM. The role of MeCP2 in brain development and neurodevelopmental disorders. *Curr Psychiatry Rep* 2010, 12:127–134.

171. Guy J, Hendrich B, Holmes M, Martin JE, Bird A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat Genet* 2001, 27:322–326.

172. Collins AL, Levenson JM, Vilaythong AP, Richman R, Armstrong DL, Noebels JL, David Sweatt J, Zoghbi HY. Mild overexpression of MeCP2 causes a progressive neurological disorder in mice. *Hum Mol Genet* 2004, 13:2679–2689.

173. Gebhard C, Benner C, Ehrich M, Schwarzfischer L, Schilling E, Klug M, Dietmaier W, Thiede C, Holler E, Andreesen R, et al. General transcription factor binding at CpG islands in normal cells correlates with resistance to de novo DNA methylation in cancer cells. *Cancer Res* 2010, 70:1398–1407.

174. Caretti G, Di Padova M, Micales B, Lyons GE, Sartorelli V. The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation. *Genes Dev* 2004, 18:2627–2638.

175. Lindroth AM, Park YJ, McLean CM, Dokshin GA, Persson JM, Herman H, Pasini D, Miro X, Donohoe ME, Lee JT, et al. Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. *Plos Genet* 2008, 4.e1000145.

176. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science (New York, NY)* 2009, 324:929–930.

177. Tahiliani M, Koh KP, Shen Y, Pastor Wa, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (New York, NY)* 2009, 324:930–935.

178. Cliffe LJ, Kieft R, Southern T, Birkeland SR, Marshall M, Sweeney K, Sabatini R. JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res* 2009, 37:1452–1462.

179. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, Schubeler D. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 2005, 37:853–862.

180. Wilson IM, Davies JJ, Weber M, Brown CJ, Alvarez CE, MacAulay C, Schubeler D, Lam WL. Epigenomics: mapping the methylome. *Cell Cycle* 2006, 5:155–158.

181. Jacinto FV, Ballestar E, Esteller M. Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. *Biotechniques* 2008, 44.35, 37, 39 passim.

182. Fraga MF, Ballestar E, Montoya G, Taysavang P, Wade PA, Esteller M. The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res* 2003, 31:1765–1774.

183. Fatemi M, Wade PA. MBD family proteins: reading the epigenetic code. *J Cell Sci* 2006, 119:3033–3037.

184. Sasai N, Defossez PA. Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. *Int J Dev Biol* 2009, 53:323–334.

185. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, Korlach J, Turner SW. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 2010, 7:461–465.

