Dynamic regulation of DNA methylation during mammalian development

DNA methylation occurs on cytosines, is catalyzed by DNA methyltransferases (DNMTs), and is present at high levels in all vertebrates. DNA methylation plays essential roles in maintaining genome integrity, but its implication in orchestrating gene-expression patterns remained a matter of debate for a long time. Recent efforts to map DNA methylation at the genome level helped to get a better picture of the distribution of this mark and revealed that DNA methylation is more dynamic between cell types than previously anticipated. In particular, these datasets showed that DNA methylation is targeted to important developmental genes and might act as a barrier to prevent accidental cellular reprogramming. In this review, we will discuss the distribution and function of DNA methylation in mammalian genomes, with particular emphasis on the waves of global DNA methylation reprogramming occurring in early embryos and primordial germ cells.

KEYWORDS: CpG island • DNA methylation • DNMT • methylome • primordial germ cell • reprogramming

Sylvain Guibert, Thierry Forné & Michael Weber†

†Author for correspondence: Institute of Molecular Genetics, CNRS UMR 5535, 1919 Route de Mende, 34293 Montpellier Cedex 5, France
Tel.: +33 467 613 631
Fax: +33 467 040 231
michael.weber@igmm.cnrs.fr

The regulation of complex genomes involves not only sequence-specific protein factors, but also additional levels of regulation such as DNA modifications, histone modifications and chromatin remodeling. These modifications are often referred to as being 'epigenetic', although most of them do not fulfill the strict definition that implies a sequence-independent inheritance during cell division [1]. The establishment of unique epigenetic profiles in different cell types is believed to be important in maintaining cellular identity. Research in this field has expanded over the last decade as novel technologies emerged to map these modifications genome-wide.

DNA methylation occurs on cytosines and is the only known epigenetic modification that directly affects the DNA molecule in eukaryotes. The addition of a methyl group occurs on the carbon 5 of the cytosine and is catalyzed by enzymes called DNA methyltransferases (DNMTs) (Figure 1A). In animals, cytosine methylation occurs almost exclusively in the context of CG dinucleotides, whereas methylated cytosines are found in various sequence contexts in plants (e.g., *Arabidopsis thaliana*) or fungi (e.g., *Neurospora crassa*) [2]. However, cytosine methylation is not found in all eukaryotes, as it is absent in classical genetic models such as *Saccharomyces cerevisiae* and *Caenorhabditis elegans*. In insects, cytosine methylation has been detected in some species but not others: it is very rare in *Drosophila melanogaster*, but has recently been reported to be abundant in the honey bee *Apis mellifera* [3]. This complex evolutionary history is one factor contributing to the fact that the function of DNA methylation is still not clearly understood.

In addition, the distribution of cytosine methylation within the genome also varies between organisms (Figure 1B). In invertebrates, fungi and plants, the most common pattern is mosaic methylation with methylation-free domains interspersed with heavily methylated sequences. For example, in *Neurospora crassa*, cytosine methylation is restricted to repeat elements. In *Arabidopsis thaliana*, methylation is specifically targeted to repeats (via a mechanism involving small interfering RNAs) and the body of certain genes [4]. In various invertebrates, such as the sea urchin *Ciona intestinalis* or the honey bee *Apis mellifera*, cytosine methylation seems to be enriched in the transcription unit of genes rather than repeat elements [3,5]. In contrast, DNA methylation in mammals is present at high levels throughout most of the genome (see 'Distribution of DNA methylation in mammalian genomes'). In fact, earlier studies using *Hpa*II digestion of genomic DNA suggested that global genome methylation is present in all vertebrates [6]. The functional significance of the transition from mosaic methylation to global genome methylation at the invertebrate/vertebrate boundary is still unclear [7].

Cytosine methylation has been associated with repressed chromatin and inhibition of promoter activity. Two models of repression

future medicine part of fsg

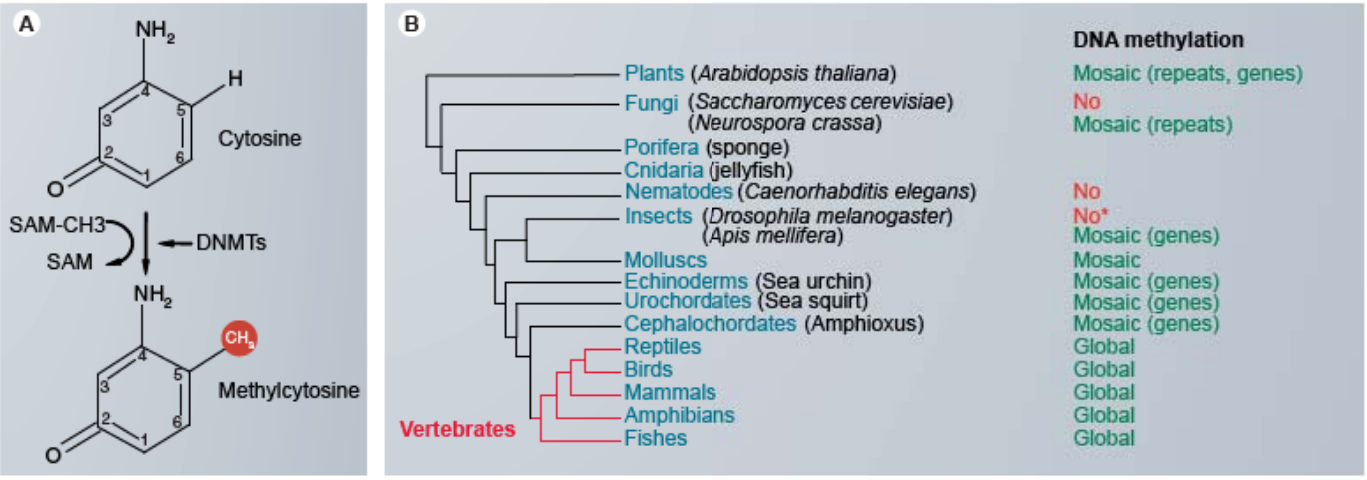

Figure 1. Cytosine methylation in eukaryotes. (A) Members of the DNMT family catalyze the transfer of a methyl group (CH_3) from SAM to the carbon-5 position of cytosine. **(B)** Schematic phylogenic tree representing the distribution of DNA methylation in eukaryotes. Presence of DNA methylation, indicated on the right, is based on analysis performed in representative members of each group shown in parenthesis. No cytosine methylation is detectable in *Saccharomyces cerevisiae* or *Caenorhabditis elegans*, while only traces of cytosine methylation have been found in *Drosophila melanogaster* (*). In plants and other invertebrates, the most characteristic pattern is mosaic methylation with cytosine methylation being restricted to repetitive and/or gene bodies. In contrast, all vertebrates tested have shown global methylation throughout the genome.
DNMT: DNA methyltransferase; SAM: S-adenosylmethionine.

have been proposed: cytosine methylation can directly prevent the binding of transcription factors; or it indirectly modulates the chromatin environment through the recruitment of methyl-binding proteins (MBPs) and cofactors such as histone deacetylases (HDACs). In addition, profiles of DNA methylation are perturbed in all types of cancer, with both genome-wide hypomethylation and locus-specific hypermethylation of tumor-suppressor genes. Details on the perturbation of DNA methylation patterns in cancer cells can be found in other recent review articles [8]. In this review, we will focus on the distribution and function of DNA methylation in mammalian genomes, with particular emphasis on the changes in DNA methylation profiles that happen during normal development and cellular differentiation.

The DNA methylome

Novel technologies to map DNA methylation

Our understanding of the biological significance of DNA methylation remained limited for a long time, partly because it was difficult to know where it is located at the genome level. In the past years, novel approaches have emerged to map DNA methylation genome-wide in an unbiased way. Some approaches directly isolate methylated DNA fragments by affinity approaches after the DNA has been fragmented. Methylated DNA can be isolated with anti-5mC antibodies, such as in the methylated DNA immunoprecipitation (MeDIP) assay [9], or binding to methyl-CpG binding proteins immobilized on a sepharose matrix, such as in the methylated-CpG island recovery assay (MIRA) [10]. Purified methylated DNA fragments can subsequently be used for microarray hybridization or high-throughput sequencing (Figure 2A). These approaches allow rapid profiling of DNA methylation, but one caveat is that they lack sensitivity to detect changes in DNA methylation at single CpGs.

The gold standard approach to study DNA methylation is to treat DNA with sodium bisulfite, a chemical compound that deaminates Cs into Ts only if the C is unmethylated. Therefore, by comparing the sequence of a defined region before and after treatment with sodium bisulfite, it is possible to have access to the methylation status of individuals CpGs. One very promising approach, termed BS-seq, is to combine conversion of the DNA with sodium bisulfite and high-throughput sequencing, which can potentially provide DNA methylation readout at single-nucleotide resolution for the entire genome (Figure 2B). The technical challenge here is on one hand to generate enough sequence reads, and on the other hand to remap to the reference genome the sequence reads that contain Cs converted into Ts, which becomes unfeasible with short sequence reads and complex genomes such as in mammals. Full-genome BS-seq has been validated for the first time in *Arabidopsis thaliana*, an organism with a small genome [11]. In mammals, BS-seq approaches have so far been applied on defined

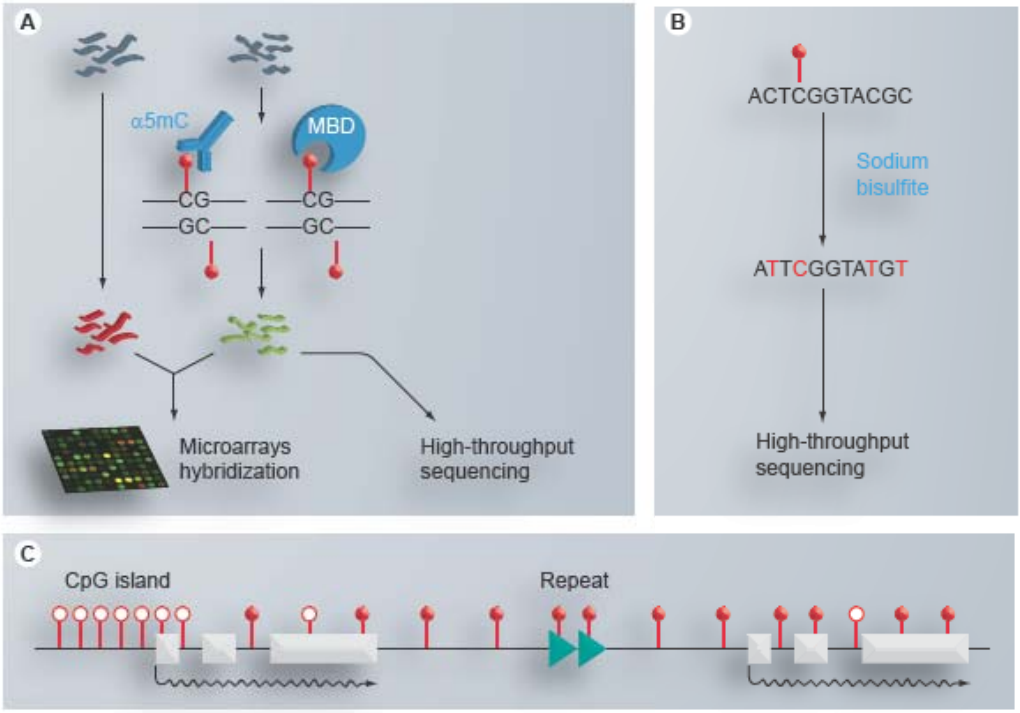

Figure 2. Genome-wide mapping of DNA methylation in mammalian genomes.
(A) Methylated DNA can be specifically isolated by binding to anti-5mC antibody, as is done in the methylated DNA immunoprecipitation (MeDIP) assay [9], or alternatively by capture with MBD proteins, as performed in the methylated-CpG island recovery assay (MIRA) [10]. Methylated DNA and total input DNA can subsequently be hybridized to any existing microarray. Methylated DNA isolated by affinity approaches can also be used as a template for high-throughput sequencing to obtain full-genome coverage. Red lollipops denote methyl groups. **(B)** Conversion of DNA with sodium bisulfite is the gold-standard approach to study DNA methylation. Sodium bisulfite converts Cs into Ts, unless the C is methylated. Combining bisulfite conversion with high-throughput sequencing can potentially give cytosine methylation readout at single base pair resolution for the entire genome. **(C)** Schematic representation of the distribution of DNA methylation in mammalian genomes. Most CpGs in gene bodies, intergenic regions and repeats are methylated. CpG-poor promoters (as for the gene on the right) can be DNA methylated even if the gene is active in the cellular population. CpG islands, which colocalize with 60% of all promoters (as for the gene on the left), are the only genomic compartment remaining mostly unmethylated. Gray boxes denote exons; triangles denote repetitive elements; CpGs are schematically represented with red (methylated) or white (unmethylated) lollipops.
5mC: 5-methylcytosine; MBD: methyl-CpG binding domain.

portions of the genome. One pioneering method, termed high-throughput reduced representation bisulfite sequencing (RRBS), involves the generation of a reduced representation library of the genome with *Msp*I digestion, followed by bisulfite conversion and sequencing on an Illumina (CA, USA) 1G analyzer, which provides methylation information for up to 5% of all CpGs in the genome [12]. Future improvements in the sequencing technology will probably help to increase the coverage for genome-scale bisulfite sequencing on large mammalian genomes.

Distribution of DNA methylation in mammalian genomes

Mammalian genomes are globally depleted in CpG dinucleotides, which means that they contain much less CpGs than one would predict if the nucleotides were organized randomly. The exceptions are short stretches of DNA, named CpG islands, which retained a high frequency of CpGs and typically contain between 30 and 200 CpGs per kilobase. CpG islands vary between 200 base pairs and a few kilobases in size, account for a very small fraction (1–2%) of the total genome, and often colocalize with gene promoters. This unequal distribution of CpGs needs to be taken into account when interpreting the repartition of 5-methylcytosine. Using recently developed technologies, several studies reported the first genome-wide profiles of DNA methylation in mammalian genomes. These maps agree that cytosine methylation is highly abundant in intergenic regions, coding regions and repeats throughout the genome, but rare at regulatory regions containing CpG islands

(FIGURE 2C). In a simplified view, virtually all CpGs in the genome show a high degree of methylation, except at CpG islands. This distribution is in striking contrast with the one observed in invertebrates, fungi and plants, where cytosine methylation is restricted to mobile elements or gene bodies (FIGURE 1B).

The distribution of DNA methylation in mammalian genomes suggests that its main functions are in the regulation of genome integrity rather than gene expression. One major role of DNA methylation is to keep repetitive DNA in a silent state in order to protect the genome from the potentially deleterious effect of mobile elements, in particular in the germ cells. In mouse, intracisternal A-particle (IAP) elements are transcriptionally reactivated in mouse embryos lacking DNA methylation [13]. IAP and LINE-1 elements are also reactivated in male germ cells lacking functional *de novo* DNA methylation [14]. DNA methylation also participates in the maintenance of genomic stability. In mouse male germ cells, hypomethylation of IAP and LINE-1 elements is associated with meiotic failure and widespread nonhomologous chromosome pairing [14]. In both human and mouse cultured cells, partial loss of global DNA methylation after depletion of DNMTs induces chromosomal abnormalities and changes in ploidy [15,16]. Many types of cancer cells also display global genome hypomethylation, which has been implicated in increased chromosomal instability and cancer progression [8]. The mechanisms by which DNA methylation protects from chromosomal rearrangements in mammals remain unknown, but one model is that unmasking of repetitive DNA results in increased illegitimate recombination between homologous repeats [17].

DNA methylation machinery

Four DNMTs sharing a conserved C-terminal catalytic domain have been identified in mammals (FIGURE 3A). The founding member, DNMT1, is responsible for the maintenance of CG methylation patterns after DNA replication. Propagation of CG methylation states is made possible because CG sites are symmetrical – that is, there are complementary CGs on both strands. This makes CG methylation a clear example of chromatin modification being faithfully inherited during cell division. The consensus model is that DNMT1 recognizes sites of hemimethylation that occur after DNA replication and methylates the CG on the newly synthesized DNA strand, therefore propagating methylation states. Indeed, DNMT1 has been shown to preferentially recognize hemimethylated substrate DNA *in vitro*, interact with **proliferating cell nuclear antigen** (PCNA) and localize at sites of DNA replication during S-phase. The function of DNMT1 in maintenance methylation is also supported by mouse genetic models. Mice with a partial loss-of-function mutation of *Dnmt1* show a reduction of genomic DNA methylation to 30% of normal levels and do not survive past mid-gestation [18], whereas the complete inactivation of DNMT1 results in more severe hypomethylation of the genome and lethality during gastrulation [19]. However, as the absence of the PCNA-binding domain induces only minor defects in DNA methylation maintenance, it has been hypothesized that additional factors help to recruit DNMT1 at sites of hemimethylation. Recently, two groups showed that ubiquitin-like, containing PHD and RING finger domains 1 (UHRF1; also known as NP95) binds hemimethylated CG through its SET and RING-associated (SRA) domain and recruits DNMT1 (FIGURE 3B) [20,21]. Importantly, *Uhrf1*$^{-/-}$ embryos show developmental arrest shortly after gastrulation and global demethylation similar to *Dnmt1*$^{-/-}$ embryos, indicating that UHRF1 is essential to promote maintenance of DNA methylation by DNMT1.

DNMT3A and DNMT3B enzymes are responsible for the establishment of DNA methylation during development [22]. Both enzymes are highly homologous (FIGURE 3A), but possess partially distinct functions. DNMT3B is already present in preimplantation stages [23,24] and *Dnmt3b*$^{-/-}$ mice die during early embryonic development at around embryonic day E10.5 [22]. DNMT3A is more prevalent in differentiated cells as its expression is detected from E9.5 onward [24], and *Dnmt3a*$^{-/-}$ mice survive until birth [22]. Both enzymes can also have different target specificities *in vivo*. In germ cells for example, DNMT3A is required for the establishment of DNA methylation at imprinting control regions (ICRs) in both male and female germline [25], whereas DNMT3B appears dispensable for the methylation of most ICRs [25,26]. Similarly, DNMT3A seems to be the sole enzyme methylating SineB1 elements in male germ cells, whereas DNMT3A and DNMT3B cooperate to methylate IAP and LINE-1 elements [25].

Another member, DNMT3L, is related in sequence to DNMT3A/B, but lacks a functional catalytic domain (FIGURE 3A). DNMT3L is abundant in early embryos, as well as in germ cells. Its absence, despite the lack of a functional

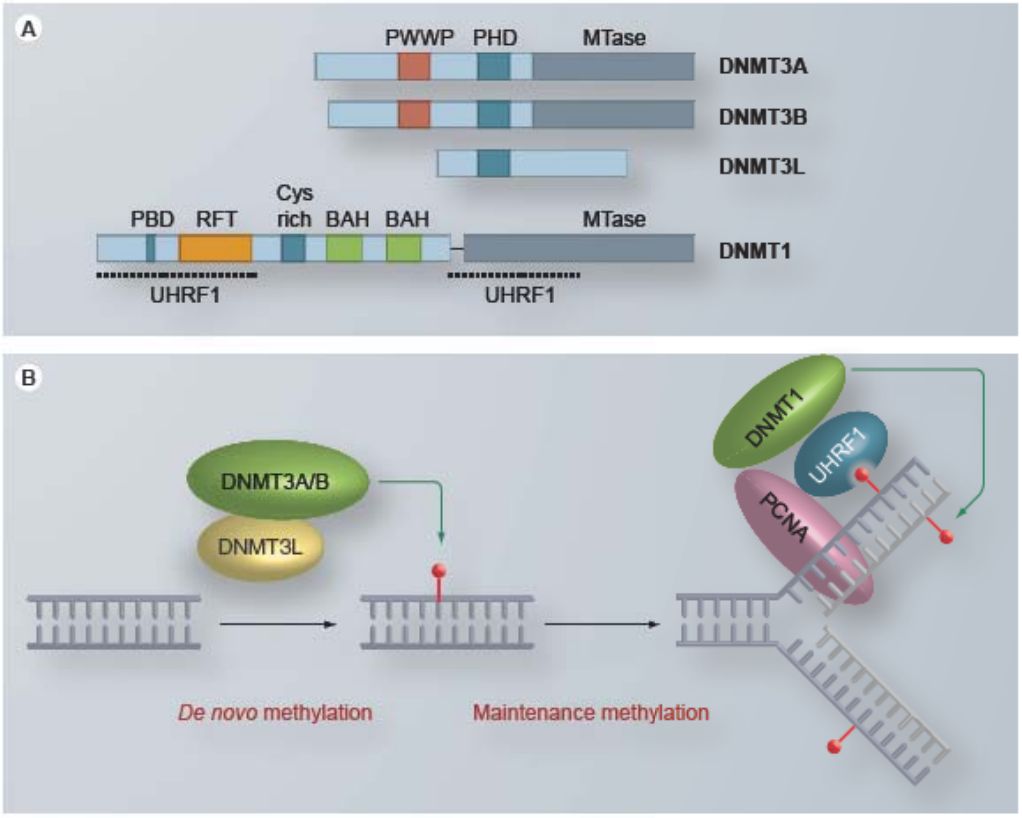

Figure 3. DNA methylation machinery in mammals. (A) Schematic structure of mouse DNMTs. DNMTs share a C-terminal catalytic domain, except DNMT3L which does not contain a functional catalytic domain but acts as a cofactor for DNMT3A/B. The N-terminal part of DNMT1 contains various regulatory domains, including interaction domains with PCNA, UHRF1 and replication foci. The dotted lines mark the domains involved in the binding to UHRF1. **(B)** *De novo* methylases DNMT3A/B are able to target previously unmethylated CpGs. DNMT3L forms a complex with DNMT3A and/or DNMT3B and acts as a factor by stimulating their *de novo* methylation activity and recruitment to chromatin. DNMT1, the maintenance methylase, preferentially recognizes sites of hemimethylation that occur after the DNA has been replicated. DNMT1 localizes at sites of hemimethylation through interactions with PCNA and UHRF1 and copies DNA methylation patterns onto the newly synthesized strand.
BAH: Brahma adjacent homology domain; Cys-rich: Cysteine-rich domain; DNMT: DNA methyltransferase; PBD: PCNA binding domain; PCNA: Proliferating cell nuclear antigen; PHD: Cysteine-rich plant homeodomain; PWWP: Proline- and tryptophane-rich domain; RFT: Replication foci targeting domain; UHRF1: Ubiquitin-like, containing PHD and RING finger domains 1.

methyltransferase domain, leads to a failure to methylate ICRs in both the male and female germline, as well as several repetitive elements in the male germline [14,25,27,28]. This phenotype is reminiscent to the one caused by the absence of DNMT3A. Biochemical and structural studies have shown that DNMT3L physically interacts with DNMT3 enzymes, which stimulates *de novo* methylation activity and promotes their recruitment to chromatin [29,30].

Proteins that specifically bind to methylated DNA (MBPs) are also important parts of the DNA methylation regulatory circuit. These proteins fall into two classes: those that share a methyl-CpG binding domain (MECP2, MBD1, MBD2 and MBD4), and those that recognize methylated DNA via zinc finger domains (KAISO, ZBTB4 and ZBTB38) [31,32]. Many of these proteins have been shown to recruit transcriptional co-repressors such as HDACs and histone methyltransferases (HMTs), thus providing a mechanism to interpret the DNA methylation signal by linking DNA modification and surrounding chromatin [31]. However, their precise function in DNA methylation-dependent gene silencing is still unclear. While chromatin immunoprecipitation experiments confirmed that MBPs bind to methylated ICRs [32,33], and that MECP2 binds to various methylated targets in human fibroblasts [34], a recent genome-wide study showed that MECP2 is mostly bound to unmethylated regions in human neurons [35].

In addition, very few genes have been shown to be deregulated in the absence of MBPs, and knockout mice show only subtle phenotypes. Mice lacking MBD1 or MBD2 are viable, but show defects in neuronal development and T-cell development, respectively [36,37]. Mice lacking the KAISO protein show no overt phenotypes [38]. *Mecp2*-deficient mice show neurological defects reminiscent of the Rett syndrome, a human neurological disorder associated with mutation in the *MECP2* gene [39]. One proposed explanation for the absence of developmental defects is that MBPs regulate a redundant set of targets. However, it has recently been shown that mice simultaneously lacking MBD2, MECP2 and KAISO still develop normally, thus leaving the developmental function of MBPs elusive [40].

DNA methylation & gene regulation: a complex story

It is a common belief that cell-type-specific-gene expression profiles are, at least in part, maintained by specific epigenetic profiles. This would imply that, as development proceeds, differentiated cells accumulate epigenetic marks that differ from those of pluripotent cells and other cell lineages. While recent genomic studies confirm that DNA methylation patterns vary to some extent between cell types, the importance of DNA methylation in orchestrating developmentally regulated gene-expression patterns is still debated.

Tissue-specific profiles of DNA methylation

Since its discovery decades ago, it has been postulated that specific DNA methylation profiles help to establish cell type specific gene expression profiles. This implies that cell-type-specific-gene expression correlates with cell-type-specific DNA methylation. Research over many years on candidate genes failed to establish such correlations, and led to the inverse dogma that CpG island promoters remain always unmethylated irrespective of gene activity [41]. There are, however, exceptions to this rule: CpG island methylation is associated with monoallelic gene expression in imprinted domains, and is also present on the inactive X chromosome in female cells. Other arguments against a widespread role of DNA methylation in gene regulation are that very few genes have been shown to be misregulated in absence of DNMTs [41–43] and that the MBPs are dispensable for mouse development [40]. Recently, several large-scale studies aimed at better understanding tissue-specific DNA methylation at the genome level, and shed new light on this important question [12,43–56]. These studies demonstrated that DNA methylation profiles are largely conserved between cell types of different developmental origin. However, despite global conservation, sites of differential methylation occur in various genomic contexts, indicating that DNA methylation profiles are more variable between cell types than previously expected.

One important observation is that tissue-specific DNA methylation appears more frequent at intergenic sites than in gene promoters [44,46,47,54,56], leaving open the possibility of a regulation by DNA methylation at distant regulatory elements. Regulation of gene expression by distant DNA methylation could be achieved through the control of distal enhancer activities [56], or via mechanisms that are commonly at work in imprinted domains – for example – regulation of distant insulators or noncoding RNAs [57]. Another provocative finding is the possible link between transcription and gene-body methylation. In apparent contrast with the repressive role of DNA methylation, studies on the X chromosome have shown that the active X chromosome is globally more methylated than the inactive one, with extra methylation being concentrated in gene bodies [9,58]. A recent study in human cells revealed that highly expressed genes are characterized by lower DNA methylation in the promoter, but higher DNA methylation in the gene body [55]. Similarly, another study in human identified several intragenic sites where DNA methylation is present only when the host gene is expressed [54]. Although gene-body methylation of active genes is also commonly seen in invertebrates, its regulatory role remains to be discovered [7].

Considering the historically unclear relationship between gene expression and DNA methylation at gene promoters, many groups decided to map DNA methylation specifically around transcription start sites. These studies revealed cell-type-specific profiles of DNA methylation at a significant fraction of gene promoters in both mouse and human [12,45–47,49–52,59]. One major conclusion from this work is that CpG-poor and CpG island promoters behave very differently in relation to DNA methylation. In the next paragraphs, we will consider these two classes of promoters separately, and explain in detail the role of DNA methylation at CpG-poor and CpG island promoters.

Methylation at CpG-poor promoters

Approximately 40% of mammalian promoters are CpG-poor, which means in practical

terms that they contain around 10–20 CpGs per kilobase. DNA methylation seems to be the default state of CpG-poor sequences in mammalian genomes, and CpG-poor promoters are no exception to this rule. Most CpG-poor promoters are found DNA methylated in a given cellular population even when the associated gene is active [12,49,50,52,55]. This suggests that a low density of cytosine methylation does not prevent gene transcription at many CpG-poor promoters. However, in apparent contradiction with this conclusion, it has also been observed that, when profiling promoter DNA methylation in differentiated cell types, most of the changes in relation to gene activity occur at CpG-poor promoters [47,51,59]. This inconsistency could be explained by the fact that CpG-poor promoters are occasionally hypomethylated as a consequence of the transcription process. For example, among the numerous copies of recombinant DNA (rDNA) genes in the human genome, only the active ones that are bound by RNA Polymerase I are hypomethylated in Hela cells, while silenced copies are methylated [60]. More recently, two groups reported that the transcriptional activation of ERα target genes (many containing CpG-poor promoters) induces transient cyclical demethylation of the promoter [61,62]. Therefore, if temporary hypomethylation marks only transcribed alleles, gene expression and DNA methylation at CpG-poor promoters would only correlate at the level of the cellular population if most alleles are transcribed simultaneously, which is rather uncommon [63]. Together these data indicate that at many CpG-poor promoters, DNA methylation might be dynamic in response to transcription and can be bypassed by activating signals. This suggests that at these promoters, low density of cytosine methylation is not causally involved in setting up gene-expression patterns, but may reinforce transcription cycles. However, this does not exclude that in other genomic regions, methylation of specific CpGs found in transcription factor binding sites is important for the spatio-temporal regulation of some genes.

Methylation at CpG island promoters

CpG island sequences are defined as having higher frequencies of CpGs, typically ranging from 30 to 200 CpGs per kilobase. In mammalian genomes, around 60% of all gene promoters colocalize with CpG islands. Genome-wide mapping confirmed that most of them remain constitutively unmethylated, even when the associated gene is not expressed [12,43,46,47,49,50,52–55]. This shows that CpG islands have the unique property to be protected from *de novo* DNA methylation, while all other genomic sequences are not. The mechanisms that protect CpG islands from the DNA methylation machinery are still not understood, and constitute one of the major puzzles in the field. One possibility is that CpG island hypomethylation is a consequence of specific chromatin features at these sequences, in particular the presence of histone H3 lysine 4 di- or tri-methylation (H3K4me2–3) (see 'Targeting of *de novo* DNA methylation').

However the same studies also revealed that a small fraction of CpG island promoters, ranging from 3 to 8% depending on the methodology and sample used, are hypermethylated in various normal somatic cells [46,49,50]. Unlike CpG-poor promoters, this high density of methylated cytosines is not compatible with gene transcription [50]. This demonstrates that, in contrast to old models, promoter-associated CpG islands can be DNA methylated outside of imprinted domains. Several lines of evidence indicate that CpG island methylation is generally acquired post-fertilization during somatic development. First, sperm has a much lower rate of CpG island methylation than any somatic cell type [46,50]. Second, the progressive increase in CpG island methylation during development can be directly visualized by profiling promoter DNA methylation in differentiating embryonic stem (ES) cells [12,45,52]. These studies show that CpG island methylation is rare in ES cells and increases upon differentiation at a few hundred promoters. Strikingly, gain of promoter methylation during stem cell differentiation is far more frequent than loss of methylation [45,52]. Therefore, although changes in expression are not linked to changes in promoter DNA methylation for most genes, promoter DNA methylation is reprogrammed at a small set of CpG-rich promoters during differentiation. As hypomethylation is the default state of CpG island sequences, this implies the existence of active mechanisms to attract DNA methylation or remove protection from DNA methylation at these targets (see 'Targeting of *de novo* DNA methylation').

Interestingly, methylated CpG islands share some common features. First, it has been observed that CpG islands promoters with moderate CpG richness (termed intermediate CpG promoters [ICPs]) are more likely to be methylated than CpG island promoters with high CpG richness (termed high CpG promoters [HCPs]) [44,50,52,55]. This suggests that CpG richness is

one factor protecting against *de novo* methylation, and that ICPs are the most likely promoters to be regulated by DNA methylation during development [50,55]. Second, many methylated CpG island promoters are associated with genes essential for development. For example, several CpG island promoters of genes regulating germ-cell development are methylated in somatic cells, but not in germ cells [49,50]. Homeobox genes and PAX family members have also been shown to be primary targets for lineage-specific CpG island methylation in human and mouse [46,64]. Finally, several genes regulating pluripotency and early patterning are *de novo* methylated during mouse ES cell differentiation, many of them associated with ICP promoters [12,45,52]. Altogether, these data suggest that CpG island methylation is primarily directed to genes important for lineage choice during development.

DNA methylation & cellular identity

The consensus arising from most recent genome-scale studies is that DNA methylation is involved in the control of specific genes only. One important question is to know whether DNA methylation is directly responsible for the transcriptional shut down at these genes, or is acquired after the gene has been repressed. Unfortunately, there are only few studies comparing the kinetics of acquisition of DNA methylation and gene repression. One study in differentiating mouse ES cells showed that acquisition of DNA methylation at the pluripotency gene *OCT4* is a late event not required for the transcriptional repression itself [65].

As presence of cytosine methylation at high density or in transcription factor-binding sites is believed to induce permanent epigenetic silencing, it is possible that late acquisition of DNA methylation protects against accidental reactivation of important genes in differentiated cells. This led to the model that DNA methylation of key genes during development helps to permanently restrict developmental potential and stabilizes cellular identity [66]. Several studies are in line with this model. It has recently been shown that trophoblast cells can be obtained from *Dnmt1*$^{-/-}$ but not wild-type ES cells, suggesting that DNA methylation is important for the stability of embryonic lineage determination [67]. In support of this, the authors demonstrated that the promoter of *Elf5*, a gene encoding an Ets transcription factor with a pivotal role in the development of the trophoblast lineage, is DNA methylated in embryonic lineages, but not in trophoblast cells. Similarly, differentiating ES cells that fail to methylate several early embryonic genes can be more easily reversed to an undifferentiated state [68], which indicates that loss of DNA methylation is a critical step to restore pluripotency. Indeed, the successful generation of induced pluripotent stem (iPS) cells from differentiated cells involves the reactivation of several pluripotency-related genes, which implies the restoration of an open chromatin state (marked by H3K4me3) and promoter DNA hypomethylation [69]. Partially reprogrammed cells show DNA hypermethylation at several pluripotency genes and inhibition of DNMT1 favors reprogramming [69]. Together, these data suggest that DNA methylation is a major barrier to prevent cellular reprogramming.

Reprogramming of DNA methylation during development

Two waves of global DNA methylation reprogramming

DNA methylation profiles are modulated during mammalian development, with both genome-wide and gene-specific changes occurring at various developmental stages. A first wave of extensive epigenetic reprogramming occurs during early embryogenesis to reach a low point of methylation at preimplantation stages (Figure 4A). After fertilization, the paternal genome is rapidly demethylated in the zygote prior to DNA replication, which implies the existence of an active demethylation process [66,70,71]. In contrast, the maternal genome is progressively demethylated owing to the absence of maintenance methylation [66]. DNA methylation patterns specific of differentiated cells are then progressively established after implantation by DNMT3A/B, and faithfully maintained during somatic cell divisions in the embryo and the adult animal through the action of DNMT1. However, the precise function of this reprogramming and the nature of the genomic sequences affected are still unclear. One can speculate that demethylation is necessary to erase possible methylation defects inherited from the gametes. Demethylation might also be required to restore an epigenetic state compatible with pluripotency in the early embryo. For example, certain key pluripotency genes such as *Oct4* and *Nanog* are DNA methylated in mature gametes, and therefore need to be epigenetically reprogrammed to allow proper transcriptional activation in preimplantation embryos [45]. Remarkably, differential methylation imprints established in the parental gametes resist this wave of demethylation, so that epigenetic marking of parental alleles is maintained in

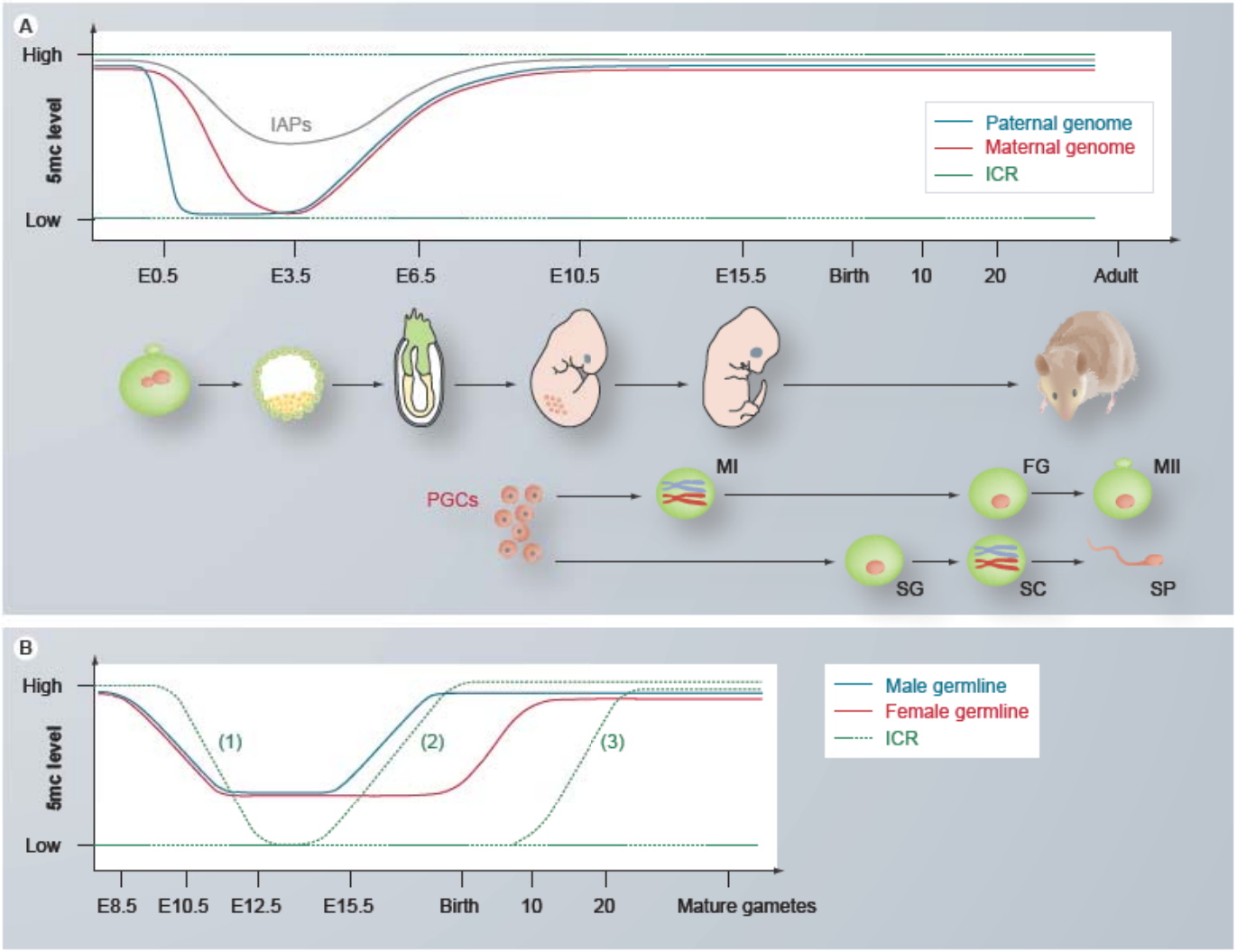

Figure 4. Reprogramming of DNA methylation during mouse development. The graphs depict the variations in global DNA methylation levels as a function of time during mouse somatic development **(A)** and germ cell development **(B)**. The timescales for somatic and germ cell development have been aligned. Global DNA methylation levels are extrapolated from immunofish and single-gene studies, and certainly not all sequences in the genome follow these dynamics. **(A)** Paternal (blue line) and maternal (red line) genomes are demethylated asynchronously during preimplantation development to reach a low point of methylation at the blastocyst stage (E3.5). IAP elements (gray line) partially resist this demethylation wave. After implantation, global DNA methylation patterns are restored and maintained throughout embryonic and adult life. ICRs (green lines) escape global reprogramming during somatic development: the methylated allele resists global demethylation in preimplantation embryos, whereas the unmethylated allele resists *de novo* methylation in post-implantation embryos. **(B)** PGCs gradually loose global DNA methylation until E12.5, while DNA methylation at ICRs (dotted green line 1) is erased in a defined window between E10.5 and E12.5. Retrotransposons, including IAP elements, are not fully demethylated at this stage. After sex determination has been initiated at E12.5, methylation patterns are re-established at different developmental time points in male and female germline. In male embryos (blue line), germ cells enter G1-phase mitotic arrest around E13.0 and methylation patterns are regained during the maturation to spermatogonia between E14.5 and birth. Methylation of paternal ICRs in the male germline (green line 2) occurs at the same time. In female embryos (red line), germ cells arrest in meiotic prophase around E13.0 and initiation of DNA methylation occurs after birth during oocyte growth. Methylation of maternal ICRs in the female germline (green line 3) occurs between 10 and 25 days after birth and is completed in fully grown oocytes.
FG: Fully grown oocyte; IAP: Intracisternal A-particle; ICR: Imprinting control region; MI: Meiosis I; MII: Meiosis II; PGC: Migrating primordial germ cell; SC: Spermatocyte; SG: Spermatogonia; SP: Spermatid.

the embryo. The specific maintenance of methylation imprints during preimplantation development requires DNMT1 [23] and additional factors such as PGC7/Stella [72] and Zfp57 [73].

A second wave of epigenetic reprogramming occurs during germ-cell development (FIGURE 4B). In mice, primordial germ cells (PGCs) become specified in the embryo around E7.25 and start migrating to the genital ridge where they settle by E11.5. Around this time, PGCs undergo extensive epigenetic reprogramming at the DNA methylation and chromatin level [74,75]. This reprogramming has been best studied in the context of genomic imprinting where parental

imprints have to be reset for the next generation. Erasure of methylation at imprinted ICRs occurs between E10.5 and E12.5, prior to sex determination, at most imprinted domains that have been tested [74,76,77]. This timing is consistent with an active demethylation process by mechanisms that are still not understood but might involve DNA repair (see 'Mechanisms of DNA demethylation'). In addition to the loss of methylation imprints, PGCs undergo a widespread loss of genomic DNA methylation to reach a low point at E12.5, as measured by immunostaining with anti-5methylcytosine antibodies [78]. Global DNA demethylation already starts in migrating PGCs after E8.0, and is accompanied by major chromatin changes with stepwise erasure of several histone modifications and exchange of histone variants between E8.5 and E12.5 [79,80]. Besides imprinted sequences, loss of DNA methylation also occurs in repetitive DNA, but it has been shown that some repetitive elements show only incomplete DNA methylation erasure [74,76,81], which might be essential to prevent the potential deleterious reactivation of mobile elements in germ cells. Hence the precise nature of the genomic sequences that lose DNA methylation in PGCs remains to be determined. Erasure of somatic methylation might contribute to returning PGCs to a more pluripotent state. Demethylation might also be required to allow the subsequent activation of genes important for the development of germ cells, as shown for three genes in the mouse [82].

Once sex determination has been initiated at E12.5, DNA methylation patterns are progressively restored, both at the global level and at imprinted genes where new methylation imprints are established according to the gender of the individual (Figure 4B). In the male germline, paternal methylation imprints are progressively established in developing spermatogonia between E14.5 and the newborn stage [25,74,77]. All repetitive elements regain full DNA methylation during the same developmental period, with a kinetic that appears more rapid than for ICRs [25,74,81]. DNA methylation of ICRs in the male germline depends on DNMT3L and DNMT3A [25,26], whereas DNA methylation of repetitive DNA requires the action of DNMT3L complemented by DNMT3A and DNMT3B which show different target specificities [14,25] (see 'DNA methylation machinery'). In the female germline, the initiation of DNA methylation occurs at various imprinted genes between 10 and 25 days after birth during the phase of oocyte growth [83,84], a process that is dependent on DNMT3L and DNMT3A [26,27]. Methylation of repetitive DNA is also believed to be restored during the phase of oocyte growth with a timing that remains to be precisely defined [74].

Mechanisms of DNA demethylation

In mammals, genome-wide demethylation occurs only at two developmental time-points: in the early embryo immediately after fertilization and in the developing PGCs. Outside of these defined developmental windows, DNA methylation patterns are believed to be faithfully transmitted through cell division and examples of DNA demethylation are very rare. Indeed, mammalian somatic development is thought to be characterized by a progressive increase in gene methylation concomitant to loss of pluripotency, with demethylation being virtually absent [45,52]. However, recent models suggesting that promoter demethylation occurs during transcription cycles might challenge this view [61,62].

In contrast to *de novo* cytosine methylation, the process that removes methylation from DNA is considerably less understood. Passive DNA demethylation – defined as the progressive dilution of DNA methylation during cell division owing to an absence of maintenance activity – cannot account for the rapid loss of DNA methylation observed in the male pronucleus or in PGCs. This implies the existence of active demethylation processes through mechanisms that are yet not elucidated. Several mechanisms have been proposed to explain active DNA demethylation, which fall into three categories [85].

The first proposed mechanism was the direct removal of the methyl group from the cytosine; however, this mechanism has been invalidated [85]. The second mechanism involves the base excision repair (BER) machinery, which would remove the 5-methylcytosine or the thymine generated by deamination of the 5-methylcytosine. This mechanism has been validated in plants with the discovery of DNA glycosylases (ROS1, DME, DML2 and DML3) that recognize and remove 5-methylcytosine from DNA. In mammals, there is no evidence for the involvement of 5-methylcytosine DNA glycosylases, but one likely scenario is the enzymatic deamination of 5-methylcytosine into thymine, followed by T-G mismatch repair (mediated by thymine DNA glycosylases TDG or MBD4) that specifically replaces thymine with cytosine. Candidate cytosine deaminases, AID and APOBEC1, are expressed at the time of epigenetic reprogramming and show a strong 5-methylcytosine

deaminase activity *in vitro* [86]. In support of this model, recent data obtained in zebrafish show a regulated DNA demethylase system involving the coupled action of the deaminase/glycosylase pair AID/Mbd4 [87]. Alternatively, it has also been proposed that 5-methylcytosine can be deaminated by DNMTs themselves in the absence of methyl donor *S*-adenylmethionine (SAM) during transcriptional cycling [61,62].

The third model involves the removal of entire nucleotides by the nucleotide excision repair (NER) machinery. Growth arrest and DNA-damage-inducible protein 45 α (GADD45A) has been proposed as one candidate involved in NER-mediated DNA demethylation [88]. Recently, it has been shown that GADD45A initiates demethylation of rRNA genes by recruiting the NER pathway [89]. However, the role of GADD45A in DNA demethylation remains debated. Other studies have reported that *Gadd45a*$^{-/-}$ mice have normal DNA methylation patterns [90], and suggest that GADD45A might act indirectly by stimulating the BER pathway [85,87].

Targeting of *de novo* DNA methylation

Understanding why specific sequences become targets for the DNA methylation machinery during development remains one of the major questions. This is particularly true for CpG-rich sequences that are by default protected from DNA methylation, which implies the existence of mechanisms to bypass this protection. DNMT3 enzymes do not seem to have specific sequence preferences, although structural studies have revealed that they methylate CpGs with an optimal interval of eight to ten base pairs [30]. Several recent studies shed new light on possible mechanisms directing DNA methylation to its targets.

One possible mechanism is the direct recruitment of DNMT enzymes by sequence-specific transcription factors (FIGURE 5A), as suggested for example, for the Ets transcription factor PU.1 [91]. In support of this model it has been shown recently that the tethering of a Krüppel-associated box (KRAB)-containing protein to a reporter construct induced promoter DNA methylation *in vivo* [92]. Interestingly, DNA methylation was induced if the KRAB protein was tethered during early development, but not in post-implantation stages. However, genetic proof is still needed to demonstrate that acquisition of DNA methylation on endogenous targets is perturbed *in vivo* in the absence of candidate transcription factors.

Recently, a link between small RNAs and silencing of repetitive DNA during spermatogenesis was discovered (FIGURE 5B). MILI and MIWI2, two members of the PIWI subfamily of Argonaute proteins, are expressed during gametogenesis. These proteins interact with a class of small RNAs called PIWI-interacting RNAs (piRNAs) that are largely derived from retrotransposons. Interestingly, MILI and MIWI2 mouse mutants show a derepression of IAP and LINE-1 elements in male germ cells, which is associated with a failure to properly *de novo* methylate these elements during spermatogenesis [93]. It is currently unclear how the piRNA pathway controls DNA methylation at retrotransposons. One model is that piRNAs act upstream of DNMT3 enzymes to serve as guides that direct DNA methylation machinery to transposons sequences [94].

Among other factors that direct DNA methylation, crosstalk between histone modifications and DNA methylation could play a major role. Although the importance of histone modifications as a prerequisite for DNA methylation has been so far demonstrated only in fungi and plants [95], similar mechanisms might be at work in mammals. Based on physical associations between HMTs and DNMTs, it has been postulated that histone H3 lysine 9 methylation (H3K9me) and lysine 27 methylation (H3K27me) could regulate DNA methylation; however, their role in setting up DNA methylation patterns during development remain to be determined. Recently, it has been shown that DNMT3L interacts through its PHD domain with histone H3 only when it is unmethylated on lysine 4 [29], suggesting that absence of H3K4me could be a signal to direct DNA methylation during development (FIGURE 5C). This is fully compatible with studies showing that unmethylated CpG islands are marked by H3K4me, and that the absence of H3K4me is the best predictor for DNA methylation at the genome level [12,50]. Another argument in favor of this hypothesis is that in somatic cells and early spermatogonia, ICRs around imprinted genes are marked by H3K4me specifically on the allele that is not DNA methylated [96]. Histone arginine methylation has also been proposed as a candidate for directing DNA methylation (FIGURE 5D). Using the human β-globin locus as a model, it has been shown that PRMT5-mediated histone H4 arginine 3 symmetrical methylation (H4R3me2s) is recognized by DNMT3A through its PHD motif and stimulates *de novo* DNA methylation [97]. All these models linking histone and DNA modifications in mammals still need to be tested genetically.

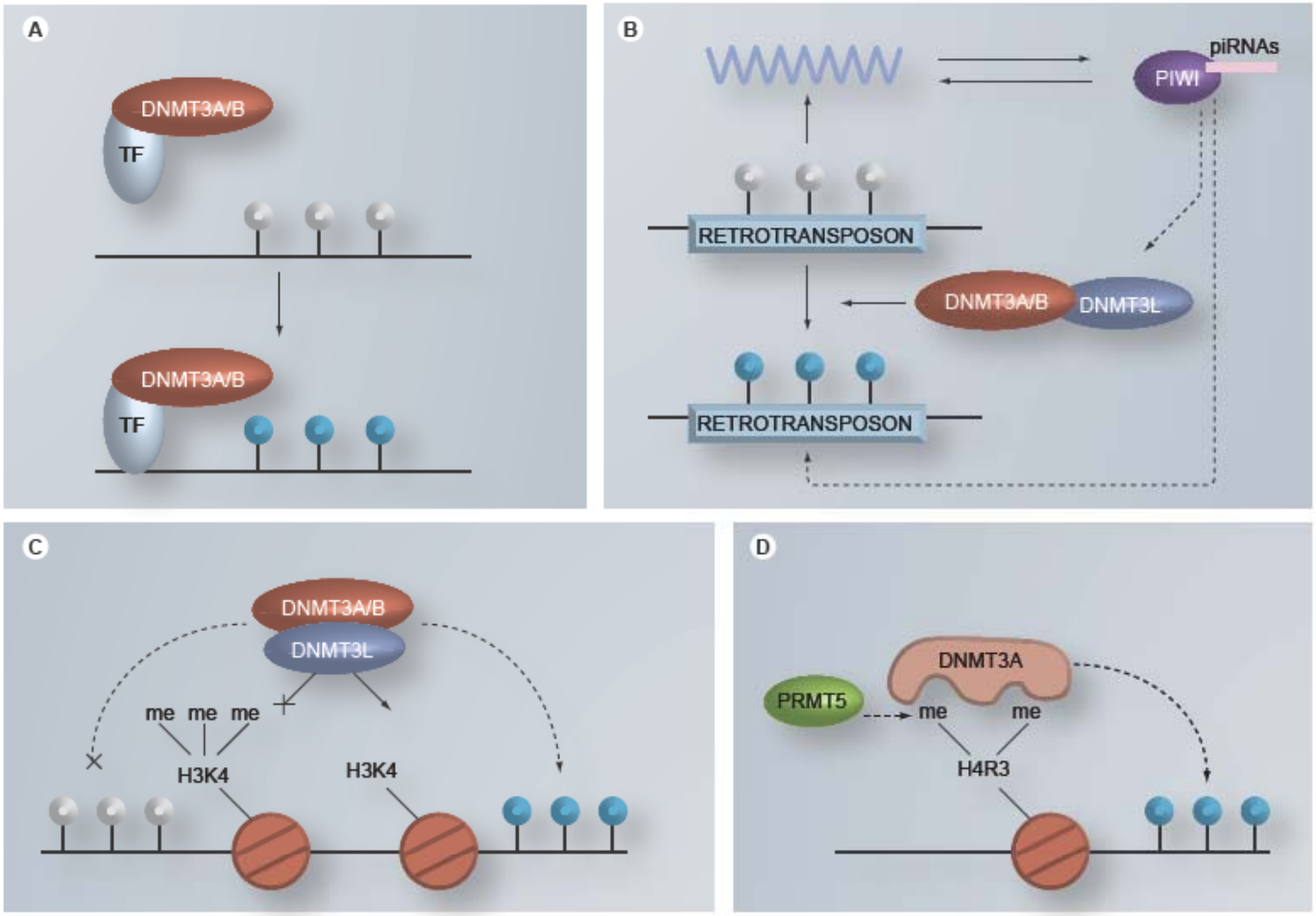

Figure 5. Models for targeting *de novo* DNA methylation in mammalian cells. (A) Some TFs have been shown to interact with DNA methyltransferases DNMT3A/B and might recruit them to their target sites. **(B)** In developing spermatogonia, DNMT3L–DNMT3A/B, complexes methylate repetitive elements. Members of the PIWI subfamily of Argonaute proteins (MILI and MIWI2) bind a class of small RNAs derived from retrotransposons called piRNAs and have been shown to be required for proper DNA methylation of retrotransposons in the germline. Whether PIWI/piRNA complexes regulate DNA methylation by targeting DNMT3 complexes or by other means remains to be determined. Adapted from [75]. **(C)** DNMT3L has been shown to interact with DNMT3A/B, as well as with the amino terminus of histone H3 only when the lysine 4 (H3K4) is unmethylated [29]. Methylation (me) of H3K4 strongly inhibits the interaction of DNMT3L with chromatin *in vitro*, suggesting that the absence of H3K4 methylation could act as a signal to attract DNA methylation in germ cells. **(D)** Symmetric methylation of histone H4 arginine 3 (H4R3) by the arginine methyltransferase PRMT5 has been shown to create a binding site for DNMT3A in human cells. Therefore, H4R3me2s could act as a signal to direct DNA methylation of adjacent CpGs. Adapted from [97]. Lollipops depict CpG dinucleotides either methylated (blue) or unmethylated (white).
DNMT: DNA methyltransferase; me: Methylation; piRNA: PIWI-interacting RNA; TF: Transcription factor.

Finally, a recent study revealed that transcription is an important factor to establish DNA methylation at imprinted genes in the germline. Using the *Gnas* locus as a model, the authors elegantly showed that the truncation of transcripts that overlap the maternally methylated ICR disrupts the acquisition of DNA methylation in the oocyte [84]. Interestingly, this might not be limited to the *Gnas* locus, as many other maternally methylated ICRs are located within transcribed units [84]. Together with increasing observations that gene transcription correlates with intragenic DNA methylation (see 'Tissue-specific profiles of DNA methylation'), this could indicate that transcription plays a general role in directing DNA methylation by mechanisms that are not yet understood.

Altogether, it appears that *de novo* methylation is regulated at many different levels. Although our current knowledge is rather incomplete, the definite mechanistic model will probably associate a conjunction of CpG periodicity, histone modifications and spatio-temporal expression of protein or RNA factors to explain the correct establishment of DNA methylation during development.

Transgenerational heritability of DNA methylation patterns

DNA methylation is heritable during somatic cell division via the maintenance activity of DNMT1. Whether DNA methylation patterns can also be inherited through generations, which would have major epidemiological implications, remains to be clarified. Transgenerational inheritance of DNA methylation patterns is thought to be prevented by global clearing of DNA methylation occurring after fertilization and during germ cell development. However, it is conceivable that certain target sequences occasionally resist global DNA methylation erasure. Imprinted genes provide one example of sequences retaining DNA methylation during early development, although they do not resist reprogramming in the germ cells.

Possible candidates for mediating epigenetic inheritance are IAP elements, which have been shown to be only partially demethylated in early embryos and germ cells, possibly to protect the genome from a complete reactivation of these mobile elements [76,81,98] (FIGURE 4). Incomplete erasure of DNA methylation at IAP elements, which can modulate neighboring genes, could therefore lead to epigenetic inheritance from one generation to the next. Studies over the past 10 years using genetic mouse models tended to support such a scenario. The best-characterized example is the epigenetic inheritance at the *Agouti viable yellow* (A^{vy}) allele in mice. Expression of this allele is under the control of a cryptic promoter present in an IAP element inserted upstream of the coding exons. Expression of the A^{vy} allele is variable among isogenic individuals and correlates with DNA methylation of the upstream IAP element. The silent A^{vy} allele is hypermethylated and produces an agouti-colored coat (termed pseudoagouti), whereas the active allele is hypomethylated and induces a yellow coat color. It has been shown that the locus displays epigenetic inheritance following maternal transmission – that is, pseudoagouti mothers produce more pseudoagouti offsprings than yellow mothers. This suggested that there is incomplete clearing of DNA methylation when the silent A^{vy} locus is passed through the female germline [99]. However, a recent study showed that, in contrast to the majority of IAPs, the A^{vy} locus is completely demethylated at the blastocyst stage, which suggests that DNA methylation is not the primary inherited epigenetic mark [100].

In humans, it has been discovered that some cases of familial colorectal cancer are associated with soma-wide promoter DNA methylation of tumor-suppressor genes such as *MLH1* or *MSH2*. This methylation is occasionally detected in successive generations, which led to the initial model of a direct transfer of abnormal DNA methylation through generations [101,102]. However, careful subsequent studies showed that there is no promoter DNA methylation at the *MLH1* gene in spermatozoa [103], which indicates that the heritability probably reflects an underlying genetic change that directs the postzygotic re-establishment of an atypical epigenetic state in each generation. Therefore, there are to date no clear examples in mammals of direct transgenerational transmission of DNA methylation states due to lack of epigenetic reprogramming. Future studies using genome-wide approaches should help to further clarify this important issue and unravel the epigenetic mechanisms responsible for cases of unusual patterns of transgenerational inheritance [99].

Future perspective

Although DNA methylation was described decades ago, our understanding of the distribution and function of this epigenetic mark was limited for a long time. One reason for this was the inability to study DNA methylation at the genome level, and the field suffered from generalizations based on studies of low numbers of genes. Novel large-scale approaches confirmed some of the pioneering observations made on single genes, but also revealed that epigenetic regulation by DNA methylation is probably more complex and dynamic than previously anticipated. High-throughput analyses also revealed novel unexpected findings; for example, the potential link between gene expression and intragenic DNA methylation [7].

More studies are needed to better understand the complex connection between DNA methylation and gene regulation during development. One important point to take into consideration from recent datasets is that cultured cells display methylation patterns that are somewhat different from their *in vivo* counterparts [12]. Therefore, future studies should aim at better characterizing dynamic DNA methylation patterns in cellular lineages *in vivo*, which is currently limited by the purification methods to isolate clean cellular populations and the limited number of cells obtained.

The mechanisms by which DNA methylation regulates gene expression largely remain to be elucidated. Further studies are needed to understand the impact of cytosine methylation on chromatin structure and the transcriptional

Executive summary

Novel technologies to map DNA methylation

- Current techniques combine isolation of methylated DNA with microarray hybridization.
- Improvements in high-throughput sequencing will help to create DNA methylation maps at single base pair resolution of mammalian genomes.

Distribution of DNA methylation in mammalian genomes

- Virtually all CpGs in mammalian genomes show a high degree of methylation, except at regions named CpG islands, which often colocalize with gene promoters.
- The distribution of cytosine methylation in mammalian genomes suggests that its major role is to participate in the maintenance of genome integrity rather than gene regulation.

The DNA methylation machinery

- UHRF1 helps to recruit DNMT1 to sites of DNA replication where it copies DNA methylation patterns onto the newly synthesized DNA strands.
- DNMT3L–DNMT3A/B complexes are involved in setting up DNA methylation patterns in developing germ cells and early embryos.

Tissue-specific profiles of DNA methylation

- Large-scale studies revealed that DNA methylation profiles are more dynamic between cell types than previously anticipated.
- Sites of cell type-specific DNA methylation appear more abundant in intergenic regions and gene bodies than in gene promoters.

Methylation at CpG-poor promoters

- Methylation at CpG-poor promoters is very common and can be bypassed by activating signals, indicating that low density of cytosine methylation does not prevent gene activation.

Methylation at CpG island promoters

- Genome-wide mapping confirmed that most CpG island promoters remain constitutively unmethylated, even when the associated gene is not expressed.
- Rare methylated CpG island promoters are associated with developmental genes and moderate CpG richness.

DNA methylation and cellular identity

- Recent work on embryonic stem cells and induced pluripotent stem cells suggests that DNA methylation of key developmental genes is a major barrier to prevent cellular reprogramming.

Two waves of global DNA methylation reprogramming

- DNA methylation is erased after fertilization, possibly to restore an epigenetic state compatible with pluripotency in the early embryo.
- A second wave of DNA methylation reprogramming occurs in primordial germ cells to reset global DNA methylation and imprinting marks for the next generation.

Mechanisms of DNA demethylation

- The mechanisms involved in erasure of DNA methylation are not yet elucidated, but most likely involve the removal of entire bases via the base excision repair pathway.

Targeting of de novo DNA methylation

- Correct targeting of *de novo* DNA methylation during development probably involves a combination of optimal CpG spacing, histone modifications, RNAs and transcription factors.

Heritability of DNA methylation patterns

- Incomplete DNA methylation clearing in early embryos and germ cells could mediate transgenerational heritability of DNA methylation.
- Examples of direct transmission of DNA methylation from one generation to the next in mammals are missing; however, future studies should help to clarify this important issue.

Future perspective

- Genome-wide mapping of DNA methylation confirmed pioneer observations on single genes, but also revealed novel unexpected aspects.
- Future studies should help to better understand the dynamics of DNA methylation in cellular lineages *in vivo*, the mechanisms by which DNA methylation impacts on gene expression, and the factors reprogramming DNA methylation during development.

machinery. Moreover, in addition to the role of DNA methylation at gene promoters, one novel question is how DNA methylation might also modulate gene expression through long-distance regulation. Addressing these functional questions is complex because of the lack of good working models. To date, it is not possible to interfere with DNA methylation at specific loci, but only globally though the inhibition of DNMT enzymes or treatment with 5-azacytidine, which makes the interpretation of results complicated. In addition, absence of DNMTs in mice leads to early embryonic lethality, which makes it difficult to study the role of DNA methylation in organogenesis unless conditional knockout strategies are used. In the future, the engineering of novel technologies to selectively modify DNA methylation of particular sequences in the genome would be extremely valuable.

Finally, particular emphasis should also be placed on studying the mechanisms that regulate and target DNA methylation in the

genome. Several pathways directing DNA methylation have been identified in plants and fungi, but they do not seem to play identical roles in mammals. This is a crucial question that might also lead to a better understanding of the deregulation of DNA methylation observed during tumorigenesis and the identification of potential novel therapeutic targets.

Acknowledgements

We thank Robert Feil and Philippe Arnaud for useful comments on the manuscript.

Financial & competing interests disclosure

Research in our laboratory is supported by funds from Centre National de la Recherche Scientifique (CNRS), Association pour la Recherche contre le Cancer (ARC contract N°4868), Agence Nationale de le Recherche (ANR-07-BLAN-0052–02) and CEFIC Long Research Initiative (LRI-EMSG49-CNRS-08). ***The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. No writing assistance was utilized in the production of this manuscript.***

Bibliography

Papers of special note have been highlighted as:
- ▪ of interest
- ▪▪ of considerable interest

1 Ptashne M: On the use of the word 'epigenetic'. *Curr. Biol.* 17, R233–R236 (2007).

2 Bird A: DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21 (2002).

3 Wang Y, Jorda M, Jones PL *et al.*: Functional CpG methylation system in a social insect. *Science* 314, 645–647 (2006).

4 Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S: Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* 39, 61–69 (2007).

5 Suzuki MM, Kerr AR, De Sousa D, Bird A: CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res.* 17, 625–631 (2007).

6 Tweedie S, Charlton J, Clark V, Bird A: Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol. Cell. Biol.* 17, 1469–1475 (1997).

7 Suzuki MM, Bird A: DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* 9, 465–476 (2008).

8 Esteller M: Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* 8, 286–298 (2007).

9 Weber M, Davies JJ, Wittig D *et al.*: Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.*37, 853–862 (2005).

10 Rauch T, Li H, Wu X, Pfeifer GP: MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Res.* 66, 7939–7947 (2006).

11 Cokus SJ, Feng S, Zhang X *et al.*: Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219 (2008).

12 Meissner A, Mikkelsen TS, Gu H *et al.*: Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766–770 (2008).

▪▪ **Describes the first application of high-throughput sequencing on mammalian bisulfite-treated DNA to generate DNA methylation maps in mouse embryonic stem cells and differentiated cells at single nucleotide resolution.**

13 Walsh CP, Chaillet JR, Bestor TH: Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* 20, 116–117 (1998).

14 Bourc'his D, Bestor TH: Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431, 96–99 (2004).

15 Chen T, Hevi S, Gay F *et al.*: Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nat. Genet.*39, 391–396 (2007).

16 Dodge JE, Okano M, Dick F *et al.*: Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *J. Biol. Chem.* 280, 17986–17991 (2005).

17 Maloisel L, Rossignol JL: Suppression of crossing-over by DNA methylation in Ascobolus. *Genes Dev.* 12, 1381–1389 (1998).

18 Li E, Bestor TH, Jaenisch R: Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915–926 (1992).

19 Lei H, Oh SP, Okano M *et al.*: *De novo* DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 122, 3195–3205 (1996).

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

21 Sharif J, Muto M, Takebayashi S *et al.*: The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450, 908–912 (2007).

22 Okano M, Bell DW, Haber DA, Li E: DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* 99, 247–257 (1999).

23 Hirasawa R, Chiba H, Kaneda M *et al.*: Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev.* 22, 1607–1616 (2008).

24 Watanabe D, Suetake I, Tada T, Tajima S: Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mech. Dev.* 118, 187–190 (2002).

25 Kato Y, Kaneda M, Hata K *et al.*: Role of the Dnmt3 family in *de novo* methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.* 16, 2272–2280 (2007).

▪ **The authors performed detailed methylation analysis in spermatogonia to show that DNMT3L has a broad function in the methylation of all tested imprinted loci and repetitive sequences, whereas DNMT3A and DNMT3B have different target specificities.**

26 Kaneda M, Okano M, Hata K *et al.*: Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 429, 900–903 (2004).

27 Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH: Dnmt3L and the establishment of maternal genomic imprints. *Science* 294, 2536–2539 (2001).

28 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* 129, 1983–1993 (2002).

29 Ooi SK, Qiu C, Bernstein E *et al.*: DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* 448, 714–717 (2007).

== **Demonstrates that DNMT3L interacts with DNMT3A/B as well as histone H3, and that interaction with histone H3 is prevented *in vitro* by methylation at lysine 4, suggesting a model whereby absence of H3K4me acts as a signal to recruit DNMT3L and induce DNA methylation by DNMT3A/B.**

30 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* 449, 248–251 (2007).

= **The crystallographic structure of human DNMT3L bound to DNMT3A shows that the interaction stimulates the DNMT3A enzymatic activity, and that the complex methylates CpGs with an optimal spacing of 8–10 base pairs.**

31 Klose RJ, Bird AP: Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* 31, 89–97 (2006).

32 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.* 26, 169–181 (2006).

33 Fournier C, Goto Y, Ballestar E *et al.*: Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. *EMBO J* 21, 6560–6570 (2002).

34 Klose RJ, Sarraf SA, Schmiedeberg L, McDermott SM, Stancheva I, Bird AP: DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell* 19, 667–678 (2005).

35 Yasui DH, Peddada S, Bieda MC *et al.*: Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. *Proc. Natl Acad. Sci. USA* 104, 19416–19421 (2007).

36 Hutchins AS, Mullen AC, Lee HW *et al.*: Gene silencing quantitatively controls the function of a developmental trans-activator. *Mol. Cell* 10, 81–91 (2002).

37 Zhao X, Ueba T, Christie BR *et al.*: Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proc. Natl Acad. Sci. USA* 100, 6777–6782 (2003).

38 Prokhortchouk A, Sansom O, Selfridge J *et al.*: Kaiso-deficient mice show resistance to intestinal cancer. *Mol. Cell. Biol.* 26, 199–208 (2006).

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

40 Martin Caballero I, Hansen J, Leaford D, Pollard S, Hendrich BD: The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation. *PLoS ONE* 4, e4315 (2009).

41 Walsh CP, Bestor TH: Cytosine methylation and mammalian development. *Genes Dev.* 13, 26–34 (1999).

42 Jackson-Grusby L, Beard C, Possemato R *et al.*: Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.*27, 31–39 (2001).

43 Fouse SD, Shen Y, Pellegrini M *et al.*: Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* 2, 160–169 (2008).

44 Eckhardt F, Lewin J, Cortese R *et al.*: DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* 38, 1378–1385 (2006).

45 Farthing CR, Ficz G, Ng RK *et al.*: Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* 4, e1000116 (2008).

46 Illingworth R, Kerr A, Desousa D *et al.*: A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol* 6, e22 (2008).

= **One of the papers (along with [49] and [50]) to identify for the first time at the genome level a large number of CpG island promoters that are hypermethylated in normal human somatic cells, including several germline-specific and developmental genes.**

47 Rakyan VK, Down TA, Thorne NP *et al.*: An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res.* 18, 1518–1529 (2008).

48 Rauch TA, Wu X, Zhong X, Riggs AD, Pfeifer GP: A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci U S A* 106, 671–678 (2009).

49 Shen L, Kondo Y, Guo Y *et al.*: Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet.* 3, 2023–2036 (2007).

= **One of the papers (along with [46] and [50]) to identify for the first time at the genome level a large number of CpG island promoters that are hypermethylated in normal human somatic cells, including several germline-specific and developmental genes.**

50 Weber M, Hellmann I, Stadler MB *et al.*: Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.*39, 457–466 (2007).

= **One of the papers (along with [46] and [49]) to identify for the first time at the genome level a large number of CpG island promoters that are hypermethylated in normal human somatic cells, including several germline-specific and developmental genes.**

51 Yagi S, Hirabayashi K, Sato S *et al.*: DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. *Genome Res.* 18, 1969–1978 (2008).

52 Mohn F, Weber M, Rebhan M *et al.*: Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* 30, 755–766 (2008).

53 Zhang Y, Rohde C, Tierling S *et al.*: DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. *PLoS Genet.* 5, e1000438 (2009).

54 Straussman R, Nejman D, Roberts D *et al.*: Developmental programming of CpG island methylation profiles in the human genome. *Nat. Struct. Mol. Biol.* 16, 564–571 (2009).

55 Ball MP, Li JB, Gao Y *et al.*: Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* 27, 361–368 (2009).

= **The authors introduce two novel approaches for DNA methylation profiling using high-throughput sequencing and show that, unexpectedly, gene-body methylation is higher in highly expressed genes.**

56 Schmidl C, Klug M, Boeld TJ *et al.*: Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res.* 19(7), 1165–1174 (2009).

57 Wan LB, Bartolomei MS: Regulation of imprinting in clusters: noncoding RNAs versus insulators. *Adv. Genet.* 61, 207–223 (2008).

58 Hellman A, Chess A: Gene body-specific methylation on the active X chromosome. *Science* 315, 1141–1143 (2007).

59 Waterland RA, Kellermayer R, Rached MT *et al.*: Epigenomic profiling indicates a role for DNA methylation in early postnatal liver development. *Hum. Mol. Genet.*, 18(16), 3026–3038 (2009).

60 Brown SE, Szyf M: Epigenetic programming of the rRNA promoter by MBD3. *Mol. Cell. Biol.* 27, 4938–4952 (2007).

61 Kangaspeska S, Stride B, Metivier R *et al.*: Transient cyclical methylation of promoter DNA. *Nature* 452, 112–115 (2008).

62 Metivier R, Gallais R, Tiffoche C *et al.*: Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 452, 45–50 (2008).

63 Osborne CS, Chakalova L, Brown KE *et al.*: Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.*36, 1065–1071 (2004).

64 Oda M, Yamagiwa A, Yamamoto S *et al.*: DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner. *Genes Dev.* 20, 3382–3394 (2006).

65 Feldman N, Gerson A, Fang J *et al.*: G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat. Cell Biol.* 8, 188–194 (2006).

66 Reik W: Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447, 425–432 (2007).

67 Ng RK, Dean W, Dawson C *et al.*: Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat. Cell Biol.* 10, 1280–1290 (2008).

•• **Mouse embryonic stem cells can generally not be differentiated into extraembryonic lineages. The authors show that *Dnmt1*$^{-/-}$ mouse ES cells can differentiate into trophoblast cells, and identify DNA methylation in the promoter of the *Elf5* gene as a candidate for mediating this epigenetic lineage restriction.**

68 Epsztejn-Litman S, Feldman N, Abu-Remaileh M *et al.*: *De novo* DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nat. Struct. Mol. Biol.* 15, 1176–1183 (2008).

69 Mikkelsen TS, Hanna J, Zhang X *et al.*: Dissecting direct reprogramming through integrative genomic analysis. *Nature* 454, 49–55 (2008).

• **Demonstrates that the reprogramming of differentiated cells into induced pluripotent stem cells involves major chromatin changes and DNA demethylation of several pluripotency-related genes, and presents evidence that DNA methylation is one barrier explaining the low efficiency of the process.**

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

71 Oswald J, Engemann S, Lane N *et al.*: Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* 10, 475–478 (2000).

72 Nakamura T, Arai Y, Umehara H *et al.*: PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* 9, 64–71 (2007).

73 Li X, Ito M, Zhou F *et al.*: A maternal-zygotic effect gene, *Zfp57*, maintains both maternal and paternal imprints. *Dev. Cell* 15, 547–557 (2008).

74 Lees-Murdock DJ, Walsh CP: DNA methylation reprogramming in the germ line. *Epigenetics* 3, 5–13 (2008).

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

76 Hajkova P, Erhardt S, Lane N *et al.*: Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* 117, 15–23 (2002).

77 Li JY, Lees-Murdock DJ, Xu GL, Walsh CP: Timing of establishment of paternal methylation imprints in the mouse. *Genomics* 84, 952–960 (2004).

78 Seki Y, Hayashi K, Itoh K, Mizugaki M, Saitou M, Matsui Y: Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* 278, 440–458 (2005).

79 Hajkova P, Ancelin K, Waldmann T *et al.*: Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* 452, 877–881 (2008).

80 Seki Y, Yamaji M, Yabuta Y *et al.*: Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* 134, 2627–2638 (2007).

81 Lees-Murdock DJ, De Felici M, Walsh CP: Methylation dynamics of repetitive DNA elements in the mouse germ cell lineage. *Genomics* 82, 230–237 (2003).

82 Maatouk DM, Kellam LD, Mann MR *et al.*: DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* 133, 3411–3418 (2006).

83 Lucifero D, Mann MR, Bartolomei MS, Trasler JM: Gene-specific timing and epigenetic memory in oocyte imprinting. *Hum. Mol. Genet.* 13, 839–849 (2004).

84 Chotalia M, Smallwood SA, Ruf N *et al.*: Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev.* 23, 105–117 (2009).

• **Demonstrates that the truncation of transcripts overlapping the imprinted *Gnas* gene disrupts the acquisition of DNA methylation at the *Gnas* gene in the oocyte, which indicates that transcription is one factor involved in the recruitment of DNA methylation at this locus and possibly also other places in the genome.**

85 Gehring M, Reik W, Henikoff S: DNA demethylation by DNA repair. *Trends Genet.* 25(2), 82–90 (2009).

86 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.* 279, 52353–52360 (2004).

87 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* 135, 1201–1212 (2008).

88 Barreto G, Schafer A, Marhold J *et al.*: Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* 445, 671–675 (2007).

89 Schmitz KM, Schmitt N, Hoffmann-Rohrer U, Schafer A, Grummt I, Mayer C: TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. *Mol. Cell* 33, 344–353 (2009).

90 Engel N, Tront JS, Erinle T *et al.*: Conserved DNA methylation in Gadd45a(-/-) mice. *Epigenetics* 4, 98–99 (2009).

91 Suzuki M, Yamada T, Kihara-Negishi F *et al.*: Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. *Oncogene* 25, 2477–2488 (2006).

92 Wiznerowicz M, Jakobsson J, Szulc J *et al.*: The Kruppel-associated box repressor domain can trigger *de novo* promoter methylation during mouse early embryogenesis. *J. Biol. Chem.* 282, 34535–34541 (2007).

93 Kuramochi-Miyagawa S, Watanabe T, Gotoh K *et al.*: DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* 22, 908–917 (2008).

94 Aravin AA, Sachidanandam R, Bourc'his D *et al.*: A piRNA pathway primed by individual transposons is linked to *de novo* DNA methylation in mice. *Mol. Cell* 31, 785–799 (2008).

95 Freitag M, Selker EU: Controlling DNA methylation: many roads to one modification. *Curr. Opin. Genet. Dev.* 15, 191–199 (2005).

96 Delaval K, Govin J, Cerqueira F, Rousseaux S, Khochbin S, Feil R: Differential histone modifications mark mouse imprinting control regions during spermatogenesis. *EMBO J.* 26, 720–729 (2007).

97 Zhao Q, Rank G, Tan YT *et al.*: PRMT5-mediated methylation of histone H4R3 recruits DNMT3A, coupling histone and DNA methylation in gene silencing. *Nat. Struct. Mol. Biol.* 16(3), 304–311 (2009).

▪ **Demonstrates that human DNMT3A binds to histone H4 methylated on arginine 3, thus providing a novel functional link between chromatin and DNA methylation.**

98 Lane N, Dean W, Erhardt S *et al.*: Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* 35, 88–93 (2003).

99 Chong S, Whitelaw E: Epigenetic germline inheritance. *Curr. Opin. Genet. Dev.* 14, 692–696 (2004).

100 Blewitt ME, Vickaryous NK, Paldi A, Koseki H, Whitelaw E: Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. *PLoS Genet.* 2, e49 (2006).

101 Chan TL, Yuen ST, Kong CK *et al.*: Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. *Nat. Genet.* 38, 1178–1183 (2006).

102 Suter CM, Martin DI, Ward RL: Germline epimutation of MLH1 in individuals with multiple cancers. *Nat. Genet.* 36, 497–501 (2004).

103 Hitchins MP, Ward RL: Erasure of MLH1 methylation in spermatozoa – implications for epigenetic inheritance. *Nat. Genet.* 39, 1289 (2007).