Biochem. J. (2013) **451**, 13–23 (Printed in Great Britain) doi:10.1042/BJ20121585



REVIEW ARTICLE

# Non-canonical functions of the DNA methylome in gene regulation

James P. REDDINGTON*, Sari PENNINGS† and Richard R. MEEHAN*[1]

*MRC Human Genetics Unit, MRC IGMM (Institute of Genetics and Molecular Medicine), University of Edinburgh, Western General Hospital, Crewe Rd, Edinburgh EH4 1SX, U.K., and †Centre for Cardiovascular Science, Queen's Medical Research Institute, University of Edinburgh, 47 Little France Crescent, Edinburgh EH16 4TJ, U.K.

Methylation of the cytosine base in DNA, DNA methylation, is an essential epigenetic mark in mammals that contributes to the regulation of transcription. Several advances have been made in this area in recent years, leading to a leap forward in our understanding of how this pathway contributes to gene regulation during embryonic development, and the functional consequences of its perturbation in human disease. Critical to these advances is a comprehension of the genomic distribution of modified cytosine bases in unprecedented detail, drawing attention to genomic regions beyond gene promoters. In addition, we have a more complete understanding of the multifactorial manner by which DNA methylation influences gene regulation at the molecular level, and which genes rely directly on the DNA methylome for their normal transcriptional regulation. It is becoming apparent that a major role of DNA modification is to act as a relatively stable, and mitotically heritable, template that contributes to the establishment and maintenance of chromatin states. In this regard, interplay is emerging between DNA methylation and the PcG (Polycomb group) proteins, which act as evolutionarily conserved mediators of cell identity. In the present paper we review these aspects of DNA methylation, and discuss how a multifunctional view of DNA modification as an integral part of chromatin organization is influencing our understanding of this epigenetic mark's contribution to transcriptional regulation.

Key words: chromatin, DNA methylation, H3K27me3, histone modification, Polycomb, Polycomb repressor complex 2 (PRC2), regulation of transcription.

## INTRODUCTION

Transcriptional regulation is achieved through the concerted action of regulatory networks of *trans*-acting factors and dynamic chromatin organization [1]. A major challenge of molecular biology is to decipher this interplay, and to establish the role that chromatin structure plays in the co-ordination of gene expression. The structure of chromatin is influenced by several layers of information encoded at the genetic and epigenetic levels [2]. One such layer of information is encoded by chemical modifications of DNA, referred to as DNA methylation, and is thought to play integral roles in many biological processes, including embryonic development and disease pathology [3]. The three most abundant forms of cytosine in mammalian genomes are 5mC (5-methylcytosine or 'DNA methylation'), unmodified cytosine and 5hmC (5-hydroxymethylcytosine) [4]. Owing to the position of the modification site within the cytosine base, these three isoforms do not show differential base pairing properties and therefore do not alter the genetic sequence. Instead, the modification state of cytosine has long been linked to transcriptional regulation, although its precise function in this process remains unclear. In the present paper we review developments in this field, focusing on those that point to previously unappreciated functions of DNA methylation in genome regulation. We first discuss insights that have arisen through studies detailing the distribution of DNA modifications throughout the genome, and how this distribution varies between tissues and in disease states. Next, we review data suggesting that DNA modifications are important modifiers of chromatin structure, and focus on the emerging interplay between DNA methylation and the Polycomb system in this regard. We discuss how this interplay has the potential to enhance our understanding of the enigmatic targeting mechanism of the Polycomb system, and also to expand upon the way that we view the role of DNA methylation during development and in human disease.

## NOVEL ASPECTS OF DNA METHYLATION: CLUES FROM THE DNA METHYLOME

A detailed knowledge of the global distribution of cytosine modifications, the DNA methylome, and how the methylome differs between cell types and individuals is a critical step towards understanding the role of DNA methylation in genome regulation [5]. Various technical advances have facilitated the mapping of modified cytosine bases at unprecedented resolution and coverage, and consequently our understanding of the DNA methylome has improved greatly in recent years. In mammals, the 5mC form of cytosine is primarily found where a cytosine is followed by a guanosine in the 5′→3′ direction (a 'CpG dinucleotide'), and it is in this context that 5mC is associated with transcriptional regulation [6–10]. Mammals, as do all vertebrates studied, have a 'global' methylome with the bulk of CpGs in the genome being found in the methylated state (5mC) [6,11]. This high level of genomic methylation is punctuated by short stretches of unmethylated DNA, many of which correspond to regions of high CpG content known as CGIs (CpG islands) [4,12]. Although the majority of CGIs are found in the unmethylated state in all tissues studied, a subset are found to be methylated in a tissue-specific manner [6,11,13]. DNA methylation patterns in general are variable between different tissues and are dynamic during cell differentiation [6,7,14]. It is this property, together with the

Abbreviations used: CGI, CpG island; CTCF, CCCTC-binding factor; ES, embryonic stem; EZH2, enhancer of zeste homologue 2; 5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine; MeCP, methyl-CpG-binding protein; PcG, Polycomb group; PMD, partially methylated domain; PRC2, Polycomb repressor complex 2; PRE, Polycomb-response element; TET, ten-eleven translocation.

[1] To whom correspondence should be addressed (email richard.meehan@igmm.ed.ac.uk).

mitotic heritability of CpG methylation patterns, and the ability of CpG methylation to modify transcription, that implicate DNA methylation as a process influencing maintenance of cell identity [15–19].

### Canonical repression by DNA methylation: gene promoter regions and repetitive elements

The most heavily studied regions of the genome with respect to DNA methylation-mediated gene regulation are those directly proximal to transcription start site-promoter elements. It is well established that if a promoter element contains a sufficient density of CpG dinucleotides, the presence of high levels of DNA methylation (5mC) is strongly associated with gene inactivity [20]. The function of DNA methylation at gene promoters has frequently been attributed to two mechanisms: (i) its negative effect on certain transcription factor–DNA interactions [21–23], and (ii) its attraction of methyl-CpG-binding proteins [4,24–26] (Figure 1). This canonical form of DNA methylation-mediated repression is essential for the parent-of-origin-specific expression of imprinted genes [27–29], and the process of X-chromosome inactivation, where many genes on one of the two X-chromosomes in female somatic cells become repressed [30,31]. Functional DNA methylation at other gene promoter regions is relatively infrequent, and is related to CpG density: CpG-poor promoters are often highly methylated irrespective of their activity status, whereas promoters with high CpG-density are commonly unmethylated, such as CGIs [20]. At high-CpG-density promoters the methylated state is tightly associated with transcriptional inactivity, whereas the unmethylated state can be associated with either actively transcribing genes, or inactive genes subject to other forms of repression [11]. Despite canonical repression now being recognized as relatively uncommon, promoter DNA methylation is an important mechanism in the control of tissue-specific gene regulation [28,32,33]. For example, a cohort of genes expressed in the mouse germline relies on promoter DNA methylation for repression in somatic cell lineages [20,34–38]. For these genes DNA methylation appears to be a primary factor regulating their expression patterns. However, in most cases methylation of promoter regions appears to represent a secondary mechanism contributing to maintenance of the repressed state [39–41]. This type of repression by DNA methylation is exemplified by its role in X-chromosome inactivation and appears to contribute mainly to long-term transcriptional repression, which in some cases needs to be maintained in a given tissue for the life of the organism. In addition to single-copy gene promoters, it is well established that DNA methylation contributes to the repression of certain repetitive elements throughout the genome, particularly IAP (intracisternal A-particle) elements [28].

### Beyond gene promoters: non-canonical roles for DNA methylation?

New methods of DNA methylation mapping with improved resolution and coverage have revealed novel aspects of the methylome, while drawing attention to genomic regions outside of gene promoter elements. These observations have highlighted the possibility that DNA methylation plays important, and as yet unclear, roles in genome regulation in addition to its canonical function at promoter elements.

#### Gene body methylation

Vertebrate methylomes are classified as ‘globally methylated’, so it comes as no surprise that genes are also methylated internally.

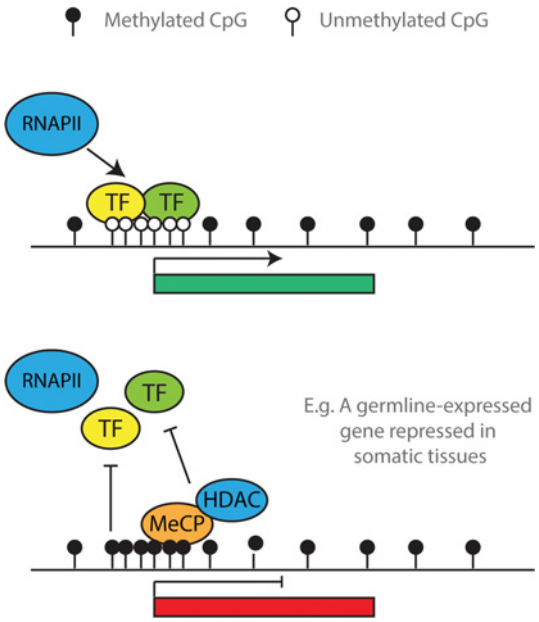


**Figure 1 Canonical functions of DNA methylation in gene regulation**

A gene that exhibits tissue-specific promoter DNA methylation is shown. When expressed, its promoter is associated with unmethylated CpGs (open lollipops), whereas repression is accompanied by promoter DNA methylation (closed lollipops). Methylation of promoter CpGs is thought to contribute to repression through two mechanisms: (i) direct inhibition of transcription factor binding which is necessary for recruitment of the transcription machinery (represented by RNA polymerase II), and (ii) attraction of MeCPs which associate with co-repressors such as histone deacetylases. HDAC, histone deacetylase; RNAPII, RNA polymerase II; TF, transcription factor.

Provocative observations suggest that gene body methylation may represent a crucial function in gene regulation. Intragenic DNA methylation appears to be evolutionarily ancient and many organisms with so-called ‘mosaic’ methylomes also exhibit abundant gene body methylation. For example, in the model flowering plant *Arabidopsis thaliana* the vast majority of DNA methylation occurs within gene bodies and transposon sequences [42]. Genes that are moderately transcribed in *Arabidopsis* contain the most 5mC in their gene body regions, whereas genes at either extreme of the expression range contain lower methylation levels [42]. Experimental reduction of DNA methylation leads to up-regulation of genes with gene body methylation, suggesting a repressive role [42]. DNA methylation has also been suggested to inhibit transcriptional elongation in the filamentous fungi *Neurospora crassa* [43]. However, in contrast with *Arabidopsis* and *Neurospora*, a positive correlation between the abundance of gene body methylation and mRNA levels has been observed in mammals [7,44–47]. In support of this, the co-transcriptionally deposited H3K36me3 (indicating the histone number modified, and the level of methylation on the lysine residue indicated, e.g. this modification is histone H3 trimethylated on Lys[36]) histone modification has been proposed to play a direct role in the targeting of DNA methylation [48,49]. In addition, many tissue-specific differentially methylated regions have been found in gene bodies in mammals, including CGIs that are found in intragenic regions [50–52]. However, the relationship between gene body methylation and transcript abundance is not universal and may be cell-type-specific [6,46,53]. It should also be noted that studies of gene body methylation are complicated by the additional presence of the 5hmC base within these regions [54], and by the inability of many of the bisulfite-based 5mC mapping methods to distinguish between the 5mC and 5hmC bases [55].

The role of gene body methylation in mammals is currently unclear, but two potential functions have been proposed (Figure 2B): (i) the regulation of co-transcriptional RNA processing such as alternative splicing, and (ii) the prevention of spurious transcription initiation within gene bodies. It is now becoming apparent that chromatin structure may play important roles in the regulation of alternative splicing and mRNA processing [56]. Associations between DNA methylation levels and alternative splicing have been noted in the honeybee, *Apis mellifera*, and also in humans, with increased DNA methylation of an exon being associated with its preferential inclusion in a mature transcript [57–59]. A direct link has been made through the multi-functional DNA-binding protein CTCF (CCCTC-binding factor) which is sensitive to CpG methylation in certain binding contexts [59]. During alternative splicing, lower affinity sites becoming more favourable if transcription elongation occurs more slowly [56]. Binding of CTCF protein downstream of a splice site was proposed to cause RNA polymerase II pausing, reduction in elongation speed and therefore an alteration in the choice of splice site in a methylation-modulated manner [59] (Figure 2Bi). This raises the possibility that tissue-specific differences in DNA methylation within gene bodies could contribute directly to differential RNA processing. A second potential function that has been speculated is the prevention of transcription initiation events within gene bodies. As the act of transcription may destabilize nucleosomes, leading to a more open chromatin conformation, this could promote new transcription initiation events producing cryptic transcripts (Figure 2Bii). Some evidence does exist that gene body methylation prevents transcriptional initiation. A tissue-specific methylated CGI was shown to act as an alternative promoter of the *Shank3* gene, in both mouse and humans; its activity was negatively regulated by DNA methylation [52]. However, a general role for gene body methylation in prevention of spurious transcription initiation within transcription units currently remains speculative.

### Orphan CpG islands

Approximately half of CGIs are not associated with the 5′ end of annotated genes, and are instead found in intragenic or intergenic locations [13,51]. Interestingly, these ‘orphan’ CGIs are more frequently found in the methylated state than their promoter counterparts and they also show a greater incidence of tissue-specific methylation [13,51]. The function of orphan CGIs and their DNA methylation state is currently unclear [12]. Unmethylated CGIs in general show many properties suited for function as regulatory elements, including an increased frequency of nucleosome-depleted regions, depleted linker histone occupancy, and histone marks that are associated with a more open chromatin configuration [4,60,61]. Therefore it is likely that many orphan CGIs represent promoters for unannotated transcripts (Figure 2Ci). In support of this idea, many unmethylated orphan CGIs show evidence of transcription initiation that varies in a tissue-specific manner [51]. Orphan CGIs may act as developmentally regulated alternative promoters for known transcripts [52]. Alternatively, they may represent promoters of functional non-coding RNAs, such as long non-coding RNAs, which may play important roles themselves in gene regulation in *cis* or *trans*. Methylation at orphan CGIs could conceivably play a role in the regulation of local chromatin configuration by modulating the binding of effector proteins that directly or indirectly modify chromatin structure [61,62] (Figure 2Cii). Intriguingly, genes with key roles in embryonic development, particularly transcription factors such as *Hox* genes, are often found in the proximity of CGI clusters and orphan CGIs that are methylated in a tissue-specific manner [13,63]. Future studies will delineate the function of orphan CGIs and the importance of their methylation status in transcriptional regulation.

### CpG ‘shores’

A set of studies noted that tissue-specific methylation occurs less frequently within CGIs and gene promoters when compared with flanking regions [64–66]. The greatest variation in DNA methylation between cell types was observed at regions immediately adjacent to CGIs, termed ‘CpG shores’. These regions also showed the most variation between human induced pluripotent stem cells and their parental fibroblasts, and between normal colon and colon cancer [64,65]. CpG shores were defined as regions immediately flanking a CGI, within 2 kb on either side, and therefore have a lower CpG content relative to their neighbouring CGI [64]. The functional consequence of CpG shore methylation is unclear, but it correlates with reduced transcription of the nearby gene [64,66]. A study of candidate genes showed that differential methylation at CpG shores is associated with the use of alternative promoters in some cases, suggesting that this process may influence the regulation of transcription and promoter choice [64]. In this regard, it has been suggested that promoter-flanking methylation can act as a docking site for MeCPs (methyl-CpG-binding proteins) such as MeCP2, and can therefore modulate the transcriptional output of a non-methylated promoter [67].

### Distal regulatory elements

DNA methylation mapping studies revealed that regulatory elements located distal to gene promoters are often differentially methylated between tissues [6,14,53,68–71]. Examples are enhancer elements, short stretches of DNA that are bound by *trans*-acting factors and act to stimulate the expression of distal target genes. DNA methylation at enhancers is often inversely correlated with their activity [6,53,68,69,71]. The function of DNA methylation at enhancers has not been extensively studied, and the cause and effect relationships are unclear (Figure 2D). Recent data from the ENCODE project shows that DNA methylation at the binding sites for a given transcription factor is negatively correlated with tissue-specific expression of that transcription factor in a panel of human cell lines, suggesting that the presence of a DNA-binding factor is sufficient to predict hypomethylation at its binding sites [69]. This idea has been directly tested by inserting a fragment of DNA containing a transcription factor motif into the genome of mammalian cells. This fragment was protected from methylation only when the DNA-binding motif was intact, suggesting that binding precludes DNA methylation, and hence leads to the observed hypomethylation at enhancer regions [68,72] (Figure 2Di). In addition, insertion of the same fragment after completely methylating it *in vitro* did not preclude binding of this particular transcription factor, suggesting that in this context CpG methylation is compatible with transcription factor occupancy [68]. However, intriguing evidence exists for the alternative hypothesis that DNA methylation can modulate enhancer activity (Figure 2Dii). The *in vitro* CpG methylation of certain tissue-specific enhancers can strongly attenuate their activity in reporter assays [71]. In addition, a study of the glucocorticoid receptor showed that DNA methylation prevents binding of this factor to an enhancer element *in vitro*, and the formation of a more open chromatin structure *in vivo* [23]. It is of course possible, and indeed likely, that any effect of DNA methylation on enhancer

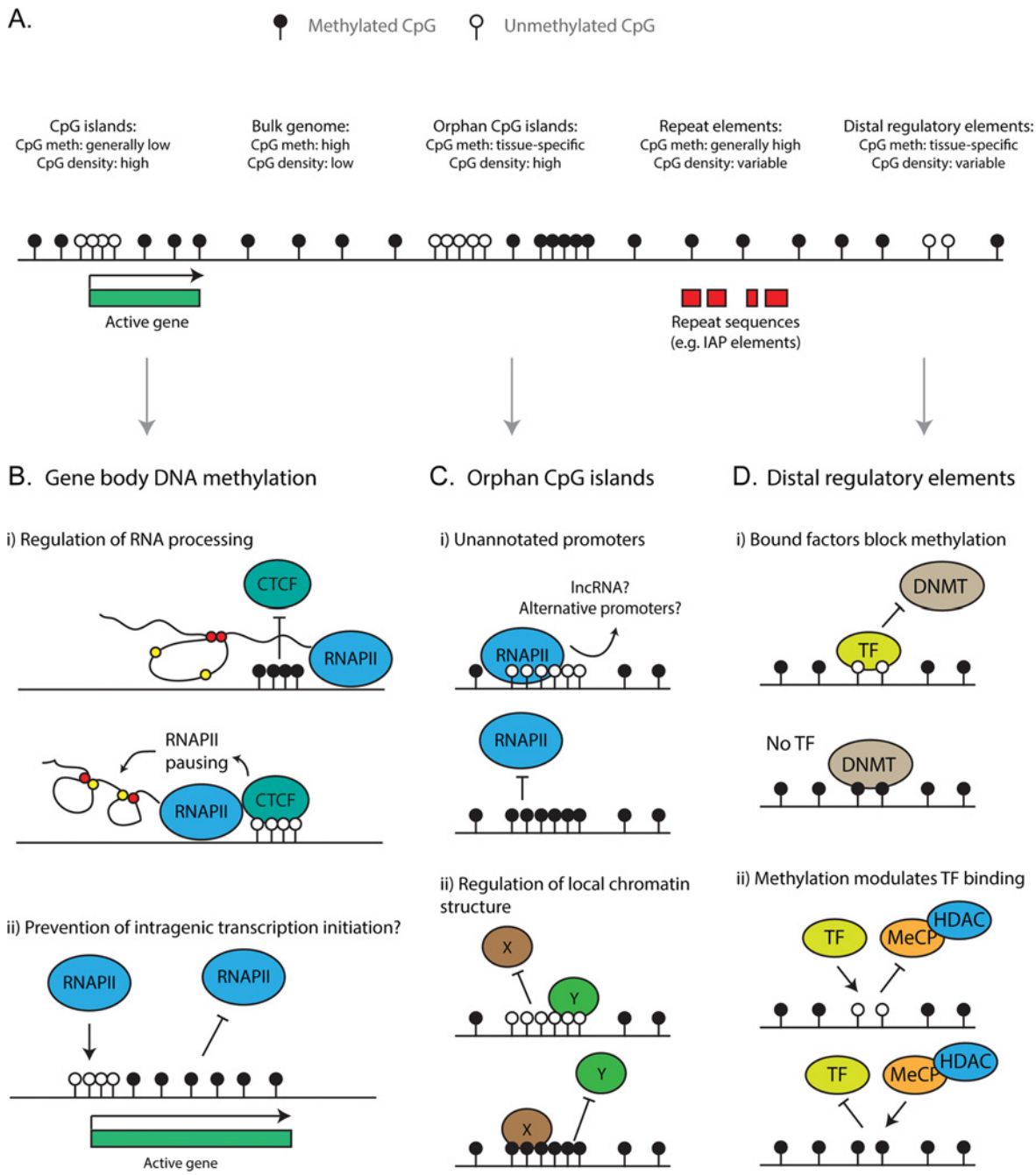


**Figure 2 Emerging non-canonical functions of the DNA methylome in gene regulation**

(**A**) A snapshot of the mammalian DNA methylome. Closed lollipops represent methylated CpGs and open lollipops represent unmethylated CpGs. For simplicity, the 5hmC base is not shown. (**B**) (i) DNA methylation within gene bodies has been linked to the regulation of RNA processing through the direct modulation of CTCF–chromatin interactions. Without CTCF binding, RNA polymerase II transits quickly, favouring assembly of splicing machinery at sites with relatively high affinity (red circles). Bound CTCF causes pausing of RNA poymerase II during transcriptional elongation, favouring protein assembly at weaker sites (yellow circles) and inclusion of an alternative exon [59]. (ii) Intragenic DNA methylation may function to prevent transcription initiation events from occurring within gene bodies. (**C**) Suggested functions for DNA methylation at orphan CGIs. (i) A proportion of orphans may be unannotated promoters for known transcripts or for non-coding transcripts such as long non-coding RNA. Here, methylation could serve to inhibit their promoter activity thereby influencing gene regulation in a tissue-specific manner. (ii) Unmethylated and methylated CpG islands are thought to recruit a different set of proteins (depicted as 'X' and 'Y') that are involved in chromatin modification [24,61,62]. Orphans may function in this way to regulate local gene expression in a methylation-modified manner. (**D**) DNA methylation at distal regulatory elements. Reduced DNA methylation has been noted at several transcription factor-binding sites. Two hypotheses exist for this observation. (i) DNA methylation loss at these elements occurs passively as access of DNA methyltransferases is inhibited by bound transcription factors. In this model, DNA methylation differences at distal regulatory elements merely reflect transcription factor occupancy and do not function to regulate element activity. (ii) DNA methylation is capable of modulating transcription factor–DNA interactions either directly or by contributing to a chromatin state that is not compatible with binding (depicted here by the recruitment of MeCPs and histone deacetylases). In this model, tissue-specific DNA methylation is a modifier of regulatory element activity. DNMT, DNA methyltransferase; HDAC, histone deacetylase; IAP, intracisternal A-particle; lncRNA, long non-coding RNA; meth, methylation; RNAPII, RNA poymerase II; TF, transcription factor.

binding is context-dependent, varying for different genomic locations and for different DNA-binding factors.

### Partially methylated domains

Global single-nucleotide resolution maps of DNA methylation have enabled the study of methylation over large genomic domains. In the first base-resolution methylome map of mammals it was noticed that large genomic regions of lower methylation stood out from the background of high DNA methylation [6]. These PMDs (partially methylated domains) had a mean methylation of less than 70 %, spanned hundreds of kilobases and covered a substantial proportion of each chromosome [6,53]. Subsequently, PMDs have been identified in other cell types and have been suggested to become further hypomethylated in cancer cell lines and tumour tissue [73–75]. However, it is currently unclear whether PMDs are a general feature of DNA methylomes or if they represent changes that occur upon adaption to culture and tumorigenesis [74]. The mechanisms that control DNA methylation at this scale, and its function in gene regulation, are currently opaque. Interestingly, PMDs often coincide with other aspects of chromatin organization, such as large domains of histone marks that are generally associated with transcriptional repression, suggesting that they may reflect aspects of higher-order chromatin structure [6,53,74,75]. Gene expression is also associated with DNA methylation level at the domain scale, with genes within PMDs exhibiting a low level of expression [6,53,73]. Whether PMDs are a cause or consequence of higher-order chromatin organization is unknown, but as CpG methylation can modulate certain aspects of chromatin structure (see later section), it is possible that PMD formation contributes to this level of genome organization. These studies, together with gene body methylation, demonstrate that the association of DNA methylation with transcription is context- and location-dependent, and goes beyond promoter elements. Gene regulation is more frequently being linked to higher-order organization of chromatin, so it will be interesting to see what future studies tell us about the functional significance of PMDs.

### 5-Hydroxymethylcytosine

Previously it was discovered that 5mC can be oxidized to 5hmC by the family of TET (ten-eleven translocation) enzymes, and that genomic 5hmC is abundant in certain mammalian tissues [76–78]. These discoveries have prompted intense research into the distribution and putative function of the 5hmC base, reviewed recently [79,80]. Mapping of the 5hmC base showed high tissue specificity, both in its global abundance and locus-specific distribution [54,81–87]. 5hmC is found throughout the genome, primarily in the sequence context of CpG; it is generally less abundant than 5mC, and is reportedly enriched at enhancer elements, promoters and intragenic regions [54,82]. The 5hmC base has been linked to the enigmatic process of DNA demethylation, where the 5mC base is converted into cytosine either through an active or passive mechanism, and this is where much of the more recent research has focused [88–97]. In this regard, 5hmC appears to be an intermediate in the various pathways that have been proposed to lead from 5mC to cytosine. It is unclear whether the 5hmC base performs functions independent of this process, and whether it is an epigenetic mark in its own right. Gene targeting in mice has shown that the TET1 and TET2 proteins are individually dispensable for embryonic development, but mutants show minor birth weight reduction and defects in the haemopoietic system respectively [98–100]. It will be interesting to see the effect of combined TET protein deficiency, where a greater decrease in 5hmC abundance would be expected. At the molecular level, 5hmC and 5mC are differentially bound by most methyl-CpG-binding proteins [96,101,102], and show differential effects on the stability of the DNA duplex [103], suggesting that conversion of 5mC into 5hmC may alleviate some of the activities of the 5mC base. Recently a mammalian protein that specifically recognizes 5hmC has been described, and corresponds to MeCP2 [102]. This observation, together with the other aspects of 5hmC biology, is likely to be an active area of research in the near future.

## DNA METHYLATION: MODULATOR OF CHROMATIN STRUCTURE

In addition to a detailed understanding of the DNA methylome, recent years have led to a more complete understanding of the molecular mechanisms by which this epigenetic mark influences transcriptional regulation. The function of DNA methylation has frequently been attributed to two mechanisms: (i) its negative effect on certain transcription factor–DNA interactions [21–23], and (ii) its attraction of methyl-CpG-binding proteins [4,24–26]. However, current evidence implies that different modification states of cytosine are interpreted by numerous mechanisms, suggesting that DNA methylation influences genome regulation in a multifactorial manner [61,62,104]. The list of proteins that show DNA-methylation-modulated binding to chromatin is growing and includes proteins with diverse functions [104]. Many of these proteins, directly or indirectly, influence different aspects of chromatin structure, suggesting that this may be a major role of DNA methylation in genome regulation (Figure 3A). Indeed, DNA methylation has been suggested to contribute to diverse aspects of chromatin organization, including the genomic distribution of histone modifications such as histone acetylation [4,105], H3K4me3 [61], H3K27me3 [106], H3K36me3 [62] and H3K9me3 [107,108]; histone variants such as H2A.Z [109,110], and the multi-functional CTCF [59] (Figure 3A).

### Polycomb repressive chromatin complements DNA methylation

During X chromosome inactivation, Polycomb proteins are initially localized to ∼150 strong sites concentrated within bivalent domains coinciding with CpG islands. PRC2 (Polycomb repressor complex 2) and histone H3 $Lys^{27}$ methylation patterns are indicative of subsequent spreading to thousands of sites, most of which are intergenic, non-bivalent and lack CpG islands [111]. Although gradient dispersal may be unique to X inactivation, it has also been proposed that PRC2 can sense the chromatin environment, and respond to local chromatin compaction upon transcription cessation. In this dense chromatin, regions of histone H3 activate PRC2 to regulate H3 $Lys^{27}$ methylation [112]. Microscopy studies have shown that PcG (Polycomb group) proteins concentrate into nuclear foci, called Pc bodies, whose number and size change upon cellular differentiation [113,114]. This suggests that PcG proteins may mediate the nuclear organization of their target genes. Several 3C (chromosome, conformation, capture)-based approaches have shown that PRC2 is involved in mediating long-range interactions in mammals, as well as *Drosophila* [115]. Whether PcG proteins are directly responsible for mediating these physical interactions is not clear; CTCF and insulators have been shown to co-localize and contribute to the nuclear organization of Pc domains [116]. *Drosophila* PcG proteins can remain associated with their DNA response elements through replication, possibly re-establishing

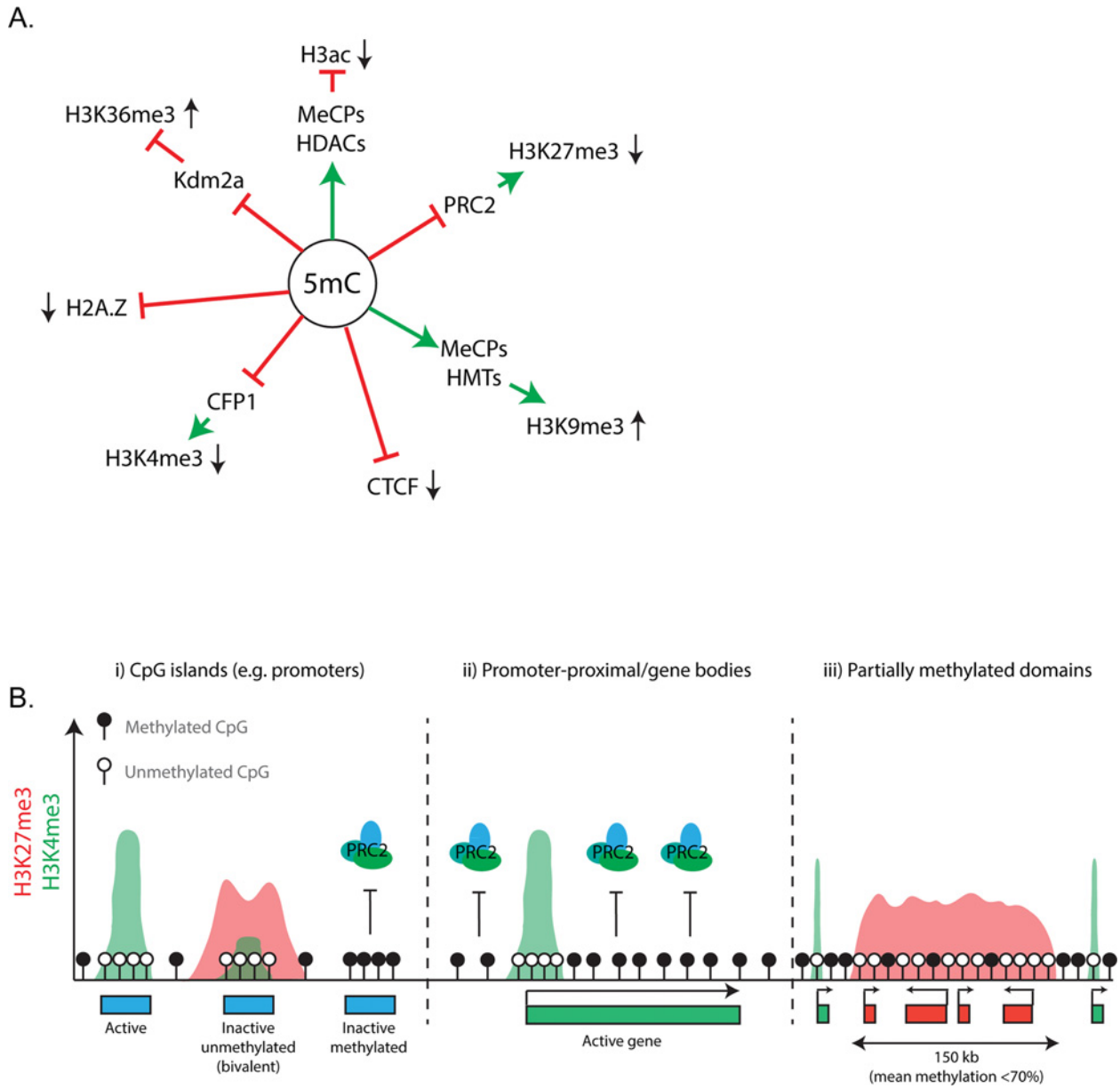


**Figure 3 DNA methylation shapes the chromatin landscape**

(**A**) Functional links between CpG methylation and aspects of chromatin structure. Red connections indicate that 5mC has a negative effect on the binding of a protein/complex to chromatin, whereas green connections indicate a positive influence of 5mC on binding. The effect of the protein/complex on an aspect of chromatin structure is indicated by red connections for negative effects (e.g. histone deacetylase on histone acetylation) and green for positive (e.g. a histone methyltransferase on histone methylation). The net effect of 5mC on each aspect of chromatin structure is indicated by the black arrows. For example, 5mC has a negative effect on PRC2 binding, which is a H3K27me3 methyltransferase, and therefore 5mC has a negative effect on H3K27me3. CFP1, CXXC-type zinc finger protein 1; HDAC, histone deactylases; HMT, histone methyltransferases; Kdm2a, lysine demethylase 2a. (**B**) DNA methylation influences Polycomb targeting. (i) PRC2 targeting has been linked to unmethylated transcriptionally inactive CGIs, and CpG methylation and the PRC2-catalysed H3K27me3 histone mark do not co-exist within CpG-rich regions of mammalian genomes [134,135]. Removal of DNA methylation from methylated CGIs results in *de novo* PRC2 targeting, suggesting that DNA methylation prevents PRC2 binding at these elements [136,138]. (ii) CpG methylation in promoter-proximal regions and gene bodies of actively transcribing genes has been shown to inhibit PRC2 binding, thereby facilitating gene expression [106]. (iii) Large regions of 'partial CpG methylation' have been identified in the methylomes of certain cells [6]. These regions often coincide with large domains of the H3K27me3 histone mark, raising the possibility that their relative hypomethylation is required for PRC2 occupancy [6,53]. An alternative hypothesis is that PRC2 occupancy inhibits DNA methylation, which has yet to be tested.

the histone modifications on newly assembled unmethylated histones, as well as their nuclear organization [117].

## DNA methylation and the Polycomb repression system

The local chromatin structures adopted by the epigenetic repression systems remain largely unknown, as functional models merely refer to their condensed nature to explain their gene repressive influence. The microscopically dense compartments in the nucleus were originally defined as heterochromatin, currently further categorized into constitutive (always condensed) heterochromatin and facultative heterochromatin, which is more dynamic and varies between tissues [3]. Pericentromeric heterochromatin and inactive X chromatin are paradigms for constitutive and facultative heterochromatin respectively. These

also contain archetypal examples for DNA methylation and Polycomb-repressed regions [111]. The epigenetic signature of pericentromeric heterochromatin is not established until late during fetal development, even though it is present in ES (embryonic stem) cells derived from embryos, indicating developmental plasticity and a role in maintenance of epigenetic states [118]. Nucleosome deposition and DNA methylation peaks at coding region boundaries suggest that DNA methylation can have effects at the local chromatin level [119,120]. PcG proteins were identified as developmental factors, but are now also known to be involved in controlling dynamics and plasticity of gene regulation, particularly during differentiation, by interacting with other components of the transcriptional apparatus [121].

DNA methylation and the Polycomb repressor system have been viewed as independent and parallel mechanisms of maintaining a cell-heritable memory of transcriptional repression [122]. However, in recent years intriguing links have been drawn between the two systems, suggesting that these epigenetic mechanisms are in fact connected. This interaction promises to enhance our understanding of each process individually, while their inter-connectivity may influence the way we view their roles in development and disease. The Polycomb repressor system comprises chromatin-associated proteins that form distinct complexes with histone-modifying activity [123]. Of central importance is the PRC2, a protein complex composed of four core components that include the histone H3 $Lys^{27}$ methyltransferase EZH2 (enhancer of zeste homologue 2) [124]. Both the PRC2 complex, and its signature histone mark H3K27me3, play an important and evolutionarily conserved role in the maintenance of transcriptional repression of a large set of genes, many of which are key regulators of embryonic development [123,125–127]. The first link between PRC2 and DNA methylation was the suggestion that the EZH2 protein may directly recruit DNA methyltransferases to establish aberrant DNA methylation in cancer [128]. Indeed, genes aberrantly hypermethylated in cancer are frequently PRC2 targets in early development, suggesting that PRC2 may somehow pre-mark genes for cancer-associated DNA methylation [129]. However, the generality of this mechanism is currently unclear as EZH2-mediated recruitment of DNA methyltransferases is not observed in all cancer cells, or in non-cancer tissues [130].

### DNA methylation modulates Polycomb targeting

How mammalian Polycomb complexes are targeted to specific genomic regions is a long-standing enigma in the field and remains an active area of research. In the fruit fly *Drosophila melanogaster*, PcG protein complexes are recruited to chromatin by DNA elements called PREs (Polycomb-response elements) [123], which mediate the inheritance of silent chromatin states during development. However, examples of elements that function like fly PREs are rare in mammalian genomes [123,131,132]. Recently, a model has been proposed in which the PRC2 component PHF19 binds H3K36me3 via its Tudor domain, forming essential contact points that allow recruitment of PRC2 and H3K36me3 demethylase activity to expressed gene loci during their transition to a Polycomb-repressed state [133]. Interestingly, it has emerged that DNA methylation influences the genomic targeting of the PRC2 complex, revealing it as an important piece of the mammalian Polycomb targeting puzzle (Figure 3B). The first clues came from studies that mapped PRC2 components in mammalian cells and revealed enriched binding at CGIs, regions that represent a large portion of the DNA methylation-free fraction of the genome [134–136] (Figure 3Bi). It was shown that *de novo* insertion of a CGI is sufficient to recruit PRC2 [136], as long as it is depleted of activating motifs [135], suggesting that Polycomb binding may be the default chromatin state for a CGI in ES cells. Importantly, endogenous CGIs containing high levels of DNA methylation are not bound by PRC2 [136]. Other studies have also revealed a negative correlation between DNA methylation and the PRC2 signature histone mark H3K27me3 in mammalian genomes, both at single loci, including mono-allelic H3K27me3 at an imprinted region [137], and at the genome-wide level [53,75,138,139]. Specifically, it appears that high levels of DNA methylation and H3K27me3 do not co-exist at CpG-rich regions of the genome [138,140]. A negative correlation between the two marks has also been noted in plant tissues, implying deep conservation of this relationship [141]. These observations led to the hypothesis that DNA methylation could be a determining factor in Polycomb targeting, by negatively modulating the binding of the PRC2 complex to chromatin. This idea was directly tested by mapping PRC2 binding and/or H3K27me3 upon perturbation of DNA methylation levels. Different laboratories have now shown in multiple contexts that experimentally reducing DNA methylation levels results in redistribution of PRC2 activity [106,136–138,142]. For example, deficiency in the *de novo* DNA methyltransferase Dnmt3a in mouse neural stem cells leads to loss of methylation from specific intergenic and intragenic regions [106]. At certain expressed genes required for neurogenesis, loss of methylation from regions surrounding their proximal promoters is associated with increased binding of the PRC2 complex, increased H3K27me3 and PRC2-dependent transcriptional silencing that results in aberrant neurogenesis [106]. This shows that an important role of DNA methylation at these genomic regions is to prevent PRC2-mediated repression of actively transcribed genes (Figure 3Bii). That study nicely demonstrated that the role of DNA methylation in transcriptional regulation is context-dependent and extends beyond its canonical role at gene promoter elements. Other studies have shown that global reduction in DNA methylation levels have a widespread effect on PRC2 targeting, with many regions showing increased PRC2 binding upon loss of DNA methylation [136,138]. The mechanistic basis for the inhibition of PRC2–chromatin binding by DNA methylation is currently unclear. PRC2 is proposed to associate with chromatin through multiple histone and DNA contacts [124] and it is feasible that one or more of these contacts are perturbed by DNA methylation. Indeed, *in vitro* experiments suggest that chromatin binding by PRC2 is directly attenuated by CpG methylation [104,106], a possibility that future experiments will need to investigate further.

This interplay between two epigenetic mechanisms raises interesting questions for future research. Is the difference in Polycomb targeting in CpG methylation-free organisms, such as *Drosophila*, due to the absence of restrictive activity of DNA methylation? Sequence-specific DNA-binding proteins have been more strongly linked to Polycomb targeting in flies than in mammals [123]. What is the consequence of disease-associated changes in DNA methylation on PRC2 distribution and gene expression? In this regard, cancer-associated DNA hypomethylation and hypermethylation have been reciprocally linked to increased and decreased H3K27me3 respectively, raising the possibility that DNA methylation changes in cancer affect gene expression and pathology in as yet unappreciated ways [75,139].

## CONCLUDING REMARKS

Collectively, the observations described above have widened the perspective for how DNA methylation contributes to gene

regulation and cell identity. They have demonstrated that DNA methylation has important regulatory functions in addition to its well-established repressive role at gene promoters, and therefore probably contributes to genome regulation in a multifactorial manner. Mechanistically, it is now clear that DNA modifications are an integral component of chromatin structure, and influence epigenetic landscapes through modulating the binding of a wide range of effector proteins to chromatin. Consequently, we are now a few steps closer to understanding the contribution that this epigenetic mark makes to the regulation of the genome during development, and the involvement of aberrant patterns of DNA modification in disease.

## ACKNOWLEDGEMENTS

We apologize to all colleagues whose work we have not cited due to space limitations. We thank Professor Javier Cáceres (MRC Human Genetics Unit, Edinburgh, U.K.) for encouragement in writing the present review.

## FUNDING

Work in the laboratory of R.M. is supported by the Medical Research Council, the Biotechnology and Biological Sciences Research Council and by the Innovative Medicine Initiative Joint Undertaking [grant number 115001 (MARCAR project)]. The S.P. laboratory acknowledges support from the Biotechnology and Biological Sciences Research Council and the British Heart Foundation.

## REFERENCES

1 Peter, I. S. and Davidson, E. H. (2011) A gene regulatory network controlling the embryonic specification of endoderm. Nature **474**, 635–639
2 Gardner, K. E., Allis, C. D. and Strahl, B. D. (2011) Operating on chromatin, a colorful language where context matters. J. Mol. Biol. **409**, 36–46
3 Cedar, H. and Bergman, Y. (2012) Programming of DNA methylation patterns. Annu. Rev. Biochem. **81**, 97–117
4 Bird, A. (2011) The dinucleotide CG as a genomic signalling module. J. Mol. Biol. **409**, 47–53
5 Zilberman, D. and Henikoff, S. (2007) Genome-wide analysis of DNA methylation patterns. Development **134**, 3959–3965
6 Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M. et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature **462**, 315–322
7 Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., Low, H. M., Kin Sung, K. W., Rigoutsos, I., Loring, J. and Wei, C. L. (2010) Dynamic changes in the human methylome during differentiation. Genome Res. **20**, 320–331
8 Li, Y., Zhu, J., Tian, G., Li, N., Li, Q., Ye, M., Zheng, H., Yu, J., Wu, H., Sun, J. et al. (2010) The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol. **8**, e1000533
9 Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P. and Jaenisch, R. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc. Natl. Acad. Sci. U.S.A. **97**, 5237–5242
10 Ziller, M. J., Muller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., Boyle, P., Epstein, C. B., Bernstein, B. E., Lengauer, T. et al. (2011) Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. PLoS Genet. **7**, e1002389
11 Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B. et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature **454**, 766–770
12 Deaton, A. M. and Bird, A. (2011) CpG islands and the regulation of transcription. Genes Dev. **25**, 1010–1022
13 Illingworth, R., Kerr, A., Desousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J. et al. (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol. **6**, e22
14 Bock, C., Beerman, I., Lien, W. H., Smith, Z. D., Gu, H., Boyle, P., Gnirke, A., Fuchs, E., Rossi, D. J. and Meissner, A. (2012) DNA methylation dynamics during *in vivo* differentiation of blood and skin stem cells. Mol. Cell **47**, 633–647
15 Ng, R. K., Dean, W., Dawson, C., Lucifero, D., Madeja, Z., Reik, W. and Hemberger, M. (2008) Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. Nat. Cell Biol. **10**, 1280–1290
16 Broske, A. M., Vockentanz, L., Kharazi, S., Huska, M. R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R. et al. (2009) DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat. Genet. **41**, 1207–1215
17 Trowbridge, J. J., Snow, J. W., Kim, J. and Orkin, S. H. (2009) DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell **5**, 442–449
18 Sen, G. L., Reuter, J. A., Webster, D. E., Zhu, L. and Khavari, P. A. (2010) DNMT1 maintains progenitor function in self-renewing somatic tissue. Nature **463**, 563–567
19 Dhawan, S., Georgia, S., Tschen, S. I., Fan, G. and Bhushan, A. (2011) Pancreatic $\beta$-cell identity is maintained by DNA methylation-mediated repression of Arx. Dev. Cell **20**, 419–429
20 Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Paabo, S., Rebhan, M. and Schubeler, D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat. Genet. **39**, 457–466
21 Santoro, R. and Grummt, I. (2001) Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. Mol. Cell **8**, 719–725
22 Watt, F. and Molloy, P. L. (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev. **2**, 1136–1143
23 Wiench, M., John, S., Baek, S., Johnson, T. A., Sung, M. H., Escobar, T., Simmons, C. A., Pearce, K. H., Biddie, S. C., Sabo, P. J. et al. (2011) DNA methylation status predicts cell type-specific enhancer activity. EMBO J. **30**, 3028–3039
24 Klose, R. J. and Bird, A. P. (2006) Genomic DNA methylation: the mark and its mediators. Trends Biochem. Sci. **31**, 89–97
25 Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. and Bird, A. P. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell **58**, 499–507
26 Meehan, R. R., Lewis, J. D. and Bird, A. P. (1992) Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. Nucleic Acids Res. **20**, 5085–5092
27 Li, E., Beard, C. and Jaenisch, R. (1993) Role for DNA methylation in genomic imprinting. Nature **366**, 362–365
28 Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E. and Jaenisch, R. (2001) Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat. Genet. **27**, 31–39
29 Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E. and Sasaki, H. (2004) Essential role for *de novo* DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature **429**, 900–903
30 Sharp, A. J., Stathaki, E., Migliavacca, E., Brahmachary, M., Montgomery, S. B., Dupre, Y. and Antonarakis, S. E. (2011) DNA methylation profiles of human active and inactive X chromosomes. Genome Res. **21**, 1592–1600
31 Sado, T., Fenner, M. H., Tan, S. S., Tam, P., Shioda, T. and Li, E. (2000) X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. Dev. Biol. **225**, 294–303
32 Lande-Diner, L., Zhang, J., Ben-Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A. G., Rechavi, G. and Cedar, H. (2007) Role of DNA methylation in stable gene repression. J. Biol. Chem. **282**, 12194–12200
33 Fouse, S. D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R. and Fan, G. (2008) 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
34 Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R. A. and Issa, J. P. (2007) Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet. **3**, 2023–2036
35 Borgel, J., Guibert, S., Li, Y., Chiba, H., Schubeler, D., Sasaki, H., Forne, T. and Weber, M. (2010) Targets and dynamics of promoter DNA methylation during early mouse development. Nat. Genet. **42**, 1093–1100
36 Velasco, G., Hube, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Pequignot, E. and Francastel, C. (2010) Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. Proc. Natl. Acad. Sci. U.S.A. **107**, 9281–9286
37 Hackett, J. A., Reddington, J. P., Nestor, C. E., Dunican, D. S., Branco, M. R., Reichmann, J., Reik, W., Surani, M. A., Adams, I. R. and Meehan, R. R. (2012) Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. Development **139**, 3623–3632

38 Hou, Y., Yuan, J., Zhou, X., Fu, X., Cheng, H. and Zhou, R. (2012) DNA demethylation and USF regulate the meiosis-specific expression of the mouse Miwi. PLoS Genet. **8**, e1002716
39 Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H. and Bergman, Y. (2008) *De novo* DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. Nat. Struct. Mol. Biol. **15**, 1176–1183
40 Sproul, D., Nestor, C., Culley, J., Dickson, J. H., Dixon, J. M., Harrison, D. J., Meehan, R. R., Sims, A. H. and Ramsahoye, B. H. (2011) Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. Proc. Natl. Acad. Sci. U.S.A. **108**, 4364–4369
41 Feng, Y. Q., Desprat, R., Fu, H., Olivier, E., Lin, C. M., Lobell, A., Gowda, S. N., Aladjem, M. I. and Bouhassira, E. E. (2006) DNA methylation supports intrinsic epigenetic memory in mammalian cells. PLoS Genet. **2**, e65
42 Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. and Henikoff, S. (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. Nat. Genet. **39**, 61–69
43 Rountree, M. R. and Selker, E. U. (1997) DNA methylation inhibits elongation but not initiation of transcription in *Neurospora crassa*. Genes Dev. **11**, 2383–2395
44 Rauch, T. A., Wu, X., Zhong, X., Riggs, A. D. and Pfeifer, G. P. (2009) A human B cell methylome at 100-base pair resolution. Proc. Natl. Acad. Sci. U.S.A. **106**, 671–678
45 Ball, M. P., Li, J. B., Gao, Y., Lee, J. H., LeProust, E. M., Park, I. H., Xie, B., Daley, G. Q. and Church, G. M. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nat. Biotechnol. **27**, 361–368
46 Aran, D., Toperoff, G., Rosenberg, M. and Hellman, A. (2011) Replication timing-related and gene body-specific methylation of active human genes. Hum. Mol. Genet. **20**, 670–680
47 Hellman, A. and Chess, A. (2007) Gene body-specific methylation on the active X chromosome. Science **315**, 1141–1143
48 Dhayalan, A., Rajavelu, A., Rathert, P., Tamas, R., Jurkowska, R. Z., Ragozin, S. and Jeltsch, A. (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J. Biol. Chem. **285**, 26114–26120
49 Xiao, T., Hall, H., Kizer, K. O., Shibata, Y., Hall, M. C., Borchers, C. H. and Strahl, B. D. (2003) Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast. Genes Dev. **17**, 654–663
50 Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V. K., Attwood, J., Burger, M., Burton, J., Cox, T. V., Davies, R., Down, T. A. et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat. Genet. **38**, 1378–1385
51 Illingworth, R. S., Gruenewald-Schneider, U., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., Smith, C., Harrison, D. J., Andrews, R. and Bird, A. P. (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet. **6**, e1001134
52 Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., D'Souza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y. et al. (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature **466**, 253–257
53 Hawkins, R. D., Hon, G. C., Lee, L. K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L. E., Kuan, S., Luu, Y., Klugman, S. et al. (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell **6**, 479–491
54 Nestor, C. E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., Katz, E., Dixon, J. M., Harrison, D. J. and Meehan, R. R. (2012) Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. Genome Res. **22**, 467–477
55 Nestor, C., Ruzov, A., Meehan, R. and Dunican, D. (2010) Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. Biotechniques **48**, 317–319
56 Brown, S. J., Stoilov, P. and Xing, Y. (2012) Chromatin and epigenetic regulation of pre-mRNA processing. Hum. Mol. Genet. **21**, R90–R96
57 Flores, K. B., Wolschin, F., Allen, A. N., Corneveaux, J. J., Huentelman, M. and Amdam, G. V. (2012) Genome-wide association between DNA methylation and alternative splicing in an invertebrate. BMC Genomics **13**, 480
58 Foret, S., Kucharski, R., Pellegrini, M., Feng, S., Jacobsen, S. E., Robinson, G. E. and Maleszka, R. (2012) DNA methylation dynamics, metabolic fluxes, gene splicing, and alternative phenotypes in honey bees. Proc. Natl. Acad. Sci. U.S.A. **109**, 4968–4973
59 Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R. and Oberdoerffer, S. (2011) CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature **479**, 74–79
60 Tazi, J. and Bird, A. (1990) Alternative chromatin structure at CpG islands. Cell **60**, 909–920
61 Thomson, J. P., Skene, P. J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A. R., Deaton, A., Andrews, R., James, K. D. et al. (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature **464**, 1082–1086
62 Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., Farcas, A. M., Park, P. J. and Klose, R. J. (2010) CpG islands recruit a histone H3 lysine 36 demethylase. Mol. Cell **38**, 179–190
63 Branciamore, S., Chen, Z. X., Riggs, A. D. and Rodin, S. N. (2010) CpG island clusters and pro-epigenetic selection for CpGs in protein-coding exons of HOX and other transcription factors. Proc. Natl. Acad. Sci. U.S.A. **107**, 15485–15490
64 Irizarry, R. A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M. et al. (2009) The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. **41**, 178–186
65 Doi, A., Park, I. H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S. et al. (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat. Genet. **41**, 1350–1353
66 Ji, H., Ehrlich, L. I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M. J., Irizarry, R. A., Kim, K. et al. (2010) Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature **467**, 338–342
67 Stancheva, I., Collins, A. L., Van den Veyver, I. B., Zoghbi, H. and Meehan, R. R. (2003) A mutant form of MeCP2 protein associated with human Rett syndrome cannot be displaced from methylated DNA by notch in *Xenopus* embryos. Mol. Cell **12**, 425–435
68 Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D. et al. (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature **480**, 490–495
69 Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E., Sheffield, N. C., Stergachis, A. B., Wang, H., Vernot, B. et al. (2012) The accessible chromatin landscape of the human genome. Nature **489**, 75–82
70 Hodges, E., Molaro, A., Dos Santos, C. O., Thekkat, P., Song, Q., Uren, P. J., Park, J., Butler, J., Rafii, S., McCombie, W. R. et al. (2011) Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. Mol. Cell **44**, 17–28
71 Schmidl, C., Klug, M., Boeld, T. J., Andreesen, R., Hoffmann, P., Edinger, M. and Rehli, M. (2009) Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. Genome Res. **19**, 1165–1174
72 Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F. and Schubeler, D. (2011) Identification of genetic elements that autonomously determine DNA methylation states. Nat. Genet. **43**, 1091–1097
73 Schroeder, D. I., Lott, P., Korf, I. and LaSalle, J. M. (2011) Large-scale methylation domains mark a functional subset of neuronally expressed genes. Genome Res. **21**, 1583–1591
74 Berman, B. P., Weisenberger, D. J., Aman, J. F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C. P., van Dijk, C. M., Tollenaar, R. A. et al. (2012) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. Nat. Genet. **44**, 40–46
75 Hon, G. C., Hawkins, R. D., Caballero, O. L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L. E. et al. (2012) Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res. **22**, 246–258
76 Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L. and Rao, A. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science **324**, 930–935
77 Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science **324**, 929–930
78 Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C. and Zhang, Y. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature **466**, 1129–1133
79 Matarese, F., Carrillo-de Santa Pau, E. and Stunnenberg, H. G. (2011) 5-Hydroxymethylcytosine: a new kid on the epigenetic block? Mol. Syst. Biol. **7**, 562
80 Wu, H. and Zhang, Y. (2011) Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes Dev. **25**, 2436–2452
81 Szulwach, K. E., Li, X., Li, Y., Song, C. X., Han, J. W., Kim, S., Namburi, S., Hermetz, K., Kim, J. J., Rudd, M. K. et al. (2011) Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. PLoS Genet. **7**, e1002154
82 Booth, M. J., Branco, M. R., Ficz, G., Oxley, D., Krueger, F., Reik, W. and Balasubramanian, S. (2012) Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science **336**, 934–937
83 Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., Marques, C. J., Andrews, S. and Reik, W. (2011) Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature **473**, 398–402
84 Song, C. X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C. H., Zhang, W., Jian, X. et al. (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat. Biotechnol. **29**, 68–72
85 Li, W. and Liu, M. (2011) Distribution of 5-hydroxymethylcytosine in different human tissues. J. Nucleic Acids **2011**, 870726

86 Ruzov, A., Tsenkina, Y., Serio, A., Dudnakova, T., Fletcher, J., Bai, Y., Chebotareva, T., Pells, S., Hannoun, Z., Sullivan, G. et al. (2011) Lineage-specific distribution of high levels of genomic 5-hydroxymethylcytosine in mammalian development. Cell Res. **21**, 1332–1342

87 Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E. and Zhang, Y. (2011) Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. **25**, 679–684

88 Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science **333**, 1300–1303

89 Chen, C. C., Wang, K. Y. and Shen, C. K. (2012) The mammalian *de novo* DNA methyltransferases Dnmt3a and Dnmt3b are also DNA 5-hydroxymethyl cytosine dehydroxymethylases. J. Biol. Chem. **287**, 33116–33121

90 Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D. et al. (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell **146**, 67–79

91 Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., Xie, Z. G., Shi, L., He, X., Jin, S. G. et al. (2011) The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature **477**, 606–610

92 Guo, J. U., Su, Y., Zhong, C., Ming, G. L. and Song, H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell **145**, 423–434

93 He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L. et al. (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science **333**, 1303–1307

94 Inoue, A. and Zhang, Y. (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. Science **334**, 194

95 Valinluck, V. and Sowers, L. C. (2007) Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res. **67**, 946–950

96 Hashimoto, H., Liu, Y., Upadhyay, A. K., Chang, Y., Howerton, S. B., Vertino, P. M., Zhang, X. and Cheng, X. (2012) Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. **40**, 4841–4849

97 Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W. and Walter, J. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nat. Commun. **2**, 241

98 Dawlaty, M. M., Ganz, K., Powell, B. E., Hu, Y. C., Markoulaki, S., Cheng, A. W., Gao, Q., Kim, J., Choi, S. W., Page, D. C. and Jaenisch, R. (2011) Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell **9**, 166–175

99 Ko, M., Bandukwala, H. S., An, J., Lamperti, E. D., Thompson, E. C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S. B. and Rao, A. (2011) Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proc. Natl. Acad. Sci. U.S.A. **108**, 14566–14571

100 Quivoron, C., Couronne, L., Della Valle, V., Lopez, C. K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M. H. et al. (2011) TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell **20**, 25–38

101 Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M. C., Antes, I. and Leonhardt, H. (2011) Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. PLoS ONE **6**, e21306

102 Mellen, M., Ayata, P., Dewell, S., Kriaucionis, S. and Heintz, N. (2012) MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell **151**, 1417–1430

103 Thalhammer, A., Hansen, A. S., El-Sagheer, A. H., Brown, T. and Schofield, C. J. (2011) Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation. Chem. Commun. (Cambridge, U.K.) **47**, 5325–5327

104 Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M. and Kouzarides, T. (2010) Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell **143**, 470–484

105 Skene, P. J., Illingworth, R. S., Webb, S., Kerr, A. R., James, K. D., Turner, D. J., Andrews, R. and Bird, A. P. (2010) Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol. Cell **37**, 457–468

106 Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y. and Sun, Y. E. (2010) Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science **329**, 444–448

107 Sarraf, S. A. and Stancheva, I. (2004) Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol. Cell **15**, 595–605

108 Fuks, F., Hurd, P. J., Wolf, D., Nan, X., Bird, A. P. and Kouzarides, T. (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J. Biol. Chem. **278**, 4035–4040

109 Zilberman, D., Coleman-Derr, D., Ballinger, T. and Henikoff, S. (2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature **456**, 125–129

110 Conerly, M. L., Teves, S. S., Diolaiti, D., Ulrich, M., Eisenman, R. N. and Henikoff, S. (2010) Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. Genome Res. **20**, 1383–1390

111 Pinter, S. F., Sadreyev, R. I., Yildirim, E., Jeon, Y., Ohsumi, T. K., Borowsky, M. and Lee, J. T. (2012) Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. Genome Res. **22**, 1864–1876

112 Yuan, W., Wu, T., Fu, H., Dai, C., Wu, H., Liu, N., Li, X., Xu, M., Zhang, Z., Niu, T. et al. (2012) Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. Science **337**, 971–975

113 Ficz, G., Heintzmann, R. and Arndt-Jovin, D. J. (2005) Polycomb group protein complexes exchange rapidly in living *Drosophila*. Development **132**, 3963–3976

114 Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U. and Cavalli, G. (2006) RNAi components are required for nuclear clustering of Polycomb group response elements. Cell **124**, 957–971

115 Tolhuis, B., Blom, M., Kerkhoven, R. M., Pagie, L., Teunissen, H., Nieuwland, M., Simonis, M., de Laat, W., van Lohuizen, M. and van Steensel, B. (2011) Interactions among Polycomb domains are guided by chromosome architecture. PLoS Genet. **7**, e1001343

116 Pirrotta, V. and Li, H. B. (2012) A view of nuclear Polycomb bodies. Curr. Opin. Genet. Dev. **22**, 101–109

117 Petruk, S., Sedkov, Y., Johnston, D. M., Hodgson, J. W., Black, K. L., Kovermann, S. K., Beck, S., Canaani, E., Brock, H. W. and Mazo, A. (2012) TrxG and PcG proteins but not methylated histones remain associated with DNA through replication. Cell **150**, 922–933

118 Wongtawan, T., Taylor, J. E., Lawson, K. A., Wilmut, I. and Pennings, S. (2011) Histone H4K20me3 and HP1$\alpha$ are late heterochromatin markers in development, but present in undifferentiated embryonic stem cells. J. Cell Sci. **124**, 1878–1890

119 Choi, J. K., Bae, J. B., Lyu, J., Kim, T. Y. and Kim, Y. J. (2009) Nucleosome deposition and DNA methylation at coding region boundaries. Genome Biol. **10**, R89

120 Pennings, S., Allan, J. and Davey, C. S. (2005) DNA methylation, nucleosome formation and positioning. Briefings Funct. Genomics Proteomics **3**, 351–361

121 Prezioso, C. and Orlando, V. (2011) Polycomb proteins in mammalian cell differentiation and plasticity. FEBS Lett. **585**, 2067–2077

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

123 Schuettengruber, B. and Cavalli, G. (2009) Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development **136**, 3531–3542

124 Margueron, R. and Reinberg, D. (2011) The Polycomb complex PRC2 and its mark in life. Nature **469**, 343–349

125 Bracken, A. P., Dietrich, N., Pasini, D., Hansen, K. H. and Helin, K. (2006) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. **20**, 1123–1136

126 Lee, T. I., Jenner, R. G., Boyer, L. A., Guenther, M. G., Levine, S. S., Kumar, R. M., Chevalier, B., Johnstone, S. E., Cole, M. F., Isono, K. et al. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. Cell **125**, 301–313

127 Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K. et al. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature **441**, 349–353

128 Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J. M. et al. (2006) The Polycomb group protein EZH2 directly controls DNA methylation. Nature **439**, 871–874

129 Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B. E. et al. (2007) Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. Nat. Genet. **39**, 232–236

130 Denis, H., Ndlovu, M. N. and Fuks, F. (2011) Regulation of mammalian DNA methyltransferases: a route to new mechanisms. EMBO Rep. **12**, 647–656

131 Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H. D. and Cordes, S. P. (2009) A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. Cell **138**, 885–897

132 Woo, C. J., Kharchenko, P. V., Daheron, L., Park, P. J. and Kingston, R. E. (2010) A region of the human HOXD cluster that confers polycomb-group responsiveness. Cell **140**, 99–110

133 Brien, G. L., Gambero, G., O'Connell, D. J., Jerman, E., Turner, S. A., Egan, C. M., Dunne, E. J., Jurgens, M. C., Wynne, K., Piao, L. et al. (2012) Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nat. Struct. Mol. Biol. **19**, 1273–1281
134 Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S. et al. (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. **4**, e1000242
135 Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., Ku, M. and Bernstein, B. E. (2010) GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. **6**, e1001244
136 Lynch, M. D., Smith, A. J., De Gobbi, M., Flenley, M., Hughes, J. R., Vernimmen, D., Ayyub, H., Sharpe, J. A., Sloane-Stanley, J. A., Sutherland, L. et al. (2012) An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J. **31**, 317–329
137 Lindroth, A. M., Park, Y. J., McLean, C. M., Dokshin, G. A., Persson, J. M., Herman, H., Pasini, D., Miro, X., Donohoe, M. E., Lee, J. T. et al. (2008) Antagonism between DNA and H3K27 methylation at the imprinted Rasgrf1 locus. PLoS Genet. **4**, e1000145
138 Brinkman, A. B., Gu, H., Bartels, S. J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A. and Stunnenberg, H. G. (2012) Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. **22**, 1128–1138
139 Gal-Yam, E. N., Egger, G., Iniguez, L., Holster, H., Einarsson, S., Zhang, X., Lin, J. C., Liang, G., Jones, P. A. and Tanay, A. (2008) Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. Proc. Natl. Acad. Sci. U.S.A. **105**, 12979–12984
140 Statham, A. L., Robinson, M. D., Song, J. Z., Coolen, M. W., Stirzaker, C. and Clark, S. J. (2012) Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res. **22**, 1120–1127
141 Weinhofer, I., Hehenberger, E., Roszak, P., Hennig, L. and Kohler, C. (2010) H3K27me3 profiling of the endosperm implies exclusion of polycomb group protein targeting by DNA methylation. PLoS Genet. **6**, e1001152
142 Wu, H., D'Alessio, A. C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y. E. and Zhang, Y. (2011) Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. Nature **473**, 389–393

Received 19 October 2012/9 January 2013; accepted 10 January 2013
Published on the Internet 14 March 2013, doi:10.1042/BJ20121585