Functions of DNA methylation: islands, start sites, gene bodies and beyond

Peter A. Jones

Abstract | DNA methylation is frequently described as a 'silencing' epigenetic mark, and indeed this function of 5-methylcytosine was originally proposed in the 1970s. Now, thanks to improved genome-scale mapping of methylation, we can evaluate DNA methylation in different genomic contexts: transcriptional start sites with or without CpG islands, in gene bodies, at regulatory elements and at repeat sequences. The emerging picture is that the function of DNA methylation seems to vary with context, and the relationship between DNA methylation and transcription is more nuanced than we realized at first. Improving our understanding of the functions of DNA methylation is necessary for interpreting changes in this mark that are observed in diseases such as cancer.

CpG islands
CpG-rich regions of DNA that are often associated with the transcription start sites of genes and that are also found in gene bodies and intergenic regions.

Bisulphite-treated DNA
Bisulphite treatment of DNA converts cytosine to uracil but leaves 5-methylcytosine intact. Thus, 5-methylcytosine patterns can be mapped by subsequent sequencing.

Insulators
DNA elements that control interactions between enhancers and promoters.

USC Norris Comprehensive Cancer Center, Keck School of Medicine of University of Southern California, Los Angeles, California 90089–99176, USA.
e-mail: pjones@med.usc.edu

Two key papers in 1975 independently suggested that methylation of cytosine residues in the context of CpG dinucleotides could serve as an epigenetic mark in vertebrates[1,2]. These papers proposed that sequences could be methylated *de novo*, that methylation can be inherited through somatic cell divisions by a mechanism involving an enzyme that recognizes hemimethylated CpG palindromes, that the presence of methyl groups could be interpreted by DNA-binding proteins and that DNA methylation directly silences genes. Although several of these key tenets turned out to be correct, the relationship between DNA methylation and gene silencing has proved to be challenging to unravel.

Most work in animals has focused on 5-methylcytosine (5mC) in the CpG sequence context. Methylation of other sequences is widespread in plants and some fungi[3,4] and has recently been reported in mammals[5]. In mammals, the function of non-CpG methylation is currently unknown. Here I primarily focus on CpG methylation in mammalian genomes, including some discussion of the differences observed in other animals and in plants.

Understanding the functions of DNA methylation requires consideration of the distribution of methylation across the genome. More than half of the genes in vertebrate genomes contain short (approximately 1 kb) CpG-rich regions known as CpG islands (CGIs), and the rest of the genome is depleted for CpGs. As 5mC can be converted to thymine by spontaneous or enzymatic deamination, it is thought that the loss of genomic CpGs is due to deamination of methylated sequences in the germline; CGIs are thought to exist because they are probably never or only transiently methylated in the germline[6]. However, there is a lot of discussion as to exactly what the definition of the CGI should be[7], and although the CpG density of promoters in mammalian genomes has a bimodal distribution, regions with intermediate CpG densities also exist[8]. Until recently, much of the work on DNA methylation focused on CGIs at transcriptional start sites (TSSs), and it is this work that has tended to shape general perceptions about the function of DNA methylation.

Recent approaches that enable genome-wide studies of the methylome (BOX 1) — for example, using bisulphite-treated DNA (which detects 5mC and hydroxymethyl-cytosine; see BOX 1) — have emphasized that the position of the methylation in the transcriptional unit influences its relationship to gene control. For example, methylation in the immediate vicinity of the TSS blocks initiation, but methylation in the gene body does not block and might even stimulate transcription elongation, and exciting new evidence suggests that gene body methylation may have an impact on splicing. Methylation in repeat regions such as centromeres is important for chromosomal stability[9] (for example, chromosome segregation at mitosis) and is also likely to suppress the expression of transposable elements and thus to have a role in genome stability. The role of methylation in altering the activities of enhancers, insulators and other regulatory elements is only just

Box 1 | Measuring DNA methylation genome-wide

Several approaches have been developed over the past few years to map 5-methylcytosine (5mC) patterns on a genome-wide scale, and their strengths and weaknesses have recently been compared[89]. These methods include enzymatic digestion with methylation-sensitive restriction enzymes and capture of 5mC by methylated DNA-binding proteins followed by next-generation sequencing. Methyl-DNA immunoprecipitation (MeDIP) is another approach in which extracted DNA is cleaved, denatured and precipitated using an antibody to 5mC, and then the precipitated fragments are sequenced[35]. Methods based on the treatment of DNA with bisulphite have become very popular. Bisulphite treatment converts unmethylated Cs to Us, which are subsequently amplified as Ts by PCR. Microarrays, such as the Illumina 450K array, have been used extensively to analyse bisulphite-treated DNA. Reduced representation bisulphite sequencing is an approach in which DNA is cleaved by methylation-sensitive restriction enzymes before bisulphite treatment. The most comprehensive coverage at single-base level is obtained by shotgun sequencing of bisulphite-treated DNA. A potential problem with the use of bisulphite sequencing is the fact that it cannot distinguish between 5mC and 5-hydroxmethylcytosine, which has recently been found in DNA. The current data may therefore have to be revisited in the future to accommodate this fact[90]. Although biases can be introduced by all of these approaches, analysing the data in conjunction with genetic polymorphisms such as SNPs can provide a calibration of expected to observed results, thus validating the results.

Ten-eleven translocation
(TET). Proteins of this type were recently shown to catalyse the conversion of 5-methylcytosine to 5-hydroxymethylcytosine.

Activation-induced cytidine deaminase
(AID). An enzyme that removes the amino group from cytosine or 5-methylcytosine. It is involved in class switch recombination and DNA demethylation.

Thymine DNA glycosylase
A protein that is involved in the repair of T:G mismatches that are often caused by 5-methylcytosine deamination and that participates in DNA demethylation.

Nucleosome-depleted regions
(NDRs). Regions of DNA that are not extensively wrapped up in nucleosomes. They can be seen at transcription start sites and other regulatory regions such as enhancers.

Polycomb proteins
Polycomb proteins participate in the silencing of genes by mechanisms that do not involve DNA methylation. They often silence genes that are key regulators of differentiation.

beginning to be appreciated. Furthermore, although there is abundant evidence that methylated CGIs at TSSs are associated with some silent genes, the timing of *de novo* methylation with respect to gene silencing is now beginning to be elucidated.

The function of DNA methylation is intrinsically linked to the mechanisms for establishing, maintaining and removing the methyl group. These mechanisms have been reviewed elsewhere[10,11], but some key points need to be borne in mind. It has been known for many years that DNA methyltransferases, including the so-called *de novo* DNA methyltransferase enzymes DNMT3A and DNMT3B, are essential for setting up DNA methylation patterns in early development. Our realization of how this happens has been helped enormously by the realization that in some cases the substrate for a *de novo* methyltransferase is nucleosomal DNA and that the modifications of histones within the nucleosome profoundly influence the ability of these enzymes to induce *de novo* methylation[12]. It was previously thought that DNMT1 by itself could maintain an established pattern of DNA methylation, but we now know that this is not true and that the ongoing participation of DNMT3A and DNMT3B is required for methylation maintenance[10]. Each of the three DNMTs is required for embryonic or neonatal development[13,14], and complete lack of methylation is incompatible with viability of somatic[15] or cancer cells[16] but not of embryonic stem cells (ESCs)[17]. DNMT3A has recently been shown to be essential for haematopoietic stem cell differentiation[18], again pointing to the fundamental role of 5mC in vertebrate differentiation. The timing of *de novo* methylation with respect to gene silencing has been an area of discussion, but the idea that DNA methylation directly silences genes *de novo* as proposed by Riggs[2] and Holliday and Pugh[1] is probably not the predominant pathway for gene silencing.

5mC must be removed by either passive or active means to establish a permissive state for subsequent gene expression. The search for DNA demethylases has been a long one and one that has been fraught with many false starts[19], but it is now more widely accepted that demethylases exist[20,21]. Recently, a plethora of papers has shown that active demethylation can be achieved, although this requires a mechanism that ultimately involves cell division or DNA repair and the excision of the base rather than the removal of the methyl group directly from the 5mC moiety[11,22]. The involvement of enzymes such as the ten-eleven translocation (TET) methylcytosine dioxygenases, activation-induced cytidine deaminase (AID) and thymine DNA glycosylase (TDG) in active and passive demethylation and in gene activation is now being elucidated[11,23–26]. Indeed, the absence of TET3 leads to a failure to demethylate CpG sites in key genes such as *Oct4* (also known as *Pou5f1*) or *Nanog* on the paternal genome and delays embryogenesis[27,28].

Alterations in DNA methylation are now known to cooperate with genetic events and to be involved in human carcinogenesis[29]. Therefore, understanding the roles of DNA methylation is essential for understanding disease processes. In this Review, I evaluate the evidence relating to the functions of DNA methylation in different genomic contexts, with a particular emphasis on the relationship with transcription (key knowns and unknowns are summarized in BOX 2). I then introduce possible mechanisms by which DNA methylation might exert its effects — for example, through altering protein binding — and I consider remaining questions.

Transcription start sites

Patterns at CpG island transcription start sites. Most CGIs remain unmethylated in somatic cells. When genes with CGIs at their TSS are active, their promoters are usually characterized by nucleosome-depleted regions (NDRs) at the TSS, and these NDRs are often flanked by nucleosomes containing the histone variant H2A.Z and are marked with trimethylation of histone H3 at lysine 4 (H3K4me3)[30] (FIG. 1). The levels of gene expression are controlled by transcription factors[31]. CGI promoters can be repressed by various mechanisms, such as repression mediated by Polycomb proteins. For example, genes encoding master regulators of embryonic development, such as myogenic differentiation 1 (*MYOD1*) or paired box 6 (*PAX6*), are suppressed by the Polycomb complex both in ESCs and in differentiated cells that are not expressing these genes; they have nucleosomes at the TSS and are marked by H3K27me3, which is generally associated with inactive genes[32].

However, some repressed genes do have methylated promoter CGIs. Methylated promoter CGIs are usually restricted to genes at which there is long-term stabilization of repressed states. Examples include imprinted genes, genes located on the inactive X chromosome and genes that are exclusively expressed in germ cells and that would presumably be inappropriate for expression in somatic cells. The stabilization of suppression by DNA methylation of CGIs can last over a 100-year lifespan and has no effect on the existence

This box summarizes the key points regarding our knowledge and lack of knowledge of DNA methylation in mammals.

Known
- Most CpG islands (CGIs) are not methylated when located at transcription start sites (TSSs).
- CGI methylation of the TSS is associated with long-term silencing (for example, X-chromosome inactivation, imprinting, genes expressed predominantly in germ cells and some tissue-specific genes).
- CGIs in gene bodies are sometimes methylated in a tissue-specific manner.
- Non-CGI methylation is more dynamic and more tissue-specific than CGI methylation.
- Methylation blocks the start of transcription not elongation (note that this is different in the fungus *Neurospora crassa*).
- Methylation of transposable elements silences these elements but allows the host gene to undergo transcriptional elongation.
- Gene body methylation contributes to cancer-causing somatic and germline mutations[53].

Unknown
- Does non-CGI methylation silence genes (that is, is it a cause or a consequence)?
- The function of methylation in the context CHG (where H is A, C or T).
- The roles of active and passive demethylation in activating genes.
- The function of 'shore' methylation.
- Does gene body methylation control splicing?
- The role of 5-hydroxymethylation in the brain and other tissues.
- The role of methylation in enhancer or insulator function.
- Does silencing always precede methylation?

of CGIs, because any deamination events within these regions in somatic cells would not be passed on through the germline to subsequent generations. We still do not completely understand why a minority of CpG islands become methylated, whereas most do not.

Patterns at non-CpG-island TSSs. In contrast to genes with CGIs at their TSSs, substantial fluctuations occur in the promoter methylation levels of genes that are CpG-poor at the TSS. Genes with non-CGI TSSs that are expressed in primordial germ cells are unmethylated at the TSS, whereas genes that are exclusively expressed in ESCs or tissue-specific genes often show methylation in sperm but not in oocytes or in expressing somatic cells[33]. Well-known examples are the genes encoding the OCT4 and NANOG transcription factors, which are essential for the maintenance of the stem cell state. Recent studies have suggested that the *OCT4* and *NANOG* promoters may undergo active demethylation by AID[11,22] and/or by TET3 (REF. 27). Some tissue-specific genes, however, show methylation in sperm and in ESCs and only show demethylation in the specific tissues in which these genes are expressed[34].

One genome-wide study postulated that no inverse relationship existed between methylation of non-CGIs and expression[35], but re-analysis of the data suggested that such a relationship between expression and methylation is in fact apparent genome-wide[36]. Because of the long-standing focus on CGIs, we still do not know the details of the role of methylation in controlling non-CGI TSSs.

Does methylation silence transcription initiation? Given the observations of methylation at some repressed TSSs described above, what is the functional relationship between DNA methylation and transcription initiation? There is incontrovertible evidence that methylated CGIs at TSSs cannot initiate transcription after the DNA has been assembled into nucleosomes[37–39]. However, the issue of whether silencing or methylation comes first has long been a discussion in the field. Early experiments by Lock *et al.*[40] clearly showed that methylation of the *Hprt* gene on the inactive X chromosome occurred after the chromosome had been inactivated. In other words, methylation appeared to serve as a 'lock' to reinforce a previously silenced state of X-linked genes. Although most CGIs on autosomal genes remain unmethylated in somatic cells, a small number of them (<10%) become methylated in normal tissues and cells[7], but the timing of the *de novo* methylation with respect to silencing has not been studied in depth. As mentioned above, recent findings regarding the role of DNMT3A in haematopoietic stem cell differentiation raise doubts about the universality of the long-term 'locking' model[18]. Because the authors of this study showed that the methylase was essential for differentiation of a fairly short-lived cell type, it seems possible that DNA methylation has a more instructive role in initiating rather than reinforcing the silencing.

Genome-wide studies in cancer cells have, however, shown that genes with CGI promoters that are already silenced by Polycomb complexes are much more likely than other genes to become methylated in cancer: that is, the silent state precedes methylation[36,41–43]. Therefore, it seems likely that silencing preceding methylation is a general mechanism, but the data are not yet mature enough to be sure. In addition to alterations at the CpG islands themselves, tissue-specific changes occur in 'shores' surrounding them[44]. However, the implications of these alterations is not yet understood. The evidence regarding the timing of DNA methylation is consistent with the idea that methylation adds an additional level of stability to epigenetic states. Intriguingly, it is not required for this purpose in some species, including *Drosophila melanogaster* and yeast.

The relationship between transcription and **de novo** *methylation.* The reasons why DNA methylation is probably not used as an initial silencing mechanism are now starting to be understood. The pioneering work of Ooi *et al.*[12] showed that the process of *de novo* methylation in cells expressing DNMT3L (which is a catalytically inactive homologue of DNMT3A and DNMT3B) is achieved by a tetrameric complex of two molecules each of DNMT3A2 and DNMT3L and requires a nucleosome. Active TSSs are depleted of nucleosomes and therefore lack this substrate for *de novo* methylation. Recently, we have directly tested the role of the nucleosome in triggering *de novo* methylation by examining the kinetics of *OCT4* silencing in embryonal carcinoma cells induced to differentiate with retinoic acid[45] (FIG. 2). These experiments showed that at the *OCT4* distal enhancer and the *NANOG* promoter, following differentiation, first a

Imprinted genes
Imprinted genes show parent-of-origin expression and are controlled by epigenetic processes, including DNA methylation.

X-chromosome inactivation
One of the two X chromosomes in female mammalian somatic cells is stably silenced by epigenetic processes, including DNA methylation, to achieve dosage compensation.
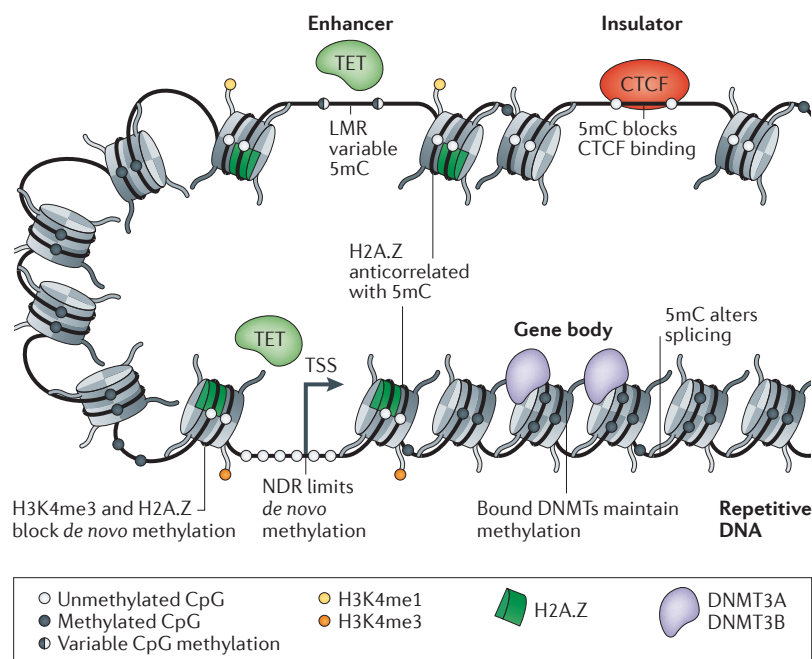
Figure 1 | **Molecular anatomy of CpG sites in chromatin and their roles in gene expression.** About 60% of human genes have CpG islands (CGIs) at their promoters and frequently have nucleosome-depleted regions (NDRs) at the transcriptional start site (TSS). The nucleosomes flanking the TSS are marked by trimethylation of histone H3 at lysine 4 (H3K4me3), which is associated with active transcription, and the histone variant H2A.Z, which is antagonistic to DNA methyltransferases (DNMTs). Downstream of the TSS, the DNA is mostly CpG-depleted and is predominantly methylated in repetitive elements and in gene bodies. CGIs, which are sometimes located in gene bodies, mostly remain unmethylated but occasionally acquire 5-methylcytosine (5mC) in a tissue-specific manner (not shown). Transcription elongation, unlike initiation, is not blocked by gene body methylation, and variable methylation may be involved in controlling splicing. Gene bodies are preferential sites of methylation in the context CHG (where H is A, C or T) in embryonic stem cells[5], but the function is not understood (not shown). DNA methylation is maintained by DNMT1 and also by DNMT3A and/or DNMT3B, which are bound to nucleosomes containing methylated DNA[99]. Enhancers tend to be CpG-poor and show incomplete methylation, suggesting a dynamic process of methylation or demethylation occurs, perhaps owing to the presence of ten-eleven translocation (TET) proteins in these regions, although this remains to be shown. They also have NDRs, and the flanking nucleosomes have the signature H3K4me1 mark and also the histone variant H2A.Z[32,100]. The binding of proteins such as CTCF to insulators can be blocked by methylation of their non-CGI recognition sequences, thus leading to altered regulation of gene expression, but the generality of this needs further exploration. The sites flanking the CTCF sites are strongly nucleosome-depleted, and the flanking nucleosomes show a remarkable degree of phasing. The figure does not show the structure of CpG-depleted promoters or silenced CGIs, although in both cases the silent state is associated with nucleosomes at the TSS. LMR, low-methylated region.

nucleosome becomes present, and then this is followed by the recruitment of DNMT3A to this nucleosome and, subsequently, *de novo* methylation occurs. Whether a similar sequence of events occurs in cells that are not expressing DNMT3L is not yet known.

Furthermore, Ooi *et al.*[12] showed that *de novo* methylation could not occur on a nucleosome bearing the H3K4me2 or H3K4me3 marks, which are associated with active genes. The nucleosomes flanking the nucleosome-depleted start site often contain both the histone mark H3K4me3 and the histone variant

H2A.Z, both of which are strongly anti-correlated with DNA methylation[46,47]. The occurrence of the H3K4me3 mark in mice is possibly maintained by the presence of CXXC finger protein 1 (CXXC1; also known as CFP1), which recruits the H3K4 methyltransferase to the region, thus ensuring that the +1 and –1 nucleosomes contain marks that are incompatible with *de novo* DNA methylation[48]. The unmethylated state of the CpG island is also presumably ensured by the presence of the TET1 protein, which is found at a high proportion of the TSSs of high-CpG-content promoters. Presumably, TET1 converts any 5mC that might be in this region into 5-hydroxymethylcytosine[49]. The molecular anatomy of active CGIs can therefore explain why they are resistant to methylation (FIG. 1).

Of course, not all CGI-promoter genes are expressed in ESCs, and many are suppressed by the Polycomb complex, so why are these not *de novo* methylated? The answer probably lies in the fact that they contain the antagonistic H3K4me3 (REF. 12) and H2A.Z marks[46,47] and are also bound by TET1, which would ensure that they remain 5mC-free. Interestingly, this protection seems to break down during immortalization[50], and these CGIs become highly susceptible to *de novo* methylation, which increases after oncogenic transformation[41–43].

This model predicts that the higher the level of expression is, the less likely it is that a CGI is to become *de novo* methylated. Direct evidence in support of this prediction has recently come from several exciting papers that have shown that monoallelic methylation of CGIs preferentially occurs on the allele that is less highly expressed. For example, Hitchins *et al.*[51] showed that an allele of the *MLH1* gene containing a single-nucleotide variant in the promoter, which was less active than the more common allele in transfection experiments, was more likely to become methylated in the somatic cells of cancer-affected families. In other words, the less active allele was the one that was more likely to acquire *de novo* methylation. An alternative scenario was shown by Boumber *et al.*[52], who found that an allele of *RIL* (also known as *PDLIM4*) bearing a polymorphism in the promoter that created an additional binding site for the transcription factor SP1 or SP3 was much less likely to become *de novo* methylated than the allele without this polymorphism. The extra SP1 site therefore confers resistance of this allele to *de novo* methylation, although the authors could not demonstrate that the extra transcription factor binding site increased gene expression.

Gene body methylation

Most gene bodies are CpG-poor, are extensively methylated and contain multiple repetitive and transposable elements. Methylation of the CpG sites in gene exons is a major cause of C→T transition mutations, leading to disease-causing mutations in the germline and cancer-causing mutations in somatic cells[53]. It is important to realize that although many CGIs are located at gene promoters, CGIs also exist within the bodies of genes[54] and within gene deserts. Although their functions here remain unknown, Adrian Bird has proposed that these regions may represent 'orphan promoters'

that might be used at early stages of development and have escaped methylation in the germline so that their high CpG density is maintained[55].

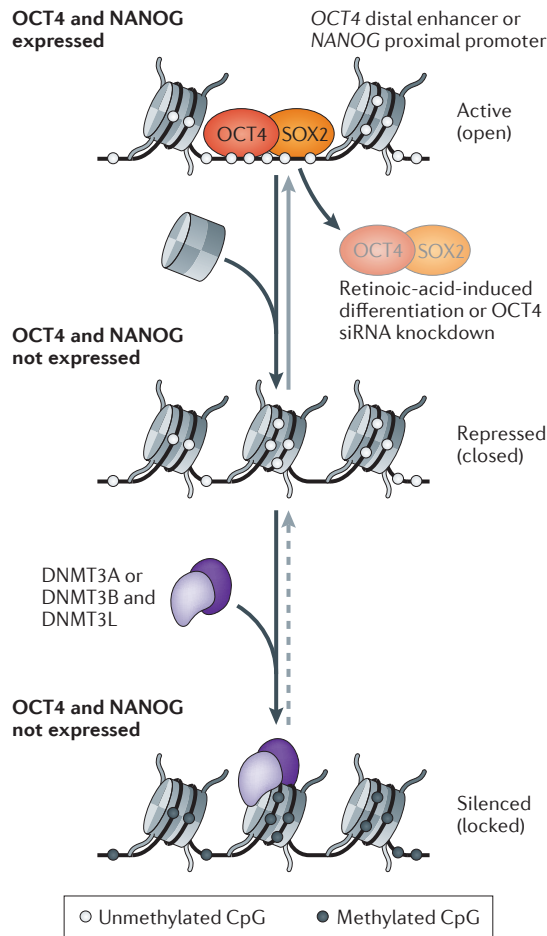*Gene body methylation is not associated with repression.* It has been known from the early days of DNA methylation research that gene body methylation is a feature of transcribed genes[56]. Extensive positive correlations between active transcription and gene body methylation have recently been confirmed on the active X chromosome[57] and by shotgun bisulphite sequencing of plant and animal genomes[3,5,58]. Most gene bodies are not CGIs, and when CGIs are situated in intragenic regions, they were, with a few exceptions[59], thought to remain unmethylated. However, recent experiments[55,60] have changed this perception: for example, as many as 34% of all intragenic CGIs are methylated in the human brain[60]. The role of this methylation, which is tissue-specific, is not yet clear. It is intriguing, especially because TSSs largely remain unmethylated. Intragenic CGIs can also be preferential sites for *de novo* methylation in cancer[61].

Even though gene body CGIs can become extensively methylated, this does not block transcription elongation. This is despite the fact that the methylated CGIs are marked by H3K9me3 and are bound by methyl-CpG-binding protein 2 (MECP2), which are chromatin features that are associated with repressed transcription when they are present at the TSS[62]. This leads to an apparent paradox in which methylation in the promoter is inversely correlated with the expression, whereas methylation in the gene body is positively correlated with expression[54]. Thus, in mammals, it is the initiation of transcription but not transcription elongation that seems to be sensitive to DNA methylation silencing. By contrast, cytosine methylation in CpG and other sequence contexts in *Neurospora crassa* blocks elongation but not initiation[4]. Therefore, it is not simply the presence of a 5mC mark itself that governs its relationship to transcription but rather the interpretation of the mark in a particular genomic and cellular context.

Possible functions of gene body methylation. What is the function of the gene body methylation outside CGIs? Initially, it was thought that this methylation was primarily a mechanism for silencing repetitive DNA elements, such as retroviruses, LINE1 elements, *Alu* elements and others, and evidence has been obtained to substantiate this idea[63]. Methylation blocks initiation of transcription at these elements while at the same time allowing transcription of the host gene to run through them.

It has also been proposed that the process of transcript elongation could itself stimulate DNA methylation and that H3K36me3, which is also associated with elongation but not initiation, might be involved in the recruitment of DNMTs[64]. However, whole-genome studies have shown that there might be alternative functions for DNA methylation in gene bodies. This work has shown that exons are more highly methylated than introns, and transitions in the degree of methylation occur at exon–intron boundaries, possibly suggesting a role for methylation in regulating splicing[65]. Indeed, genome-wide nucleosome-positioning data suggest that exons also show increased nucleosome occupancy levels compared to introns[66] and nucleosomes are preferential sites for DNA methylation[67]. A recent study has suggested that binding of CTCF (which can be regulated

OCT4 and NANOG expressed

OCT4 distal enhancer or NANOG proximal promoter

Active (open)

Retinoic-acid-induced differentiation or OCT4 siRNA knockdown

OCT4 and NANOG not expressed

Repressed (closed)

DNMT3A or DNMT3B and DNMT3L

OCT4 and NANOG not expressed

Silenced (locked)

○ Unmethylated CpG ● Methylated CpG

Figure 2 | **Silencing precedes DNA methylation.** Active promoters and enhancers have nucleosome-depleted regions (NDRs) that are often occupied by transcription factors and chromatin remodellers. Loss of factor binding — for example, during differentiation — leads to increased nucleosome occupancy of the regulatory region, providing a substrate for *de novo* DNA methylation. DNA methylation subsequently provides added stability to the silent state and is likely to be a mechanism for more accurate epigenetic inheritance during cell division. The example given is for the *OCT4* and *NANOG* genes[45], and its generality is not yet known, but inactive genes are often more susceptible to *de novo* methylation than their more active counterparts (REFS 36,40–43,51,52). In the figure, OCT4 binding is shown and NANOG binding is not shown, although its expression is required. Recent experiments have demonstrated that the methylation must be removed by active and/or passive processes to reactivate the gene. DNMT3A, DNA methyltransferase 3A; siRNA, small interfering RNA.

by DNA methylation; see below) results in the pausing of RNA polymerase II (RNAPII) and, as the kinetics of RNAPII movement influences splicing, this might link DNA methylation to splicing[68]. These observations suggest a previously unrecognized role for DNA methylation at the transcriptional level, possibly resulting in alternative splicing. Therefore, it seems likely that DNA methylation in gene bodies will have outcomes beyond the recognized function in the silencing of intragenic repetitive DNA sequences.

When is a body a start site? It is often assumed that TSSs and gene bodies are two separate genomic features. However, most genes have at least two TSSs, so the downstream start sites are within the 'bodies' of the transcriptional units of the upstream promoters. These alternative promoters can be CGIs or non-CGIs, or there can be combinations of an upstream non-CGI and a downstream CGI, or vice versa. These alternative start sites complicate the interpretation of experiments linking expression to methylation, because probes that are used to measure expression often detect the output of all of the promoters, yet only one might be active in a given cell type. Methylation of a downstream promoter would only block transcription from that promoter — it would allow the elongation of a transcript that emanates from an upstream promoter[62] — leading to an apparent discordance between methylation and expression. Indeed, DNA methylation may well be a mechanism for controlling alternative promoter usage[60].

Other regulatory sites

Methylation at enhancers. Enhancers are situated at variable distances from promoters and are key to controlling gene expression in development and cell function. They are mostly CpG-poor, and their methylation status has been examined by whole-methylome analysis (in plants and mammals)[3,5,69]. In general, these regions tend to have fairly variable methylation. Indeed, Stadler *et al.*[70] identified enhancers in the mouse genome on the basis that they are regions that are not 100% methylated or unmethylated and termed these 'low-methylated regions' (LMRs). Because a given cytosine can either be completely methylated or unmethylated, 'variable methylation' is the outcome of averaging these binary states. This might suggest that the CpG sites are in a dynamic state and that at a given time some are methylated and others are not, owing to competing methylation and demethylation events. Alternatively, the DNA methylation status of each CpG might not be accurately maintained during cell division, and so the LMR state might be due to inefficient inheritance. In different subsets of T cells, Schmidl *et al.*[71] also found a large number of differentially methylated regions (DMRs) within the enhancers of differentiation specific genes. In terms of function, this study showed that methylation of these CpG sites could result in reduced activity of the enhancer in reporter assays.

The idea that the methylation status of an enhancer and enhancer function are closely connected is supported by several observations of proteins that modulate methylation at these regions. For example, analysis of the binding of the glucocorticoid receptor to distal regulatory elements showed that CpGs could become demethylated and that the enhancer could be activated by the presence of this receptor[72]. Similar findings had originally been reported more than 25 years ago by Saluz *et al.*[73], who demonstrated demethylation of the overlapping oestradiol and glucocorticoid receptor binding sites in roosters that had been treated with oestradiol. In addition, 5-hydroxymethylcytosine and the TET proteins can be detected at these elements[74–78]. However, the relationship between CpG methylation and transcription factor binding is complex (see below), and so we have a long way to go before we understand the mechanisms by which the methylation of CpG-poor enhancers is involved in the regulation of these regions.

Methylation at insulators. Insulators can be defined as elements that block the interaction between an enhancer and a promoter. The most well-studied examples are DNA sequences bound by the CTCF protein, which binds to a somewhat heterogeneous sequence motif. A well-studied case is CFCF binding to a site within the imprinted *IGF2–H19* locus, at which the presence or absence of CTCF binding controls enhancer–promoter interactions. It has been shown that methylation of a CTCF-binding site at this locus blocks the binding of CTCF, so DNA methylation has an important role in controlling this locus[79]. More recent studies have likewise shown that CTCF binding to exon 5 of the gene encoding CD45 is inhibited by DNA methylation, leading to effects on splicing[68]. However, global studies in mouse ESCs and differentiated cells have suggested that CTCF binding within CpG-poor regions is generally not affected by the methylation status of the binding sites, but rather that the binding itself initiates local demethylation[70]. Therefore, is it possible that there are no universal rules for the effects of methylation on CTCF sites (which tend to be degenerate) and binding? In this regard, it is important to note that there are seven potential CTCF-binding sites in the human *H19* promoter, and only one of them shows differential parent-of-origin methylation[80].

Possible mechanisms

The mechanisms by which an inactive CGI promoter is held in a stably repressed state by DNA methylation are fairly well understood and have been extensively reviewed[7]. The methylated promoter has nucleosomes at the TSS[81] that bear the repressive H3K9me3 mark and that are stabilized by methylated DNA-binding proteins, which in turn recruit histone deacetylases to the region[82].

The issue of causality of methylation changes in stabilizing inactive gene expression states of non-CGI promoters has been the subject of much controversy, and the issue has not yet been adequately resolved. Because transcription factors can bind strongly to methylated DNA sequences, subsequently resulting in the passive

REVIEWS

Box 3 | DNA methylation and disease

Methylation of cytosine strongly increases the rate of C→T transition mutations and is thought to be responsible for about one-third of all disease-causing mutations in the germline[91]. In somatic cells, gene body methylation is a major cause of cancer gene mutations in tumour suppressor genes, such as *TP53*, which encodes p53 (REF. 53). The transcriptional start sites of many genes encoding tumour suppressors, such as retinoblastoma-associated protein 1 (RB1), MLH1, p16 and BRCA1, among others, lie within CGIs. Factors that reduce expression of these genes increase the likelihood of *de novo* methylation and irreversible silencing. These gene promoters have been found to be extensively methylated in a large number of tumours, such as retinoblastoma, colon, lung and ovarian cancers[51,92,93]. The methylation changes are present in tumours before they are placed into culture but become enhanced during passaging *in vitro*[94]. In general, the tumour genome is hypomethylated, yet the potential role of methylation of CpG-poor promoters, enhancers, insulators, repetitive elements and gene bodies in cancer is almost completely unknown. The relevance of DNA methylation to human cancer has become even more evident with the identification of mutations in DNMT3A[95] and isocitrate dehydrogenase 1 (IDH1) and IDH2 in leukaemias[96]. Several diseases, such as fragile X syndrome[97] and immunodeficiency, centromere instability and facial anomalies syndrome[98], also have a demonstrable link to DNA methylation. The new discoveries discussed in this Review are stimulating efforts to assess more accurately the contribution of DNA methylation to cancer and other human diseases.

---

demethylation of these regions[83], it is not always clear whether the methylation changes are a result of transcription or whether they stabilize transcriptionally incompetent states. Methylation of non-CGI regions can have a direct impact on the binding of transcription factors to target sites. Indeed, it has been known for some time that the binding of MYC to its cognate sequence is directly inhibited by the presence of 5mC[84], whereas the binding of SP1 does not show such a relationship[85]. However, methylation of transcription factor binding sites — for example, in the laminin beta 3 (*LAMB3*), runt-related transcription factor 2 (*RUNX2*)[34] or *Oct4* (REF. 86) promoters — can decrease gene expression in transfection experiments.

More recent genome-wide studies have confirmed that transcription factor binding can be strongly influenced by methylation of CpG sites within their recognition sequences[87]. These experiments point to a cause-and-effect relationship between CpG methylation at the TSS and gene repression, but there are still issues relating to the probable mechanisms. For example, a puzzling observation is our finding that in human ESCs, there is almost always no OCT4 bound at OCT4 target sites when there is DNA methylation within 100 bp on each side of the target sequence[45]. As the OCT4 recognition sequence does not contain a CpG sequence, it is difficult to postulate a mechanism that could explain this strong correlation between CpG methylation in the vicinity of the site and a lack of binding. Nevertheless, the fact that methylation at the CpG sites flanking the OCT4-binding sites is inherently more variable than at CGIs, the neighbourhood could be a fruitful area for investigation in the future and has largely been overlooked until now.

Conclusions

Whole-genome approaches have given us a detailed view of the methylome and have shown that methylation patterns beyond TSSs are far more dynamic than was previously appreciated. Now that we have these global patterns, we need to resolve their potential roles and mechanisms — methylated sites beyond TSSs clearly are not simply 'passengers' as had been previously assumed (FIG. 1). Compared to CGIs at TSSs, which are generally unmethylated and seem mostly to be methylated to ensure long-term silencing, the patterns of modification in the CpG-depleted regions of the genome are more interesting. Although we clearly do not understand the detailed mechanisms by which methylation of enhancers, insulators and gene bodies influence the binding and function of regulatory proteins, there seems to be little doubt that this is crucial to development, differentiation and indeed cellular viability. However, some species are able to survive without methylation, yet mammals require three DNMTs; finding out why this is the case will help to explain its function. The identification of the TET genes and the localization of the TET proteins to regulatory regions is highly suggestive of a dynamic turnover of 5mC, which is consistent with it having a control function in gene expression.

The potential involvement of methylation beyond CGI promoters in human disease has been largely overlooked because of the focus on abnormal CGI methylation in cancer. Future work that aims to provide detailed maps of epigenomes in normal and diseased states is crucial to our understanding of many human diseases (BOX 3). This will be essential if we are to develop strategies and drugs to target the epigenome and to treat these diseases[88].

1. Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science* **187**, 226–232 (1975).
2. Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* **14**, 9–25 (1975).
3. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008). **This was the first paper to provide single-base resolution of DNA methylation genome-wide.**
4. Rountree, M. R. & Selker, E. U. DNA methylation inhibits elongation but not initiation of transcription in *Neurospora crassa. Genes Dev.* **11**, 2383–2395 (1997).
5. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009). **This was the first report of a human methylome at single-base resolution.**
6. Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genet.* **43**, 811–814 (2011).
7. Illingworth, R. S. & Bird, A. P. CpG islands—'a rough guide'. *FEBS Lett.* **583**, 1713–1720 (2009).
8. Takai, D. & Jones, P. A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl Acad. Sci. USA* **99**, 3740–3745 (2002).

9. Moarefi, A. H. & Chedin, F. ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated *de novo* DNA methylation. *J. Mol. Biol.* **409**, 758–772 (2011).
10. Jones, P. A. & Liang, G. Rethinking how DNA methylation patterns are maintained. *Nature Rev. Genet.* **10**, 805–811 (2009).
11. Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042–1047 (2010).
12. Ooi, S. K. *et al.* DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448**, 714–717 (2007).
    This paper provided a structural basis to the mechanisms of *de novo* methylation and showed how active histone marks could exclude methylation of DNA.
13. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
14. Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell* **99**, 247–257 (1999).
    This was a key paper in defining the need for DNA cytosine methylation in mammals.
15. Jackson-Grusby, L. *et al.* Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature Genet.* **27**, 31–39 (2001).
16. Chen, T. *et al.* Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nature Genet.* **39**, 391–396 (2007).
17. Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805–814 (2006).
18. Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nature Genet.* **44**, 23–31 (2011).
19. Ooi, S. K. & Bestor, T. H. The colorful history of active DNA demethylation. *Cell* **133**, 1145–1148 (2008).
20. Wu, H. & Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* **25**, 2436–2452 (2011).
21. Branco, M. R., Ficz, G. & Reik, W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nature Rev. Genet.* **13**, 7–13 (2012).
22. Popp, C. *et al.* Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105 (2010).
23. Inoue, A. & Zhang, Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334**, 194 (2011).
24. Iqbal, K., Jin, S. G., Pfeifer, G. P. & Szabo, P. E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl Acad. Sci. USA* **108**, 3642–3647 (2011).
25. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
26. Cortazar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423 (2011).
27. Gu, T. P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).
28. Wossidlo, M. *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature Commun.* **2**, 241 (2011).
29. Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome — biological and translational implications. *Nature Rev. Cancer* **11**, 726–734 (2011).
30. Kelly, T. K. *et al.* H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. *Mol. Cell* **39**, 901–911 (2010).
31. Gal-Yam, E. N. *et al.* Constitutive nucleosome depletion and ordered factor assembly at the GRP78 promoter revealed by single molecule footprinting. *PLoS Genet.* **2**, e160 (2006).
32. Taberlay, P. C. *et al.* Polycomb-repressed genes have permissive enhancers that initiate reprogramming. *Cell* **147**, 1283–1294 (2011).
33. Farthing, C. R. *et al.* Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* **4**, e1000116 (2008).
34. Han, H. *et al.* DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Hum. Mol. Genet.* **20**, 4299–4310 (2011).
35. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genet.* **37**, 853–862 (2005).
36. Gal-Yam, E. N. *et al.* Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc. Natl Acad. Sci. USA* **105**, 12979–12984 (2008).
37. Hashimshony, T., Zhang, J., Keshet, I., Bustin, M. & Cedar, H. The role of DNA methylation in setting up chromatin structure during development. *Nature Genet.* **34**, 187–192 (2003).
38. Kass, S. U., Landsberger, N. & Wolffe, A. P. DNA methylation directs a time-dependent repression of transcription initiation. *Curr. Biol.* **7**, 157–165 (1997).
39. Venolia, L. & Gartler, S. M. Comparison of transformation efficiency of human active and inactive X-chromosomal DNA. *Nature* **302**, 82–83 (1983).
    This is a key paper that unequivocally established that the covalent application of methyl groups to DNA could result in silencing and is involved in X-chromosome inactivation.
40. Lock, L. F., Takagi, N. & Martin, G. R. Methylation of the *Hprt* gene on the inactive X occurs after chromosome inactivation. *Cell* **48**, 39–46 (1987).
    This paper unexpectedly showed that methylation of cytosine was not the primary silencing mechanism for X inactivation.
41. Ohm, J. E. *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature Genet.* **39**, 237–242 (2007).
42. Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nature Genet.* **39**, 232–236 (2007).
43. Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nature Genet.* **39**, 157–158 (2007).
44. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genet.* **41**, 178–186 (2009).
45. You, J. S. *et al.* OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. *Proc. Natl Acad. Sci. USA* **108**, 14497–14502 (2011).
46. Zilberman, D., Coleman-Derr, D., Ballinger, T. & Henikoff, S. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456**, 125–129 (2008).
    This paper showed the importance of histone variants in relation to DNA methylation. Previously, most of the focus was on histone modification.
47. Conerly, M. L. *et al.* Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. *Genome Res.* **20**, 1383–1390 (2010).
48. Thomson, J. P. *et al.* CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082–1086 (2010).
49. Williams, K., Christensen, J. & Helin, K. DNA methylation: TET proteins—guardians of CpG islands? *EMBO Rep.* **13**, 28–35 (2011).
50. Jones, P. A. *et al.* *De novo* methylation of the MyoD1 CpG island during the establishment of immortal cell lines. *Proc. Natl Acad. Sci. USA* **87**, 6117–6121 (1990).
51. Hitchins, M. P. *et al.* Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5′UTR. *Cancer Cell* **20**, 200–213 (2011).
    This study demonstrated that single-nucleotide variants that decrease promoter activity can lead to preferential allele-specific methylation.
52. Boumber, Y. A. *et al.* An Sp1/Sp3 binding polymorphism confers methylation protection. *PLoS Genet.* **4**, e1000162 (2008).
53. Rideout, W. M., I. I. I., Coetzee, G. A., Olumi, A. F. & Jones, P. A. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science* **249**, 1288–1290 (1990).
54. Jones, P. A. The DNA methylation paradox. *Trends Genet.* **15**, 34–37 (1999).
55. Illingworth, R. S. *et al.* Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* **6**, e1001134 (2010).
56. Wolf, S. F., Jolly, D. J., Lunnen, K. D., Friedmann, T. & Migeon, B. R. Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. *Proc. Natl Acad. Sci. USA* **81**, 2806–2810 (1984).
57. Hellman, A. & Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141–1143 (2007).
58. Feng, S. *et al.* Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl Acad. Sci. USA* **107**, 8689–8694 (2010).
59. Larsen, F., Solheim, J. & Prydz, H. A methylated CpG island 3′ in the apolipoprotein-E gene does not repress its transcription. *Hum. Mol. Genet.* **2**, 775–780 (1993).
60. Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253–257 (2010).
61. Nguyen, C. *et al.* Susceptibility of nonpromoter CpG islands to *de novo* methylation in normal and neoplastic cells. *J. Natl Cancer Inst.* **93**, 1465–1472 (2001).
62. Nguyen, C. T., Gonzales, F. A. & Jones, P. A. Altered chromatin structure associated with methylation-induced gene silencing in cancer cells: correlation of accessibility, methylation, MeCP2 binding and acetylation. *Nucleic Acids Res.* **29**, 4598–4606 (2001).
63. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
    This paper pointed out the crucial role of 5mC in suppressing the transcription of transposable elements.
64. Hahn, M. A., Wu, X., Li, A. X., Hahn, T. & Pfeifer, G. P. Relationship between gene body DNA methylation and intragenic H3K9me3 and H3K36me3 chromatin marks. *PLoS ONE* **6**, e18844 (2011).
65. Laurent, L. *et al.* Dynamic changes in the human methylome during differentiation. *Genome Res.* **20**, 320–331 (2010).
66. Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon-intron structure. *Nature Struct. Mol. Biol.* **16**, 990–995 (2009).
67. Chodavarapu, R. K. *et al.* Relationship between nucleosome positioning and DNA methylation. *Nature* **466**, 388–392 (2010).
68. Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).
69. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
70. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).
71. Schmidl, C. *et al.* Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res.* **19**, 1165–1174 (2009).
72. Wiench, M. *et al.* DNA methylation status predicts cell type-specific enhancer activity. *EMBO J.* **30**, 3028–3039 (2011).
73. Saluz, H. P., Jiricny, J. & Jost, J. P. Genomic sequencing reveals a positive correlation between the kinetics of strand-specific DNA demethylation of the overlapping estradiol/glucocorticoid-receptor binding sites and the rate of avian vitellogenin mRNA synthesis. *Proc. Natl Acad. Sci. USA* **83**, 7167–7171 (1986).
74. Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. & Jacobsen, S. E. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* **12**, R54 (2011).
75. Szulwach, K. E. *et al.* Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* **7**, e1002154 (2011).
76. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
77. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).
78. Wu, S. C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nature Rev. Mol. Cell. Biol.* **11**, 607–620 (2010).
79. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**, 482–485 (2000).
    This was a key paper showing how methylation of CTCF sites could alter insulator function by directly blocking binding of CTCF.

REVIEWS

80. Takai, D., Gonzales, F. A., Tsai, Y. C., Thayer, M. J. & Jones, P. A. Large scale mapping of methylcytosines in CTCF-binding sites in the human H19 promoter and aberrant hypomethylation in human bladder cancer. *Hum. Mol. Genet.* **10**, 2619–2626 (2001).

81. Lin, J. C. *et al.* Role of nucleosomal occupancy in the epigenetic silencing of the MLH1 CpG island. *Cancer Cell* **12**, 432–444 (2007).

82. Wade, P. A. & Wolffe, A. P. ReCoGnizing methylated DNA. *Nature Struct. Biol.* **8**, 575–577 (2001).

83. Hsieh, C. L. Dynamics of DNA methylation pattern. *Curr. Opin. Genet. Dev.* **10**, 224–228 (2000).

84. Prendergast, G. C. & Ziff, E. B. Methylation-sensitive sequence-specific DNA binding by the c-Myc basic region. *Science* **251**, 186–189 (1991).

85. Harrington, M. A., Jones, P. A., Imagawa, M. & Karin, M. Cytosine methylation does not affect binding of transcription factor Sp1. *Proc. Natl Acad. Sci. USA* **85**, 2066–2070 (1988).

86. Simonsson, S. & Gurdon, J. DNA demethylation is necessary for the epigenetic reprogramming of somatic cell nuclei. *Nature Cell Biol.* **6**, 984–990 (2004).

87. Chen, P. Y., Feng, S., Joo, J. W., Jacobsen, S. E. & Pellegrini, M. A comparative analysis of DNA methylation across human embryonic stem cell lines. *Genome Biol.* **12**, R62 (2011).

88. American Association for Cancer Research Human Epigenome Task Force & European Union, Network of Excellence & Scientific Advisory Board. Moving AHEAD with an international human epigenome project. *Nature* **454**, 711–715 (2008).

89. Harris, R. A. *et al.* Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nature Biotech.* **28**, 1097–1105 (2010).
   This is a good criticial review of sequencing-based methods for studying 5mC patterns.

90. Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* **5**, e8888 (2010).

91. Cooper, D. N. & Youssoufian, H. The CpG dinucleotide and human genetic disease. *Hum. Genet.* **78**, 151–155 (1988).
   This paper clearly highlighted the important role of 5mC in generating disease-causing mutations.

92. The Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**, 609–615 (2011).

93. Stirzaker, C. *et al.* Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. *Cancer Res.* **57**, 2229–2237 (1997).

94. Markl, I. D. *et al.* Global and gene-specific epigenetic patterns in human bladder cancer genomes are relatively stable *in vivo* and *in vitro* over time. *Cancer Res.* **61**, 5875–5884 (2001).

95. Ley, T. J. *et al.* DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363**, 2424–2433 (2010).

96. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).

97. Bell, M. V. *et al.* Physical mapping across the fragile X: hypermethylation and clinical expression of the fragile X syndrome. *Cell* **64**, 861–866 (1991).

98. Xu, G. L. *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187–191 (1999).

99. Sharma, S., De Carvalho, D. D., Jeong, S., Jones, P. A. & Liang, G. Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. *PLoS Genet.* **7**, e1001286 (2011).

100. Ong, C. T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Rev. Genet.* **12**, 283–293 (2011).