Review

Exposing the DNA methylome iceberg

'Matladi N. Ndlovu*, Hélène Denis* and François Fuks

Laboratory of Cancer Epigenetics, Faculty of Medicine, Université Libre de Bruxelles, 808 route de Lennik, 1070 Brussels, Belgium

DNA methylation was the first epigenetic modification discovered. Until recently, comprehensive coverage of the composition and distribution of methylated cytosines across the genome was lacking. Technological advances, however, are providing methylation maps that can reveal the genomic distribution of DNA methylation in different cell states or phenotypes. The emerging picture includes extensive gene body methylation that is highly conserved in eukaryotes, the presence of DNA methylation in previously unappreciated sequence contexts, and the discovery of another modified DNA base, 5-hydroxymethylcytosine. These new data point to the role of DNA methylation both in gene silencing and gene activation; reconciliation of these seemingly contradictory roles will be essential to fully unravel the biological function of DNA methylation in eukaryotes. Here we review how these recently exposed features of the DNA methylome are challenging previously held dogmas in the field.

DNA methylation: a brief background

The genetic information in a cell is encoded in DNA, which is tightly packaged around an octamer of eight core histones within the nucleus, forming a dynamic structure known as chromatin [1]. In eukaryotic organisms including plants, animals and fungi, chromatin is subjected to enzymatic modifications at both the DNA and protein levels, which impart an additional layer of heritable information on the DNA code that is important for regulating proper gene expression. For instance, correct deployment of developmental programs and the maintenance of cell fates rely on precise and timely activation or repression of gene expression. It is now evident that multiple dynamic modifications, commonly known as epigenetic modifications, are involved in modulating chromatin structure such that it becomes suitable for gene transcription or silencing. Such epigenetic mechanisms include DNA methylation, post-translational histone modifications, and chromatin remodelling [2]. Notably, numerous human malignancies exhibit alterations in epigenetic processes, and aberrant DNA methylation is currently the best-characterised epigenetic hallmark for a number of pathologies including cancer; cancer cells are characterised by both global and gene-specific loss of DNA methylation, as well as hypermethylation of specific promoters [3–5]. DNA methylation, therefore, is vital for gene expression during normal development and in disease aetiology. It plays a crucial role in genomic imprinting, X-chromosome inactivation, genome stability, and the long-term silencing of retrotransposons and repetitive elements [6]. The heritable covalent addition of a methyl group at the fifth carbon of cytosine residues in DNA ($5^{me}C$), referred to as DNA methylation, is catalysed by DNA methyltransferases (DNMTs) primarily in the sequence context of CpG dinucleotides where it impacts gene expression and activity in a heritable manner [7]. In vertebrates, 70–80% of CpG dinucleotides in the genome are methylated, with the exception of short conspicuous clusters of CpG dinucleotides, called CpG islands (CGIs), which are at least 200 bp long and have a high GC content. These regions of unmethylated CpG dinucleotides are generally characterised by transcriptionally permissive chromatin and punctuate the genome at transcription start sites and at approximately 70% of gene promoters [8,9]. The mechanisms that establish and maintain unmethylated CGIs remain largely unknown, although DNMTs, histone modifications, transcription factors and chromatin-modifying complexes are known to be involved [8–12]. Although the majority of promoters are associated with unmethylated CGIs, some methylated tissue-specific CGI promoters have been identified during early development and in differentiated tissues. Overall, CGI promoter methylation is associated with silencing of gene expression; examples include X-chromosome inactivation, genomic imprinting and germline-specific methylation. Gene silencing by DNA methylation is achieved through different mechanisms: (i) the presence of $5^{me}C$ can directly inhibit the specific binding of transcription factors or (ii) methyl-binding domain proteins (MBDs) directly recognise methylated DNA and recruit repressive chromatin-modifying complexes [13]. Despite over 50 years of research, our knowledge of the scope of the functional importance of DNA methylation in eukaryotic systems has repeatedly been described as being only the tip of the iceberg. Many questions remain regarding how DNA methylation regulates gene expression, ultimately influencing both normal cellular differentiation and disease aetiology. Ingenious technologies developed in recent years have triggered a cascade of new insights into the secrets of DNA methylation in gene regulation. Before the development of these high-throughput technologies, DNA methylation could only be determined for specific loci either via assays that use methyl-sensitive restriction enzymes or via chemical deamination with sodium bisulfite followed by Sanger sequencing [14]. These methods have been adapted to enable genome-scale DNA methylation analysis by combining microarrays or high-throughput DNA sequencing with: (i) enrichment of methylated genomic DNA fragments using either antibodies specific for $5^{me}C$ or methyl-binding proteins (e.g. MeDIP-seq and MethylCap-seq) [15–18]; (ii) bisulfite-based methods that selectively deaminate unmethylated cytosine but not $5^{me}C$, thereby promoting

Corresponding author: Fuks, F. (ffuks@ulb.ac.be)
* These authors contributed equally to this work.

  doi:10.1016/j.tibs.2011.03.002

discrimination of the methylated cytosine (e.g. Infinium assay and reduced representation bisulfite sequencing, RRBS) [6,19]; or (iii) digestion with methylation-sensitive restriction enzymes for fractionation of only methylated or unmethylated DNA (e.g. comprehensive high-throughput arrays for relative methylation, CHARM; and HpaII tiny fragment enrichment by ligation-mediated PCR, HELP) [20–23]. Although each of these methods can accurately translate DNA methylation measurements, each has different strengths and weaknesses with regard to genome coverage, resolution, reproducibility, sensitivity and cost [24,25].

In this review we highlight the huge leaps made towards unravelling the DNA methylation landscape, with a focus on several surprising recent discoveries. Notably, the emerging picture of the highly conserved yet diverse genomic distribution of eukaryotic DNA methylation in which intragenic DNA methylation plays a major role in mammalian gene regulation is discussed [26,27]. These findings challenge the previously held view that DNA methylation regulates mammalian gene expression primarily through 5′ promoters. In addition, identification of the presence of non-CpG methylation in mammals and the discovery of 5-hydroxymethylcytosine have been surprising and point to the multiple levels through which DNA methylation regulates gene expression [28–30]. This is in stark contrast to the long-held notion that CpGs are the only known methylated dinucleotides in mammals.

The initial explosion of DNA methylome data

Completion of the Human Genome Project provided a comprehensive platform of genomic maps and catalogues that permitted the accurate inference of gene structure and detection of changes across the genome [31]. Indeed, the genome proved to be far more complex than expected. Consequently, the study of distinctive functional domains associated with specific epigenetic patterns in the genome was recognised as being of equal importance [31]. High-throughput, next-generation deep sequencing technologies have facilitated in-depth mapping of the global distribution of DNA methylation and chromatin modifications at single-base resolution throughout the genome, thus allowing precise measurement of their extent and location at specific loci [25,32]. Notably, the combination of powerful modern genomics with the study of the complete set of DNA methylation modifications of a cell, coined the DNA methylome [33], has provided an important platform from which to extract DNA methylation information from complex genomes. The precision afforded by these latest technological advances has revealed surprising new insights into the dynamics of DNA methylation and, most importantly, has led to an explosion of data that can ultimately be layered on top of genome sequences to provide a more descriptive genomic physiology for distinct cell types. First, the presence of extensive $5^{me}C$ in a non-CpG sequence context, as confirmed by bisulfite ultra-deep sequencing methods [29,30,34], has reinvigorated the field. These studies have shown that

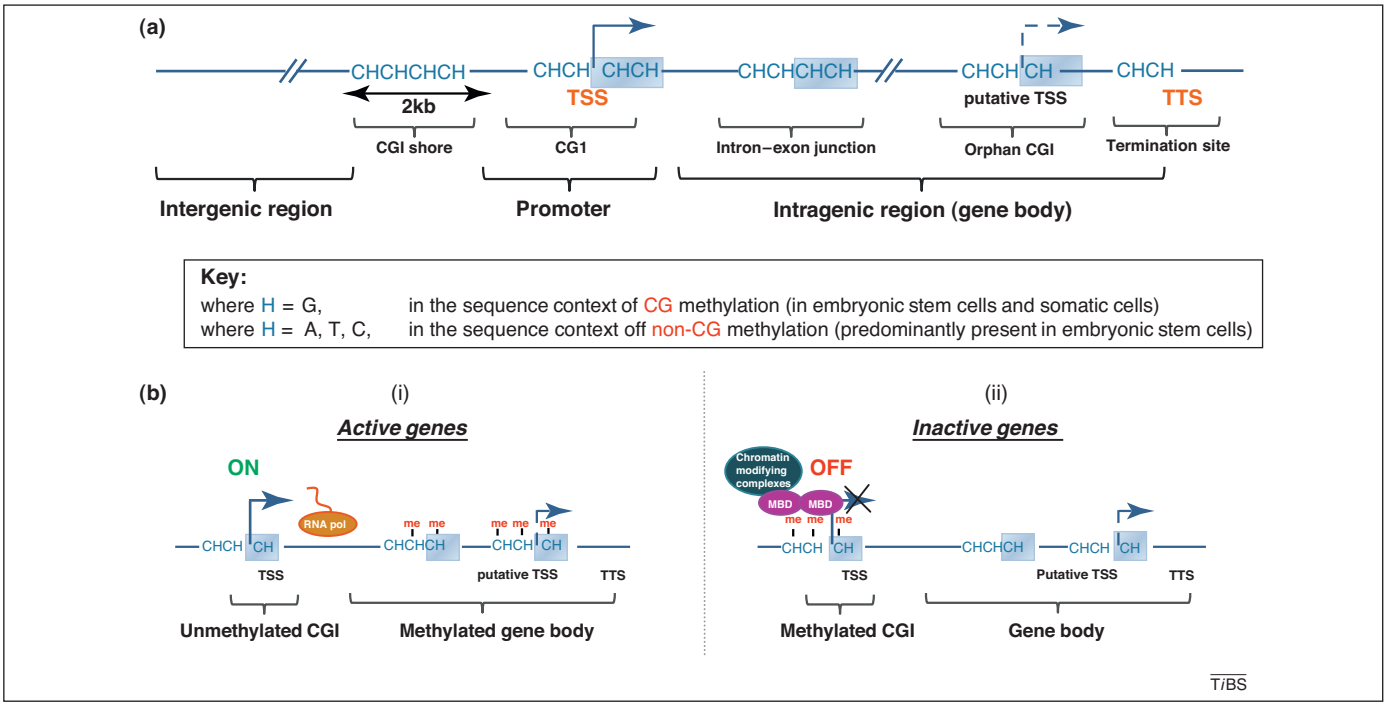

Figure 1. Schematic representation of the mammalian DNA methylome. The average DNA methylation distribution is mapped to a gene model indicating distinct DNA methylation patterns. The relationship of DNA methylation to transcription at the promoter and gene body has been broadly corroborated and is also highlighted. **(a)** Gene model indicating annotated genomic features in the neighbourhood of genes including a promoter, intragenic region (gene body) and intergenic region. TSS denotes transcriptional start site and TTS, transcriptional termination site. To delineate regions with subtle methylation features: CpG island (CGI) promoters, CGI shores, which are up to 2 kb away from a promoter, orphan CGIs and intron–exon junctions are represented. CGI shores are characterised mainly by differential methylation in most tissues, and thus lead to differential gene expression. **(b)** Predominant impact of DNA methylation on gene transcription; 'me' denotes methylation, ON indicates gene expression and OFF indicates gene repression. (i) In active genes, CGI promoters are normally unmethylated, which facilitates transcription. Methylation at the gene body is positively correlated with gene expression, whereby it might prevent spurious transcription initiation. (ii) In inactive genes, CGI promoter methylation is primarily associated with gene repression through the recruitment of methyl-binding domain (MBD) proteins and associated chromatin complexes; the intensity of transcriptional repression is determined by the overall promoter CG content. The remaining challenge will be to expose the different tissue- and cell-type sequence contexts and the methylation machinery responsible for these differences.

Box 1. 5-Hydroxymethylcytosine, a DNA modification closely related to 5-methylcytosine

The traditional view of methylated cytosine being the only modified DNA residue in mammalian genomes has been challenged by recent confirmation of the presence of 5-hydroxymethylcytosine [36,37]. Intriguingly, the presence of $5^{hm}C$ in mammalian DNA was first described approximately 40 years ago [70], but because of technical limitations the results could not be corroborated until recently with the use of high-precision mass spectrometry [36,37]. The ten–eleven translocation (TET) family of enzymes catalyses the oxidation of $5^{me}C$ (in CG dinucleotides) to $5^{hm}C$, and thus fuels the hypothesis that this mechanism might regulate gene expression and/or represent an intermediate form of an active DNA demethylation process [71]. To date, $5^{hm}C$ has only been implicated in passive DNA demethylation through interference with binding of MBD proteins and their associated chromatin-modifying complexes [72]. Owing to the great similarities between $5^{hm}C$ and $5^{me}C$, the widely used bisulfite-based technologies have been incapable of distinguishing these two modifications and consequently no functional distinctions have been made [73,74]. Therefore, there is great need for sensitive and unbiased whole-genome approaches that can distinguish between these two closely related modifications. New techniques and technologies that can specifically detect $5^{hm}C$ are currently being developed, and some have already provided new insights into $5^{hm}C$ genomic distribution [35,75]. For instance, the newly described direct detection technology through single-molecule real-time (SMRT) sequencing has predicted that $5^{hm}C$ might also be present in non-CpG contexts [35]. The jury is still out on the function of $5^{hm}C$; however, the observation that $5^{hm}C$ seems to be present at physiologically relevant levels in a tissue-specific manner implies that this modification could potentially have great biological significance.

methylation of CpHpG and CpHpH trinucleotides (H=A, T, C), although predominant in human embryonic stem cells (hESCs), is also present in differentiated cells (Figure 1a). Specifically, of the 20% of methylated cytosines in a non-CpG context in hESCs, 5^{me}CpA was the most prevalent, with a similar profile to that of 5^{me}CpG across the genome [29]. In addition, the methylation profile for 5^{me}CpA was conserved in differentiated cells and included the same sites as in hESCs. These data raise intriguing questions regarding the functional significance of non-CpG DNA methylation and how the DNA methylation machinery is directed to these specific sequences. The challenge, therefore, is to determine both the function of non-CpG methylation in gene regulation and the principal components of the methylation machinery that are essential for establishing these complex DNA methylation patterns.

A second confounding variable in DNA modifications of the genome is the identification of a novel DNA nucleotide, 5-hydroxymethylcytosine ($5^{hm}C$), which is generated from $5^{me}C$ within a CpG context (Box 1) [35–37]. Collectively, DNA methylation maps have undoubtedly highlighted the crucial interplay of dynamic mechanisms that regulate DNA methylation at CpG and non-CpG sequence contexts and through hydroxymethylation; however, technical limitations must still be overcome to fully understand these interactions [38,39].

Gene-body DNA methylation: an ancient eukaryotic genomic landmark

One of the most exciting and revolutionary new discoveries in the field relates to the abundant levels of DNA methylation found within the bodies of active genes in protein-coding regions. Although genome-wide methylation of gene bodies was first described in *Arabidopsis thaliana,* where it was associated with active genes [38], there had been previous reports of CpG methylation downstream of a transcription site, albeit at a single locus, in mammalian cells, where it did not silence gene expression [40]. Strikingly, genome-scale methylation profiles based on bisulfite treatment and methylation-sensitive technologies have shown, with few exceptions, that DNA methylation at gene bodies is positively correlated with elevated expression [20,26,29,38,41–45]. It is noteworthy that the extent of DNA methylation varies significantly in different regions of the gene body (Figure 1a). Specifically, the use of next-generation bisulfite sequencing technology to closely examine methylation levels at intragenic regions revealed high levels of methylation throughout most of the gene body, with a characteristic sharp decrease and increase in methylation levels across exon–intron junctions and at transcription termination sites; only subtle fluctuations have been detected in intergenic regions [29,30]. These observations suggest a putative functional role for DNA methylation in tissue- and cell type-specific regulation via differential mRNA splicing and the use of alternative promoters, for instance at transcribable regions on the active human X chromosome [26,41]. Unbiased next-generation bisulfite sequencing at single-base-pair resolution has further shown that the newly discovered non-CpG methylation is also present in gene bodies in both human ESCs and differentiated tissues such as brain [29]. Furthermore, experiments using MeDIP-seq and MRE-seq technologies, which map the methylated and unmethylated fraction of the genome, respectively, showed that 34% of CGIs in intragenic regions were methylated. This is a surprising finding considering that only a small fraction of CGIs at gene promoters (2%) are methylated [26]. Taken together, these data suggest that gene body methylation plays a major role in regulating context-specific gene expression.

Interestingly, genome-wide methylation profiles across 20 eukaryotic genomes have further revealed that gene-body methylation is an ancestral form of DNA methylation predating the divergence of plants and animals approximately 1.6 billion years ago [28,46]. These reports uncovered important conserved and divergent features of DNA methylation, in particular that gene-body methylation is a general phenomenon that is highly conserved and extensive in diverse organisms. The fact that these features have been 'conserved', despite the nearly 1.6 billion years since the plant and animal lineages diverged, is said to be highly indicative of gene-body methylation being crucial for the survival of organisms [26,28,46]. Intriguingly, DNA methyltransferase 1 (DNMT1) and the methylation binding co-factor ubiquitin-like containing PHD and ring finger domains (UHRF1) share a conserved function in maintaining gene-body methylation in various organisms [28]. Collectively, mapping experiments have shown that DNA methylation seems to be a widespread epigenetic regulatory mechanism. However, several questions remain regarding how gene-body DNA methylation is targeted to specific gene regions and the function(s) of this conserved genic methylation.

CpG islands: a distinctive feature in vertebrate genomes

The bulk of the human genome is depleted of CpG dinucleotides owing to the high mutation rate of methylated cytosine [47]. It is believed that CGIs are specifically protected from *de novo* DNA methylation based on their location within the genome, the local CpG density and their association with chromatin-modifying complexes [10,48]. Therefore, a relatively high density of CpG dinucleotides and elevated GC content characterise mammalian CGIs [8]. To gain further insight into CpG dinucleotide distribution at specific genomic loci, methylation profiling by MeDIP, combined with microarray technology and supplemented by a stringent new algorithm, has been carried out on the promoter regions of 24 134 human genes [48,49]. Strikingly, the pattern of CpGs across the vertebrate genome revealed the presence of a characteristic bimodal distribution of CpG content in promoters; this bimodal distribution was found to be indicative of distinct populations of high, intermediate and low CpG-content promoters [48–50]. The substantial overlap between the high and low CpG-content populations of promoters, corresponding to promoters with an intermediate CpG frequency, was later identified as a dynamic region mediating soma-specific methylation [48]. Consequently, promoters have been categorised into three classes based on their CpG ratio, GC content and the length of the CpG-rich region [48]. Whereas the high CpG-content (HCP) promoter class is linked to CGIs and is predominantly hypomethylated, the low CpG-content (LCP) promoter class, representing non-CGI promoters, is primarily methylated. The intermediate CpG-content (ICP) promoter class falls short of the accepted CGI standard criteria, namely CpG islands have an average ratio of observed to expected CpG dinucleotides of 0.5 [51] and thus were identified as weak CGI promoters [48]. Correlation of gene activity and DNA methylation revealed that the frequency of promoter activity varied between promoter classes depending on the CpG content, with HCPs and ICPs more prone to differential regulation by DNA methylation than LCPs [48]. Intriguingly, earlier reports using *in vitro* assays had suggested that methylation-mediated gene regulation is a function of the density of methylated CpGs, their location relative to the promoter and promoter strength (Figure 1b) [52]. Interestingly, although DNA methylation is sufficient to inactivate strong CGI promoters (HCPs), the vast majority of the HCP promoter class remains methylation-free, regardless of the expression of associated genes [10,48,53]. Indeed, the fact that transcription does not seem to be a requirement for the conspicuous absence of methylation on HCP promoters [48] illustrates the complexity involved in determining the functional impact of promoter CGI methylation on transcriptional control. An emerging body of evidence supports the possibility that promoters associated with CGIs, namely HCPs and ICPs, mainly mediate transcription through chromatin architecture and act as platforms for additional levels of epigenetic and transcriptional control [54–57]. Taken together, these reports challenge the long-standing view that DNA methylation at promoter CGIs primarily defines transcription start sites of gene promoters as being silent.

To further delineate the role of CGIs in vertebrates on a genome-wide scale, a comprehensive map of all CGIs in the human and mouse genomes was compiled using CXXC Affinity Purification-seq (CAP-seq) enrichment of unmethylated CpG fragments [58]. Notably, this method identified nearly equivalent numbers of CGIs in the human and mouse genomes, contrary to computational predictions. These experiments showed that both human and mouse genomes have approximately 25 000 CGIs, and that approximately 60% of gene promoters are associated with these CGIs. However, more than half of all CGIs in both genomes, provisionally named orphan CGIs because they are remote from annotated promoters, are evenly embedded within genes (i.e. in the gene body) and between coding regions; of these, approximately 42% are associated with sites of transcriptional initiation (Figure 1a) [58]. Interestingly, these orphan CGIs are associated with novel transcripts and could represent promoters of non-annotated genes or non-coding RNAs [27]. Collectively, these data suggest that the majority of DNA methylation-specific, tissue-restricted gene expression is mediated primarily by orphan CGIs [58] or intragenic methylation [26,29]. The recent identification of ZF-CxxC domain-containing proteins, including CXXC finger protein 1 (CFP1) and lysine (K)-specific demethylase 2A (KDM2A), that specifically bind unmethylated CGIs and are components of histone-modifying complexes, clearly demonstrates that CGIs intrinsically influence the local chromatin modification state [55,59]. This is the first time it was shown that unmethylated CGIs are epigenetically differentiated from the bulk of chromatin and non-CGI promoters [55,59]. From an evolutionary standpoint, it can be deduced from these initial reports that ZF-CxxC domain recognition of CGIs could have contributed to the maintenance of DNA methylation-free zones within the unevenly distributed CGIs within a genome that is globally depleted of CpG dinucleotides and heavily methylated [55,59].

Recent studies indicate that 76% of tissue-specific differential methylation is located in regions adjacent to CGIs, denoted CpG Island shores (CGI shores) and defined as regions within 2000 bp of the promoter but outside of a CGI (Figure 1a) [42,60]. For example, between haematopoetic populations, CGI shores exhibit the greatest variability in DNA methylation and this variability is correlated with changes in gene expression [61]. Indeed, many questions remain regarding details of how the extent and specific distribution of $5^{me}C$ residues upstream of promoters affect their activity. Interestingly, differentially methylated CGI shores that distinguish specific tissues are conserved between human and mouse genomic regions [42]. These results point to a need to study beyond the promoter to identify the biological influence of CGI methylation on gene regulation; they also allude to a broader network of mechanisms involved in the regulation of DNA methylation during development and pathologies.

Taken together, recent state-of-the art reports have undoubtedly altered our perceptions of how DNA methylation regulates mammalian cells. These new genome mapping reports on the previously unknown distribution and location of DNA methylation have earmarked DNA methylation for more research on its biological functions. In addition, there are still some key regions of the DNA methylome that remain indeterminate and obscure

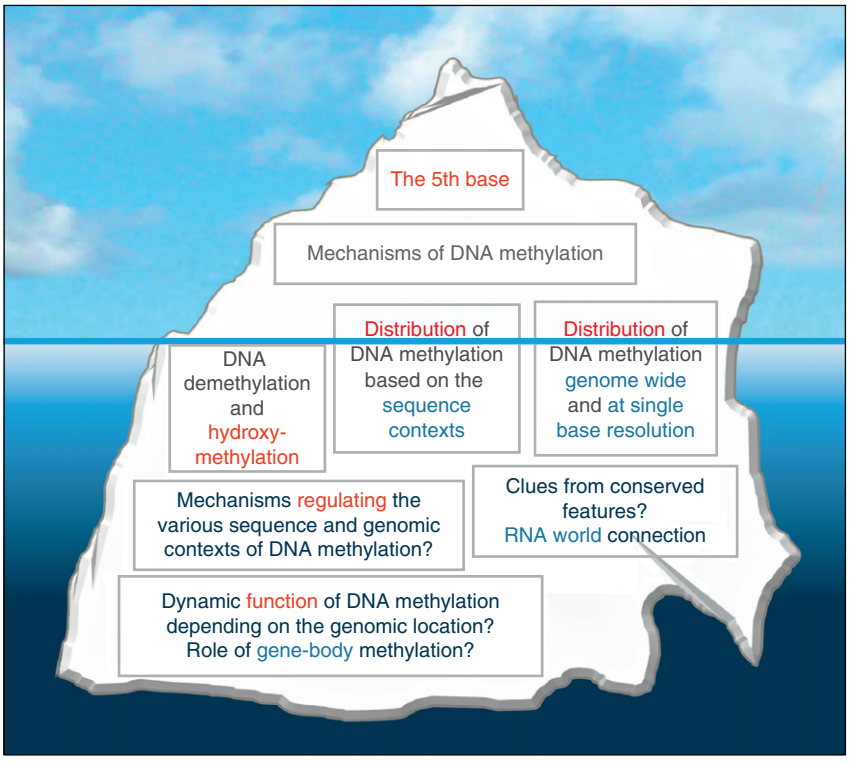

Figure 2. Exposing the DNA methylome iceberg. Next-generation technologies are providing a glimpse into the emerging landscape of the DNA methylome. This figure depicts the ways in which 5-methyl deoxycytidine, the fifth base, has emerged as an important nucleotide in DNA. DNA methylation has been studied in the context of many fields of research, including developmental and cancer biology. Since the scientific community was alerted to its presence in *Tubercle bacillus* and calf thymus DNA in 1925 and 1948, respectively, there have been intensive biochemical studies related to DNA methylation mechanisms. Although much information has been amassed regarding the key enzymes and mechanisms that methylate DNA, numerous questions regarding how DNA methylation is related to different cell phenotypes and gene regulation remain to be answered. Indeed, next-generation technologies have begun to unravel the impact/extent of DNA methylation by revealing new sequence contexts and an unexpected genomic distribution of $5^{me}C$. Subtle yet distinct features of DNA methylation have also been uncovered. The fact that some of these recently revealed features are conserved across eukaryotes could provide hints to their function and establishment. Subsequent experiments that ask how non-CG and CG methylation at distinct genomic locations is interconnected should reveal new insights into the dynamic function in complex biological processes.

(Figure 2) and require urgent attention for a truly comprehensive understanding of the functional methylome.

Concluding remarks and future perspectives

We have highlighted how the explosion of the DNA methylome data has revealed intricate details of the DNA methylation landscape. Recent advances in genome-wide DNA methylation analysis have enabled unprecedented resolution and identified subtle features in DNA methylation [24,32]. Distinctive fluctuations at annotated genomic features (Figure 1a), which could be crucial for fully unravelling the key role of DNA methylation in normal and diseased states, have been identified [26,29,30,34,37,43,44,55,58,59]. From these genome-wide studies emerges a picture of highly conserved features of DNA methylation among eukaryotes, especially regarding CGI location and gene-body methylation, thus emphasising the urgent need for in-depth analyses of the machinery involved in targeting DNA methylation to specific genomic regions. Furthermore, the fact that tissue-specific DNA methylation changes frequently occur at weak CGI promoters in both gene coding and intergenic regions, which coincide with DNA sequences that are highly conserved between mouse and human genomes, raises the possibility that they reflect the sequence characteristics of the larger DNA domains of which they are part (Figure 2) [62,63].

Given that developmental stages are reflected in both the extent of non-CpG methylation and the level of global methylation [29,30], it will be crucial to determine the functional significance of specific variations in the methylation of non-CG dinucleotides during development and differentiation. There is also an intriguing possibility of cross-talk between CG and non-CG methylation as a direct or indirect effect of their interaction with the methylation machinery.

Although the overall picture of the DNA methylome is still under construction, a milestone has undoubtedly been reached: single-base-resolution global DNA methylation maps are divulging context-specific information for essentially all cell types during development, in disease and under different environmental conditions [64]. Given the complexity involved in cellular regulation and the fact that epigenetic changes are inherently more plastic and dynamic than genetic changes, mapping of DNA methylation and other chromatin features (e.g. histone modifications and noncoding RNAs) simultaneously in the same cell types will be essential to gain a better understanding of the cross-talk among epigenetic regulatory mechanisms. With respect to health and disease, large-scale epigenome maps of different cell types in relation to different environmental exposure, disease states and individual genomic variations will also be essential [65]. The total volume of distinct

epignomes that could be analysed is astronomical; therefore, to maximise the exploration of epigenomic features in a wide array of cell types and diseases while minimising duplication, major concerted actions that integrate different disciplines, resources and initiatives are recommended [31,66]. There are some concerns that the larger-scale epigenome projects divert funds from investigator-initiated grants and ignore the role of *cis*-regulatory elements and transcriptional networks in controlling epigenetic changes [67]. It seems, however, that the potential scientific benefit of large-scale epigenomics projects in eventually overlaying hundreds of thousands of epigenomic marks on top of genome sequences could provide a launching pad for individual investigations aimed at identifying the transcription factors, regulatory genomic elements and pathways that integrate diverse signals and modulate epigenomic features [66,68]. In addition, epigenome maps could potentially enable individual investigators to identify genes with similar patterns of epigenetic features that might suggest coordinated regulation of gene expression in particular cell types [69]. Therefore, large-scale epigenomics projects have the potential to enable the creation of comprehensive reference maps and provide a framework to integrate genome-wide data from genome profiling and gene expression studies with epigenetic mechanisms in order to unravel factors underlying disease development as well as different or complex phenotypes [6,25,45]. Synergy between large-scale community projects, interdisciplinary collaborations and individual investigations will be crucial in the next decade to modify the current dogma that DNA methylation functions predominantly at specific sequences and genomic regions to irreversibly silence transcription.

Collectively, these new data call for intensive research efforts devoted to a reappraisal of the entrenched dogma on the role of DNA methylation in mammalian systems. Changes to the dogma must encompass the substantial differences observed in the sequence context of cytosines targeted for DNA methylation during development and in different cell types [30,39,64], the biological role of $5^{hm}C$, tissue-specific distinctions based on the location of methylated cytosines and the impact of gene-body methylation on gene expression [26,42,48,58]. Tackling some of the interesting questions posed by the new data will inevitably lead to the development of new hypotheses regarding the biological consequences of DNA methylation.

Acknowledgements

M.N.N. is supported by E.U. grant CANCERDIP FP7. H.D. is supported by the Brussels Region programme BruBreast. F.F. is a Senior Research Associate from the Belgian Fonds de la Recherche Scientifique.

References

1 Luger, K. (2006) Dynamic nucleosomes. *Chromosome Res.* 14, 5–16
2 Tost, J. (2009) DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Methods Mol. Biol.* 507, 3–20
3 Jones, P.A. and Baylin, S.B. (2007) The epigenomics of cancer. *Cell* 128, 683–692
4 Esteller, M. (2007) Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* 8, 286–298
5 Portela, A. and Esteller, M. (2010) Epigenetic modifications and human disease. *Nat. Biotechnol.* 28, 1057–1068
6 Beck, S. and Rakyan, V.K. (2008) The methylome: approaches for global DNA methylation profiling. *Trends Genet.* 24, 231–237
7 Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* 74, 481–514
8 Illingworth, R.S. and Bird, A.P. (2009) CpG islands – 'a rough guide'. *FEBS Lett.* 583, 1713–1720
9 Straussman, R. *et al.* (2009) Developmental programming of CpG island methylation profiles in the human genome. *Nat. Struct. Mol. Biol.* 16, 564–571
10 Blackledge, N.P. and Klose, R.J. (2011) CpG island chromatin: a platform for gene regulation. *Epigenetics* 6, 147–152
11 Fuks, F. (2005) DNA methylation and histone modifications: teaming up to silence genes. *Curr. Opin. Genet. Dev.* 15, 490–495
12 Brenner, C. and Fuks, F. (2007) A methylation rendezvous: reader meets writers. *Dev. Cell* 12, 843–844
13 Klose, R.J. and Bird, A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* 31, 89–97
14 Fraga, M.F. and Esteller, M. (2002) DNA methylation: a profile of methods and applications. *Biotechniques* 33, 632 634, 636–649
15 Brinkman, A.B. *et al.* (2010) Whole-genome DNA methylation profiling using MethylCap-seq. *Methods* 52, 232–236
16 Cross, S.H. *et al.* (1994) Purification of CpG islands using a methylated DNA binding column. *Nat. Genet.* 6, 236–244
17 Down, T.A. *et al.* (2008) A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat. Biotechnol.* 26, 779–785
18 Serre, D. *et al.* (2010) MBD-isolated genome sequencing provides a high-throughput and comprehensive survey of DNA methylation in the human genome. *Nucleic Acids Res.* 38, 391–399
19 Pomraning, K.R. *et al.* (2009) Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods* 47, 142–150
20 Ball, M.P. *et al.* (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat. Biotechnol.* 27, 361–368
21 Irizarry, R.A. *et al.* (2008) Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res.* 18, 780–790
22 Oda, M. *et al.* (2009) High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. *Nucleic Acids Res.* 37, 3829–3839
23 Khulan, B. *et al.* (2006) Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res.* 16, 1046–1055
24 Bock, C. *et al.* (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat. Biotechnol.* 28, 1106–1114
25 Laird, P.W. (2010) Principles and challenges of genome-wide DNA methylation analysis. *Nat. Rev. Genet.* 11, 191–203
26 Maunakea, A.K. *et al.* (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466, 253–257
27 Medvedeva, Y.A. *et al.* (2010) Intergenic, gene terminal, and intragenic CpG islands in the human genome. *BMC Genomics* 11, 48
28 Feng, S. *et al.* (2010) Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8689–8694
29 Laurent, L. *et al.* (2010) Dynamic changes in the human methylome during differentiation. *Genome Res.* 20, 320–331
30 Lister, R. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322
31 Lander, E.S. (2011) Initial impact of the sequencing of the human genome. *Nature* 470, 187–197
32 Gupta, R. *et al.* (2010) Advances in genome-wide DNA methylation analysis. *Biotechniques* 49, iii–xi
33 Feinberg, A.P. (2001) Methylation meets genomics. *Nat. Genet.* 27, 9–10
34 Ramsahoye, B.H. *et al.* (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
35 Flusberg, B.A. *et al.* (2010) Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465
36 Kriaucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930
37 Tahiliani, M. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935

38 Cokus, S.J. *et al.* (2008) Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219
39 Zhang, Y. *et al.* (2009) DNA methylation analysis of chromosome 21 gene promoters at single base pair and single allele resolution. *PLoS Genet.* 5, e1000438
40 Jones, P.A. (1999) The DNA methylation paradox. *Trends Genet.* 15, 34–37
41 Hellman, A. and Chess, A. (2007) Gene body-specific methylation on the active X chromosome. *Science* 315, 1141–1143
42 Irizarry, R.A. *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
43 Rauch, T.A. *et al.* (2009) A human B cell methylome at 100-base pair resolution. *Proc. Natl. Acad. Sci. U.S.A.* 106, 671–678
44 Shann, Y.J. *et al.* (2008) Genome-wide mapping and characterization of hypomethylated sites in human tissues and breast cancer cell lines. *Genome Res.* 18, 791–801
45 Zilberman, D. *et al.* (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* 39, 61–69
46 Zemach, A. *et al.* (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919
47 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
48 Weber, M. *et al.* (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* 39, 457–466
49 Saxonov, S. *et al.* (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1412–1417
50 Mohn, F. and Schubeler, D. (2009) Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends Genet.* 25, 129–136
51 Takai, D. and Jones, P.A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3740–3745
52 Boyes, J. and Bird, A. (1992) Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J.* 11, 327–333
53 Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21
54 Ayton, P.M. *et al.* (2004) Binding to unmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. *Mol. Cell Biol.* 24, 10470–10478
55 Blackledge, N.P. *et al.* (2010) CpG islands recruit a histone H3 lysine 36 demethylase. *Mol. Cell* 38, 179–190
56 Lee, J.H. and Skalnik, D.G. (2005) CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. *J. Biol. Chem.* 280, 41725–41731
57 Roh, T.Y. *et al.* (2005) Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* 19, 542–552
58 Illingworth, R.S. *et al.* (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.* 6, e1001134
59 Thomson, J.P. *et al.* (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464, 1082–1086
60 Doi, A. *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
61 Ji, H. *et al.* (2010) Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* 467, 338–342
62 Eckhardt, F. *et al.* (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* 38, 1378–1385
63 Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* 9, 465–476
64 Brunner, A.L. *et al.* (2009) Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res.* 19, 1044–1056
65 Satterlee, J.S. *et al.* (2010) Tackling the epigenome: challenges and opportunities for collaboration. *Nat. Biotechnol.* 28, 1039–1044
66 Henikoff, S. *et al.* (2008) Epigenomics: a roadmap to chromatin. *Science* 322, 853
67 Madhani, H.D. *et al.* (2008) Epigenomics: a roadmap, but to where? *Science* 322, 43–44
68 Berger, S.L. *et al.* (2009) An operational definition of epigenetics. *Genes Dev.* 23, 781–783
69 Ernst, J. and Kellis, M. (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* 28, 817–825
70 Penn, N.W. *et al.* (1972) The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* 126, 781–790
71 Ito, S. *et al.* (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466, 1129–1133
72 Valinluck, V. *et al.* (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.* 32, 4100–4108
73 Huang, Y. *et al.* (2010) The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* 5, e8888
74 Jin, S.G. *et al.* (2010) Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. *Nucleic Acids Res.* 38, e125
75 Song, C.X. *et al.* (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* 29, 68–72