REVIEW

doi:10.1038/nature14192

Function and information content of DNA methylation

Dirk Schübeler[1,2]

Cytosine methylation is a DNA modification generally associated with transcriptional silencing. Factors that regulate methylation have been linked to human disease, yet how they contribute to malignances remains largely unknown. Genomic maps of DNA methylation have revealed unexpected dynamics at gene regulatory regions, including active demethylation by TET proteins at binding sites for transcription factors. These observations indicate that the underlying DNA sequence largely accounts for local patterns of methylation. As a result, this mark is highly informative when studying gene regulation in normal and diseased cells, and it can potentially function as a biomarker. Although these findings challenge the view that methylation is generally instructive for gene silencing, several open questions remain, including how methylation is targeted and recognized and in what context it affects genome readout.

Methylation of nucleotides provides a molecular means to reversibly mark genomic DNA. Bacteria can methylate adenosine or cytosine to identify and degrade invading DNA and to track mismatch repair and the progress of genome duplication before cell division[1,2]. In eukaryotes, DNA methylation only occurs at cytosine residues. Since the discovery that *in vitro* methylated DNA is transcriptionally inactive when transfected into *Xenopus* oocytes[3] or cultured mammalian cells[3,4], methylation has been functionally linked to gene repression. Importantly, methylation is not always essential to eukaryotic gene regulation as it is absent in many organisms. These include metazoans such as dipteran insects or *Caenorhabditis elegans*, illustrating that transcriptional changes during development do not necessarily require the organism to methylate DNA. Although DNA methylation is obligatory in many clades, its prevalence and genomic distribution varies widely, suggesting that there are distinct modes of targeting and function[5–7].

Methylation of DNA can change the functional state of regulatory regions, but it does not change the Watson–Crick base pairing of cytosine. It thus presents the classic 'epigenetic' mark and is functionally involved in many forms of stable epigenetic repression, such as imprinting, X chromosome inactivation and silencing of repetitive DNA[8]. In vertebrates, heritable methylation only occurs at the CpG dinucleotide. Sequence symmetry of CpGs enables propagation of the methyl mark through cell division[9]; in 1975, this was proposed as a pathway for cellular memory of transcriptional states[10,11]. This potential for inheritance coupled with the fact that DNA methylation patterns change during development and disease[8,12] partially explains the interest in DNA methylation as a memory module[13].

Technological advances have resulted in genomic maps of DNA methylation at unprecedented resolution, revealing that regulatory sequences are indeed unmethylated when active[14–18]. Despite these striking but correlative observations, our ability to correctly assign a function to the local presence of DNA methylation at particular genes has remained surprisingly limited. This Review summarizes recent advances in our understanding of the regulation and function of DNA methylation in mammals, and discusses the utility of DNA methylation as a cellular marker in basic biology and biomedicine.

Conservation of methylation patterns in non-vertebrates

Among the many clades that methylate their genome, vertebrates are unique in that cytosine methylation occurs throughout the entire genome (Fig. 1). By contrast, plants and invertebrates that have been analysed so far show 'mosaic' methylation patterns because only specific genomic elements are targeted[5] (Fig. 1). More specifically, repetitive DNA and actively transcribed sequences are methylated[6,7]. In the case of repetitive DNA, it is evident that DNA methylation is used for repressing expression and preventing further expansion of these elements[19–22]. In plants, this targeting occurs through short RNAs derived from repeat transcripts that guide *de novo* methylation to this class of elements[23]. A similar PIWI-interacting RNA (piRNA)-guided process seems to occur in the male mammalian germ line when DNA methylation is re-established[12,24]. Furthermore, DNA methylation of pro-viruses has been shown to depend, in part, on the presence of repressive histone methylation at lysine 9 of histone H3 (H3K9me)[25], illustrating distinct RNA and chromatin pathways that guide DNA methylation to repeats.

The second canonical target is methylation of actively transcribed genes, a process that does not seem to primarily regulate these genes[6]. This is prominent in organisms with mosaic DNA methylation patterns and can even occur in the absence of repeat methylation, as in the case of the invertebrate chordate *Ciona intestinalis*[26,27]. Despite its strong evolutionary conservation, genic methylation remains surprisingly poorly understood at both the molecular and functional level. Current models suggest that it helps to counteract the disruption of chromatin, such as nucleosome displacement, which is caused by elongating RNA polymerase[8,26,27]. This speculation builds on established chromatin pathways such as the marking of transcribed sequences by methylation at lysine 36 of histone H3 (H3K36me)[28]. Work in baker's yeast (*Saccharomyces cerevisiae*), which lacks DNA methylation, revealed that H3K36me recruits enzymes such as histone deacetylases, resulting in more densely packed chromatin. In the absence of this pathway, transcription creates a more open chromatin structure, leading to spurious activation of cryptic start sites[29]. Such a process might be even more relevant in the genomes of multicellular organisms, which harbour substantially larger genes. Indeed, DNA methylation has been suggested to suppress intragenic promoters in mammalian cells[30], and remethylation of genes in cancer cells after treatment with an inhibitor of methylation maintenance occurs much faster at actively transcribed genes and seems to be required for their proper expression[31]. Notably, retroviruses preferentially integrate into actively transcribed genes[32], and a process that would methylate histones and DNA and potentially silence them after genic integration thus seems

[1]Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, CH-4058 Basel, Switzerland. [2]University of Basel, Faculty of Science, Petersplatz 1, CH-4003 Basel, Switzerland.

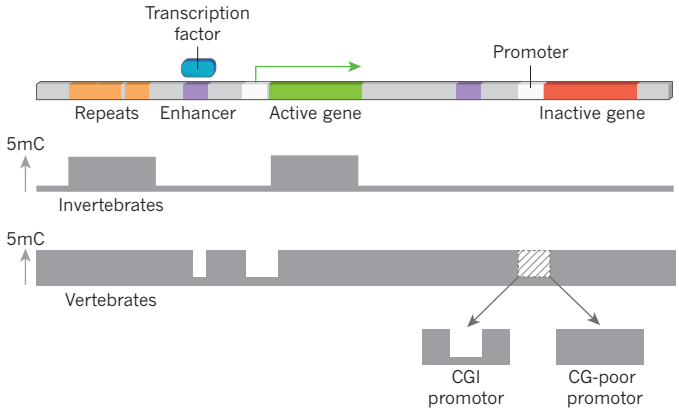

Figure 1 | Genomic distribution of methylated cytosine in a typical invertebrate and vertebrate genome. The representative genomic region includes an example of an active and an inactive gene with proximal (promoter) and distal (enhancer) regulatory regions. The height of the bar indicates the relative proportion of DNA methylation (5-methylcytosine, 5mC) that is observed in each region. CpG islands (CGIs), which often overlap with promoter regions, generally remain unmethylated, whereas CG-poor promoters are methylated when not active.

plausible. Furthermore, genic methylation has been linked to splicing in mammalian cells, as subtle methylation differences were observed between introns and exons[33,34]. However, experimental evidence to support this link remains scarce[35], and an explanation for how the elongating polymerase or splicing machinery could be affected by the comparatively small differences in methylation levels is lacking.

Methylation in vertebrates

The widespread methylation of vertebrate genomes suggests that their methylation might be a default state. This causes CpG depletion over evolutionary time through inefficient base-excision repair. Although spontaneous deamination of cytosine results in uracil, a base efficiently removed by uracil-DNA glycosylase (UDG), deamination of a methylated cytosine results in thymine, a proper genomic base. The resulting mismatch frequently manifests in a C to T transition, despite the presence of thymine-DNA glycosylase (TDG) and methyl-CpG-binding domain protein 4 (MBD4): two glycosylases that are thought to target this particular mismatch within methylated CGs and lead to the removal of thymine[36]. Notably, a C to T transition at CpG dinucleotides is the most frequent mutation observed in human diseases and between closely related mammals, arguing that genome-wide and genic DNA methylation comes at the price of an increased mutational load. Among vertebrates, CG depletion is particularly prominent in mammalian genomes, but less obvious in those of fish or *Xenopus*[37]. Exceptions to this general loss of CGs reside in genomic elements termed CpG islands (CGIs)[38]. CGIs predominantly overlap with promoter regions and retain their expected CG content, as they generally remain unmethylated in the germ line, with notable exceptions[39]. Importantly, methylation of CGIs causes robust transcriptional repression and CGI methylation is required for all reported examples of long-term mono-allelic silencing, including X inactivation[40] and genomic imprinting[41]. Although CGI methylation can result in stable repression of the linked gene (for example, see ref. 42), surprisingly few CGIs change DNA-methylation state during normal development. Prominent examples include a set of germline-specific genes that require promoter methylation for their repression in somatic cells[43]. Importantly, however, inactive CGI promoters do not, in general, acquire DNA methylation but become methylated at lysine 27 of histone H3 (H3K27me3), a mark set by the Polycomb system[44,45]. Interestingly, CGIs marked by H3K27me3 are nevertheless more susceptible to DNA methylation during differentiation and in disease states such as cancer[46,47]. Furthermore, CGIs harbour H3K4 methylation independently of their activity, but only in the absence of DNA methylation[48]. This modification can repulse *de novo* methyltransferases *in vitro* and thus is thought to be causally involved in maintaining the hypomethylated state of CGIs[49].

Recent genome-wide mapping of DNA methylation at single-base resolution revealed that CG-poor regulatory regions generally acquire a low methylation state when occupied by transcription factors[14–18]. Such variable DNA methylation in mammals closely reflects changes in gene regulation and, as a result, methylome data provide a rich source of information about ongoing gene activity (see 'Utility of DNA methylation'). Although the functional relevance of reduced methylation at CG-poor regulatory regions is still unclear, it is tempting to imply an instructive role for DNA methylation in distal gene regulation. This popular model assumes a generally repressive effect of DNA methylation regardless of the position and densities of CGs within a particular regulatory region. Experimental evidence does not, however, generally support this presumption. For example, several factors have been shown to bind methylated CG-poor sequences and, in turn, lead to their demethylation[14,50,51]. In this scenario, changes in DNA methylation occur downstream of transcription-factor binding to their target sequence, arguing against a generally instructive role. It is likely that other transcription factors are more sensitive to DNA methylation, in particular those that contain a CG in their binding motif[52] (Fig. 2). Potential effects might also only apply to certain binding sites such as those with lower affinity, at which DNA methylation might further reduce the likelihood of binding. Nevertheless, there is limited evidence at present that DNA methylation at CG-poor regulatory regions is generally instructive. This does not mean that methylation is irrelevant because tissue-specific deletions of, for example, DNA methyltransferases show very distinct phenotypes, including clear changes in gene expression, and mutations in these enzymes can contribute to disease[12]. However, it remains to be determined whether these phenotypes result from local differential methylation of regulatory regions or global perturbations, including reactivation of repeats and/or methylated CGIs.

Setting and removing DNA methylation

In the textbook scenario, *de novo* DNA methyltransferases DNMT3A and DNMT3B[53] in combination with DNMT3L[54] establish a pattern of methylation that is then faithfully maintained through cell division by the maintenance methyltransferase DNMT1 (ref. 55) and associated proteins[56]. This model of stable DNA methylation propagation has recently been revised in several ways. It was long known that DNA

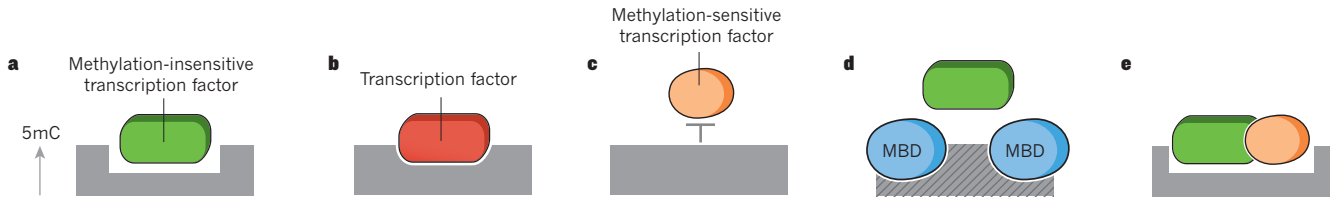

Figure 2 | Potential scenarios for the interplay between cytosine methylation (shown by level of 5-methylcytosine) and transcription-factor binding. **a**, A methylation-insensitive transcription factor causes reduced methylation after binding. **b**, A transcription factor binds specifically to the methylated state of its binding site. **c**, A methylation-sensitive transcription factor is blocked by 5-methylcytosine (5mC). **d**, Methyl-CpG-binding domain (MBD) proteins bind to the methylated state, leading to indirect repression, which probably requires high local density of CGs (shading). **e**, A methylation-insensitive transcription factor functions as a pioneer factor and creates a site of reduced methylation that allows a methylation-sensitive factor to bind.

methylation can be lost passively through imperfect maintenance[57], but the recent discovery of the ten-eleven translocation (TET) family of proteins provided a convincing path for catalysed active demethylation in vertebrates[58]. The three members of this protein family have since been implicated in development, meiosis, maintenance of imprinting and stem-cell reprogramming[59–64]. TET proteins convert 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC), a modified base that was first described more than 30 years ago[65]. Further iterative oxidations catalysed by TET result in 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[66,67] (Fig. 3), which can be efficiently removed by TDG[68,69] — a pathway referred to as 'active' demethylation. These modifications create an opportunity to identify sites of TET action because they can be specifically detected[69–73], but their detection and quantification remains challenging owing to their low genomic abundance compared with 5mC. Current data argue that active regulatory regions and transcribed sequences show increased levels of hydroxymethylation, which is indicative of a higher turnover[69,73–76]. Loss of all three TET proteins leads to increased DNA methylation at enhancer regions in stem cells and subtle expression changes of linked genes[77], suggesting that active demethylation could contribute to the activity state of distal regulatory regions[78]. Notably, active DNA demethylation could occur in most cell types because TET proteins are promiscuously expressed at varying levels in non-dividing somatic cells. Proper quantification of DNA-methylation-turnover kinetics is crucial not only for interpreting genomic maps of this mark but also for concepts of epigenetic memory that invoke DNA methylation. Memory implies stability of the modified base, which could be impaired by pathways of active demethylation. However, a conclusive picture regarding how turnover is regulated is lacking. Current models of binding and local activity for both demethylases and methyltransferases, which are largely based on *in vitro* interaction data, suggest that histone modifications and RNA–DNA hybrids have a role in repelling and recruiting these enzymes[12,74,79,80]. Recruitment probably constitutes just one level of regulation. For example, TET activity can be enhanced by the addition of vitamin C to the culture medium of stem cells[81], suggesting that cofactors modulate enzyme activity.

CG-specific readers

The CG dinucleotide provides distinct signals based on its methylation state that can be differentially recognized by specific protein domains. The methylated form of CG is recognized by functional MBD-containing proteins *in vitro* and *in vivo*[82–84], which are thought to indirectly contribute to methylation-mediated repression[85] (Figs 2d, 3b). This concept of indirect repression might explain why a high local concentration of methylated CpGs, such as at methylated CGIs, do confer efficient repression. Importantly, individual deletions of genes encoding MBD proteins do not result in reactivation of methylated CGIs and thus MBD proteins are thought to function in a redundant fashion. This model awaits testing through comprehensive deletions of the MBD protein family. CXXC domains, however, can specifically recognize unmethylated CGs and can target proteins, such as CXXC finger protein 1 (CFP1)[86] or the histone demethylases KDM2A and KDM2B to unmethylated CGIs[87,88] (Fig. 3b). It has been suggested that this recruitment might help to maintain an unmethylated state similar to that seen with classic transcription factors, but evidence for this model is still missing[87]. Importantly, the presence of a functional CXXC domain does not necessarily predict targeting of proteins to unmethylated CGs. In the case of DNMT1, the CXXC domain seems to be used for proofreading[9].

A crucial question is whether specific binders exist for cytosine oxidation derivatives, which would suggest that these bases function as distinct signals. This is a topic of active research, and several candidate proteins have been proposed[74,79,89]. However, the low frequency of these modifications has so far stymied efforts to convincingly show selective binding *in vivo*[87]. It also remains to be determined whether methylation of cytosines outside the CpG context, which seems to be a frequent event in embryonic stem cells, subsets of neurons[17] and oocytes[90], is specifically recognized.

Genetic determination and environmental influence

How much do our methylomes vary, and how can this variation be explained? Recent genome-wide association studies compared DNA sequence variation within human cohorts with changes in chromatin modifications[91–93] or DNA methylation[94]. The data showed that genetic variation explains a large part of the observed changes in histone modifications or DNA methylation[91–94]. The likely explanation is that mutations within regulatory regions affect binding of transcription factors, which in turn influence DNA methylation and histone modifications (Fig. 4). Alternatively, a mutation in a transcription factor could impair its ability to bind a specific sequence, leading to increased methylation. This transcription-factor dependency is similar to that observed during development, in which expression of cell-type-specific transcription factors coincides with reduced methylation of their binding sites (Fig. 4). These genome-wide correlations are also supported by gene-insertion experiments showing that local DNA sequence can be the primary determinant of DNA methylation state[95,96]. Indeed a role for transcription factors in establishing DNA methylation states was observed 20 years ago[50,97,98]. Thus several lines of evidence reinforce the notion that the underlying DNA sequence, through its recognition by transcription factors, seems to account for a substantial part of the observed DNA methylation and chromatin patterns and their variation between individuals or cell types.

How transcription factors influence chromatin and DNA methylation and which factors are instructive or sensitive to chromatin states remains largely unexplored (Fig. 2). These are crucial questions, because understanding this genetic and epigenetic crosstalk will be essential to correctly interpret and assign function to patterns of DNA methylation in development and disease. It should help in identifying sites that are more susceptible to sporadic change and environmental exposure. Indeed, differences in DNA methylation are observed in isogenic subjects, such as identical

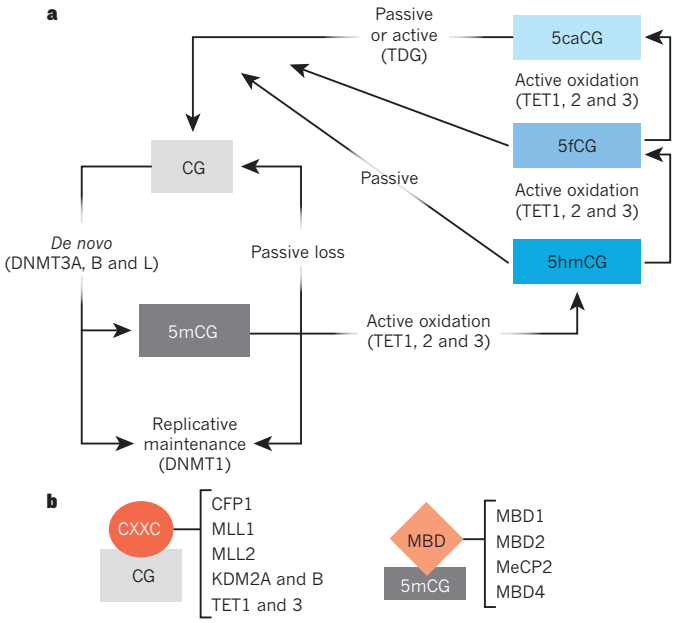

Figure 3 | Setting, erasing and recognizing cytosine methylation. **a**, Different methylation states of the CG dinucleotide and the enzymatic pathways that set, maintain and erase the mark. The pathways leading from the oxidized forms to the unmethylated state are under debate[74]. DNMT, DNA methyltransferase; TDG, thymine-DNA glycosylase. **b**, A subset of CXXC-domain-containing proteins are listed that can specifically bind to the unmethylated CG dinucleotide and could potentially reinforce the unmethylated state or recruit regulatory proteins to unmethylated CG islands. Methyl-CpG-binding domain (MBD) proteins specifically bind to the methylated CG with little or no further sequence sensitivity, potentially mediating transcriptional repression, which would be strongest in methylated CG islands. Readers of oxidized forms are not shown owing to ongoing debate about proposed candidates[87].

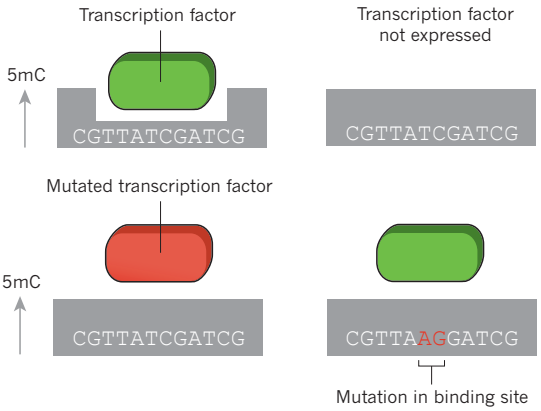

Figure 4 | Potential DNA sequence determinants of cytosine methylation at CG-poor regions. In a simplified model, transcription-factor binding causes reduced methylation at its binding site. Loss of expression of the respective transcription factor in development or disease will cause increased methylation. Mutations to the transcription factor that affect its binding preference will influence genomic methylation patterns. Mutation in the DNA binding site will abolish binding even in a cell expressing the transcription factor, indicating how genetic variation can result in methylation differences between individuals.

twins, but the actual level is under debate[99]. Furthermore, developing mouse embryos display methylation changes after severe undernutrition *in utero*[100]. The question remains as to whether these differences occur downstream of perturbed gene activity or are causally involved in the resulting phenotype. In the case of the undernutrition model, the metabolic phenotype can even propagate across generations, despite the lack of differences in DNA methylation[100]. Generally, although a role for DNA methylation in trans-generational inheritance is clearly appealing, there is limited evidence for this pathway in mammals (reviewed in ref. 101). At the same time, stably inherited epialleles clearly exist in plants[101]. One explanation for this difference is that mammals show waves of DNA-methylation gain and loss in the germ line and early development, leading to the erasure of most acquired methylation. This important aspect of DNA methylation biology is not the focus of this Review and has recently been discussed in detail[12,13,24].

New disease links

Perturbations of DNA methylation patterns are frequently observed in disease, particularly in cancer. These perturbations include methylation of CGI promoters for tumour suppressor genes, implying a functional role[8]. At the time of the initial observations, support for this model in the form of recurring mutations, in tumours, of proteins that regulate DNA methylation was missing. This changed with the discovery of the enzymatic activity of TET proteins — TET2 mutations have been linked to many myeloid malignancies[102,103] — and through the observation that DNMT3A is one of the most frequently mutated genes in acute myeloid leukaemia (AML) (mutated in 25% of adults with the disease)[104]. Although the actual mechanism in both cases remains to be determined, it is striking that TET2 and DNMT3A loss of function seem to be primary events. Mutations in DNMT3A are pre-leukaemic and can occur in blood stem cells, but only in combination with subsequent 'driver mutations' do they lead to leukaemia[105,106].

In fact, recurrent mutations of many genes encoding chromatin components have now been identified in diverse cancer types[107]. It is unclear how these proteins actually contribute to disease, but one hypothesis is that chromatin perturbations make the regulatory landscape more vulnerable to subsequent mutations[105]. Such concepts of genetic buffering by chromatin pathways have already been proposed based on genetic interaction screens between hypomorphic mutants, for example in *C. elegans*[108]. In this study, mutations in chromatin modifiers enhanced the phenotype of mutations in many different transcription factors. One interpretation of these findings is that chromatin pathways globally contribute to the stability of transcriptional regulation rather than directing activity to specific genes. It is tempting to speculate that DNA methylation might also contribute in this way to genome regulation.

Utility of DNA methylation

Although the function of DNA methylation at CG-poor regulatory regions remains unclear, it is evident that patterns of this mark have high information content about the ongoing activity of transcription factors and, thus, can help to identify cell-type-specific aspects of gene regulation[109]. For this reason, DNA methylation profiles can provide insight into many aspects of biology, but also function as biomarkers in medicine. It is important to keep in mind that the value of a biomarker is only defined by its ability to predict disease state or treatment response regardless of whether the measured parameter is causally involved in the disease. Part of the appeal of DNA methylation is the feasibility of the analysis. Bisulphite conversion of unmethylated cytosines followed by DNA sequencing requires no live cells and limited amounts of DNA, which can even be low quality, making DNA methylation less sensitive to specimen handling compared with RNA or proteins[110]. As a result, bisulphite sequencing, which, importantly, cannot distinguish between 5mC and 5hmC, can even be conducted on DNA isolated from small amounts of fixed tissue[111]. Any primary sample from which DNA can be sequenced should be suitable for bisulphite sequencing, a technique readily performed in any laboratory that is set-up for genome sequencing. Given that technical issues do not represent a significant hurdle, the crucial question is how much medically relevant information can be obtained from such DNA methylation analysis. Potential pitfalls include cellular heterogeneity within disease samples, confounding genetic variation between individuals or intrinsic variability, which could limit the utility of methylation measurements.

Nevertheless, in the case of medulloblastoma, a detailed analysis of methylation levels in regulatory regions enabled identification of disease- and tumour-subtype-specific changes[112]. These changes reflect not only transcription-factor activity but also putative new markers for disease states[112]. The aforementioned study adds to a list of examples for which DNA methylation can assist in identifying tumour subtypes[8].

Moreover, in most disease settings, cells from the affected tissue cannot be easily obtained, and it is questionable whether those that are available will be informative (for example, blood cells). Several studies have now shown that non-affected tissue can be informative[113], even leading to treatment decisions[114] and stratification of clinical cohorts[115]. Further insights into the functionality of DNA methylation will come from ongoing large cohort studies, leading towards a more comprehensive view of whether and how disease phenotypes can be associated with specific DNA methylation patterns[116]. The link between gene activity and DNA methylation, as well as the accessibility of this base modification, justify further exploration. One intriguing possibility is that DNA methylation might be prognostic of disease onset or personal risk in combination with genetic predisposition and environmental exposure.

The potential use of DNA methylation is not limited to biomedical applications. Measurements at the resolution of individual molecules and cells promise to substantially advance our understanding of regulation in biology. They provide more quantitative information and reveal heterogeneity within the cell population or dynamic changes at the level of individual cells[117]. Bisulphite conversion followed by high-throughput sequencing represents a single-molecule measure and has already been successfully applied in stem-cell biology to define population heterogeneity and the degree of pluripotency after induced reprogramming[118,119]. For example, subpopulations of mouse embryonic stem cells that differ in the level of the transcription factor Rex1 can be separated by their DNA methylation patterns[120]. Based on chemical inhibition, DNA methylation was argued to directly contribute to the switch between these populations[120]. Studies that tackle questions of cell identity and origin will further benefit from the recent development of approaches that allow genome-wide measurements at the level of single cells[121].

Tracking methylation patterns over time can also reveal insights into

the fidelity of DNA methylation maintenance. A comparison of the heterogeneity of DNA methylation at specific genomic loci suggested that human stem cells in culture undergo high turnover of DNA methylation. By contrast, fibroblast cells seem to show more stable inheritance of frequent sporadic changes, implying that somatic cells show more faithful maintenance and thus pass on sporadic changes in DNA methylation[122]. If the same applies *in vivo*, high-coverage DNA methylation measurements from single cells should provide information on lineage trees. This approach, when applied to an *in vitro* model of cellular transformation, has revealed insights into the kinetics of progressive changes in regions that are commonly hypermethylated in cancer[123]. Although it remains to be seen whether this principle is relevant to carcinogenesis *in vivo*, it highlights how DNA methylation can be used to track cellular states over time.

Outlook

The emerging picture is that genomic DNA methylation in mammals reflects, to a large extent, cell-intrinsic regulation encoded within the DNA sequence in the form of CG density and binding motifs for transcription factors. This genetic dependency is very reminiscent of histone modifications, which are also an integral part of the activation of regulatory regions and the process of transcription. It explains why chromatin and DNA modifications are informative indicators of underlying regulatory activity, but it does not reveal what their actual impact is on the regulation of individual genes. This is likely to be highly contextual at the level of DNA sequence and binding factors, as shown by the efficient repression of CGIs by DNA methylation, which might represent the exception rather than the rule because dynamics in DNA methylation occur elsewhere in the genome. This exemplifies our limited understanding of how local levels of DNA methylations are set, turned over by demethylation and, in turn, recognized by transacting factors. A better understanding of this regulation is required to define these potential functions and determine how misregulation contributes to disease. ■

Received 17 September; accepted 19 November 2014.

1. Arber, W. & Dussoix, D. Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J. Mol. Biol.* **5,** 18–36 (1962).
2. Wion, D. & Casadesus, J. N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nature Rev. Microbiol.* **4,** 183–192 (2006).
3. Vardimon, L., Kressmann, A., Cedar, H., Maechler, M. & Doerfler, W. Expression of a cloned adenovirus gene is inhibited by *in vitro* methylation. *Proc. Natl Acad. Sci. USA* **79,** 1073–1077 (1982).
4. Stein, R., Razin, A. & Cedar, H. *In vitro* methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. *Proc. Natl Acad. Sci. USA* **79,** 3418–3422 (1982).
5. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nature Rev. Genet.* **9,** 465–476 (2008).
6. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328,** 916–919 (2010).
7. Feng, S. *et al*. Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl Acad. Sci. USA* **107,** 8689–8694 (2010).
8. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Rev. Genet.* **13,** 484–492 (2012).
9. Song, J., Rechkoblit, O., Bestor, T. H. & Patel, D. J. Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331,** 1036–1040 (2011).
10. Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* **14,** 9–25 (1975).
11. Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science* **187,** 226–232 (1975).
12. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nature Rev. Genet.* **14,** 204–220 (2013).
13. Lee, H. J., Hore, T. A. & Reik, W. Reprogramming the methylome: erasing memory and creating diversity. *Cell Stem Cell* **14,** 710–719 (2014).
14. Stadler, M. B. *et al*. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495 (2011).
15. Hon, G. C. *et al*. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature Genet.* **45,** 1198–1206 (2013).
16. Ziller, M. J. *et al*. Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500,** 477–481 (2013).
17. Lister, R. *et al*. Global epigenomic reconfiguration during mammalian brain development. *Science* **341,** 1237905 (2013).
18. Hodges, E. *et al*. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol. Cell* **44,** 17–28 (2011).
19. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13,** 335–340 (1997).
20. Selker, E. U. Epigenetic phenomena in filamentous fungi: useful paradigms or repeat-induced confusion? *Trends Genet.* **13,** 296–301 (1997).
21. Lippman, Z. *et al*. Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430,** 471–476 (2004).
22. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retro viruses is constrained by cytosine methylation. *Nature Genet.* **20,** 116–117 (1998). **This study showed that absence of DNA methylation in the developing mouse embryo leads to reactivation of retroviruses.**
23. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Rev. Genet.* **11,** 204–220 (2010).
24. Hill, P. W., Amouroux, R. & Hajkova, P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics* **104,** 324–333 (2014).
25. Liu, S. *et al*. Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev.* **28,** 2041–2055 (2014).
26. Suzuki, M. M., Kerr, A. R., De Sousa, D. & Bird, A. CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res.* **17,** 625–631 (2007).
27. Simmen, M. W. *et al*. Nonmethylated transposable elements and methylated genes in a chordate genome. *Science* **283,** 1164–1167 (1999).
28. Keogh, M. C. *et al*. Cotranscriptional set2 methylation of histone H3 lysine 36 recruits a repressive Rpd3 complex. *Cell* **123,** 593–605 (2005).
29. Carrozza, M. J. *et al*. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123,** 581–592 (2005).
30. Maunakea, A. K. *et al*. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466,** 253–257 (2010).
31. Yang, X. *et al*. Gene body methylation can alter gene expression and is a therapeutic target in cancer. *Cancer Cell* **26,** 577–590 (2014).
32. Mitchell, R. S. *et al*. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol*. **2,** e234 (2004).
33. Maunakea, A. K., Chepelev, I., Cui, K. & Zhao, K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* **23,** 1256–1269 (2013).
34. Chodavarapu, R. K. *et al*. Relationship between nucleosome positioning and DNA methylation. *Nature* **466,** 388–392 (2010).
35. Shukla, S. *et al*. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479,** 74–79 (2011).
36. Cortázar, D. *et al*. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470,** 419–423 (2011).
37. Long, H. K. *et al*. Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *eLife* **2,** e00348 (2013).
38. Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321,** 209–213 (1986).
39. Smallwood, S. A. *et al*. Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genet.* **43,** 811–814 (2011).
40. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33** (Suppl.), 245–254 (2003).
41. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366,** 362–365 (1993). **This work established that imprinted gene expression requires DNA methylation.**
42. Velasco, G. *et al*. Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl Acad. Sci. USA* **107,** 9281–9286 (2010).
43. Borgel, J. *et al*. Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genet.* **42,** 1093–1100 (2010).
44. Tanay, A., O'Donnell, A. H., Damelin, M. & Bestor, T. H. Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc. Natl Acad. Sci. USA* **104,** 5521–5526 (2007).
45. Lynch, M. D. *et al*. An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. *EMBO J.* **31,** 317–329 (2012).
46. 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).
47. Mohn, F. *et al*. Lineage-specific polycomb targets and *de novo* DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30,** 755–766 (2008).
48. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genet.* **39,** 457–466 (2007).
49. Ooi, S. K. *et al*. DNMT3L connects unmethylated lysine 4 of histone H3 to *de novo* methylation of DNA. *Nature* **448,** 714–717 (2007).
50. Han, L., Lin, I. G. & Hsieh, C. L. Protein binding protects sites on stable episomes and in the chromosome from *de novo* methylation. *Mol. Cell. Biol*. **21,** 3416–3424 (2001).
51. Kress, C., Thomassin, H. & Grange, T. Active cytosine demethylation triggered by a nuclear receptor involves DNA strand breaks. *Proc. Natl Acad. Sci. USA* **103,** 11112–11117 (2006).
52. Liu, Y., Zhang, X., Blumenthal, R.M. & Cheng, X. A common mode of recognition for methylated CpG.*Trends Biochem. Sci.* **38,** 177–183 (2013).
53. 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).
54. Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. & Bestor, T. H. Dnmt3L and the establishment of maternal genomic imprints. *Science* **294,** 2536–2539 (2001).

55. Yoder, J. A. & Bestor, T. H. A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum. Mol. Genet.* **7,** 279–284 (1998).
56. Sharif, J. *et al*. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450,** 908–912 (2007).
57. Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23,** 5594–5605 (2003).
58. Tahiliani, M. *et al*. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).
This work established TET enzymes as active DNA demethylases and 5hmC as their product.
59. Gu, T. P. *et al*. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477,** 606–610 (2011).
60. Quivoron, C. *et al*. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell* **20,** 25–38 (2011).
61. Ficz, G. *et al*. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473,** 398–402 (2011).
62. Yamaguchi, S. *et al*. Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492,** 443–447 (2012).
63. Dawlaty, M. M. *et al*. Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* **24,** 310–323 (2013).
64. Wu, H. *et al*. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473,** 389–393 (2011).
65. Penn, N. W., Suwalski, R., O'Riley, C., Bojanowski, K. & Yura, R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* **126,** 781–790 (1972).
66. He, Y. F. *et al*. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333,** 1303–1307 (2011).
67. Ito, S. *et al*. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303 (2011).
68. Zhang, L. *et al*. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nature Chem. Biol.* **8,** 328–330 (2012).
69. Yu, M. *et al*. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149,** 1368–1380 (2012).
70. Williams, K. *et al*. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473,** 343–348 (2011).
71. Pastor, W. A. *et al*. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473,** 394–397 (2011).
72. Booth, M. J. *et al*. Quantitative sequencing of 5-methylcytosine and 5-hydroxy methylcytosine at single-base resolution. *Science* **336,** 934–937 (2012).
73. Song, C. X. *et al*. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153,** 678–691 (2013).
74. Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nature Rev. Mol. Cell Biol.* **14,** 341–356 (2013).
75. 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).
76. Feldmann, A. *et al*. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.* **9,** e1003994 (2013).
77. Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev.* **28,** 2103–2119 (2014).
78. Hon, G. C. *et al*. 5mC Oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol. Cell* **56,** 286–297 (2014).
79. Wu, H. & Zhang, Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156,** 45–68 (2014).
80. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I. & Chedin, F. R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45,** 814–825 (2012).
81. Blaschke, K. *et al*. Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* **500,** 222–226 (2013).
82. Baubec, T., Ivanek, R., Lienert, F. & Schubeler, D. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* **153,** 480–492 (2013).
83. Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. & Bird, A. P. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* **58,** 499–507 (1989).
This study reported the identification of the first protein that specifically binds the methylated CG dinucleotide.
84. Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* **18,** 6538–6547 (1998).
85. Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31,** 89–97 (2006).
86. Thomson, J. P. *et al*. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464,** 1082–1086 (2010).
87. Hahn, M. A., Szabo, P. E. & Pfeifer, G. P. 5-Hydroxymethylcytosine: a stable or transient DNA modification? *Genomics* **104,** 304–323 (2014).
88. Long, H. K., Blackledge, N. P. & Klose, R. J. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem. Soc. Trans.* **41,** 727–740 (2013).
89. Spruijt, C. G. *et al*. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152,** 1146–1159 (2013).
90. Tomizawa, S. *et al*. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* **138,** 811–820 (2011).
91. Kasowski, M. *et al*. Extensive variation in chromatin states across humans. *Science* **342,** 750–752 (2013).
92. Kilpinen, H. *et al*. Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* **342,** 744–747 (2013).
93. McVicker, G. *et al*. Identification of genetic variants that affect histone modifications in human cells. *Science* **342,** 747–749 (2013).
94. Gutierrez-Arcelus, M. *et al*. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *eLife* **2,** e00523 (2013).
95. Lienert, F. *et al*. Identification of genetic elements that autonomously determine DNA methylation states. *Nature Genet.* **43,** 1091–1097 (2011).
96. Krebs, A., Dessus-Babus, S., Burger, L. & Schubeler, D. High-throughput engineering of a mammalian genome reveals building principles of methylation states at CG rich regions. *eLife* **3,** e04094 (2014).
97. Brandeis, M. *et al*. Sp1 elements protect a CpG island from *de novo* methylation. *Nature* **371,** 435–438 (1994).
98. Macleod, D., Charlton, J., Mullins, J. & Bird, A. P. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev.* **8,** 2282–2292 (1994).
99. Bell, J. T. & Spector, T. D. DNA methylation studies using twins: what are they telling us? *Genome Biol.* **13,** 172 (2012).
100. Radford, E. J. *et al*. *In utero* undernourishment perturbs the adult sperm methylome and intergenerational metabolism. *Science* **345,** 1255903 (2014).
101. Heard, E. & Martienssen, R. A. Transgenerational epigenetic inheritance: myths and mechanisms. *Cell* **157,** 95–109 (2014).
102. Ko, M. *et al*. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468,** 839–843 (2010).
103. Ono, R. *et al*. LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res.* **62,** 4075–4080 (2002).
104. Ley, T. J. *et al*. DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* **363,** 2424–2433 (2010).
105. Shlush, L. I. *et al*. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506,** 328–333 (2014).
106. Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc. Natl Acad. Sci. USA* **111,** 2548–2553 (2014).
107. Plass, C. *et al*. Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer. *Nature Rev. Genet.* **14,** 765–780 (2013).
108. Lehner, B., Crombie, C., Tischler, J., Fortunato, A. & Fraser, A. G. Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nature Genet.* **38,** 896–903 (2006).
109. Burger, L., Gaidatzis, D., Schubeler, D. & Stadler, M. B. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res.* **41,** e155 (2013).
110. Heyn, H. & Esteller, M. DNA methylation profiling in the clinic: applications and challenges. *Nature Rev. Genet.* **13,** 679–692 (2012).
111. Schillebeeckx, M. *et al*. Laser capture microdissection-reduced representation bisulfite sequencing (LCM-RRBS) maps changes in DNA methylation associated with gonadectomy-induced adrenocortical neoplasia in the mouse. *Nucleic Acids Res.* **41,** e116 (2013).
112. Hovestadt, V. *et al*. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510,** 537–541 (2014).
113. Rakyan, V. K., Down, T. A., Balding, D. J. & Beck, S. Epigenome-wide association studies for common human diseases. *Nature Rev. Genet.* **12,** 529–541 (2011).
114. Weller, M. *et al*. MGMT promoter methylation in malignant gliomas: ready for personalized medicine? *Nature Rev. Neurol.* **6,** 39–51 (2010).
115. Jacquemont, S. *et al*. Epigenetic modification of the FMR1 gene in fragile X syndrome is associated with differential response to the mGluR5 antagonist AFQ056. *Sci. Transl. Med.* **3,** 64ra61 (2011).
116. Bock, C. Analysing and interpreting DNA methylation data. *Nature Rev. Genet.* **13,** 705–719 (2012).
117. Junker, J. P. & van Oudenaarden, A. Every cell is special: genome-wide studies add a new dimension to single-cell biology. *Cell* **157,** 8–11 (2014).
118. Ma, H. *et al*. Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature* **511,** 177–183 (2014).
119. Xie, W. *et al*. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153,** 1134–1148 (2013).
120. Singer, Z. S. *et al*. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Mol. Cell* **55,** 319–331 (2014).
121. Smallwood, S. A. *et al*. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature Methods* **11,** 817–820 (2014).
122. Shipony, Z. *et al*. Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature* **513,** 115–119 (2014).
123. Landan, G. *et al*. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nature Genet.* **44,** 1207–1214 (2012).

Acknowledgements I apologize to colleagues, whose work I could not cite or only discuss in a limited context owing to space limitations. I thank in particular P. Ginno and further T. Baubec., M. Lorincz and N. Thomae for critical input on the manuscript. Work in my laboratory is supported by the Novartis research foundation, the European Union (NoE EpiGeneSys FP7- HEALTH-2010-257082 and the Blueprint consortium FP7-282510), the European Research Council (EpiGePlas), the SNF Sinergia programme and the Swiss Initiative in Systems Biology (RTD Cell Plasticity).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of this paper at go.nature.com/hdj5ba. Correspondence should be addressed to D.S. (dirk@fmi.ch).