DNA methylation landscapes: provocative insights from epigenomics

Miho M. Suzuki and Adrian Bird

Abstract | The genomes of many animals, plants and fungi are tagged by methylation of DNA cytosine. To understand the biological significance of this epigenetic mark it is essential to know where in the genome it is located. New techniques are making it easier to map DNA methylation patterns on a large scale and the results have already provided surprises. In particular, the conventional view that DNA methylation functions predominantly to irreversibly silence transcription is being challenged. Not only is promoter methylation often highly dynamic during development, but many organisms also seem to target DNA methylation specifically to the bodies of active genes.

Imprinted gene
A gene that is expressed or silenced depending on which parent contributed it to the zygote. In a mouse cell, for example, the paternal insulin-like growth factor allele is expressed, but the maternal allele is not. In some cases, imprinting depends on differential DNA methylation of gene regulatory regions.

The Wellcome Trust Centre for Cell Biology, The University of Edinburgh, Michael Swann Building, The King's Buildings, Edinburgh EH9 3JR, UK.
Correspondence to A.B
e-mail: a.bird@ed.ac.uk
doi:10.1038/nrg2341
Published online 8 May 2008

The genomes of eukaryotes carry chemical marks that are added to either DNA or chromatin proteins. This epigenetic information is not uniform, but is applied regionally, and it signals or preserves local activity states, such as gene transcription or silencing[1]. The sum total of all epigenetic information is termed the 'epigenome'. If we are to understand the biological and biomedical significance of epigenetic phenomena, it is obviously important to map the epigenome in some detail. However, unlike the genome, the epigenome is highly variable between cells and fluctuates in time according to conditions even within a single cell. There are therefore at least as many epigenomes as there are cell types. Despite this challenge, a number of projects have started to put epigenetic flesh on the bare bones of the genome. The focus in this Review is on studies that have begun to describe the large-scale distribution of one epigenetic mark — DNA methylation — in normal (that is, non-cancerous) tissues and cell types. Although it is essentially descriptive, this work has turned up surprising findings that call for a re-assessment of prevailing views about the significance of methyl groups on genomic DNA.

In eukaryotes ranging from plants to humans, DNA methylation is found exclusively at cytosine residues. This post-synthetic modification has important roles. For example, it is essential for mammalian embryonic development as shown by early lethality in mice that lack DNA methyltransferases (DNMTs)[2,3]. *Dnmt*-null mice have reduced DNA methylation levels, but the precise reasons for death during development are unclear. Defects in repression of the inactivated X chromosome in female cells and in the establishment and maintenance of allele-specific expression of imprinted genes have been observed[4–6], as has elevated expression of transposon RNA in embryos[7]. These findings, and numerous other studies over the past decades, have led to the generalization that cytosine DNA methylation functions to maintain the repressed chromatin state and therefore stably silence promoter activity[8].

Many studies of DNA methylation in animals have been carried out in mammalian systems, in which genomic DNA methylation is found throughout the genome with the conspicuous exception of short unmethylated regions called CpG islands (CGIs)[9,10] (FIG. 1). It is important to bear in mind, however, that the global DNA methylation pattern seen in vertebrates is by no means ubiquitous among eukaryotes (TABLE 1). Several well-studied model systems have no recognizable *Dnmt*-like genes and are devoid of DNA methylation (for example, the yeast *Saccharomyces cerevisiae* and the nematode worm *Caenorhabditis elegans*). In fungi that have genomic 5-methylcytosine (m5C), only repetitive DNA sequences are methylated[11] (FIG. 1a). The most frequent pattern in invertebrate animals is 'mosaic methylation', comprising domains of heavily methylated DNA interspersed with domains that are methylation free[12,13] (FIG. 1c). The highest levels of DNA methylation among all eukaryotes have been observed in plants, with up to 50% of cytosine being methylated in some species[14]. In maize, for example, such high levels seem to be due to large numbers of transposons, the degenerate relics of which dominate intergenic regions and are targeted for methylation[15,16] (FIG. 1e). However, other plants, such as *Arabidopsis thaliana*, display a mosaic DNA methylation pattern that is reminiscent of invertebrate animals (FIG. 1b).

Despite similarities in DNA methylation landscapes, there are important differences between DNA methylation in animals and plants. Most significant is the presence of non-CpG methylation in plants that is targeted to transposable elements by a mechanism that depends upon small interfering RNAs (siRNAs)[17–19]. So far, there is no convincing evidence for a parallel mechanism in animals.

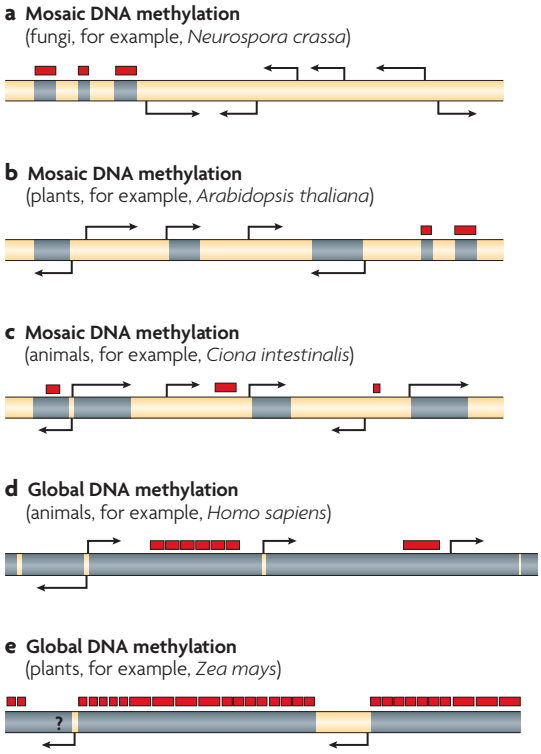

Figure 1 | **DNA methylation landscapes in fungi, animals and plants. a** | Mosaic DNA methylation, whereby stable methylated (grey) and unmethylated (yellow) domains are interspersed, is seen in certain fungi owing to the efficient targeted methylation of transposable elements (red boxes). **b** | The plant *Arabidopsis thaliana* has a small genome and illustrates a mosaic methylation pattern that is due to gene-body methylation, as seen in invertebrates. Unlike animals, transposons and repetitive elements are subject to targeted methylation by an RNA-mediated mechanism of genome defence. **c** | Mosaic methylation is also characteristic of most tested invertebrates, but has only been mapped in detail in the sea squirt *Ciona intestinalis*. Gene-body methylation affects over half of all genes, but the remainder are embedded within unmethylated DNA. Transposable elements are frequently unmethylated and match the methylation status of the surrounding DNA. **d** | Vertebrate genomes are globally methylated, with only CpG islands being unmethylated. Transposable elements are methylated, as are gene bodies and intergenic DNA. **e** | The DNA methylation landscape of plants with large genomes, such as maize, has not been mapped in detail, but it is evident that genes are separated by long tracts of DNA that contain transposable elements and their relics[16]. Genes tend to be unmethylated, but the existence of gene body-targeted methylation has not yet been investigated.

CpG island
(CGI). A DNA patch of approximately 1,000 bp, within which the dinucleotide CpG occurs at close to its expected frequency. This contrasts with the majority of the vertebrate genome, in which CpG is depleted. Despite the abundance of CpGs that could potentially be methylated, CGIs are unmethylated in germ cells and most are also DNA methylation free in somatic cells. In mammals, CGIs are GC-rich in base composition (~65%) compared with the genome as a whole (~40%).

The apparent similarities and differences between epigenomes within and between eukaryotic groups prompt the question of whether there is a common underlying mechanism at work, or whether the DNA methylation system been co-opted to distinct biological roles in different organismal groups. A precondition for answering this question is a thorough understanding of the distribution of cytosine methylation throughout the genomes of a variety of species. High-throughput methodologies have recently evolved to the point that global analysis of DNA methylation landscapes has become feasible. This Review will discuss results emerging from these studies that cast a new and refreshing light on the quest for an understanding of DNA methylation. Over the past decade, a consensus view has taken hold that sees DNA methylation primarily as a mediator of irrevocable transcriptional silencing. Its potential role in choreographing the complex changes in gene expression that occur during development — once the primary motivation for many scientists studying DNA methylation — are currently seen as limited. However, studies of large numbers of promoters have revealed many at which DNA methylation varies significantly according to the cell type. Although, so far, there is scant evidence for a causal role in modulating gene expression, dynamic patterns of promoter methylation provide a pretext for revisiting the possibility of a developmental role. In addition, large-scale studies of plant and invertebrate genomic methylation patterns have uncovered an entirely unexpected spatial relationship between DNA methylation and genes. Seemingly at odds with its role in gene silencing, evidence from diverse systems reveals DNA methylation that is targeted to the transcription units of actively transcribed genes. By highlighting the limits to our understanding of this epigenetic system, the new approaches are invigorating DNA methylation research.

Mapping global DNA methylation patterns

The gold-standard technology for detection of m5C is bisulphite genomic sequencing, which maps sites at single base-pair resolution[20]. This method depends on the finding that, following prolonged incubation with sodium bisulphite, cytosines in single-stranded DNA are deaminated to give uracil. The modified nucleoside m5C is immune to this transformation and therefore any cytosines that remain in bisulphite-treated DNA must have been methylated, as outlined in TABLE 2. Normally, bisulphite-treated DNA is amplified by PCR using locus-specific primers, and multiple subcloned fragments are then sequenced. Large-scale bisulphite DNA sequencing has been successfully initiated[21,22], but this is a time- and resource-intensive task, as outlined in TABLE 3. Therefore, attempts to map DNA methylation on a genome-wide scale have so far relied on less direct methods. Approaches based on the sensitivity of restriction enzymes to CpG methylation within their cleavage recognition site[23] are comparatively low resolution, but they are useful when combined with genomic microarrays[24,25]. Alternatively, recent

Table 1 | **Examples of genomic methylation patterns in various eukaryotic phyla**

Species	Overall pattern	Methylated sequences*	Transposon methylation	Targeted transposon methylation	Gene-body methylation	Detection method of gene-body methylation	Refs
Plants							
Arabidopsis thaliana	Mosaic	CG, CNG and CNN	Yes	Yes (RdDM)	Yes	Genome-wide methylation mapping by microarray	17–19, 28, 35
Zea mays	Mosaic	CG, CNG and CHH	Yes	–	See footnote‡	Methylation-sensitive restriction enzyme mapping of several unmethylated genes, filtration of unmethylated genomic DNA	43, 82, 83
Oryza sativa	Mosaic	CG, CNG and CHH	Yes	–	Yes	Restriction enzyme analysis plus Bioinformatics	81, 84
Fungi							
Saccharomyces cerevisiae, *Schizosaccharomyces pombe*	–	–	No	–	No	–	–
Neurospora crassa	Mosaic	CNN	Yes	Yes (RIP)	No	Methyl-CpG affinity chromatography	11, 86, 87
Ascobolus immersus	Mosaic	CNN	Yes	Yes (MIP)	–	–	85, 88, 89
Invertebrates: insects							
Drosophila melanogaster	See footnote§	CT and CA	Yes‖	–	Yes	MedIP and bisulphite sequencing	90
Apis mellifera	Mosaic	CG	No	–	Yes	Bisulphite sequencing of 6 genes	42
Myzus persicae	Mosaic	CG	–	–	Yes	Bisulphite sequencing of the elastase gene[41]	41
Invertebrates: deuterostomes							
Echinus esculentus	Mosaic	CG	Yes	–	–	–	12
Strongylocentrotus purpuratus	Mosaic	CG	–	–	Yes	Southern blots with methylation-sensitive restriction enzymes	13
Ciona intestinalis	Mosaic	CG	Yes	No	Yes	Bisulphite sequencing plus bioinformatics	38, 39
Vertebrates							
Danio rerio	Global	CG	Yes	–	Yes	Methylation-sensitive restriction enzyme mapping and bisulphite sequencing	91
Xenopus laevis	Global	CG	Yes	–	Yes	Bisulphite sequencing of a few genes	92
Homo sapiens	Global	CG	Yes	See footnote¶	Yes	Bisulphite sequencing of exonic sequences	21, 93, 94

*H could be A, T or C; N could be A, C, G or T; ‡Only the existence of unmethylated genes has been shown. §0.4% of the cytosine residues are methylated in embryos. ‖Putative substrates of the DNA methyltransferase gene dDNMT2. ¶See the text for further information. MedIP, methylated DNA immunoprecipitation; MIP, methylation induced pre-meiotically; RdDM, RNA-directed DNA methylation; RIP, repeat-induced point mutation.

high-throughput studies have used protein affinity to enrich methylated sequences as probes for genomic microarrays. Methylated DNA fragments are affinity purified with either an anti-m5C antibody (methylated DNA immunoprecipitation; MedIP[26]) or by using the DNA-binding domain of a methyl-CpG-binding protein (methyl-binding domain affinity purification; MAP[27]). A comparison between these enrichment methods indicated that they give comparable results[28]. Both require a relatively high density of DNA methylation, such as when CGIs become methylated. This is an important constraint, as bulk genomic DNA from mammals contains one methyl-CpG site on average every 150 bp, which would not be efficiently recovered. An method to enrich specifically for unmethylated DNA using CXXC affinity purification (CAP; X represents any residue) was also recently introduced[29]. The sample pretreatment methods described above are summarized in TABLE 2. Samples enriched in these ways can be interrogated using DNA microarrays or by direct large-scale sequencing techniques, as summarized in TABLE 3.

High-throughput approaches have been used to analyse DNA methylation patterns across the whole *A. thaliana* genome as well as in the mouse and human genomes (see below and TABLE 4). The large sizes of mammalian genomes (~3.3 x 10^9 bp) compared with that of *A. thaliana* (1.1 x 10^8 bp) makes comprehensive profiling a significant technical challenge. As a result, studies in mammals so far have either surveyed much of the genome at low resolution or have focused in detail on a small genomic fraction. However, the potential

Table 2 | **Current methods for high-throughput DNA methylation analysis: sample pretreatment**

Pretreatment method	General basis	Resolution	Advantages	Disadvantages	Refs
Bisulphite conversion	Sodium bisulphite converts unmethylated cytosine to uracil, whereas methylated cytosines are protected from conversion	High: single base resolution	Applicable to any samples	Complete conversion is essential	20
Methylation-sensitive restriction enzyme methods					
RLGS; HELP assay	DNA is differentially fragmented with a methylation-sensitive restriction enzyme. Following size fractionation, this method enriches methylated DNA	Moderate	Relatively simple	Analysis limited to methylation at restriction sites	25, 95
McrBC digestion	DNA digestion with a methylation-specific restriction enzyme, McrBC. Following size fractionation, this method enriches unmethylated DNA	Moderate	Effective in degrading most methylated DNA	–	37,96, 97
Affinity purification methods					
Methylated DNA immunoprecipitation (MedIP)	Immunoprecipitate DNA containing methylated cytosines using a monoclonal antibody	Moderate	The antibody is commercially available. Precipitates methylated cytosines in all contexts	High m5C density required	26,28, 35,72, 98
MBD affinity purification (MAP)	Immunoprecipitate DNA containing methylated CpG using an MBD column	Moderate	Only methylated CpGs are recovered	High m5CpG density required	27–29, 99
CXXC affinity purification (CAP)*	Immunoprecipitate DNA containing unmethylated CpG using a CXXC-domain column	Moderate	A direct method to extract unmethylated DNA	High CpG density required	29

*X could be any residue. HELP, HpaII tiny fragment enrichment by ligation-mediated PCR; m5C, 5-methyl cytosine; m5CGI, CGI containing m5C; MBD, methyl-binding domain; RLGS, restriction landmark genome scanning.

for rapid data acquisition is growing fast owing to techniques such as BeadArray (manufactured by Illumina)[30] — in which a large number of samples can be assayed simultaneously — and large-scale sequencing technologies (TABLE 3). In addition, 454 sequencing has been used for the parallel sequencing of bisulphite-treated DNA instead of the standard subcloning and sequencing method[31,32]. In the future, the huge number of reads offered by these high-throughput sequencing technologies offers the realistic prospect of analysing DNA methylation across the whole mammalian genome[33]. A recent report describes the successful application of Solexa bisulphite sequencing to the whole *A. thaliana* genome[34]. Pilot experiments suggested that this approach might also be applicable to the entire mammalian genome.

Methylation in gene bodies

***The DNA methylation landscape of* A. thaliana.** The first genome-wide map of DNA methylation was reported for the flowering plant *A. thaliana* by probing methylated DNA, which was affinity purified using MedIP, against tiled arrays of genomic DNA[28,35]. In the mosaically methylated *A. thaliana* genome, repetitive DNA is a major target of DNA methylation by an RNA-dependent DNA methylation system[17]. These studies showed almost 20% of the genome to be densely methylated in the adult plant, including transcriptionally inactive heterochromatin such as centromeres, pericentromeric heterochromatin and the heterochromatic knob on chromosome 4. As expected, repetitive DNA sequences and regions, the transcripts of which can be recovered as siRNA, are greatly enriched in these methylated domains[36]. In a mutant plant that lacks the DNA methyltransferase MET1, over 60% of the methylated regions became demethylated and this was accompanied by transcriptional activation of transposons and pseudogenes residing in heterochromatin[28]. These data support the conclusion that MET1-mediated DNA methylation is mainly responsible for the silencing of heterochromatic regions of the plant genome.

More unexpected were the observations concerning DNA methylation in transposon-free euchromatin of *A. thaliana*[28,35]. Some euchromatic methylated domains corresponded to pseudogenes and to a small proportion of promoters, in line with the view that DNA methylation associates with transcriptional silencing. The surprising result, however, was that a large fraction of all genes (33%) were covered by CpG methylation in their transcribed regions. DNA methylation in these cases was clearly biased away from gene ends, such that neither the 5′ end nor the 3′ ends of transcription units were methylated. Gene-body methylation of this kind does not shut off expression of the gene — the average expression level of affected genes was significantly higher than that of either promoter-methylated or entirely unmethylated genes (62% of all expressed genes). Overall, genes displaying gene-body methylation were characterized by a moderate level of expression in many tissue types. Many could be broadly classified as 'housekeeping genes', the products of which are necessary for basic processes required by all cell types. Surprisingly, different ecotypes of *A. thaliana* show differences in the DNA methylation status of many gene bodies, suggesting that this epigenetic feature can be variable within the same species[37].

454 sequencing and Solexa bisulphite sequencing
Independent proprietary high-throughput DNA-sequencing technologies that both use massively parallel sequencing-by-synthesis approaches. These new methods allow an increase in generated sequence per run of about two orders of magnitude compared with conventional Sanger sequencing technologies, and therefore allow rapid comprehensive sequence screening of large genomic fractions or whole genomes.

Heterochromatic knob
A chromosomal region that can be identified microscopically as being darkly stained compared with surrounding chromatin. DNA sequence analysis has shown that knobs often contain highly repeated DNA sequences. They were described initially in the 1930's by McClintock during her studies of maize chromosome structure.

Table 3 | **Current methods for high-throughput DNA methylation analysis: readout**

Readout method	Sample pretreatment method	General basis	Resolution	Other features	Uses	Refs
DNA microarrays						
Oligonucleotide arrays	Bisulphite conversion, methylation-sensitive restriction enzyme or affinity purification methods	Short (25-mer) or long (60-mer) oligonucleotide array	Moderate	–	Tiling genomic arrays, promoter arrays and custom arrays	28,35, 100–102, 105
SNP arrays		SNP selective probe array	Moderate	–	Detection of allele-specific DNA methylation	103
BeadArray (Illumina)	Bisulphite conversion	Ratio of the methylated and unmethylated PCR products is determined at single CpG sites	High: single-base resolution, quantitative	A large set of primers needs to be designed	Detection of methylation polymorphisms (96 samples assayed in parallel)	30
Sequencing						
Standard sequencing	Bisulphite conversion	Sanger sequencing	High: single-base resolution, quantitative	–	Expensive and labour intensive for genome-wide analysis	–
Direct large-scale sequencing	Bisulphite conversion, methylation-sensitive restriction enzyme or affinity purification methods	Short-read sequencing (Solexa sequencing: 40 million reads of 25–35 bases; 454 sequencing: 400,000 reads of >100 bases)	High: single-base resolution, quantitative	High-quality reference sequence is required	Fast and relatively inexpensive. Genotype information can be obtained simultaneously	34,104

Gene-body methylation is evolutionarily ancient. The finding of gene-body methylation in plants provides large-scale evidence for a phenomenon that had been noted previously at specific genes in several invertebrate genomes. Gene body-specific methylation was initially mapped using DNA methylation-sensitive restriction enzymes in the invertebrate chordate *Ciona intestinalis* (sea squirt), which possess a mosaic DNA methylation pattern comprising both methylated and unmethylated DNA in roughly equal proportions[38]. Bisulphite sequencing of a ~100 kb region of the *C. intestinalis* genome, together with verified computational prediction of DNA methylation status, that covered ~1 Mb of the genome showed that gene-body methylation is widespread in this genome[39]. About 60% of all *C. intestinalis* genes show evidence of gene-body methylation, and this apparently accounts for the majority of all DNA methylation in this species. The characteristics of *C. intestinalis* body-methylated genes were similar to those observed in *A. thaliana* because most were housekeeping genes, whereas highly expressed genes tended to be unmethylated. Additionally, gene-body methylated genes in *C. intestinalis* tend to be more evolutionarily conserved than other genes. Interestingly, repetitive sequences, including transposable elements, are not preferentially methylated in *C. intestinalis*, but seem to mimic the methylation status of the surrounding DNA domain. This suggests that the elements are not active targets for *de novo* DNA methylation, but might acquire their methylation status passively.

In addition to *A. thaliana* and *C. intestinalis*, bisulphite sequencing in two insect species shows comparable intragenic CpG methylation. The first evidence for CpG methylation in an insect was established for the amplified esterase E4 gene of the aphid *Myzus persicae*[40]. Bisulphite sequencing detected CpG methylation within the active gene, but not at 5′ and 3′ regions of the transcription unit[41]. Recently, several honeybee genes similarly showed CpG methylation within the transcription units but not at their extremities[42]. An early survey of invertebrate genomic DNA methylation patterns suggested that mosaic methylation is the most common configuration among invertebrates and emphasized that methylation of housekeeping gene bodies is widespread[13] (TABLE 1). Based on the above examples, it seems that, in animals, mosaicism is predominantly due to the presence of methylated gene bodies separated by unmethylated DNA. How the evolutionary transition from mosaic to global methylation was accomplished remains a mystery, but we speculate that the change benefited the innate immune system (BOX 1).

Gene-body methylation in mammals. Mammalian genomes, like those of all vertebrates tested so far, are globally methylated in the sense that all categories of DNA sequence (genes, transposons and intergenic DNA) are targets for CpG methylation[21,43,44]. Thus, unlike mosaically methylated genomes, in which methylated and unmethylated domains coexist in approximately equal proportions, mammalian genomes are dominated by methylated DNA. Unmethylated domains (that is, most CGIs) account for a small fraction (1–2%) of the total[10,29,45]. Because the vast majority of DNA is methylated to a high level, it follows that gene bodies are also methylated in vertebrates, and this has been confirmed by numerous studies[21,43,44]. However, ubiquitous DNA methylation makes it difficult to determine whether the methylation is targeted specifically to gene sequences or is a default state that happens to affect genes as well as most other sequences.

Table 4 | **Recent large-scale methylation studies done in mammals**

Authors	Year	Region studied	Samples	Method	Scale	Refs
Eckhardt *et al.*	2006	Human chromosomes 6, 20 and 22, selected 5′ UTRs, evolutionarily conserved regions, introns, exons and others	43 samples from 12 tissues from different individuals and primary cells	Bisulphite conversion then standard sequencing	2,524 amplicons	21
Rollins *et al.*	2006	Randomly selected human genomic sequences	Human brain tissue	Methylation-sensitive restriction enzyme then standard sequencing	3,073 unmethylated and 2,565 methylated domains	22
Schumacher *et al.*	2006	~12 Mb of human chromosomes 21 and 22	Human brain tissue from 8 individuals	Methylation-sensitive restriction enzyme then oligonucleotide array	Tiling array with probes spaced on average every 35 bp	24
Khulan *et al.*	2006	6.2 Mb of the mouse genome	Mouse brain tissue and spermatogenic cells	Methylation-sensitive restriction enzyme then oligonucleotide array	HpaII fragment tilling array with average 15mer frequency	25
Keshet *et al.*	2006	Human promoter array	Normal lymphoblasts and colon cancer cells	MedIP then oligonucleotide array	13,000 promoters of human genes	98
Weber *et al.*	2007	Human promoter array	Primary fibroblasts, and sperm cells	MedIP then oligonucleotide array	16,000 promoters of human genes	72
Rauch *et al.*	2008	~140 Mb of human chromosome 7 and 8 and human CGI array	Normal lung and lung cancer tissues from 4 individuals	MAP then oligonucleotide array	Whole-genome tiling arrays at 100 bp resolution plus 27,800 CGIs	105
Illingworth *et al.*	2008	Human CGI array	Blood, brain, muscle and spleen tissues	MAP then probe CGI array	14,000 CGIs	29
Hellman and Chess	2007	Human SNP mapping array	Human embryonic stem cells and B-lymphocyte cells	Methylation-sensitive restriction enzyme then SNP array	500,000 SNPs	48
Bibikova *et al.*	2006	371 human genes	Normal lung and lung cancer samples	Bisulphite conversion then SNP array	1,536 CGIs	106
Ladd-Acosta *et al.*	2007	807 human genes	76 brain tissue samples from 43 individuals	Bisulphite conversion then SNP array	1,505 CGIs	75

CAP, CXXC affinity purification (X could be any of the four bases); CGI, CpG island; MAP, methyl-binding domain (MBD) affinity purification; MedIP, methylated DNA immunoprecipitation.

Data derived from the human X chromosome has provided specific evidence that gene-body methylation in mammals, like that of plants and invertebrates, is associated with transcriptional activity (FIG. 2). Compensation for the differing dosage of the X chromosome in males and females is achieved in placental mammals by shutting down most genes on one of the female's X chromosomes. DNA methylation is implicated in this gene silencing, and early evidence showed that promoter CGIs on the inactive X chromosome (X_i) are hypermethylated and causally involved in maintaining silencing[46,47]. However, a recent study confirmed earlier hints that X_i is in fact less methylated than the active X chromosome (X_a) over much of the chromosome[26]. Using SNPs to distinguish homologous X chromosomes on microarrays, Hellman and co-workers[48] reported more than twice as much methylation on X_a as on X_i. Significantly, extra methylation on X_a was concentrated within gene bodies. Did the difference arise because X_a had become unusually densely methylated compared with autosomes, or was X_i abnormally demethylated? To answer this question, DNA methylation was examined in a cell line in which X chromosomes are biallelically active, representing a stage prior to X inactivation. Both X chromosomes were methylated in these cells, suggesting that hypomethylation of X_i arises by demethylation relative to the normal state[48].

Is this phenomenon peculiar to X chromosomes, or does the profound difference in transcriptional activity between X_a and X_i allow detection of a DNA methylation pattern that also affects genes on other chromosomes? In other words, is the gene-body methylation that is detected on X_a also a feature of mammalian autosomes? It is tempting to conclude that X_a resembles the normal methylation status of autosomes because gene bodies on autosomes are clearly methylated (see REFS 21,43,44 for examples). A common feature of gene-body methylation in plants and invertebrates is that the 5′ and 3′ extremities of genes are significantly less methylated. Mammalian CGI-associated genes partially conform to this generalization, as the unmethylated domain usually extends from the 5′ end into the gene body by several hundred base pairs. Reduced CpG methylation at the 3′ end of mammalian genes has not been reported. We do not yet have an answer to the general question of whether gene-body methylation in mammals is evolutionarily and functionally equivalent to that seen in other taxonomic groups. An answer awaits a functional assay for this phenomenon.

The origin of gene-body methylation. Plants and animals diverged about 1.6 billion years ago, yet the evidence described above suggests that similar patterns of DNA methylation in the bodies of active genes are

Box 1 | **The immune system and the transition from mosaic to global DNA methylation**

Mosaic methylation of the genome is characteristic of a wide range of animal phyla, but has not been seen in vertebrates[13]. It is therefore reasonable to postulate that mosaic methylation was ancestral to vertebrate global methylation, although the steps by which unmethylated domains could become methylated without disastrous phenotypic consequences are unclear. Regardless of the precise mechanism, we speculate that innate immunity has been enhanced by this transition and might have provided a selective pressure. Dendritic cells are known to express a range of Toll-like receptors that, following stimulation, trigger the innate immune response[78]. One of the receptors expressed by plasmacytoid dendritic cells and B cells, Toll-like receptor 9 (TLR9), detects genomes of invading bacterial pathogens by recognizing DNA that is rich in unmethylated CpG moieties[78]. The globally methylated, CpG-deficient, vertebrate host genome is unlikely to activate this response, thereby preventing auto-immunity. A mosaic methylated genome, on the other hand, comprises about 50% unmethylated CpG-rich DNA and would run the risk of initiating an auto-immune response. We propose that the transition from mosaic to global methylation was a prerequisite for the evolution of CpG DNA immunity. Compatible with this hypothesis, TLR9 has not been detected in any invertebrate genome and seems to have first evolved with the vertebrate lineage. For example, the genomic sequence of the sea urchin *Strongylocentrotus purpuratus* revealed a vast repertoire of 222 Toll-like receptors (many more than in humans), but no TLR9 family member was found[79]. Therefore, the ability to detect pathogens by their CpG-rich DNA seems to have gone hand in hand with an expansion of DNA methylation to eliminate almost all genomic DNA that might trigger this response. Only unmethylated CpG islands are exempt. Is it possible that these CpG-rich sequences, which amount to less than 2% of the genome, can, under certain circumstances, trigger human auto-immunity?

present in both groups. This implies that gene-body methylation reflects a primary and ancestral function of DNA methylation in animals (FIG. 3). What might this role be and how does it square with current perceptions of the role of DNA methylation? Evidence from many sources implicates DNA methylation as an agent of transcriptional silencing. Methylation of gene promoters on X_i, at imprinted genes and at various genes in cancers or cell lines imposes gene silencing that can be reversed by artificial demethylation[9]. In the light of this evidence, the notion that DNA methylation is a reliable feature of transcriptionally active genes seems heretical. A suggested function that preserves the idea that DNA methylation is a transcriptional repressor posits that intragenic methylation prevents transcriptional interference owing to spurious initiation within an active transcription unit[35,38]. To explain the absence of methylation at many genes in genomes that show a mosaic pattern of methylation, it is proposed that the relatively weak promoters of housekeeping genes are more susceptible to such interference than are highly transcribed genes.

Although these speculations have not yet been tested experimentally, there are intriguing parallels with the occurrence of intragenic repressive histone marks in eukaryotes. In particular, methylation of histone H3 lysine 9 (H3K9), once thought of as diagnostic of constitutive heterochromatin, is reported to occur within actively transcribed genes[49]. In addition, the histone deacetylation that is triggered by methylation of histone H3K36 within yeast transcription units is required to prevent spurious intragenic transcriptional initiation[50]. Elongating forms of RNA polymerase II are biochemically implicated in recruitment of this histone-modifying activity in yeast. Indirect evidence has raised the possibility that gene-body DNA methylation is also recruited by RNA polymerase II activity. Specifically, Zilberman and colleagues[35] noted that the methylated regions of gene bodies in *A. thaliana* corresponded with regions of polymerase elongation, whereas the DNA methylation-free 5′ and 3′ extremities of genes often had high RNA polymerase II densities in either the initiation or termination modes. Only expressed genes showed lack of methylation at the extremities of the transcription unit, as *A. thaliana* pseudogenes did not exhibit this phenomenon. A speculative scenario is that transcriptional elongation somehow reinforces methylation of the underlying DNA. There is currently no evidence for a mechanistic connection between DNA methylation and the transcription process.

An alternative explanation for the presence of DNA methylation in gene bodies is that RNA-mediated gene silencing in plants, which triggers DNA methylation at repeated sequences, provides the link between transcription and *de novo* methylation. According to this scenario, gene-body DNA methylation might be caused by antisense transcription within an active gene. However, in-depth sequencing of small RNAs that can act as intermediates in *de novo* methylation failed to detect sequences corresponding to methylated gene bodies[28,34]. Similarly, it has been argued that DNA methylation triggered by RNAi is unlikely to exist in animals and would therefore be an unlikely source of their gene-body methylation[51]. In spite of these reservations, an RNA-mediated origin for gene-body methylation remains possible at this stage[52], as do other mechanisms.

Revisiting the function of DNA methylation. The widespread occurrence of intragenic DNA methylation calls for a reassessment of our understanding of the biological significance of DNA methylation, particularly in the case of animals. Two common perceptions deserve scrutiny: that DNA methylation contributes to the formation of heterochromatin; and that a primary role of DNA methylation is to defend the genome against transposons. Heterochromatin is a word of declining usefulness, as there is no coherent view of what it describes. Nevertheless, all would agree that it does not refer to transcriptionally active genes. Yet active genes are the sites of gene-body CpG methylation, which accounts for the majority of genomic DNA methylation in *C. intestinalis* and other mosaically

methylated invertebrate genomes. The independence of DNA methylation from heterochromatin is also obvious in organisms that form apparently normal heterochromatin (that is, condensed chromosomal regions, often including tandemly repeated DNA sequences) yet lack CpG methylation (for example, *Drosophila melanogaster*). Even in the mouse, in which densely methylated repetitive DNA sequences form easily visible heterochromatic blocks surrounding centromeres, the absence of DNA methylation leaves heterochromatic foci visible by microscope, albeit with a somewhat altered composition[53].

The idea that DNA methylation is primarily a mechanism of genome defence has received robust support from the analysis of fungal and plant genomes, in which transposable elements are evidently specific targets and are prevented from transposition by this modification[54] (FIGS 1,3). In animals, however, the case is inconclusive. Methylation maps in organisms as diverse as *C. intestinalis*[38,39] and the bee[42] indicate that genes, rather than transposons, are targets of CpG methylation. In the mammalian genome, it is less easy to determine if transposons are actively targeted or if they become methylated passively, as almost all chromosomal DNA (with the exception of CGIs) is methylated (FIG. 1d). There is, however, a further prediction of the genome defence hypothesis: hypomethylation should lead to increased transposition. So far, neither DNA methyltransferase gene mutants nor naturally hypomethylated cells, such as tumour cells, have betrayed evidence of enhanced transposition[55]. Current data therefore sustains the view that CpG methylation exerts its function at genes rather than elsewhere in the genome. Methylation of promoters leads to stable gene silencing, whereas it is conceivable that intragenic methylation helps to dampen transcriptional noise[56].

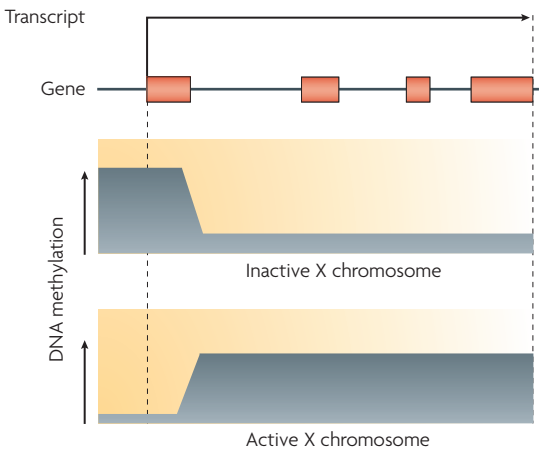

Figure 2 | **Gene-body methylation on the human active X chromosome.** Comparison of DNA methylation levels on the active (X_a) and inactive (X_i) X chromosomes showed reduced methylation specifically over gene bodies on X_i. Therefore, the DNA methylation patterns are inverted on these two chromosomes: promoter CpG islands are methylated on X_i but unmethylated on X_a.

The mammalian DNA methylation landscape

Large-scale studies of DNA methylation patterns in mammals have so far focused mainly on humans because comprehensive DNA methylation maps from both normal and diseased human cell types is of both biological and biomedical interest[57]. Earlier research on individual DNA sequences suggested the generalization that the mammalian genome is globally methylated, with the exception of CGIs. In line with this conclusion, analysis of the distribution of small DNA fragments derived from genomic DNA by digestion with DNA methylation-sensitive restriction endonucleases confirmed that long contiguously methylated domains are occasionally interrupted by unmethylated regions. These unmethylated regions were usually at promoters and CGIs in a 6.2-Mb segment of the mouse genome[25]. A similar landscape was deduced from a combination of global computational analysis of patterns of CpG depletion and direct sequencing of enriched unmethylated and methylated domains from human brain DNA[22]. Again, unmethylated domains were enriched in the 5′ regions of genes, promoters, CGIs and first exons.

The added detail provided by bisulphite sequencing has allowed useful generalizations about global human DNA methylation. An initial study examined the histocompatibility locus (including 90 genes) and, more recently, another study examined 1.9 million CpG sites on human chromosomes 6, 20 and 22 (including 873 genes) in twelve tissues[21,44]. The results showed that the majority of the analyzed regions were either hypomethylated (less than 30% of CpG sites) or hypermethylated (more than 70% of CpG sites). Thus, there was not a continuum of CpG methylation levels at these loci, many of which were CGIs. This suggests two alternative states: silent (heavily methylated) and potentially active (essentially unmethylated), although the biological rationale for this switch-like behaviour remains to be elucidated. Eckhardt *et al.*[21] also noted an unmethylated core region of about 1,000 bp centred at the transcriptional start site (TSS); this was also found at the TSS of plant genes. These hypomethylated sites might be passive footprints showing where DNA methyltransferases have been excluded by bound factors[58]. Alternatively, localized promoter hypomethylation might be required for gene expression to take place efficiently.

CGI methylation in normal human tissues. CGIs represent a discrete fraction of the genome in several respects. They correspond to short regions of DNA that lack methylation, at least in the germ line, and this ensures that they do not suffer the mutational loss of CpGs that affects the rest of the genome[10]. Also, in mammals and birds, CGIs have a GC-rich base composition compared with bulk genomic DNA, which is AT rich. They have an average length of ~1,000 bp and are often associated with genes; for example, approximately 56% of human genes have CGI promoters[59]. Unmethylated promoters are also present in amphibians and fish[60], and in invertebrates with methylated genomes[39], but here they tend not to differ in base composition from the surrounding DNA[61].

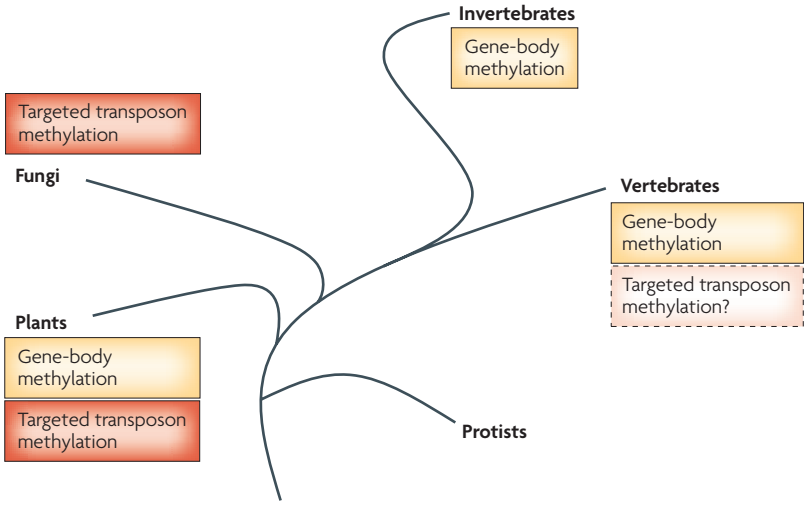

Figure 3 | **Evolution of eukaryotic DNA methylation patterns.** There is strong evidence for targeting of DNA methylation to repetitive elements in fungi and plants, but no evidence for an equivalent process in invertebrate animals. Vertebrates are problematic; the elements are methylated, as is most of the genome, but it is not clear that this is due to specific targeting. Gene-body methylation is reported in plants as well as invertebrate and vertebrate animals, suggesting an ancient origin. Fungi do not show gene-body methylation; indeed, intragenic methylation inhibits transcriptional elongation[80].

Identification of mammalian CGIs usually depends upon computational prediction. Most commonly, the criteria require a GC content of at least 55% and a ratio of observed to expected CpG frequency of at least 0.6 (REF. 62). The length parameter is crucial. The original algorithm[63], devised before genome sequences were available, used 200 bp as the criterion and this became the norm, but recent studies have indicated a vast excess (~10-fold) of false positives using this method[29]. Increasing the minimum length over which the base compositional and CpG frequency criteria must apply to 500 bp eliminates most false positives and has become accepted as standard. A different approach to CGI identification has recently been introduced, which is based on sequencing of DNA fragments that were isolated from human blood DNA using an affinity reagent that specifically binds clusters of unmethylated CpG[29]. This criterion takes account of CpG clustering, but, unlike the computational methods, also requires absence of CpG methylation. Most of the DNA fragments obtained by this method matched those predicted by the algorithm, but a fraction of these fragments (~20%) were novel. Interestingly, about half of all CGIs were found at the TSS of an annotated gene, the remainder being downstream or in intergenic regions. The functional significance of intergenic CGIs remains unclear, but their existence at the promoters of the non-coding RNAs *Xist* and *Air*, both of which regulate gene expression[64,65], raises the intriguing possibility that at least some CGIs correspond to the promoters of regulatory RNAs.

Although most CGIs remain unmethylated throughout development regardless of expression state[66], a minority become methylated during development[9], and this correlates with transcriptional silencing of the associated gene. The classic example is X chromosome inactivation, during which hundreds of CGIs on X_i become heavily methylated, ensuring transcriptional silence of the associated genes, as discussed above. Other examples of natural CGI methylation have been seen at imprinted genes and at genes that are exclusively expressed in the germ line[67,68]. Interestingly, the post-migratory silencing of several genes that are expressed in migrating primordial germ cells has recently been shown to depend upon DNA methylation[69]. There has long been evidence that CGI methylation can occur at other loci in normal somatic cells, but until recently this has been qualified by uncertainty about the bioinformatic criteria for CGI identification (see the discussion section in REF. 70). Using stringent criteria, a PCR-based methylation analysis of predicted CGIs on human chromosome 21 indicated that 31 out of 149 were fully methylated in peripheral blood[71]. In other studies, large-scale bisulphite sequencing[21] detected 9.2% of 511 CGIs to be methylated in a variety of tissues, promoter microarrays detected 3% of CGIs as somatically methylated[72], and a microarray analysis of 14,000 CGIs isolated by CpG affinity detected ~12% of CGI methylation in human blood, brain, muscle and spleen[29].

These studies make it abundantly clear that CGI methylation is a widespread phenomenon in human somatic tissues. Apart from gene silencing associated with X chromosome inactivation or imprinting, we have little idea about its biological significance, although intriguing clues are starting to emerge. Illingworth and colleagues[29] noted that differentially methylated CGIs preferentially included genes that have central roles in development, such as homeobox (*HOX*) genes and paired box (*PAX*) genes and their relatives. Does this signify a role for differential CGI methylation in development? This study also noted that CGIs not associated with TSSs (that is, those within or between recognized genes) were significantly more likely to be methylated than those at gene promoters (7% versus 16%). Unravelling the significance of distal CGI methylation with respect to gene expression and development is an evident priority.

Variable methylation outside CpG islands. A large-scale analysis of mammalian DNA methylation using microarrays focused exclusively on sequences surrounding the TSS of 16,000 annotated genes, which are predicted to include regulatory and promoter DNA sequences[26,72]. CGI promoters predominantly remained unmethylated regardless of expression, as suggested by studies of specific loci, whereas CpG-deficient promoters often retained methylation that did not seem to interfere with expression. Most dynamic with respect to DNA methylation, however, were promoters with an intermediate CpG density (that is, an average ratio of observed to expected CpG of 0.5), which frequently acquired DNA methylation in somatic tissues. Bisulphite data supports the view that differentially methylated regions are over-represented within the non-CGI category of promoters[21]. Dynamic DNA methylation changes within the so-called weak CpG island category raise interesting questions. Are weak CpG islands discrete, like CGIs, or do they reflect

the sequence characteristics of the larger DNA domains of which they are part? Are there shared features of these sequences or their associated genes that might account for their susceptibility to *de novo* methylation?

Comparison between human tissue types and between individuals by bisulphite seqeuencing has begun to address in detail the issue of human variation with respect to DNA methylation[21]. Interestingly, levels of DNA methylation as a whole were not significantly different between unrelated individuals, even when disparate age groups were compared (26±4 years old versus 68±8 years old). The homogeneity of DNA methylation levels in this large sample indicates that this DNA mark is subject to restricted interindividual heterogeneity. Different tissues, on the other hand, showed marked local differences in DNA methylation. For example, 7.1% of all genomic CpGs in 2,524 amplicons showed differential methylation between $CD4^+$ lymphocytes and dermal fibroblasts. Such tissue-specific methylated regions were detected in gene-coding regions as well as in intergenic regions, raising the speculative possibility that they correspond to *cis*-regulatory regions involved in the control of gene expression. Their potential importance is emphasized by the observation that they preferentially coincide with DNA sequences that are highly conserved between the mouse and human genomes. The divergence of DNA methylation patterns between cell types within one individual contrasts with the conservation seen between individuals, and implies that differences in methylation are involved in, or result from, changes that arise during differentiation.

Conclusions and future directions

Studies of short individual DNA segments provided useful examples of DNA methylation patterns, but we have for too long been ignorant of their generality. Now that high-throughput analyses are being applied, some of the generalizations are holding up, but new and unexpected phenomena are also being detected. Most surprisingly, the bodies of active genes are specifically targeted by DNA methylation in plants and invertebrates, and in some organisms this seems to be the predominant source of genomic m5C. There is tantalizing evidence for a parallel phenomenon in mammals, raising the possibility that this role is conserved in diverse life forms. At the same time, studies of global genomic methylation in mammalian genomes, particularly the human genome, are rejuvenating the idea that DNA methylation plays a part in development and differentiation, as apparently specific variations in methylation of both CGI and non-CGI promoters are repeatedly documented. Many of these changes are not coincident with annotated genes, raising the possibility that distal regions of the genome can influence genome activity — for example, as promoters of non-coding RNAs. These new findings might herald a reappraisal of conventional wisdom concerning the functional significance of CpG methylation.

Biomedical interest in DNA methylation centres on the possibility that epigenetic variation between individuals can have repercussions for health[73], but there is currently relatively little evidence for this. One prominent study found significant DNA methylation differences between monozygotic twins that became prominent with age[74]. Recently, evidence for interindividual variation in brain DNA methylation has also emerged[75,76]. By contrast, large-scale bisulphite sequencing failed to detect significant differences in DNA methylation between unrelated individuals of widely disparate ages[21]. Although it might be argued that stably methylated regions were chosen by chance for the sequencing study, future studies are needed to address this apparent discrepancy.

The role of aberrant DNA methylation in cancer has been persuasively argued[77]. More recently, other human diseases have been hypothetically linked to abnormalities in DNA methylation[57], but causality is notoriously difficult to establish. The recent history of complex genetic traits is an interesting parallel in this respect, as for many years results were relatively disappointing. Technical advances, however, have led to an explosion of new data that promises to revolutionize our understanding of human disease. Arguably, epigenetic theories of complex disease rose to prominence within the vacuum that was caused by the dearth of genetic information. Now that the vacuum is being rapidly filled, it is time to replace speculation with hard experimental data. A leap in the scale of analysis will be crucial. Fortunately, emerging high-throughput DNA-sequencing technologies can potentially enable this leap to be made, allowing us, in time, to compare and contrast complete DNA methylation maps.

1. Bird, A. Perceptions of epigenetics. *Nature* **447**, 396–398 (2007).
2. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
3. 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).
4. Chen, T. & Li, E. Structure and function of eukaryotic DNA methyltransferases. *Curr. Top. Dev. Biol.* **60**, 55–89 (2004).
5. Karpf, A. R. & Matsui, S. Genetic disruption of cytosine DNA methyltransferase enzymes induces chromosomal instability in human cancer cells. *Cancer Res.* **65**, 8635–8639 (2005).
6. Dodge, J. E. *et al.* Inactivation of *Dnmt3b* in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *J. Biol. Chem.* **280**, 17986–17991 (2005).
7. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genet.* **20**, 116–117 (1998).
8. Bird, A. P. & Wolffe, A. P. Methylation-induced repression — belts, braces, and chromatin. *Cell* **99**, 451–454 (1999).
9. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
10. Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209–213 (1986).
11. Selker, E. U. *et al.* The methylated component of the *Neurospora crassa* genome. *Nature* **422**, 893–897 (2003).
12. Bird, A. P., Taggart, M. H. & Smith, B. A. Methylated and unmethylated DNA compartments in the sea urchin genome. *Cell* **17**, 889–901 (1979).
13. Tweedie, S., Charlton, J., Clark, V. & Bird, A. Methylation of genomes and genes at the invertebrate–vertebrate boundary. *Mol. Cell Biol.* **17**, 1469–1475 (1997).
14. Montero, L. M. *et al.* The distribution of 5-methylcytosine in the nuclear genome of plants. *Nucleic Acids Res.* **20**, 3207–3210 (1992).
15. Palmer, L. E. *et al.* Maize genome sequencing by methylation filtration. *Science* **302**, 2115–2117 (2003).
16. SanMiguel, P. *et al.* Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**, 765–768 (1996).
17. Chan, S. W., Henderson, I. R. & Jacobsen, S. E. Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nature Rev. Genet.* **6**, 351–360 (2005).
18. Chan, S. W. *et al.* RNA silencing genes control *de novo* DNA methylation. *Science* **303**, 1336 (2004).
19. Mette, M. F., Aufsatz, W., van der Winden, J., Matzke, M. A. & Matzke, A. J. Transcriptional silencing and promoter methylation triggered by double-stranded RNA. *EMBO J.* **19**, 5194–5201 (2000).

20. Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl Acad. Sci. USA* **89**, 1827–1831 (1992).
21. Eckhardt, F. *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature Genet.* **38**, 1378–1385 (2006).
In this paper, large-scale bisulphite sequence analysis of human chromosomal regions reveals greater variation in methylation patterns between tissues in a given individual than between individuals. Methylated regions were highly conserved, raising the possibility that they correspond to regulatory sequences.
22. Rollins, R. A. *et al.* Large-scale structure of genomic methylation patterns. *Genome Res.* **16**, 157–163 (2006).
23. Bird, A. P. Use of restriction enzymes to study eukaryotic DNA methylation. II: the symmetry of methylated sites supports semi-conservative copying of the methylation pattern. *J. Mol. Biol.* **118**, 48–60 (1978).
24. Schumacher, A. *et al.* Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res.* **34**, 528–542 (2006).
25. Khulan, B. *et al.* Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res.* **16**, 1046–1055 (2006).
26. Weber, M. *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nature Genet.* **37**, 853–862 (2005).
27. Cross, S. H., Charlton, J. A., Nan, X. & Bird, A. P. Purification of CpG islands using a methylated DNA binding column. *Nature Genet.* **6**, 236–244 (1994).
28. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**, 1189–1201 (2006).
This reference and reference 35 are two of the first genome-wide analyses of DNA methylation, in which gene body-specific methylation is described.
29. Illingworth, R. *et al.* A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *PLoS Biol.* **6**, e22 (2008).
This paper presents a large-scale analysis of human CpG islands that are purified using an unmethylated CpG-affinity column. Half of all CGIs are not coincident with annotated promoters and a significant fraction becomes methylated in normal somatic tissues, particularly at genes involved in development.
30. Fan, J. B. *et al.* Illumina universal bead arrays. *Methods Enzymol.* **410**, 57–73 (2006).
31. Korshunova, Y. *et al.* Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA. *Genome Res.* **18**, 19–29 (2008).
32. Taylor, K. H. *et al.* Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. *Cancer Res.* **67**, 8511–8518 (2007).
33. Zilberman, D. & Henikoff, S. Genome-wide analysis of DNA methylation patterns. *Development* **134**, 3959–3965 (2007).
34. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).
35. Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. & Henikoff, S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genet.* **39**, 61–69 (2007).
36. Henderson, I. R. & Jacobsen, S. E. Epigenetic inheritance in plants. *Nature* **447**, 418–424 (2007).
37. Vaughn, M. W. *et al.* Epigenetic natural variation in *Arabidopsis thaliana*. *PLoS Biol.* **5**, e174 (2007).
38. Simmen, M. W. *et al.* Non-methylated transposable elements and methylated genes in a chordate genome. *Science* **283**, 1164–1167 (1999).
39. 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).
In this paper, analysis of DNA methylation in the invertebrate chordate *C. intestinalis* establishes that about half of all genes are methylated and show gene-body methylation.
40. Field, L. M., Devonshire, A. L., Ffrench-Constant, R. H. & Forde, B. G. Changes in DNA methylation are associated with loss of insecticide resistance in the peach-potato aphid *Myzus persicae* (Sulz.). *FEBS Lett.* **243**, 323–327 (1989).
41. Field, L. M. Methylation and expression of amplified esterase genes in the aphid *Myzus persicae* (Sulzer). *Biochem. J.* **349**, 863–868 (2000).
42. Wang, Y. *et al.* Functional CpG methylation system in a social insect. *Science* **314**, 645–647 (2006).
43. Rabinowicz, P. D. *et al.* Genes and transposons are differentially methylated in plants, but not in mammals. *Genome Res.* **13**, 2658–2664 (2003).
44. Rakyan, V. K. *et al.* DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.* **2**, e405 (2004).
45. Bird, A., Taggart, M., Frommer, M., Miller, O. J. & Macleod, D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**, 91–99 (1985).
46. Mohandas, T., Sparkes, R. S. & Shapiro, L. J. Reactivation of an inactive human X-chromosome: evidence for X-inactivation by DNA methylation. *Science* **211**, 393–396 (1981).
47. Venolia, L. *et al.* Transformation with DNA from 5-azacytidine-reactivated X chromosomes. *Proc. Natl. Acad. Sci. USA* **79**, 2352–2354 (1982).
48. Hellman, A. & Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141–1143 (2007).
This study shows that genes on the active X chromosome are in fact more methylated than those on the inactive X chromosome, and that methylation is high within gene bodies.
49. 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).
50. Vakoc, C. R., Mandat, S. A., Olenchock, B. A. & Blobel, G. A. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol. Cell* **19**, 381–391 (2005).
51. Schaefer, C. B., Ooi, S. K., Bestor, T. H. & Bourc'his, D. Epigenetic decisions in mammalian germ cells. *Science* **316**, 398–399 (2007).
52. Gehring, M. & Henikoff, S. DNA methylation dynamics in plant genomes. *Biochim. Biophys. Acta* **1769**, 276–286 (2007).
53. Ma, Y. *et al.* DNA CpG hypomethylation induces heterochromatin reorganization involving the histone variant macroH2A. *J. Cell Sci.* **118**, 1607–1616 (2005).
54. Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nature Rev. Genet.* **8**, 272–285 (2007).
55. Wilson, A. S., Power, B. E. & Molloy, P. L. DNA hypomethylation and human diseases. *Biochim. Biophys. Acta* **1775**, 138–162 (2007).
56. Bird, A. P. Gene number, noise reduction and biological complexity. *Trends Genet.* **11**, 94–100 (1995).
57. Robertson, K. D. DNA methylation and human disease. *Nature Rev. Genet.* **6**, 597–610 (2005).
58. Hsieh, C.-L. Evidence that protein binding specifies sites of DNA demethylation. *Mol. Cell Biol.* **19**, 46–56 (1999).
59. Antequera, F. & Bird, A. Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. USA* **90**, 11995–11999 (1993).
60. Cross, S., Kovarik, P., Schmidtke, J. & Bird, A. P. Non-methylated islands in fish genomes are GC-poor. *Nucleic Acids Res.* **19**, 1469–1474 (1991).
61. Glass, J. L. *et al.* CG dinucleotide clustering is a species-specific property of the genome. *Nucleic Acids Res.* **35**, 6798–6807 (2007).
62. Takai, D. & Jones, P. A. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl Acad. Sci. USA* **99**, 3740–3745 (2002).
63. Gardiner-Gardner, M. & Frommer, M. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**, 261–282 (1987).
64. Panning, B. & Jaenisch, R. DNA hypomethylation can activate *Xist* expression and silence X-linked genes. *Genes Dev.* **10**, 1991–2002 (1996).
65. Sleutels, F., Zwart, R. & Barlow, D. P. The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810–813 (2002).
66. Antequera, A. & Bird, A. in *DNA Methylation: Molecular Biology and Biological Significance* (eds J. P. Jost & H. P. Saluz) 169–185 (Birkhäuser, Basel, 1993).
67. Stoger, R. *et al.* Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. *Cell* **73**, 61–71 (1993).
68. Sutcliffe, J. S. *et al.* Deletions of a differentially methylated CpG island at the *SNRPN* gene define a putative imprinting control region. *Nature Genet.* **8**, 52–58 (1994).
69. Maatouk, D. M. *et al.* DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133**, 3411–3418 (2006).
70. Song, F. *et al.* Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proc. Natl Acad. Sci. USA* **102**, 3336–3341 (2005).
71. Yamada, Y. *et al.* A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. *Genome Res.* **14**, 247–266 (2004).
72. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genet.* **39**, 457–466 (2007).
This paper presents a large-scale analysis of DNA methylation at human promoters using microarrays. Variable methylation was particularly apparent at moderately CpG-rich promoters, suggesting a role in regulation of gene expression.
73. Petronis, A. Human morbid genetics revisited: relevance of epigenetics. *Trends Genet.* **17**, 142–146 (2001).
74. Fraga, M. F. *et al.* Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA* **102**, 10604–10609 (2005).
75. Ladd-Acosta, C. *et al.* DNA methylation signatures within the human brain. *Am. J. Hum. Genet.* **81**, 1304–1315 (2007).
76. Siegmund, K. D. *et al.* DNA methylation in the human cerebral cortex is dynamically regulated throughout the life span and involves differentiated neurons. *PLoS ONE* **2**, e895 (2007).
77. Jones, P. A. & Baylin, S. B. The epigenomics of cancer. *Cell* **128**, 683–692 (2007).
78. Krieg, A. M. Therapeutic potential of Toll-like receptor 9 activation. *Nature Rev. Drug Discov.* **5**, 471–484 (2006).
79. Hibino, T. *et al.* The immune gene repertoire encoded in the purple sea urchin genome. *Dev. Biol.* **300**, 349–365 (2006).
80. Rountree, M. R. & Selker, E. U. DNA methylation inhibits elongation but not initiation of transcription in *Neurospora crassa*. *Genes Dev.* **11**, 2383–2395 (1997).
81. Takata, M. *et al.* Rice transposable elements are characterized by various methylation environments in the genome. *BMC Genomics* **8**, 469 (2007).
82. Bennetzen, J. L., Schrick, K., Springer, P. S., Brown, W. E. & SanMiguel, P. Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. *Genome* **37**, 565–576 (1994).
83. Rabinowicz, P. D. *et al.* Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. *Nature Genet.* **23**, 305–308 (1999).
84. Ashikawa, I. Gene-associated CpG islands in plants as revealed by analyses of genomic sequences. *Plant J.* **26**, 617–625 (2001).
85. Goyon, C., Rossignol, J. L. & Faugeron, G. Native DNA repeats and methylation in *Ascobolus*. *Nucleic Acids Res.* **24**, 3348–3356 (1996).
86. Galagan, J. E. & Selker, E. U. RIP: the evolutionary cost of genome defense. *Trends Genet.* **20**, 417–423 (2004).
87. Selker, E. U. Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu. Rev. Genet.* **24**, 579–613 (1990).
88. Faugeron, G. Diversity of homology-dependent gene silencing strategies in fungi. *Curr. Opin. Microbiol.* **3**, 144–148 (2000).
89. Rhounim, L., Rossignol, J. L. & Faugeron, G. Epimutation of repeated genes in *Ascobolus immersus*. *EMBO J.* **11**, 4451–4457 (1992).

90. Salzberg, A., Fisher, O., Siman-Tov, R. & Ankri, S. Identification of methylated sequences in genomic DNA of adult *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun.* **322**, 465–469 (2004).
91. Macleod, D., Clark, V. H. & Bird, A. Absence of genome-wide changes in DNA methylation during development of the zebrafish. *Nature Genet.* **23**, 139–140 (1999).
92. Stancheva, I., El-Maarri, O., Walter, J., Niveleau, A. & Meehan, R. R. DNA methylation at promoter regions regulates the timing of gene activation in *Xenopus laevis* embryos. *Dev. Biol.* **243**, 155–165 (2002).
93. Estecio, M. R. *et al.* LINE-1 hypomethylation in cancer is highly variable and inversely correlated with microsatellite instability. *PLoS ONE* **2**, e399 (2007).
94. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**, 335–340 (1997).
95. Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H. & Mukai, T. A genomic scanning method for higher organisms using restriction sites as landmarks. *Proc. Natl Acad. Sci. USA* **88**, 9523–9527 (1991).
96. Irizarry, R. A. *et al.* Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res.* (2008).
97. Lippman, Z., Gendrel, A. V., Colot, V. & Martienssen, R. Profiling DNA methylation patterns using genomic tiling microarrays. *Nature Methods* **2**, 219–224 (2005).
98. Keshet, I. *et al.* Evidence for an instructive mechanism of *de novo* methylation in cancer cells. *Nature Genet.* **38**, 149–153 (2006).
99. Jorgensen, H. F., Adie, K., Chaubert, P. & Bird, A. P. Engineering a high-affinity methyl-CpG-binding protein. *Nucleic Acids Res.* **34**, e96 (2006).
100. Dalma-Weiszhausz, D. D., Warrington, J., Tanimoto, E. Y. & Miyada, C. G. The affymetrix GeneChip platform: an overview. *Methods Enzymol.* **410**, 3–28 (2006).
101. Nuwaysir, E. F. *et al.* Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.* **12**, 1749–1755 (2002).
102. Reinders, J. *et al.* Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. *Genome Res.* **18**, 469–476 (2008).
103. Yuan, E. *et al.* A single nucleotide polymorphism chip-based method for combined genetic and epigenetic profiling: validation in decitabine therapy and tumor/normal comparisons. *Cancer Res.* **66**, 3443–3451 (2006).
104. Bentley, D. R. Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* **16**, 545–552 (2006).
105. Rauch, T. A. *et al.* High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc. Natl Acad. Sci. USA* **105**, 252–257 (2008).
106. Bibikova, M. *et al.* High-throughput DNA methylation profiling using universal bead arrays. *Genome Res.* **16**, 383–393 (2006).

Acknowledgements
This work was supported by a grant from the Wellcome Trust. We thank A. Deaton for constructive criticism.

FURTHER INFORMATION
Adrian Bird laboratory homepage: http://www.homepages.ed.ac.uk/dmac/Bird_Lab/birdlab.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF