APPLICATIONS OF NEXT-GENERATION SEQUENCING

Principles and challenges of genome-wide DNA methylation analysis

Peter W. Laird

Abstract | Methylation of cytosine bases in DNA provides a layer of epigenetic control in many eukaryotes that has important implications for normal biology and disease. Therefore, profiling DNA methylation across the genome is vital to understanding the influence of epigenetics. There has been a revolution in DNA methylation analysis technology over the past decade: analyses that previously were restricted to specific loci can now be performed on a genome-scale and entire methylomes can be characterized at single-base-pair resolution. However, there is such a diversity of DNA methylation profiling techniques that it can be challenging to select one. This Review discusses the different approaches and their relative merits and introduces considerations for data analysis.

Transposons
Mobile DNA elements that can relocate within the genome of their hosts.

Restriction–modification system
A set of enzymes found in many bacteria and archaea that protects the host genome from genomic parasites. Restriction–modification systems consist of sequence-specific restriction endonucleases, which target invading DNA, and associated DNA methyltransferases with similar recognition sequences, which protect the host genome from the action of the endonucleases.

Mismatch repair
A DNA-repair pathway that removes mismatched bases and corrects the insertion or deletion of short stretches of (repeated) DNA.

USC Epigenome Center, University of Southern California, Keck School of Medicine, 1450 Biggy Street, Room G511B, Los Angeles, California 90089-9601, USA.
e-mail: plaird@usc.edu
doi:10.1038/nrg2732
Published online
2 February 2010

The phenotype of a cell is primarily determined by its expression profile and its response to environmental cues. Epigenetics provides stability and diversity to the cellular phenotype through chromatin marks that affect local transcriptional potential and that are preserved or regenerated during cell division. Methylation of DNA cytosine residues at the carbon 5 position ($5^{me}C$) is a common epigenetic mark in many eukaryotes and is often found in the sequence context CpG or CpHpG (H = A, T, C). When located at gene promoters, DNA methylation is usually a repressive mark. However, CpG DNA methylation is increased in the gene bodies of actively transcribed genes in plants[1–3] and mammals[4–6]. Plant CpHpG methylation is found in non-expressed transposons[7]. Bacteria and archaea also have $5^{me}C$, along with *N*-4-methylcytosine and *N*-6-methyladenine, and these modified bases participate in restriction–modification systems and mismatch repair strand discrimination, among other roles.

DNA methylation is laid down by dedicated DNA methyltransferases with highly conserved catalytic motifs. In eukaryotes, usually only a subset of potential target sequences in the genome are methylated, therefore the distribution of methyl marks can convey epigenetic information by demarcating regions of transcriptional silence or transcriptional potential. The delineation of regional DNA methylation patterns, and broader DNA methylation profiles, has important implications for understanding why certain regions of the genome can be expressed in specific developmental contexts and how epigenetic changes might enable aberrant expression patterns and disease.

DNA methylation information is erased by standard molecular biology techniques, such as cloning in bacteria and PCR, and it is not revealed by hybridization as the methyl group is located in the major groove of DNA rather than at the hydrogen bonds. Therefore, methylation-dependent pretreatments of DNA were developed to reveal the presence or absence of the methyl group at cytosine residues. The techniques based on these pretreatments were initially restricted to relatively localized regions of the genome, but many have now been scaled up to enable DNA methylation analysis on a genome-wide scale.

Although initial forays into broad DNA methylation profiling were made with two-dimensional gel electrophoresis[8–13], the era of epigenomics truly took off with the adaptation of microarray hybridization techniques from the gene expression and genomic fields to the profiling of histone modifications and DNA methylation patterns[14–21]. The current revolution in sequencing technology has recently opened the door to single-base-pair resolution whole-genome DNA methylation analysis[22–24]. Many of the high-resolution genome-wide DNA methylation profiling techniques were pioneered in model organisms with small genomes, such as *Arabidopsis thaliana*[22,24], but are now being applied to organisms with larger genomes, including mammals[6,23,25,26]. These recent developments in the rapidly evolving panoply of DNA methylation analysis techniques demand an updated synthesis of the main principles and an overview of the advantages and disadvantages of the various approaches.

This Review covers the general principles of DNA methylation analysis, with a particular emphasis on

CpG islands
In eukaryotic genomes, regions of several hundred base pairs that are not depleted of CpGs by 5-methylcytosine deamination owing to them being unmethylated in the germ line. They often overlap transcription start sites. Most definitions of CpG islands set a minimum length (for example, 200 or 500 bp), a minimum observed:expected CpG ratio (for example, greater than 0.6 or 0.65) and a minimum GC content (for example, 50% or 55%).

genome-scale DNA methylation analysis technologies. The relative merits of the different techniques are discussed, and bioinformatic challenges that are unique to DNA methylation analysis are outlined. Although some of the methods covered in this Review rely on species-specific arrays, most of the principles could be applied to any organism with $5^{me}C$, including bacteria and archaea. Some of the methods described in this Review may be applicable to 5-hydroxymethylcytosine, which was recently confirmed to be present in mammalian cells[27,28].

Distribution and detection of DNA methylation

As DNA samples are usually derived from a collection of cells, which may vary in their DNA methylation patterns, the distribution of $5^{me}C$ in a DNA sample is complex. Measurements can be made either of the pattern of methylated target sequences along individual DNA molecules or as an average methylation level at a single genomic locus across many DNA molecules[29]. Analysis of DNA methylation is further complicated by the uneven distribution of methylation target sequences, such as CpG, across the genome. As a consequence of the frequent mutation of $5^{me}C$ to thymine, these targets are depleted throughout most of the genome but are maintained in specific regions, such as CpG islands. This non-uniform distribution is an important consideration in DNA methylation analysis, as discussed below. As noted above, $5^{me}C$ is not readily distinguished from unmethylated cytosine by hybridization-based methods and, as DNA methyltransferases are not present during PCR or in biological cloning systems, DNA methylation information is erased during amplification. Some investigators have suggested that it could be feasible to maintain the pattern of methylation during PCR if an appropriate DNA methyltransferase were present in the PCR reaction. This approach would require a thermostable DNA methyltransferase with very high efficiency and maintenance fidelity and a complete lack of *de novo* methyltransferase activity, and so has not been realized to date. Therefore, almost all sequence-specific DNA methylation analysis techniques rely on a methylation-dependent treatment of the DNA before amplification or hybridization[29–32]. There are three main approaches: endonuclease digestion, affinity enrichment and bisulphite conversion. After genomic DNA has been treated with one of the methylation-dependent steps, various molecular biology techniques, including probe hybridization and sequencing, can be used to reveal the location of the $5^{me}C$ residues. The combination of different types of pretreatment followed by different analytical steps has resulted in a plethora of techniques for determining DNA methylation patterns and profiles[29,32–42]. In the following sections I discuss techniques based on the three main approaches, and a summary is provided in TABLE 1. Expression profiling of cells treated with DNA methyltransferase inhibitors and/or histone deacetylase inhibitors has also been used as a discovery tool for epigenetically silenced genes. However, it is prone to false-positive and false-negative results and is not considered to be a reliable gauge of DNA methylation at a given locus, and is therefore not discussed in this Review.

Endonuclease digestion

Restriction endonucleases are such powerful tools in molecular biology that their biological role in restriction–modification systems in bacteria and archaea is sometimes overlooked. Each sequence-specific restriction enzyme has an accompanying DNA methyltransferase that protects the endogenous DNA from the restriction defence system by methylating bases in the recognition site. Some restriction enzymes are inhibited

Table 1 | **Main principles of DNA methylation analysis**

Pretreatment	Analytical step			
	Locus-specific analysis	**Gel-based analysis**	**Array-based analysis**	**NGS-based analysis**
Enzyme digestion	• *Hpa*II-PCR	• Southern blot • RLGS • MS-AP-PCR • AIMS	• DMH • MCAM • HELP • MethylScope • CHARM • MMASS	• Methyl–seq • MCA–seq • HELP–seq • MSCC
Affinity enrichment	• MeDIP-PCR		• MeDIP • mDIP • mCIP • MIRA	• MeDIP–seq • MIRA–seq
Sodium bisulphite	• MethyLight • EpiTYPER • Pyrosequencing	• Sanger BS • MSP • MS-SNuPE • COBRA	• BiMP • GoldenGate • Infinium	• RRBS • BC–seq • BSPP • WGSBS

AIMS, amplification of inter-methylated sites; BC–seq, bisulphite conversion followed by capture and sequencing; BiMP, bisulphite methylation profiling; BS, bisulphite sequencing; BSPP, bisulphite padlock probes; CHARM, comprehensive high-throughput arrays for relative methylation; COBRA, combined bisulphite restriction analysis; DMH, differential methylation hybridization; HELP, *Hpa*II tiny fragment enrichment by ligation-mediated PCR; MCA, methylated CpG island amplification; MCAM, MCA with microarray hybridization; MeDIP, mDIP and mCIP, methylated DNA immunoprecipitation; MIRA, methylated CpG island recovery assay; MMASS, microarray-based methylation assessment of single samples; MS-AP-PCR, methylation-sensitive arbitrarily primed PCR; MSCC, methylation-sensitive cut counting; MSP, methylation-specific PCR; MS-SNuPE, methylation-sensitive single nucleotide primer extension; NGS, next-generation sequencing; RLGS, restriction landmark genome scanning; RRBS, reduced representation bisulphite sequencing; –seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing.

by 5^{me}C in the sequence context CpG, so the patterns of cutting by such enzymes can provide a read-out of DNA methylation.

The most widely used methylation-sensitive restriction enzymes for DNA methylation studies are *Hpa*II and *Sma*I, in part because they each have an isoschizomer (*Msp*I for *Hpa*II) or neoschizomer (*Xma*I for *Sma*I) that is not inhibited by CpG methylation. That is not to say that these isoschizomers are not methylation-sensitive. For example, *Msp*I is inhibited by non-CpG methylation of the first cytosine in the recognition sequence (CCGG). Therefore, *Msp*I can be used to screen for ^{me}CCG in plants[43].

The first locus-specific DNA methylation analyses in the late 1970s and early 1980s relied on digestion with methylation-sensitive restriction enzymes followed by gel electrophoresis and hybridization on Southern blots[44–47]. This method is still applicable for some locus-specific studies that require linkage of DNA methylation information across multiple kilobases, either between CpGs or between a CpG and a genetic polymorphism. Methylation-sensitive restriction digestion followed by PCR across the restriction site is a very sensitive technique that is still used in some applications today. However, it is extremely prone to false-positive results caused by incomplete digestion for reasons other than DNA methylation.

Genome-scale approaches. Over the past decade, several endonuclease-dependent genome-scale DNA methylation analysis approaches have been developed. Restriction landmark genome scanning (RLGS) was the first technique for broad DNA methylation profiling. Differences in methylation are detected as differences in the pattern of restriction fragments generated by digestion with a methylation-sensitive restriction enzyme separated by two-dimensional gel electrophoresis. RLGS has been widely used to identify imprinted loci and sites that are methylated in a cancer- or tissue-specific manner[8–13]. Simpler but less comprehensive gel electrophoretic techniques include methylation-sensitive arbitrarily primed PCR (MS-AP-PCR)[48] and amplification of inter-methylated sites (AIMS)[49], which rely on the differential patterns of arbitrarily primed PCR products generated with methylation-sensitive or -resistant restriction enzyme-digested genomic DNA. The use of RLGS, MS-AP-PCR and AIMS is decreasing as these techniques are replaced by methods that are less labour-intensive and that do not rely on gel electrophoresis.

Array hybridization. Many different techniques have been developed that couple enzymatic methods to array-based analysis. One of these is methylated CpG island amplification (MCA), which uses the differential methylation sensitivities and cutting behaviours of *Sma*I and its neoschizomer *Xma*I. It is often combined with representational difference analysis (RDA) and, more recently, with array hybridization (when it is known as MCAM)[50–54]. However, MCA provides lower resolution coverage than other enzyme-based techniques that rely on pools of restriction enzymes or on enzymes with 4-bp recognition sequences (as opposed to the 6-bp recognition sequences of *Sma*I and *Xma*I).

An alternative approach is differential methylation hybridization (DMH), which involves digestion of one pool of genomic DNA with a methylation-sensitive restriction enzyme and mock digestion of another pool. This produces parallel DNA pools that are amplified and labelled with different fluorescent dyes for two-colour array hybridization[17,18,21,55]. The relative fluorescent signal intensities can be used to extract DNA methylation information at the corresponding loci on the array. DMH is one of several techniques that incorporate pre-cutting of genomic DNA with the restriction enzyme *Mse*I. The recognition site of *Mse*I is AATT, so it cuts frequently in the genome but relatively rarely within CpG islands[56]. This digestion step is often incorporated into protocols to deplete the sample of CpG-poor DNA before further processing, such as adaptor ligation. Small genomes, such as the *A. thaliana* genome, do not require an amplification step and can be directly labelled for hybridization[43,57].

An adaptation of DMH is to use the methylation-dependent endonuclease *Mcr*BC. This provides greater sensitivity to densely methylated regions[58,59] than using a methylation-sensitive enzyme. A further modification, known as MethylScope, is to use *Mcr*BC to cut randomly sheared DNA (instead of *Mse*I-digested DNA)[60,61]. Optimization of tiling array design and data processing has resulted in improved performance for *Mcr*BC digestion-based techniques. The optimized workflow is referred to as comprehensive high-throughput arrays for relative methylation (CHARM)[62]. A further variation of DMH is to use a cocktail of methylation-sensitive restriction enzymes to digest one pool of DNA and *Mcr*BC to digest the other pool[63–65]. This is referred to as microarray-based methylation assessment of single samples (MMASS)[63]. MMASS has been used to generate libraries for Sanger sequencing[65]. Another method, known as *Hpa*II tiny fragment enrichment by ligation-mediated PCR (HELP), uses ligation-mediated PCR for the amplification of *Hpa*II or *Msp*I genomic restriction fragments followed by array hybridization[62,66,67].

Sequencing approaches. Restriction enzyme enrichment techniques are currently being adapted so that the read-out can be obtained by next-generation sequencing techniques instead of array hybridization. Sequence-based analysis is more flexible and powerful as it allows for allele-specific DNA methylation analysis, does not require an appropriately designed microarray, can cover more of the genome with less input DNA and avoids hybridization artefacts, although it is still subject to sequence library biases. Next-generation sequencing has been used to analyse the output of the HELP assay[67]. Sequencing-by-synthesis of libraries constructed from size-fractionated *Hpa*II or *Msp*I digests that are compared with randomly sheared fragments[68] is known as Methyl–seq; sequence-based analysis of *Hpa*II digestion followed by the use of a flanking cut with a type-IIS restriction enzyme (*Mme*I) and adaptor ligation is known as methylation-sensitive cut counting (MSCC)[6,41].

Isoschizomers
Pairs of structurally distinct restriction enzymes with the same recognition sequence and the same cleavage positions.

Neoschizomers
Pairs of structurally distinct restriction enzymes with the same recognition sequence but with different cleavage positions.

Imprinted
A locus with monoallelic expression determined by the parental origin of the allele.

Affinity enrichment

Chromatin immunoprecipitation (ChIP) followed by microarray hybridization (ChIP–chip) or next-generation sequencing (ChIP–seq) has proven to be a particularly useful technique for genome-wide studies of histone modifications[69–74]. Similarly, affinity enrichment of methylated regions using antibodies specific for 5^{me}C (in the context of denatured DNA) or using methyl-binding proteins with affinity for methylated native genomic DNA are proving to be particularly powerful tools for comprehensive profiling of DNA methylation in complex genomes. Affinity purification of methylated DNA was first demonstrated with the methyl-binding protein MECP2 (REF. 56). Enrichment of methylated regions by immunoprecipitation of denatured genomic DNA with an antibody specific for methylated cytosine[75], followed by hybridization to either a tiling array or to a feature microarray, such as a CpG island array, is referred to as MeDIP[76,77], mDIP[78] or mCIP[1]. These techniques have been widely used to explore the methylomes of plant[1,2], mouse[75,79–81] and human[62,76–78,82–88] cells. Confusingly, the term MCIp has also been used to describe affinity purification of methylated native DNA by a recombinant polypeptide containing the methyl-binding domain of human methyl-CpG-binding-domain protein 2 (MBD2) fused to the Fc tail of human immunoglobulin G[89–91]. More recently, approaches have been developed that use higher affinity methyl-binding proteins, including multimerized MBD1 domains[92] and protein complexes that contain the short isoform of MBD2 (MBD2b) and MBD3L1 (the latter approach is called the methylated CpG island recovery assay (MIRA)[93–95]). Also, ChIP with an antibody specific for native MBD proteins has been used as an indirect measure of the distribution of genomic methylcytosine[75,96].

These affinity-enrichment methods have mostly been combined with array hybridization, in which the input DNA and enriched DNA are labelled with different fluorescent dyes. As with other array-based analyses, methylcytosine affinity-enrichment techniques are now rapidly shifting to analysis by next-generation sequencing techniques[88]. Affinity-based methods allow for rapid and efficient genome-wide assessment of DNA methylation, but they do not yield information on individual CpG dinucleotides and require substantial experimental or bioinformatic adjustment for varying CpG density at different regions of the genome.

Chromatin immunoprecipitation
A technique that is used to identify the location of DNA-binding proteins and epigenetic marks in the genome. Genomic sequences containing the mark of interest are enriched by binding soluble DNA chromatin extracts (complexes of DNA and protein) to an antibody that recognizes the mark. Related techniques — such as methylated DNA immunoprecipitation — use antibodies to recognize DNA modifications directly.

Bisulphite conversion

The discovery that treatment of denatured genomic DNA with sodium bisulphite chemically deaminates unmethylated cytosine residues much more rapidly than methylated cytosines[97,98] spurred a revolution in DNA methylation analysis in the 1990s[99,100]. This chemical treatment of DNA effectively turns an epigenetic difference into a genetic difference — unmethylated Cs are converted to Ts (by uracil) — thereby enabling many new DNA methylation detection and analysis techniques[29,32,34–42]. Analysis of bisulphite-converted DNA was initially done by Sanger sequencing of cloned PCR products from single loci[99,101]. Many enhancements have since been developed, including the quantitative direct Sanger sequencing of PCR products[100] and the highly automated application of this approach[102].

Array hybridization. Bisulphite genomic sequencing excels at producing base-pair resolution DNA methylation information, but bisulphite-based methods are not easily adapted to array-hybridization techniques and so, until recently, were rarely used for genome-scale DNA methylation analysis. With the exception of 5^{me}C residues, bisulphite-treated DNA is comprised of three different bases instead of four. This reduced sequence complexity, and therefore greater sequence redundancy, results in decreased hybridization specificity. Hybridization of bisulphite-converted DNA either requires dedicated arrays based on the bisulphite-converted genome or must allow for substantial mismatches in the hybridization.

The first mammalian hybridization arrays for analysing bisulphite-treated DNA required the amplification of individual regions of the genome before hybridization to a dedicated oligonucleotide array[103,104] and were therefore, in essence, modestly scaled-up locus-specific assays. Smaller genomes, such as that of *A. thaliana*, can be hybridized to oligonucleotide arrays after whole-genome amplification using random tetranucleotide primers. This method is referred to as bisulphite methylation profiling (BiMP)[105]. It is worth noting that this approach relies on microarrays developed for non-bisulphite-converted DNA. Therefore, as a result of mismatches caused by the conversion of unmethylated cytosine residues, the overall hybridization signal is low, both within and outside potential methylation target sequences. Regions with dense cytosine methylation are least affected — as they retain more Cs — and yield a relatively strong signal[105]. Therefore, BiMP is only applicable to methylation-dense regions of small genomes.

Illumina has adapted its GoldenGate BeadArray technology to interrogate DNA methylation in human genomic DNA samples. Multiplexed methylation-specific primer extension of bisulphite-converted DNA at up to 1,536 different CpG sites is performed using primers that are specific for methylated and unmethylated sequences at each site[106–108]. The primers for the two different methylation states are labelled with different fluorescent dyes and the products are hybridized to bead arrays containing approximately 30 beads per CpG site. The current standard implementation for the human genome is the GoldenGate Methylation Cancer Panel I, which covers 1,505 CpG sites selected from 807 genes, but custom arrays can also be designed[109]. The GoldenGate assay can be performed on 96 samples simultaneously and is therefore well suited for the profiling of large numbers of specimens[109–116]. Illumina has also adapted its more comprehensive Infinium platform to DNA methylation analysis[68,117,118]. The Infinium platform incorporates a whole-genome amplification step after bisulphite conversion, which is followed by fragmentation and hybridization of the sample to methylation-specific DNA oligomers that are linked

to individual bead types. Each bead type corresponds to a specific DNA CpG site and methylation state. The current implementation of the Illumina Infinium assay for DNA methylation analysis (known as the HumanMethylation27 DNA Analysis BeadChip) interrogates 27,578 CpG sites from 14,495 protein-coding gene promoters and 110 microRNA gene promoters[117].

Sequencing approaches. Although the adaptation of bisulphite-converted DNA to array hybridization has been challenging, bisulphite-converted DNA is particularly well suited for sequencing-based approaches and is now enjoying a resurgence thanks to the application of next-generation sequencing platforms. Ultra-deep sequencing of a limited number of loci has been achieved by direct pyrosequencing of PCR products[119] and by sequencing of more than 100 PCR products in a single run at an average coverage of more than 1,600 reads per locus on the Roche 454 platform[120].

The main challenge in sequencing bisulphite-converted DNA arises from its low sequence complexity. Reduced representation bisulphite sequencing (RRBS) was introduced to reduce sequence redundancy by selecting only some regions of the genome for sequencing by size-fractionation of DNA fragments after *Bgl*II digestion[121] or after *Msp*I digestion[26]. These choices of restriction enzymes enrich for CpG-containing segments of the genome but do not target specific regions of interest in the genome. Targeting has been accomplished by array capture or padlock capture before sequencing. Targeted capture on fixed arrays or by solution hybrid selection can enrich for sequences targeted by a library of DNA or RNA oligonucleotides and can be performed before or after bisulphite conversion. The advantage of capture before bisulphite conversion is that a standard genomic array can be used and the methylation status of the region does not influence the capture. However, array capture is a relatively inefficient process, particularly for GC-rich target regions such as CpG islands, which are of particular interest in DNA methylation studies. Targeted capture requires relatively large quantities of native genomic DNA or amplification before hybridization, which would erase DNA methylation information in native genomic DNA. Solution hybrid selection is more efficient but allows complex cross-hybridization structures, resulting in more enrichment of off-target sequences.

Alternatively, array capture after bisulphite conversion and adaptor-mediated PCR amplification provides sufficient DNA for constructing a sequencing library (referred to here as bisulphite conversion followed by capture and sequencing (BC–seq))[122]. Although this involves small amounts of input DNA, it requires multiple permutations of capture oligonucleotides to reflect different methylation states and is subject to potential measurement error caused by varying capture efficiency of different methylation states[122].

Padlock capture provides improved enrichment efficiency by combining the increased annealing specificity of two tethered probes, and subsequent amplification with universal primers allows for a more uniform representation than amplification with locus-specific primers. Several groups have combined padlock capture of mammalian DNA with next-generation sequencing to achieve 90–99% target specificity after pooling tens of thousands of probes for bisulphite-converted DNA in a technique called bisulphite padlock probes (BSPP)[6,25,41,123]. However, as capture occurs after bisulphite conversion, there is again the concern that the methylation state could influence capture efficiency and therefore distort the DNA methylation measurements, even if all methylation permutations are included in the capture probes. Alternatively, capture probes can be designed to avoid CpG dinucleotides in organisms in which DNA methylation is largely restricted to CpGs. However, this design approach severely limits the comprehensive capture of CpG-rich regions, such as CpG islands.

The ultimate comprehensive single-base-pair resolution DNA methylation analysis technique is whole-genome bisulphite sequencing. Whole-genome shotgun bisulphite sequencing (WGSBS) has been achieved on the Illumina Genome Analyzer platform for small eukaryotic genomes, such as *A. thaliana*[22,24], and for mammalian DNA[23]. Increased read lengths and paired-end sequencing strategies have aided the implementation of WGSBS[124–126], although approximately a tenth of the CpG dinucleotides in the mammalian genome remain refractory to alignment of bisulphite-converted reads.

Other approaches. An alternative approach to sequencing or hybridization is detection by matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry[127–129]. Although this approach requires gene-specific amplification, and should therefore be considered a candidate gene method, it is amenable to automation, as implemented in the EpiTYPER platform developed by Sequenom, which relies on gene-specific amplification of bisulphite-treated DNA followed by *in vitro* transcription, base-specific RNA cleavage and MALDI-TOF analysis[128,129]. Although it is not a genome-wide technology, it can provide quantitatively accurate results at multiple CpG dinucleotides for hundreds of gene loci[128,130] and can be reliably applied to pooled DNA samples to obtain group averages for valuable samples[130].

Relative merits of different approaches

A straightforward comparison of DNA methylation analysis methods is hampered by the complexity and diversity of the various techniques. Many approaches have competing strengths and weaknesses. The choice of technique will be influenced by the number of samples and the quality and quantity of DNA, as well as the desired coverage and resolution. Also, it is necessary to take into account the organism that is being studied: array-based analysis requires that a suitable array for the species of interest is available, whereas sequence-based analyses are generally applicable to any species for which a reference genome exists. A summary of features and potential sources of bias is shown in TABLE 2.

Array capture
A method for enriching whole genomic DNA for many regions of interest by hybridization to an array containing RNA or DNA sequences complementary to the regions of interest.

Padlock capture
A method for simultaneously capturing and amplifying large numbers of regions of interest from whole genomic DNA. Each padlock probe has two complementary oligonucleotide sequences that flank a region of interest. The sequences are joined by a loop of DNA that ensures efficient joint hybridization and contains sequences for PCR with universal primers.

Solution hybrid selection
A method for enriching whole genomic DNA for many regions of interest by hybridization to a complex library of RNA or DNA sequences in solution, followed by retrieval of the annealed hybrids.

Table 2 | **Features and sources of bias for various techniques**

Technology	Features							Potential sources of bias							
	Unambiguous identification of CpG measured	In *cis* co-methylation information	Non-CpG methylation information	Allele-specific measurement capability	Good coverage of regions with low CpG density	Compatible with low amounts of input DNA	Full repeat-masked genome coverage	Copy-number variation bias	Fragment size bias	Incomplete bisulphite conversion bias	Bisulphite PCR bias	Cross-hybridization bias	DNA methylation status bias	GC content bias	CpG density bias
Infinium	(•)					•				•	•	•			
Enzyme–chip	(•)	(•)			(•)				•			•		•	
MeDIP–chip							•	•				•		•	•
BSPP	•	•	•	•						•	•		•		
BC–seq	•	•	•	•						•	•		•		
RRBS	•	•	•	•		•				•	•				
Enzyme–seq	•	•		•	(•)	•			•						
MeDIP–seq				•			•	•						•	•
WGSBS	•	•	•	•	•	•	•			•	•				

'•' indicates that the method has this feature or potentially has this bias; '(•)' indicates that the method has this feature to a limited extent or in some circumstances. BC–seq, bisulphite conversion followed by capture and sequencing; BSPP, bisulphite padlock probes; –chip, followed by microarray; MeDIP, methylated DNA immunoprecipitation; RRBS, reduced representation bisulphite sequencing; –seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing.

Sample requirements. The many DNA methylation analysis techniques differ vastly in sample requirements. Some techniques, such as RLGS, require in excess of 2 μg of high-purity, high-molecular-weight genomic DNA. Others, such as the Illumina GoldenGate technology, are compatible with degraded DNA extracted from formalin-fixed, paraffin-embedded samples[131]. As a general rule, methods based on methylation-sensitive or methylation-dependent endonuclease treatment tend to require DNA of high purity, quantity and integrity, although adaptations of these techniques can reduce sample quantity requirements[67]. Affinity-enrichment techniques are more tolerant of DNA impurity and integrity, but many require substantial quantities of input genomic DNA to produce sufficient enriched output DNA. The ratio of input DNA to affinity reagent can affect the enrichment efficiencies of regions with varying methylcytosine density and should be carefully considered when making adjustments to the amount of input DNA in affinity-enrichment methods. The fairly high DNA quantity and quality requirements of many enzyme-dependent and affinity-enrichment methods have resulted in the use of cell lines, as opposed to primary tissues, in many early mammalian profiling studies. Low quantities of input DNA can be mitigated by amplification methods, such as PCR and whole-genome amplification, but only if amplification is implemented after the methylation-dependent step, otherwise the methylation information will be erased. For example, the Illumina Infinium platform incorporates whole-genome amplification after bisulphite conversion and as a result can profile 27,578 CpG sites using less than 1 μg of input genomic DNA.

MeDIP and bisulphite-based techniques both use ssDNA and are therefore compatible with DNA samples that have been previously denatured. Enzyme-based techniques and MIRA both act on dsDNA. This has the potential advantage that these methods could in principle be applied to native DNA before purification — for example, they could be used to detect tumour-derived methylated DNA in serum or plasma.

Bisulphite treatment not only requires DNA denaturation before treatment but also causes substantial DNA degradation, and purification is needed to remove the sodium bisulphite. For these reasons, input DNA for many bisulphite-based methods can be of low purity and integrity. In fact, crude lysis without subsequent purification can be sufficient preparation of samples for bisulphite-based analysis.

PCR amplification is used in many locus-specific applications of bisulphite-treated DNA. It is common practice in PCR to use dUTP during amplification and to pretreat subsequent samples with uracil DNA glycosylase to remove carry-over PCR products from prior reactions. This quality-assurance step is incompatible with standard bisulphite conversion methods, as the bisulphite conversion deaminates unmethylated cytosines to uracil, which would lead to degradation of the bisulphite-converted genomic DNA by the uracil DNA glycosylase. Bisulphite deamination without desulphonation before uracil DNA glycosylase treatment can resolve this incompatibility[132,133].

Sample throughput. Sample throughput is generally limited by the cost of labour and reagents. Highly automated techniques have low labour costs but are often

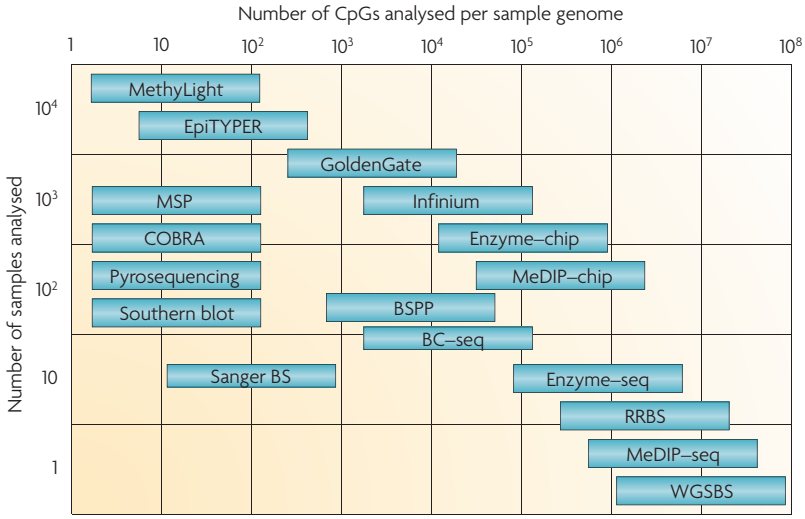

Figure 1 | **Sample throughput versus genome coverage.** A plot of sample throughput against genome coverage for various DNA methylation techniques. Throughput is determined by the number of samples that can be analysed per experiment, based on large eukaryotic genomes. Coverage is determined by the number of CpGs in the genome that can be analysed per experiment. BC–seq, bisulphite conversion followed by capture and sequencing; BS, bisulphite sequencing; BSPP, bisulphite padlock probes; –chip, followed by microarray; COBRA, combined bisulphite restriction analysis; MeDIP, methylated DNA immunoprecipitation; MSP, methylation-specific PCR; RRBS, reduced representation bisulphite sequencing; –seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing.

reagent-intensive. The labour involved in many of the current enzyme-based and affinity-enrichment methods precludes the processing of large numbers of samples. Most studies that have used array-hybridization or next-generation sequencing applications have reported results for just a few samples at most. By contrast, semi-automated bisulphite-based technologies, such as the Sequenom EpiTYPER and Illumina GoldenGate platforms, excel at characterizing large numbers of samples, and studies involving in excess of 100 primary human samples have been reported[109–111,114,115]. Cost, sample throughput and coverage are interdependent. FIG. 1 compares throughput with genome coverage (see also below).

Genome coverage and resolution. The concept of resolution is not straightforward when comparing various technologies with differing DNA fragmentation sizes and non-uniform spacing of targeted regions. As an example, bisulphite Sanger sequencing has been used to assess approximately 40,000 different CpG dinucleotides located on 2,524 different amplicons[102], whereas the Illumina Infinium DNA methylation platform has been used to interrogate 27,578 different CpG dinucleotides located at 14,495 different gene promoters[68]. As the methylation states of CpG dinucleotides located within the same amplicon are highly correlated[102], the number of independent epigenomic data points obtained on the Illumina platform is much higher (14,495 loci versus 2,524) than in the Sanger sequencing example, even though the total number of CpG dinucleotides assessed in the Sanger sequencing example is slightly higher (40,000 versus 27,578).

The coverage and resolution of enzyme-based methods are closely linked to the distribution of recognition sequences throughout the genome. Techniques relying on *Hpa*II/*Msp*I sites (*Hpa*II and *Msp*I are isoschizomers) can assess up to 98.5% of CpG islands and 91.1% of NCBI Reference Sequence (RefSeq) promoters in the human genome, but at modest resolution within each locus as a result of the average distance between *Hpa*II/*Msp*I sites within CpG islands[67]. The low-specificity of the recognition site for *Mcr*BC ($R^{me}C…R^{me}C$, in which R is A or G) suggests that *Mcr*BC-based analysis can reveal DNA methylation at almost any methylated CpG island in the genome, but the requirement of two methylated cytosines within a restricted distance from each other in *cis* (40–3,000 bp) results in a nonlinear relationship between DNA methylation density and cutting efficiency. The effective coverage and resolution of enzyme-based array hybridization techniques is a function of the distribution of potential cleavage sites and the composition of the hybridization array[62]. Likewise, the coverage and resolution of affinity methods depends on the genomic distribution of potential affinity targets (for example, CpG) and array composition. Using next-generation sequencing instead of microarrays to generate data output from enzyme-, affinity- and bisulphite-based methods eliminates some of the limitations that are attributed to array composition. However, the base-composition and fragment-size biases of sequencing platforms can introduce new coverage limitations and/or measurement error[134].

Furthermore, the sequence context of the methylcytosine residue needs to be considered when selecting an appropriate technique. A brief summary of these considerations is provided in BOX 1.

Accuracy and reproducibility. DNA methylation profiling methods vary in their accuracy, reproducibility, types of bias and ability to detect in-*cis* methylation patterns, as well as the extent to which they are influenced by target sequence density or DNA methylation density (TABLE 2). Methylation-sensitive restriction enzyme-based methods are able to resolve methylation differences in low-CpG-density regions, whereas *Mcr*BC-based and affinity-based methods perform better for CpG-rich regions[62]. Many of the enzyme-based techniques reveal DNA methylation information by DNA fragment length differences. Fragment length can influence array hybridization efficiency, but is particularly problematic in sequence-based analysis, in which library construction efficiency is strongly influenced by fragment length[62,134].

Next-generation sequencing is often used differently in affinity or enzyme pretreatment experiments and bisulphite genomic sequencing experiments. In affinity or enzyme pretreatment, short-read sequencing is used to determine the prevalence of different regions of the genome in the enriched sample by counting the number of reads that are uniquely alignable to regions of the genome. By contrast, bisulphite sequencing extracts the alignment and DNA methylation information from the sequence itself. The read count methods are prone to sources of bias, such as GC content, fragment size and

Box 1 | **Considerations for detection of methylation in different sequence contexts**

Of the approaches for detecting DNA methylation discussed in this Review, enzyme-based methods are the most narrowly confined by the sequence context of the 5-methylcytosine ($5^{me}C$) residues. Most techniques that rely on enzymes have been designed for either an ^{me}CpG context (in the case of restriction enzymes) or an $R^{me}C$ (R = A or G) sequence context (in the case of *Mcr*BC). Therefore, enzyme-based techniques will tend to be oblivious to ^{me}C residues in other sequence contexts.

Most affinity-based methods will enrich for $5^{me}C$ regardless of sequence context. However, the array hybridization or sequencing used subsequently for analysis will not be able to distinguish between CpG and non-CpG methylation. Bisulphite sequencing is the most reliable method for identifying ^{me}C in any sequence context. However, verification following more extensive bisulphite treatment and/or recurrent observation in independent samples is recommended to distinguish apparent CpH (H = A, T, C) methylation from incomplete bisulphite conversion.

Hydroxymethylcytosine is best detected using thin-layer chromatography, high-performance liquid chromatography and/or mass spectrometry[27,28]. None of these techniques is well-suited for locus-specific analysis. Hydroxymethylcytosine is likely to be indistinguishable from $5^{me}C$ in most restriction enzyme-based techniques and in bisulphite conversion techniques. It remains to be seen whether affinity-based methods show cross-affinity to hydroxymethylcytosine.

copy-number variations in the source DNA, that affect the likelihood that a particular region is included in the sequenced fragments[134]. The bisulphite-based methods are also subject to these effects, but they do not influence the DNA methylation measurement itself, as this information is extracted from the sequence.

Bisulphite-based methods tend to be fairly accurate and reproducible. The major sources of bias and measurement error are incomplete bisulphite conversion and differential PCR efficiency for methylated versus unmethylated versions of the same sequence[135]. Incomplete bisulphite conversion can arise from incomplete denaturation before bisulphite treatment or reannealing during the bisulphite conversion. Bisulphite conversion destroys the self-complementarity of DNA. Therefore, repeated denaturation cycles during the bisulphite conversion process are usually sufficient to ensure near complete conversion, as reannealing becomes progressively less likely. The completion of bisulphite conversion can be monitored with assays that are specifically designed to detect incomplete conversion[136] by monitoring the conversion of spiked DNA controls or by retention of non-target sequence cytosine dinucleotides, such as CpH in most mammalian genomes, with the caveat that native CpH methylation may exist, particularly in embryonic stem cells[23,26,120]. It should be noted that over-treatment with bisulphite not only degrades DNA but can also lead to an increased incidence of methylated cytosines converting to thymine residues[98], which results in under-reporting of DNA methylation. This can also be monitored by spiked methylated DNA controls.

Allele specificity, polymorphisms and copy-number variation. CpG dinucleotides are common sites of polymorphism, owing to deamination of methylated cytosines during evolution. The deamination event can be present in the reference genome or in the experimental genome, and the polymorphism is indistinguishable from bisulphite-induced deamination of an unmethylated cytosine on one strand of the experimental genome. The high frequency of CpG polymorphisms can result in errors in DNA methylation measurement. Indeed, polymorphisms in Illumina GoldenGate BeadArray primers can masquerade as sample-specific epigenetic variation[110]. In sequence-based analysis, aside from flagging known SNPs, the standard method for discriminating between evolutionary and bisulphite-induced deamination is to resequence the non-bisulphite-converted experimental genome. However, a more efficient method relies on the recognition that a SNP caused by evolutionary deamination of ^{me}C to T will have been propagated on the opposing DNA strand as an A, whereas bisulphite deamination of an unmethylated cytosine will leave the G on the opposing strand unaffected[137]. Sequencing of both DNA strands of bisulphite-converted DNA can therefore discriminate between a CpG SNP and an unmethylated CpG without the need to resequence the non-bisulphite converted genome. Sufficient coverage can resolve heterozygotes and randomly hemimethylated states. Interestingly, the methylation state of CpG dinucleotides can be associated with *cis*-linked SNPs[138]. Therefore, sequence-based technologies that can provide allele-specific DNA methylation information are preferable to allele-agnostic methylation assays.

Enzyme-based analyses and enrichment techniques both rely on the relative enrichment or depletion of regions of the genome. The occurrence of aneuploidy in cancer cells can lead to inaccuracies in methylation measurements if copy-number alterations are not appropriately controlled. Enzyme-based methods that rely on the balanced measurement of the ratio between methylated and unmethylated versions of a sequence are less prone to measurement error caused by copy-number alterations than methods that rely on unbalanced enrichment for either just methylated or just unmethylated sequences. Affinity-based methods are particularly prone to measurement error caused by copy-number alterations, as they are oblivious to the unmethylated version of a sequence. The inclusion of a control produced from the experimental sample, such as DNA artificially methylated by M.*Sss*I[84] or digested with *Msp*I[67], can be used to normalize copy number variations. Bisulphite-based methods, such as the Illumina GoldenGate platform, are less susceptible to copy-number variation[139].

Sensitivity. Genome-scale DNA methylation profiling techniques are not particularly well suited for the detection of low-frequency DNA methylation states in a DNA

Hemimethylated
Methylation of a residue on one strand within a palindromic target sequence but not of the corresponding residue within the palindromic target sequence on the complementary DNA strand. Not be confused with monoallelic methylation, in which one allele of a locus is methylated in a diploid organism.

Table 3 | **Bioinformatic resources**

Resource	Purpose	URL	Refs
Batman	MeDIP DNA methylation analysis tool	http://td-blade.gurdon.cam.ac.uk/software/batman	88
BDPC	DNA methylation analysis platform	http://biochem.jacobs-university.de/BDPC	153
BSMAP	Whole-genome bisulphite sequence mapping	http://code.google.com/p/bsmap	154
CpG Analyzer	Windows-based program for bisulphite DNA	-	155
CpGcluster	CpG island identification	http://bioinfo2.ugr.es/CpGcluster	156
CpGFinder	Online program for CpG island identification	http://linux1.softberry.com	-
CpG Island Explorer	Online program for CpG Island identification	http://bioinfo.hku.hk/cpgieintro.html	157
CpG Island Searcher	Online program for CpG Island identification	http://cpgislands.usc.edu	158
CpG PatternFinder	Windows-based program for bisulphite DNA	-	159
CpG Promoter	Large-scale promoter mapping using CpG islands	http://www.cshl.edu/OTT/html/cpg_promoter.html	160
CpG ratio and GC content Plotter	Online program for plotting the observed:expected ratio of CpG	http://mwsross.bms.ed.ac.uk/public/cgi-bin/cpg.pl	-
CpGviewer	Bisulphite DNA sequencing viewer	http://dna.leeds.ac.uk/cpgviewer	161
CyMATE	Bisulphite-based analysis of plant genomic DNA	http://www.gmi.oeaw.ac.at/en/cymate-index/	162
EMBOSS CpGPlot/ CpGReport/Isochore	Online program for plotting CpG-rich regions	http://www.ebi.ac.uk/Tools/emboss/cpgplot/index.html	-
Epigenomics Roadmap	NIH Epigenomics Roadmap Initiative homepage	http://nihroadmap.nih.gov/epigenomics	-
Epinexus	DNA methylation analysis tools	http://epinexus.net/home.html	-
MEDME	Software package (using R) for modelling MeDIP experimental data	http://espresso.med.yale.edu/medme	163
methBLAST	Similarity search program for bisulphite-modified DNA	http://medgen.ugent.be/methBLAST	164
MethDB	Database for DNA methylation data	http://www.methdb.de	165–168
MethPrimer	Primer design for bisulphite PCR	http://www.urogene.org/methprimer	169
methPrimerDB	PCR primers for DNA methylation analysis	http://medgen.ugent.be/methprimerdb	164
MethTools	Bisulphite sequence data analysis tool	http://www.methdb.de	170
MethyCancer Database	Database of cancer DNA methylation data	http://methycancer.psych.ac.cn	171
Methyl Primer Express	Primer design for bisulphite PCR	http://www.appliedbiosystems.com/methylprimerexpress	-
Methylumi	Bioconductor package for analysing DNA methylation data from Illumina platforms	http://www.bioconductor.org/packages/bioc/html/methylumi.html	-
Methylyzer	Bisulphite DNA sequence visualization tool	http://ubio.bioinfo.cnio.es/Methylyzer/main/index.html	-
mPod	Genome-wide DNA methylation viewer integrated with the Ensembl genome browser	http://www.compbio.group.cam.ac.uk/Projects/p4meth.html	172
PubMeth	Database of DNA methylation literature	http://www.pubmeth.org	173
QUMA	Quantification tool for methylation analysis	http://quma.cdb.riken.jp	174
TCGA Data Portal	Database of TCGA DNA methylation data	http://cancergenome.nih.gov/dataportal	-

BDPC, Bisulphite Sequencing Data Presentation and Compilation; BSMAP, Bisulphite Sequence Mapping Program; CyMATE, Cytosine Methylation Analysis Tool for Everyone; EMBOSS, European Molecular Biology Open Software Suite; MeDIP, methylated DNA immunoprecipitation; MEDME, Modelling Experimental Data with MeDIP Enrichment; NIH, US National Institutes of Health; QUMA, Quantification Tool For Methylation Analysis; TCGA, The Cancer Genome Atlas.

sample, as sensitivity is generally a function of sequencing depth. Therefore, it is relatively expensive to achieve sensitive detection of low-abundance DNA methylation patterns. The introduction of next-generation sequencing has opened the door to efficient deep sequencing of bisulphite PCR products. For example, ultra-deep sequencing on the Roche 454 pyrosequencing platform can generate thousands of reads per locus, therefore providing the potential for the detection of rare DNA methylation variants[119,120].

Bisulphite genomic sequencing can provide DNA methylation information from individual DNA molecules. This feature can be used to investigate cell lineages[140] and DNA–protein interactions[141]. Analysis of single molecules was initially achieved by subcloning of PCR products[99,100] but has more recently been adapted to the analysis of digital PCR products by limiting dilution[142–144]. This digitization of bisulphite PCR can be used to eliminate bias caused by the different amplification efficiencies of methylated and unmethylated sequences[144].

Epigenomic bioinformatics

The development of computational tools and resources for DNA methylation analysis is accelerating rapidly[145]. TABLE 3 provides a list of some available resources. However, the issues related to bioinformatic analysis are complex and are not discussed comprehensively here. The many considerations for hybridization-based techniques include image and scanning artefacts, background correction, batch and array normalization and adjustments for GC content and CpG density[88]. Sequence-based analyses involve alignment to a reference genome, collapsing of clonal reads, read counts or bisulphite-based analysis[23], and further data analysis (which for affinity-based methods might incorporate read direction). Several important statistical considerations that are specific to DNA methylation measurements are highlighted in BOX 2.

Conclusions and future directions

DNA methylation analysis has undergone a veritable revolution over the past decade. We are on the verge of witnessing a deluge of sequence-based DNA methylation data, along with other epigenomic and genomic data types. This data production will be coordinated in national and international consortia[146,147]. The bottleneck in DNA methylation advances will increasingly

β distribution
A continuous probability distribution with an interval between 0 and 1. Two positive parameters, α and β, are used to define β distributions.

Median absolute deviation
A measure of statistical dispersion that is less influenced by outliers and extreme values than standard deviation. It is defined as the median of the collection of absolute deviations from the data set's median.

Quantile normalization
A method for equalizing the total signal intensities and distributions of probe signal strengths among arrays or among colour channels on an array. It sorts all probes by signal strength and then matches probes at each rank position among arrays and forces the values at each rank position to be equal. An identical distribution of probe signal strengths among the arrays or colour channels is obtained.

LOESS normalization
A computationally intensive method in which a polynomial regression is fitted to each point in the data and more weight is given to data nearer the point of interest. It is often applied to hybridization array data to remove differences in global signal intensity among data sets or colour channels.

MA plot
A representation of microarray data in which M (vertical axis) is the intensity ratio between the red (R) and green (G) colour channels (M = log(R/G)) and A (horizontal axis) is the mean intensity (A = (logR + logG)/2). This representation is often used as a basis for normalizing microarray data, with the underlying assumptions that dye bias is dependent on signal intensity, that the majority of probes do not have very different signal intensities among channels and that approximately the same number of probes in each channel have signal intensities that are stronger than the equivalent probes in the other channel.

Box 2 | An introduction to statistical issues in DNA methylation analysis

DNA samples are generally derived from mixtures of cell populations with heterogeneous DNA methylation profiles. Bisulphite sequencing-based approaches can provide a discrete DNA methylation pattern corresponding to a single original DNA molecule. However, most techniques provide an average measurement across the sampled DNA molecules for a particular locus or CpG dinucleotide. For biological and historical reasons, this is usually expressed as a fraction or percentage methylation of the total molecules assessed. For most platforms, DNA methylation measurements represent absolute measurements for a given sample, whereas gene expression measurements are usually expressed as a differential comparison between samples. Usually, strand-specific DNA methylation or hemimethylation is not considered, although monoallelic methylation — including, but not restricted to, imprinted regions — does occur *in vivo*. The resulting measurement scale is therefore 0 to 1, or 0 to 100%, with 0 indicating that no methylated molecules were identified and 1 or 100% indicating that all identified molecules were methylated. The fraction is calculated as M/(M + U), in which M represents the signal for methylated molecules and U the signal for unmethylated molecules.

It is important to note that this is a finite scale (β distribution) that has different statistical properties to the infinite scale that is commonly used in gene expression array analysis. For example, β-distributed DNA methylation measurements are not normally distributed and the variance of measurements within a finite scale is influenced by the mean of the measurements; the variance of measurements with a mean near to the middle of the range can be much larger than the variance of measurements with a mean close to the limits (0 and 1). Therefore, sorting features (for example, genes or probes) by standard deviation will result in a bias towards features with mean methylation in the middle of the range. Reducing the number of features by selecting probes with high standard deviation or median absolute deviation is a common step in unsupervised analyses of microarray data. Therefore, variance-stabilizing data transformations or selection of probes based on a different metric should be considered. The behaviour of the distance metrics used to compare measurements across samples (such as fold-change, log-ratio or simple subtraction) is different for β-distributed fractions or percentages in DNA methylation measurements than for infinite ratio scales. This should be given careful consideration when selecting the most appropriate metric. Clustering and partitioning methods used to identify subgroups of samples with similar DNA methylation profiles are being developed for β-distributed DNA methylation data[115].

An alternative method is to report the ratio of methylated to unmethylated molecules for a particular locus (M/U), usually as a $\log_2$(M/U) ratio[62,152]. This has gained acceptance with the increased use of microarrays in DNA methylation analysis[62], and has the advantage that many of the tools developed for gene expression data can be readily applied. However, several points should be considered. First, many data normalization methods for gene expression data assume that many genes are not expressed and that most genes are not differentially expressed; similar assumptions cannot be made for DNA methylation measurements. More importantly, M and U are generally not independent, which violates an assumption of many of the statistical approaches for gene expression microarrays. M and U are biologically inversely correlated, but many DNA methylation platforms show them to be positively correlated if signal strength is strongly influenced by genomic location or by probe sequence — that is, if M and U are both derived from either a strongly or weakly hybridizing region. Furthermore, the M/U ratio may be inappropriate for situations in which the platform is measuring across multiple CpG dinucleotides. For example, if two CpGs are being measured and each CpG is methylated at 10%, M/U will equal 0.1 if the platform assesses each CpG independently, but will equal $(0.1 \times 0.1)/(0.9 \times 0.9) = 0.01$ if the platform only registers methylation when both CpGs are methylated and the absence of methylation when both CpGs are unmethylated. As the number of locally grouped CpGs increases, these distortions become quite pronounced.

An important distinction between DNA methylation measurements and gene expression measurements is that the total amount of CpG methylation can differ substantially among samples. Therefore, normalization methods, such as quantile normalization and LOESS normalization (which is often applied to data represented on MA plots), that assume similar total signal across samples can remove real biological signal. It is important to note that the fluorescent dyes in the Infinium DNA methylation assay do not correspond to DNA methylation states. Therefore, normalization algorithms based on dye channel comparisons cannot be applied to Infinium DNA methylation data without modification.

DNA methylation bioinformatics, biostatistics and computational biology are fertile areas of research that are under rapid development. TABLE 3 lists several resources for DNA methylation data analysis that are currently available.

Targeted indexing
Indexing refers to the incorporation of short sequences as tagged codes during the construction of a sequencing library, followed by the simultaneous parallel sequencing of libraries from many sources. The source of the DNA sequence for each read can be deduced from the index. This technique can be combined with targeted sequencing of regions of interest enriched by hybrid selection.

shift from data production to data analysis. Enzyme-based and affinity enrichment-based DNA methylation analysis techniques have enjoyed a resurgence recently, and have benefited by being easier to adapt to hybridization arrays than bisulphite-based methods. However, as array-based approaches are replaced by sequence-based analysis, I anticipate that the single-base-pair resolution of bisulphite sequencing and the fact that the methylation information is extracted from the sequence rather than from read counts will shift the balance back towards bisulphite-based analysis. Projects that require the analysis of large numbers of samples will still rely on array-based approaches for several more years, although targeted indexing will allow efficient sequence-based analysis of larger numbers of samples. As sequencing costs drop, it may become less expensive to lift out regions of interest bioinformatically from whole-genome bisulphite sequences rather than to target regions of interest experimentally by array capture or padlock amplification. Single-molecule sequencing approaches[148,149] are particularly well suited for bisulphite-based analysis, as they avoid bias introduced by differential amplification of methylation-derived states. Nanopore sequencing offers the potential for direct sequencing of $5^{me}C$ without bisulphite treatment[150,151], and this might herald the next revolution in high-throughput DNA methylation analysis.

1. Zhang, X. *et al.* Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* **126**, 1189–1201 (2006).
This study provided the first comprehensive DNA methylation analysis of a eukaryotic genome using whole-genome tiling arrays on affinity-enriched DNA.
2. 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).
3. Zhang, X., Shiu, S., Cal, A. & Borevitz, J. O. Global analysis of genetic, epigenetic and transcriptional polymorphisms in *Arabidopsis thaliana* using whole genome tiling arrays. *PLoS Genet.* **4**, e1000032 (2008).
4. Jones, P. A. The DNA methylation paradox. *Trends Genet.* **15**, 34–37 (1999).
5. Hellman, A. & Chess, A. Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141–1143 (2007).
6. Ball, M. P. *et al.* Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature Biotech.* **27**, 361–368 (2009).
7. Miura, A. *et al.* An *Arabidopsis* jmjC domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J.* **28**, 1078–1086 (2009).
8. Costello, J. F. *et al.* Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nature Genet.* **24**, 132–138 (2000).
9. Allegrucci, C. *et al.* Restriction landmark genome scanning identifies culture-induced DNA methylation instability in the human embryonic stem cell epigenome. *Hum. Mol. Genet.* **16**, 1253–1268 (2007).
10. Kawai, J. *et al.* Methylation profiles of genomic DNA of mouse developmental brain detected by restriction landmark genomic scanning (RLGS) method. *Nucleic Acids Res.* **21**, 5604–5608 (1993).
11. Plass, C. *et al.* Identification of *Grf1* on mouse chromosome 9 as an imprinted gene by RLGS-M. *Nature Genet.* **14**, 106–109 (1996).
12. 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).
13. Hayashizaki, Y. *et al.* Restriction landmark genomic scanning method and its various applications. *Electrophoresis* **14**, 251–258 (1993).
This study demonstrated the principle of genome-scale DNA methylation analysis using RLGS.
14. Hatada, I. *et al.* A microarray-based method for detecting methylated loci. *J. Hum. Genet.* **47**, 448–451 (2002).
15. Balog, R. P. *et al.* Parallel assessment of CpG methylation by two-color hybridization with oligonucleotide arrays. *Anal. Biochem.* **309**, 301–310 (2002).
16. van Steensel, B., Delrow, J. & Henikoff, S. Chromatin profiling using targeted DNA adenine methyltransferase. *Nature Genet.* **27**, 304–308 (2001).
17. Yan, P. S. *et al.* CpG island arrays: an application toward deciphering epigenetic signatures of breast cancer. *Clin. Cancer Res.* **6**, 1432–1438 (2000).
18. Huang, T. H., Perry, M. R. & Laux, D. E. Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet.* **8**, 459–470 (1999).
19. El-Osta, A. & Wolffe, A. P. Analysis of chromatin-immunopurified MeCP2-associated fragments. *Biochem. Biophys. Res. Commun.* **289**, 733–737 (2001).
20. Beck, S., Olek, A. & Walter, J. From genomics to epigenomics: a loftier view of life. *Nature Biotech.* **17**, 1144 (1999).
21. Yan, P. S. *et al.* Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res.* **61**, 8375–8380 (2001).
22. Cokus, S. J. *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).
23. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
24. Lister, R. *et al.* Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
References 22–24 provided the first single-base-pair resolution WGSBS of the *A. thaliana* (references 22 and 24) and human (reference 23) genomes.
25. Deng, J. *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature Biotech.* **27**, 353–360 (2009).
26. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 (2008).
This study provided the first genome-scale single-base-pair resolution DNA methylation map of mammalian genomes by RRBS.
27. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
28. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
29. Laird, P. W. The power and the promise of DNA methylation markers. *Nature Rev. Cancer* **3**, 253–266 (2003).
30. Schones, D. E. & Zhao, K. Genome-wide approaches to studying chromatin modifications. *Nature Rev. Genet.* **9**, 179–191 (2008).
31. Fraga, M. F. & Esteller, M. DNA methylation: a profile of methods and applications. *Biotechniques* **33**, 632–649 (2002).
32. Pomraning, K. R., Smith, K. M. & Freitag, M. Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods* **47**, 142–150 (2009).
33. Callinan, P. A. & Feinberg, A. P. The emerging science of epigenomics. *Hum. Mol. Genet.* **15**, R95–R101 (2006).
34. Beck, S. & Rakyan, V. K. The methylome: approaches for global DNA methylation profiling. *Trends Genet.* **24**, 231–237 (2008).
35. Ushijima, T. Detection and interpretation of altered methylation patterns in cancer cells. *Nature Rev. Cancer* **5**, 223–231 (2005).
36. Hatada, I. Emerging technologies for genome-wide DNA methylation profiling in cancer. *Crit. Rev. Oncog.* **12**, 205–223 (2006).
37. *Wilson*, I. M. *et al.* Epigenomics: mapping the methylome. *Cell Cycle* **5**, 155–158 (2006).
38. Lister, R. & Ecker, J. R. Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res.* **19**, 959–966 (2009).
39. Lieb, J. D. *et al.* Applying whole-genome studies of epigenetic regulation to study human disease. *Cytogenet. Genome Res.* **114**, 1–15 (2006).
40. Jacinto, F. V., Ballestar, E. & Esteller, M. Methyl-DNA immunoprecipitation (MeDIP): hunting down the DNA methylome. *Biotechniques* **44**, 35–43 (2008).
41. Berman, B. P., Weisenberger, D. J. & Laird, P. W. Locking in on the human methylome. *Nature Biotech.* **27**, 341–342 (2009).
42. Jeddeloh, J. A., Greally, J. M. & Rando, O. J. Reduced-representation methylation mapping. *Genome Biol.* **9**, 231 (2008).
43. Tompa, R. *et al.* Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr. Biol.* **12**, 65–68 (2002).
44. van der Ploeg, L. H. & Flavell, R. A. DNA methylation in the human γδβ-globin locus in erythroid and nonerythroid tissues. *Cell* **19**, 947–958 (1980).
45. Waalwijk, C. & Flavell, R. A. DNA methylation at a CCGG sequence in the large intron of the rabbit β-globin gene: tissue-specific variations. *Nucleic Acids Res.* **5**, 4631–4634 (1978).
46. Kaput, J. & Sneider, T. W. Methylation of somatic vs germ cell DNAs analyzed by restriction endonuclease digestions. *Nucleic Acids Res.* **7**, 2303–2322 (1979).
47. Gautier, F., Bunemann, H. & Grotjahn, L. Analysis of calf-thymus satellite DNA: evidence for specific methylation of cytosine in C-G. sequences. *Eur. J. Biochem.* **80**, 175–183 (1977).
48. Liang, G., Gonzalgo, M. L., Salem, C. & Jones, P. A. Identification of DNA methylation differences during tumorigenesis by methylation-sensitive arbitrarily primed polymerase chain reaction. *Methods* **27**, 150–155 (2002).
49. Frigola, J., Ribas, M., Risques, R. A. & Peinado, M. A. Methylome profiling of cancer cells by amplification of inter-methylated sites (AIMS). *Nucleic Acids Res.* **30**, e28 (2002).
50. Estecio, M. R. *et al.* High-throughput methylation profiling by MCA coupled to CpG island microarray. *Genome Res.* **17**, 1529–1536 (2007).
51. Toyota, M. *et al.* Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res.* **59**, 2307–2312 (1999).
52. Chung, W. *et al.* Identification of novel tumor markers in prostate, colon and breast cancer by unbiased methylation profiling. *PLoS ONE* **3**, e2079 (2008).
53. Omura, N. *et al.* Genome-wide profiling of methylated promoters in pancreatic adenocarcinoma. *Cancer Biol. Ther.* **7**, 1146–1156 (2008).
54. Shen, L. *et al.* Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc. Natl Acad. Sci. USA* **104**, 18654–18659 (2007).
55. Yan, P. S., Potter, D., Deatherage, D. E., Huang, T. H. & Lin, S. Differential methylation hybridization: profiling DNA methylation with a high-density CpG island microarray. *Methods Mol. Biol.* **507**, 89–106 (2009).
56. 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).
The first demonstration of affinity enrichment of methylated DNA.
57. Tran, R. K. *et al.* DNA methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. *Curr. Biol.* **15**, 154–159 (2005).

58. Pietrobono, R. *et al.* Quantitative analysis of DNA demethylation and transcriptional reactivation of the *FMR1* gene in fragile X cells treated with 5- azadeoxycytidine. *Nucleic Acids Res.* **30**, 3278–3285 (2002).
59. Nouzova, M. *et al.* Epigenomic changes during leukemia cell differentiation: analysis of histone acetylation and cytosine methylation using CpG island microarrays. *J. Pharmacol. Exp. Ther.* **311**, 968–981 (2004).
60. Ordway, J. M. *et al.* Identification of novel high-frequency DNA methylation changes in breast cancer. *PLoS ONE* **2**, e1314 (2007).
61. Ordway, J. M. *et al.* Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets. *Carcinogenesis* **27**, 2409–2423 (2006).
62. Irizarry, R. A. *et al.* Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res.* **18**, 780–790 (2008).
63. Ibrahim, A. E. *et al.* MMASS: an optimized array-based method for assessing CpG island methylation. *Nucleic Acids Res.* **34**, e136 (2006).
64. Schumacher, A. *et al.* Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res.* **34**, 528–542 (2006).
65. Rollins, R. A. *et al.* Large-scale structure of genomic methylation patterns. *Genome Res.* **16**, 157–163 (2006).
66. Khulan, B. *et al.* Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res.* **16**, 1046–1055 (2006).
67. Oda, M. *et al.* High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. *Nucleic Acids Res.* **37**, 3829–3839 (2009).
68. Brunner, A. L. *et al.* Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res.* **19**, 1044–1056 (2009).
69. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007).
70. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
71. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
72. Guccione, E. *et al.* Methylation of histone H3R2 by PRMT6 and H3K4 by an MLL complex are mutually exclusive. *Nature* **449**, 933–937 (2007).
73. Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R. & Young, R. A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
74. Robertson, G. *et al.* Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nature Methods* **4**, 651–657 (2007).
75. Mukhopadhyay, R. *et al.* The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res.* **14**, 1594–1602 (2004).
76. Weber, M. *et al.* Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genet.* **39**, 457–466 (2007).
77. 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).
78. Keshet, I. *et al.* Evidence for an instructive mechanism of *de novo* methylation in cancer cells. *Nature Genet.* **38**, 149–153 (2006).
References 76–78 provided the first genome-wide analyses of mammalian genomes using affinity enrichment of methylated DNA.
79. Farthing, C. R. *et al.* Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *PLoS Genet.* **4**, e1000116 (2008).
80. Fouse, S. D. *et al.* Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* **2**, 160–169 (2008).
81. Dindot, S. V., Person, R., Strivens, M., Garcia, R. & Beaudet, A. L. Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. *Genome Res.* **19**, 1374–1383 (2009).
82. Hayashi, H. *et al.* High-resolution mapping of DNA methylation in human genome using oligonucleotide tiling array. *Hum. Genet.* **120**, 701–711 (2007).
83. Cheng, A. S. *et al.* Epithelial progeny of estrogen-exposed breast progenitor cells display a cancer-like methylome. *Cancer Res.* **68**, 1786–1796 (2008).
84. Gal-Yam, E. N. *et al.* Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. *Proc. Natl Acad. Sci. USA* **105**, 12979–12984 (2008).
85. Smith, A. E. *et al.* Epigenetics of human T cells during the $G_0 \rightarrow G_1$ transition. *Genome Res.* **19**, 1325–1337 (2009).
86. Koga, Y. *et al.* Genome-wide screen of promoter methylation identifies novel markers in melanoma. *Genome Res.* **19**, 1462–1470 (2009).
87. Straussman, R. *et al.* Developmental programming of CpG island methylation profiles in the human genome. *Nature Struct. Mol. Biol.* **16**, 564–571 (2009).
88. Down, T. A. *et al.* A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nature Biotech.* **26**, 779–785 (2008).
89. Gebhard, C. *et al.* Rapid and sensitive detection of CpG-methylation using methyl-binding (MB)-PCR. *Nucleic Acids Res.* **34**, e82 (2006).
90. Gebhard, C. *et al.* Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. *Cancer Res.* **66**, 6118–6128 (2006).
91. Schmidl, C. *et al.* Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res.* **19**, 1165–1174 (2009).
92. Jorgensen, H. F., Adie, K., Chaubert, P. & Bird, A. P. Engineering a high-affinity methyl-CpG-binding protein. *Nucleic Acids Res.* **34**, e96 (2006).
93. Rauch, T. & Pfeifer, G. P. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Lab. Invest.* **85**, 1172–1180 (2005).
94. Rauch, T. A. & Pfeifer, G. P. The MIRA method for DNA methylation analysis. *Methods Mol. Biol.* **507**, 65–75 (2009).
95. 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).
96. Ballestar, E. *et al.* Methyl-CpG binding proteins identify novel sites of epigenetic inactivation in human cancer. *EMBO J.* **22**, 6335–6345 (2003).
97. Hayatsu, H. Discovery of bisulfite-mediated cytosine conversion to uracil, the key reaction for DNA methylation analysis — a personal account. *Proc. Jpn Acad. Ser. B Phys. Biol. Sci.* **84**, 321–330 (2008).
98. Wang, R. Y., Gehrke, C. W. & Ehrlich, M. Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. *Nucleic Acids Res.* **8**, 4777–4790 (1980).
99. 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).
Although differential deamination of methylated and unmethylated cytosine residues had been described previously, this study provided a practical demonstration of the technique for the analysis of DNA methylation at the single-base-pair level using PCR amplification.
100. Clark, S. J., Harrison, J., Paul, C. L. & Frommer, M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* **22**, 2990–2997 (1994).
101. Paul, C. L. & Clark, S. J. Cytosine methylation: quantitation by automated genomic sequencing and GENESCAN analysis. *Biotechniques* **21**, 126–133 (1996).
102. Eckhardt, F. *et al.* DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature Genet.* **38**, 1378–1385 (2006).
The first example of 'brute force' bisulphite Sanger sequencing of many targets in mammalian genomes.
103. Adorjan, P. *et al.* Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res.* **30**, e21 (2002).
104. Gitan, R. S., Shi, H., Chen, C. M., Yan, P. S. & Huang, T. H. Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res.* **12**, 158–164 (2002).
105. Reinders, J. *et al.* Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. *Genome Res.* **18**, 469–476 (2008).
106. Bibikova, M. *et al.* High-throughput DNA methylation profiling using universal bead arrays. *Genome Res.* **16**, 383–393 (2006).
107. Bibikova, M. *et al.* Human embryonic stem cells have a unique epigenetic signature. *Genome Res.* **16**, 1075–1083 (2006).
108. Bibikova, M. & Fan, J. B. GoldenGate assay for DNA methylation profiling. *Methods Mol. Biol.* **507**, 149–163 (2009).
109. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
110. Byun, H. M. *et al.* Epigenetic profiling of somatic tissues from human autopsy specimens identifies tissue- and individual-specific DNA methylation patterns. *Hum. Mol. Genet.* **18**, 4808–4817 (2009).
111. Ladd-Acosta, C. *et al.* DNA methylation signatures within the human brain. *Am. J. Hum. Genet.* **81**, 1304–1315 (2007).
112. Katari, S. *et al.* DNA methylation and gene expression differences in children conceived *in vitro* or *in vivo*. *Hum. Mol. Genet.* **18**, 3769–3778 (2009).
113. Martinez, R. *et al.* A microarray-based DNA methylation study of glioblastoma multiforme. *Epigenetics* **4**, 255–264 (2009).
114. Christensen, B. C. *et al.* Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet.* **5**, e1000602 (2009).
115. Houseman, E. A. *et al.* Model-based clustering of DNA methylation array data: a recursive- partitioning algorithm for high-dimensional data arising as a mixture of β distributions. *BMC Bioinformatics* **9**, 365 (2008).
116. Hinoue, T. *et al.* Analysis of the association between CIMP and $BRAF^{V600E}$ in colorectal cancer by DNA methylation profiling. *PLoS ONE* **4**, e8357 (2009).
117. Bibikova, M. *et al.* Genome-wide DNA methylation profiling using Infinium assay. *Epigenomics* **1**, 177–200 (2009).
118. Steemers, F. J. & Gunderson, K. L. Whole genome genotyping technologies on the BeadArray platform. *Biotechnol. J.* **2**, 41–49 (2007).
119. 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).
120. 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).
121. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).
122. Hodges, E. *et al.* High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res.* 6 Jul 2009 (doi:10.1101/gr.095190.109).
123. Li, J. B. *et al.* Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. *Genome Res.* **19**, 1606–1615 (2009).
124. Dunn, J. J., McCorkle, S. R., Everett, L. & Anderson, C. W. Paired-end genomic signature tags: a method for the functional analysis of genomes and epigenomes. *Genet. Eng. (NY)* **28**, 159–173 (2007).
125. Dempsey, M. P. *et al.* Paired-end sequence mapping detects extensive genomic rearrangement and translocation during divergence of *Francisella tularensis* subsp. *tularensis* and *Francisella tularensis* subsp. *holarctica* populations. *J. Bacteriol.* **188**, 5904–5914 (2006).
126. Korbel, J. O. *et al.* Paired-end mapping reveals extensive structural variation in the human genome. *Science* **318**, 420–426 (2007).
127. Tost, J., Schatz, P., Schuster, M., Berlin, K. & Gut, I. G. Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. *Nucleic Acids Res.* **31**, e50 (2003).
128. Ehrich, M. *et al.* Cytosine methylation profiling of cancer cell lines. *Proc. Natl Acad. Sci. USA* **105**, 4844–4849 (2008).
129. Ehrich, M. *et al.* Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl Acad. Sci. USA* **102**, 15785–15790 (2005).
130. Docherty, S. J., Davis, O. S., Haworth, C. M., Plomin, R. & Mill, J. Bisulfite-based epityping on pooled genomic DNA provides an accurate estimate of average group DNA methylation. *Epigenetics Chromatin* **2**, 3 (2009).

131. Killian, J. K. *et al.* Large-scale profiling of archival lymph nodes reveals pervasive remodeling of the follicular lymphoma methylome. *Cancer Res.* **69**, 758–764 (2009).
132. Tetzner, R. Prevention of PCR cross-contamination by UNG treatment of bisulfite-treated DNA. *Methods Mol. Biol.* **507**, 357–370 (2009).
133. Tetzner, R., Dietrich, D. & Distler, J. Control of carry-over contamination for PCR-based DNA methylation quantification using bisulfite treated DNA. *Nucleic Acids Res.* **35**, e4 (2007).
134. Dohm, J. C., Lottaz, C., Borodina, T. & Himmelbauer, H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* **36**, e105 (2008).
135. Warnecke, P. M. *et al.* Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res.* **25**, 4422–4426 (1997).
136. Campan, M., Weisenberger, D. J., Trinh, B. & Laird, P. W. MethyLight. *Methods Mol. Biol.* **507**, 325–337 (2009).
137. Weisenberger, D. J. *et al.* Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res.* **33**, 6823–6836 (2005).
138. Kerkel, K. *et al.* Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nature Genet.* **40**, 904–908 (2008).
139. Houseman, E. A. *et al.* Copy number variation has little impact on bead-array-based measures of DNA methylation. *Bioinformatics* **25**, 1999–2005 (2009).
140. Siegmund, K. D., Marjoram, P., Woo, Y. J., Tavare, S. & Shibata, D. Inferring clonal expansion and cancer stem cell dynamics from DNA methylation patterns in colorectal cancers. *Proc. Natl Acad. Sci. USA* **106**, 4828–4833 (2009).
141. Fatemi, M. *et al.* Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Res.* **33**, e176 (2005).
142. Weisenberger, D. J. *et al.* DNA methylation analysis by digital bisulfite genomic sequencing and digital MethyLight. *Nucleic Acids Res.* **36**, 4689–4698 (2008).
143. Li, M. *et al.* Sensitive digital quantification of DNA methylation in clinical samples. *Nature Biotech.* **27**, 858–863 (2009).
144. Chhibber, A. & Schroeder, B. G. Single-molecule polymerase chain reaction reduces bias: application to DNA methylation analysis by bisulfite sequencing. *Anal. Biochem.* **377**, 46–54 (2008).
145. Bock, C. & Lengauer, T. Computational epigenetics. *Bioinformatics* **24**, 1–10 (2008).
146. Pennisi, E. Research funding. Are epigeneticists ready for big science? *Science* **319**, 1177 (2008).
147. Jones, P. A. *et al.* Moving AHEAD with an international human epigenome project. *Nature* **454**, 711–715 (2008).
148. Pushkarev, D., Neff, N. F. & Quake, S. R. Single-molecule sequencing of an individual human genome. *Nature Biotech.* **27**, 847–852 (2009).
149. Eid, J. *et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009).
150. Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nature Biotech.* **26**, 1146–1153 (2008).
151. Clarke, J. *et al.* Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnol.* **4**, 265–270 (2009).
152. Model, F., Adorjan, P., Olek, A. & Piepenbrock, C. Feature selection for DNA methylation based cancer classification. *Bioinformatics* **17**, S157–S164 (2001).
153. Rohde, C. *et al.* Bisulfite sequencing Data Presentation and Compilation (BDPC) web server — a useful tool for DNA methylation analysis. *Nucleic Acids Res.* **36**, e34 (2008).
154. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10**, 232 (2009).
155. Xu, Y. H., Manoharan, H. T. & Pitot, H. C. CpG Analyzer, a Windows-based utility program for investigation of DNA methylation. *Biotechniques* **39**, 656–662 (2005).
156. Hackenberg, M. *et al.* CpGcluster: a distance-based algorithm for CpG-island detection. *BMC Bioinformatics* **7**, 446 (2006).
157. Wang, Y. & Leung, F. C. An evaluation of new criteria for CpG islands in the human genome as gene markers. *Bioinformatics* **20**, 1170–1177 (2004).
158. Takai, D. & Jones, P. A. The CpG Island Searcher: a new WWW resource. *In Silico Biol.* **3**, 235–240 (2003).
159. Xu, Y. H., Manoharan, H. T. & Pitot, H. C. CpG PatternFinder: a Windows-based utility program for easy and rapid identification of the CpG methylation status of DNA. *Biotechniques* **43**, 334–342 (2007).
160. Ioshikhes, I. P. & Zhang, M. Q. Large-scale human promoter mapping using CpG islands. *Nature Genet.* **26**, 61–63 (2000).
161. Carr, I. M., Valleley, E. M., Cordery, S. F., Markham, A. F. & Bonthron, D. T. Sequence analysis and editing for bisulphite genomic sequencing projects. *Nucleic Acids Res.* **35**, e79 (2007).
162. Hetzl, J., Foerster, A. M., Raidl, G. & Mittelsten Scheid, O. CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J.* **51**, 526–536 (2007).
163. Pelizzola, M. *et al.* MEDME: an experimental and analytical methodology for the estimation of DNA methylation levels based on microarray derived MeDIP-enrichment. *Genome Res.* **18**, 1652–1659 (2008).
164. Pattyn, F. *et al.* methBLAST and methPrimerDB: web-tools for PCR based methylation analysis. *BMC Bioinformatics* **7**, 496 (2006).
165. Grunau, C., Renault, E., Rosenthal, A. & Roizes, G. MethDB — a public database for DNA methylation data. *Nucleic Acids Res.* **29**, 270–274 (2001).
166. Grunau, C., Renault, E. & Roizes, G. DNA Methylation Database 'MethDB': a user guide. *J. Nutr.* **132**, 2435S–2439S (2002).
167. Amoreira, C., Hindermann, W. & Grunau, C. An improved version of the DNA methylation database (MethDB). *Nucleic Acids Res.* **31**, 75–77 (2003).
168. Negre, V. & Grunau, C. The MethDB DAS server: adding an epigenetic information layer to the human genome. *Epigenetics* **1**, 101–105 (2006).
169. Li, L. C. & Dahiya, R. MethPrimer: designing primers for methylation PCRs. *Bioinformatics* **18**, 1427–1431 (2002).
170. Grunau, C., Schattevoy, R., Mache, N. & Rosenthal, A. MethTools — a toolbox to visualize and analyze DNA methylation data. *Nucleic Acids Res.* **28**, 1053–1058 (2000).
171. He, X. *et al.* MethyCancer: the database of human DNA methylation and cancer. *Nucleic Acids Res.* **36**, D836–D841 (2008).
172. Rakyan, V. K. *et al.* An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res.* **18**, 1518–1529 (2008).
173. Ongenaert, M. *et al.* PubMeth: a cancer methylation database combining text-mining and expert annotation. *Nucleic Acids Res.* **36**, D842–D846 (2008).
174. Kumaki, Y., Oda, M. & Okano, M. QUMA: quantification tool for methylation analysis. *Nucleic Acids Res.* **36**, W170–W175 (2008).

Acknowledgements

I am grateful to K. Siegmund and to members of the University of Southern California Epigenome Center for many helpful discussions. P.W.L.'s research is supported by National Cancer Institute grants R01-CA118699 and U24-CA143882 and by the Norris Foundation, the Ovarian Cancer Research Fund, the Canary Foundation, the Entertainment Industry Foundation and the Riley Foundation.

Competing interests statement

The author declares competing financial interests: see Web version for details.

DATABASES

UniProt: http://www.uniprot.org
MBD1 | MBD2 | MBD3L1 | MECP2

FURTHER INFORMATION

Peter W. Laird's homepage: http://www.usc.edu/programs/pibbs/site/faculty/laird_p.htm
The Cancer Genome Atlas: http://cancergenome.nih.gov
NCBI Reference Sequence Collection: http://www.ncbi.nlm.nih.gov/RefSeq
University of Southern California Epigenome Center: http://epigenome.usc.edu

ALL LINKS ARE ACTIVE IN THE ONLINE PDF