Development 134, 3959-3965 (2007) doi:10.1242/dev.001131

Genome-wide analysis of DNA methylation patterns

Daniel Zilberman[1] and Steven Henikoff[2,3]

Cytosine methylation is the most common covalent modification of DNA in eukaryotes. DNA methylation has an important role in many aspects of biology, including development and disease. Methylation can be detected using bisulfite conversion, methylation-sensitive restriction enzymes, methyl-binding proteins and anti-methylcytosine antibodies. Combining these techniques with DNA microarrays and high-throughput sequencing has made the mapping of DNA methylation feasible on a genome-wide scale. Here we discuss recent developments and future directions for identifying and mapping methylation, in an effort to help colleagues to identify the approaches that best serve their research interests.

Introduction

DNA methylation is a unique and noteworthy process because it involves the covalent modification of a cell's genetic material (Chan et al., 2005; Freitag and Selker, 2005; Gehring and Henikoff, 2007; Goll and Bestor, 2005; Klose and Bird, 2006; Richards, 2006). At gene promoters, methylation generally leads to transcriptional silencing. Complex multicellular eukaryotes (plants and animals) primarily methylate DNA at cytosines within CG dinucleotides. Following each round of DNA replication, a DNA methyltransferase [from the Dnmt1 family (Goll and Bestor, 2005)] fills in the missing methylation on the newly synthesized strand, allowing faithful maintenance of DNA methylation patterns through many rounds of cell division and, at least in plants, through multiple generations (Chan et al., 2005; Soppe et al., 2000). For this reason, differential methylation is a process that most closely approximates genetic differences between cell types (or organisms) with identical DNA sequence.

There is abundant evidence that aberrant DNA methylation can preclude normal development. Knockout mutations of any one of the three mouse genes that encode DNA methyltransferases (*Dnmt1*, *Dnmt3a* and *Dnmt3b*) are lethal (Goll and Bestor, 2005). Loss-of-function of *MET1*, the *Arabidopsis thaliana* ortholog of *Dnmt1*, leads to developmental abnormalities such as delayed flowering and reduced fertility, which become very severe when additional methyltransferase genes (*CMT3* and/or *DRM2*) are mutated (Xiao et al., 2006; Zhang and Jacobsen, 2006). Depletion of Dnmt1 in zebrafish embryos causes defects in terminal differentiation of the intestine, exocrine pancreas and retina (Rai et al., 2006). Either the loss or gain of methylation at specific genes (FWA, SUP) can lead to developmental abnormalities in *Arabidopsis* (Gehring and Henikoff, 2007; Jacobsen et al., 2000; Soppe et al., 2000). In plants and mammals, DNA methylation has a central role in genomic imprinting, the monoallelic expression of a gene from either the maternal or the paternal copy (Gehring and Henikoff, 2007; Goll and Bestor, 2005). X-chromosome inactivation in female mammals is also dependent on DNA methylation (Heard and Disteche, 2006). The high failure rate of cloning by somatic nuclear transfer has been attributed to improper reprogramming of DNA methylation patterns in the donor nucleus (Meissner and Jaenisch, 2006).

Despite the clear importance of DNA methylation, the extent to which changes in somatic DNA methylation are involved in mammalian gene regulation is unclear (Goll and Bestor, 2005; Walsh and Bestor, 1999). This is largely owing to our limited knowledge of DNA methylation patterns. A study published in early 2006 estimated that DNA methylation of less than 0.1% of the human genome has been analyzed in detail (Schumacher et al., 2006). A number of recent reports have considerably expanded our knowledge of eukaryotic DNA methylation (Bibikova et al., 2006a; Eckhardt et al., 2006; Hellman and Chess, 2007; Keshet et al., 2006; Khulan et al., 2006; Rollins et al., 2006; Weber et al., 2007; Yuan et al., 2006; Zhang et al., 2006; Zilberman et al., 2007). Nonetheless, we are just beginning to unravel genomic methylation patterns, particularly in the complex genomes of vertebrates. Fortunately, technological advances in high-density microarray technology and high-throughput DNA sequencing should allow the comprehensive analysis of DNA methylation to become a routine technique. Here we focus on the most promising new methodologies and their suitability for addressing outstanding questions about the role of DNA methylation in development and disease.

Methodologies for detection of DNA methylation

Many methods of DNA methylation analysis have been developed over the years and are described in detail in a number of recent reviews (Brena et al., 2006; Callinan and Feinberg, 2006; Laird, 2003; Lieb et al., 2006; Ushijima, 2005). All of these approaches are based on one of three techniques: bisulfite conversion, digestion with methylation-sensitive restriction enzymes, and affinity purification of methylated DNA.

Bisulfite conversion

Methylated cytosine has roughly the same base-pairing characteristics as unmethylated cytosine, and is thus indistinguishable by standard sequencing approaches. To overcome this, genomic DNA can be treated with sodium bisulfite (Clark et al., 1994; Clark et al., 2006). Under appropriate conditions, this treatment causes deamination of unmethylated cytosine to uracil, while leaving methylated cytosine intact (Fig. 1). PCR amplification of converted DNA replaces the uracil with thymine. Analysis of the PCR product by Sanger sequencing (Eckhardt et al., 2006), pyrosequencing (Tost and Gut, 2006), or mass spectrometry (Ehrich et al., 2006; Ehrich et al., 2005; Schatz et al., 2004; Schatz et al., 2006; Tost et al., 2003), can be used to quantify the extent of methylation at each cytosine. A potential issue with bisulfite analysis is that it depends on the complete conversion of unmethylated cytosines. In animal DNA, a sure sign of incomplete conversion is abundant methylation at cytosines that are not in CG dinucleotides. In plant DNA, this problem can be more difficult to detect, but will frequently manifest itself as continuous stretches of heavily methylated cytosines in all sequence contexts. Spiking the reaction

[1]University of California, 211 Koshland Hall, Berkeley, CA 94720, USA. [2]Fred Hutchinson Cancer Research Center and [3]Howard Hughes Medical Institute, 1100 Fairview Avenue North, Seattle, WA 98109, USA.

e-mail: steveh@fhcrc.org

DEVELOPMENT

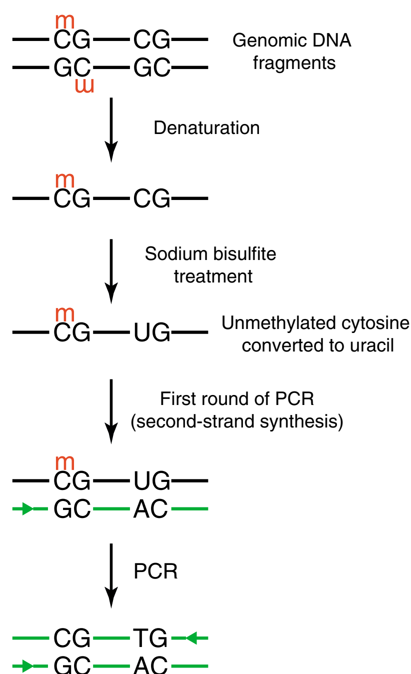

Fig. 1. Bisulfite conversion. DNA is denatured and then treated with sodium bisulfite to convert unmethylated cytosine to uracil, which is converted to thymine by PCR. An important point is that following bisulfite conversion, the DNA strands are no longer complementary, and primers are designed to assay the methylation (m) status of a specific strand.

with known unmethylated DNA, such as yeast genomic DNA, can be used as a control. It is essential to ensure that bisulfite-treated samples have been completely converted before utilizing them in high-throughput applications.

Methylation-sensitive restriction enzymes

Methylation-sensitive restriction endonucleases are classic tools of DNA methylation analysis (Bird et al., 1985; Bird and Southern, 1978; Lindsay and Bird, 1987). Most of these are inhibited by methylation of their recognition site, whereas some, most notably McrBC, specifically digest methylated DNA. Many variations of restriction enzyme-based methods have been used in conjunction with genomic analysis (Khulan et al., 2006; Lippman et al., 2005; Rollins et al., 2006; Schumacher et al., 2006; Tompa et al., 2002; Yuan et al., 2006). Here we briefly describe the general logic behind all such approaches.

Restriction enzyme-based methods either enrich for methylated DNA or unmethylated DNA (Fig. 2). Generally, comparisons are made in one of three ways: between a sample treated with an enzyme or a cocktail of enzymes and an untreated control; between a sample treated with a methylation-sensitive enzyme compared with a control treated with a methylation-insensitive isoschizomer (*Hpa*II and *Msp*I, see below); or between two test samples, such as two tissue types or mutant and wild-type samples, both treated with the same enzyme. The ability to enrich unmethylated DNA, by digesting away methylated DNA or by isolating smaller fragments generated by methylation-inhibited enzymes, is particularly useful for analyzing large, heavily methylated genomes (as discussed in more detail below). In the human genome, over 60% of CG sites are methylated (Goll and Bestor, 2005), so enriching unmethylated DNA significantly reduces the complexity of the sample. An important limitation is that all restriction enzyme-based techniques are limited to analysis of methylation within recognition sites.

The most commonly used restriction enzymes are the isoschizomers *Hpa*II and *Msp*I, which recognize the sequence CCGG. *Hpa*II is blocked by methylation of either cytosine, whereas *Msp*I is blocked only by methylation of the outer C (Korch and Hagblom, 1986; Waalwijk and Flavell, 1978). In mammalian genomes, where methylation is almost exclusively in CG sites (Goll and Bestor, 2005), *Hpa*II is inhibited and *Msp*I is not. In plant genomes, where methylation of cytosines in the CNG context is also common, *Msp*I can be used to detect CNG methylation. Another useful enzyme employed in genomic studies is McrBC (Lippman et al., 2004; Rollins et al., 2006; Schumacher et al., 2006). McrBC is an *E. coli* endonuclease that cleaves between two methylated cytosines in the context (G/A)^{met}C (Sutherland et al., 1992). The two sites can be separated by up to 3 kb, but the optimal separation is 55-100 bp (Gowher et al., 2000; Zhou et al., 2002). For this reason, McrBC is an excellent tool for the removal of densely methylated DNA. Although less of an issue with McrBC, sequence polymorphisms between samples can mimic methylation differences if they affect the enzyme recognition site. Therefore, it is safest to use restriction enzymes to compare samples that have no or little polymorphism, such as different tissues from the same organism. Alternatively, the *Msp*I/*Hpa*II isoschizomer pair can be used to control for polymorphic sites (Khulan et al., 2006).

Affinity purification

The most recent and simplest way to enrich methylated DNA is by affinity purification (Fig. 3). One approach is to take advantage of the methyl-binding domain (MBD), which binds methylated CG sites. A tagged MBD domain expressed in *E. coli* is affinity purified and the MBD column is subsequently used to purify methylated DNA (Cross et al., 1994; Selker et al., 2003; Zhang et al., 2006). Alternatively, a commercially available monoclonal antibody that specifically recognizes methylated cytosine can be used to immunoprecipitate methylated DNA (Keshet et al., 2006; Reynaud et al., 1992; Weber et al., 2005; Weber et al., 2007; Zhang et al., 2006; Zilberman et al., 2007). For plant researchers, a potential advantage is that the MBD method purifies only CG-methylated DNA, whereas the antibody will work against DNA methylated in any context. However, as almost all the methylated loci that have been characterized in plants have CG methylation, the results obtained with the two methods should be broadly similar. For most researchers, the commercial availability of the antibody combined with wide utilization of immunoprecipitation will make this the method of choice for enriching methylated DNA.

An important point regarding all affinity-based techniques is that they measure the density of methylation in a given region. Therefore, a methylated stretch of DNA where methylation target sites (CG sites in animals) are sparse might be difficult to differentiate from an unmethylated region. This is particularly an issue with mammalian genomes, where CG density is generally low and CG-dense sequences are typically unmethylated (Weber et al., 2007).

A potential twist on the affinity-based approach is to enrich for unmethylated DNA by isolating the unbound fraction from either affinity method. The ratio of antibody (or MBD domain) to DNA would have to be carefully optimized to ensure that essentially all methylated DNA is removed. Alternatively, unmethylated DNA prepared by McrBC digestion could be further enriched by the removal of residual methylated DNA by affinity reagents. This approach would be especially suitable for analyses of mammalian

DEVELOPMENT

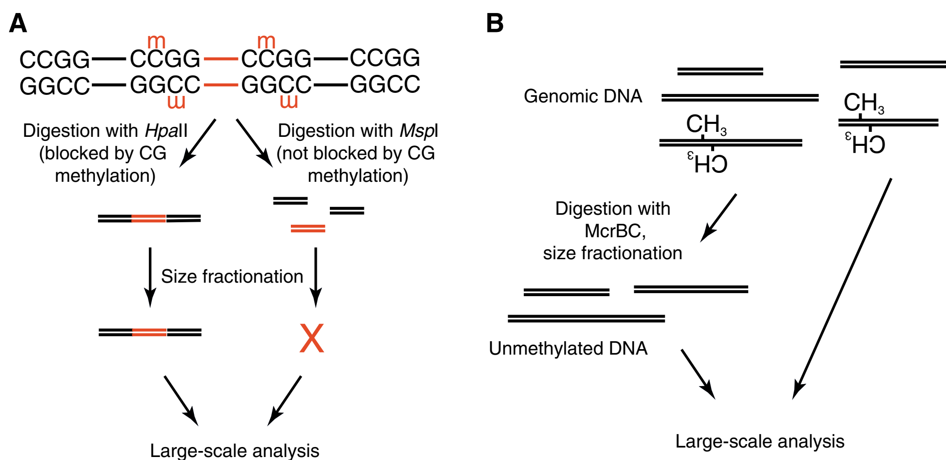

Fig. 2. Detecting DNA methylation with methylation-sensitive restriction enzymes. (**A**) A methylated (m) region of genomic DNA digested with *Hpa*II (left) or *Msp*I (right). Smaller fragments are discarded (red X), enriching for methylated DNA in the *Hpa*II-treated sample, relative to the *Msp*I-treated. (**B**) Genomic DNA is treated with McrBC, which cuts methylated (CH_3) DNA. Smaller fragments are discarded, enriching for unmethylated DNA. There are many variations on these techniques (see Khulan et al., 2006; Lippman et al., 2004; Schumacher et al., 2006; Tompa et al., 2002).

and other heavily methylated genomes because it would substantially reduce sample complexity and would overcome the limitation imposed by restriction enzyme recognition sites.

Genome-wide analysis of DNA methylation

A number of approaches exist that enable the large-scale analysis of DNA methylation. The Human Epigenome Project (www.epigenome.org) has used standard sequencing approaches to sequence a massive amount of bisulfite-converted DNA from human tissues and primary cells, and has identified a substantial number of tissue-specific differentially methylated regions (DMRs) (Eckhardt et al., 2006). Another study used restriction enzymes and standard cloning and sequencing to analyze almost 14 Mb of unmethylated human DNA and over 8 Mb of methylated DNA (Rollins et al., 2006). These approaches, although highly informative, are expensive and labor-intensive ventures that are beyond the capabilities of most laboratories. Here, we discuss two approaches that, either currently or in the near future, can be used by any laboratory to perform genome-wide DNA methylation analyses: DNA microarrays and high-throughput DNA sequencing.

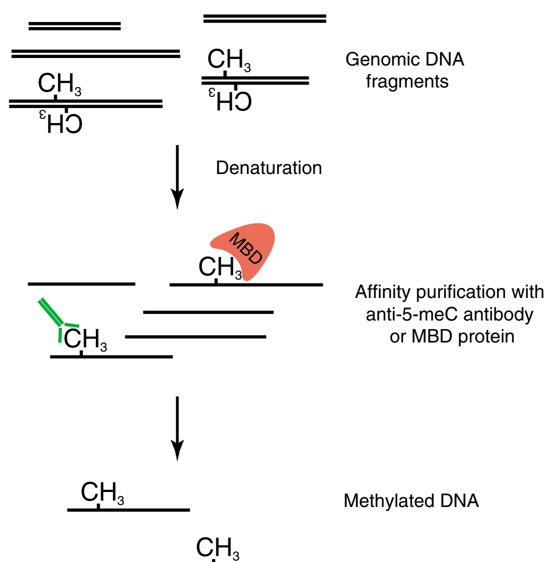

Fig. 3. Affinity enrichment of methylated DNA. Genomic DNA is denatured and then affinity purified with either an antibody (green) or a methyl-binding domain (MBD, red) protein that can be attached to a column (Cross et al., 1994; Weber et al., 2005).

DNA microarrays

Early microarray studies of DNA methylation utilized spotted arrays made by individual laboratories or microarray facilities (Lippman et al., 2004; Tompa et al., 2002). Fortunately, high-quality commercial oligonucleotide arrays – bead arrays made by Illumina, lithographic arrays made by Affymetrix, adaptive lithographic arrays made by NimbleGen, and inkjet arrays manufactured by Agilent – are widely available. Both the design and the technology behind an array influence their utility in methylation analysis. For example, the Illumina arrays are designed to analyze bisulfite-converted DNA, whereas the other arrays are well suited for restriction enzyme- and affinity-based assays (Fig. 4).

Bead arrays (Illumina)

The bead array-based analysis of DNA methylation developed by Illumina is an outgrowth of their genotyping method (Bibikova et al., 2006b; Fan et al., 2006), which is designed to provide single-base resolution, although two or more closely spaced cytosines may have to be analyzed together. Bisulfite-converted DNA is assayed with two primers, each labeled with a different fluorescent dye. One primer is designed to hybridize if the cytosine is methylated (and unconverted), whereas the other will only hybridize to a converted sequence. The two primers are used in a PCR reaction with a locus-specific methylation-insensitive primer. The ratio of the PCR products is ascertained using Illumina's Sentrix Array Matrix bead array platform, which can assay up to 1536 sites in 96 samples in one experiment. This approach provides less coverage than other array-based methods, and necessitates the development and evaluation of a large set of selective primers, thus limiting its utility for de novo genome analysis. The strength of the technique is that it provides quantitative evaluation of specific cytosines and can process many samples in parallel. Therefore, this method is well suited to compare a set of known methylated loci among a large number of cell lines or individuals to ascertain methylation polymorphisms (Fig. 4). Using this approach, a set of methylation markers that could distinguish lung carcinoma samples from normal tissue was identified (Bibikova et al., 2006b). A subsequent study identified diagnostic differences between human embryonic stem (ES) cell lines and differentiated cells (Bibikova et al., 2006a).

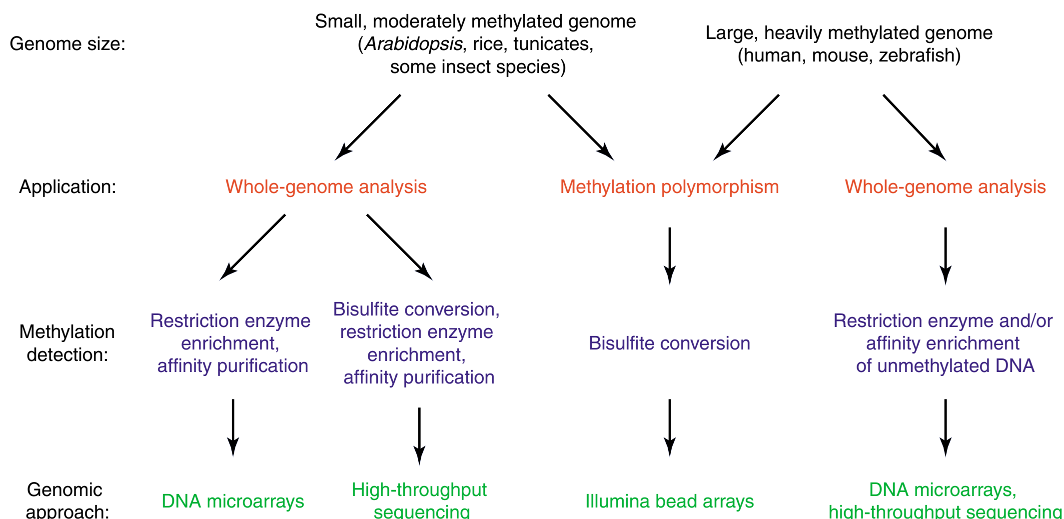

Fig. 4. Genomic analysis of DNA methylation. The suitability of a given technique of DNA methylation analysis depends on the genome size of the organism and the intended application. Bisulfite conversion coupled to Illumina bead arrays is well suited to the measurement of methylation polymorphism between multiple samples. A comprehensive analysis of methylation in organisms with small genomes can be accomplished with any one of a variety of techniques, whereas enrichment of unmethylated DNA is the preferred method when analyzing most vertebrate genomes.

Short oligonucleotide arrays (Affymetrix)

Affymetrix GeneChip arrays are produced using photolithographic technology to achieve very high feature density, with millions of probes per chip (Dalma-Weiszhausz et al., 2006). Each feature consists of 25-mer oligonucleotides. These short probes provide good specificity, but suffer from decreased sensitivity and increased random signal variation (noise) compared with longer probes (Kreil et al., 2006). Each chip is designed for 'single channel' hybridization – they are hybridized with one sample at a time. To compare samples, such as two cell lines, each sample is hybridized to a separate array and the resulting signals are compared. Generally, each sample is hybridized at least three times to allow statistical treatment of the data to identify significant differences. For methylation analysis, a tiling design is most useful, with equidistantly spaced probes across portions of a genome or an entire genome. Tiling arrays are available for the human, mouse and *Arabidopsis* genomes, as well as for human and mouse promoters (*Drosophila*, *C. elegans*, *S. cerevisiae* and *S. pombe* arrays are also available, but these organisms lack DNA methylation). Restriction enzyme-enriched unmethylated DNA from human brain tissue has been analyzed using Affymetrix tiling arrays covering chromosomes 21 and 22 (Schumacher et al., 2006). This study found that most of the unmethylated sites were close to the 5′ end of genes, consistent with the need to keep promoters free of methylation. The *Arabidopsis* array has been successfully used to profile methylated DNA enriched by MBD and antibody affinity purification to yield a high-resolution methylation map of the entire *Arabidopsis* genome (Zhang et al., 2006). Both purification methods produced comparable results. Most academic microarray facilities are set up to handle Affymetrix arrays, making them a convenient resource for researchers. However, the lithographic mask technology makes custom arrays prohibitively expensive, so that most researchers are effectively limited to the standard array designs.

Long oligonucleotide arrays (NimbleGen and Agilent)

NimbleGen arrays are synthesized using an adaptive photolithographic method (Nuwaysir et al., 2002). Tiling arrays for chromatin immunoprecipitation and methylation analysis consist of ~380,000 (soon to be expanded to over 2 million) 60-mer oligonucleotide probes. Agilent manufactures microarrays consisting of ~240,000 60-mers using inkjet technology (Wolber et al., 2006). Both arrays are dual channel – two samples are labeled with different fluorescent dyes, such as immunoprecipitated test DNA and control DNA, and are hybridized on a single chip. Hybridizing the test and control samples on the same array controls for between-array variation and thus reduces the need for replicates. The major disadvantage of these arrays versus the Affymetrix array is reduced oligonucleotide probe density. However, the longer probes provide a better balance between specificity, sensitivity and noise than the 25-mers on the Affymetrix array (Kreil et al., 2006). This translates into array data that require less statistical manipulation. We routinely use raw data from NimbleGen arrays with good results (Mito et al., 2005; Mito et al., 2007; Penterman et al., 2007; Zilberman et al., 2007). Both NimbleGen and Agilent manufacturing methods also allow the production of custom arrays. This allows for flexibility in experimental design, as well as in the analysis of DNA methylation in organisms other than mammals and *Arabidopsis*.

A restriction enzyme-based comparison of mouse spermatogenic and brain cells on custom-designed NimbleGen arrays identified over 200 DMRs in ~6.2 Mb of the mouse genome (Khulan et al., 2006). Analysis of immunoprecipitated DNA using customized NimbleGen tiling arrays produced genome-wide *Arabidopsis* DNA methylation mapping data that were broadly similar to the genome-wide DNA methylation profile generated using the Affymetrix platform (Zilberman et al., 2007). In another application of this approach with NimbleGen arrays, the whole-genome DNA methylation profile of wild-type *Arabidopsis* was compared with that of plants with loss-of-function mutations in the DNA demethylase genes *ROS1*, *DML2* and *DML3*. This approach accurately revealed nearly 200 small methylation differences (Penterman et al., 2007). The profiling of immunoprecipitated methylated DNA from human fibroblasts and sperm on NimbleGen promoter arrays has also identified a number of promoters that are methylated specifically in fibroblasts, including many germline-specific promoters (Weber et al., 2007).

DEVELOPMENT

Table 1. Overview of available platforms for large-scale analysis of DNA methylation

Technology	Applications	Advantages	Limitations	References
Illumina bead arrays	Methylation polymorphism discovery and analysis	Quantitative Rapid analysis of up to 96 samples	Requires design of a primer library Only 1536 sites can be assayed simultaneously	(Bibikova et al., 2006a; Bibikova et al., 2006b; Fan et al., 2006)
Affymetrix arrays	Whole-genome methylation mapping	High feature density Readily available mouse, human and *Arabidopsis* arrays Accessible to researchers with access to a microarray facility Reasonable price	Short oligonucleotide probes produce noisier data 'Single channel' hybridization Custom arrays prohibitively expensive	(Dalma-Weiszhausz et al., 2006; Schumacher et al., 2006; Zhang et al., 2006)
NimbleGen arrays	Whole-genome methylation mapping	Long oligonucleotide probes produce cleaner data 'Dual channel' hybridization Inexpensive custom arrays Hybridization available as a service at a reasonable price	Lower feature density than Affymetrix	(Khulan et al., 2006; Nuwaysir et al., 2002; Weber et al., 2007; Zilberman et al., 2007)
Agilent arrays	Large-scale methylation mapping	Long oligonucleotide probes produce cleaner data 'Dual channel' hybridization	Substantially lower feature density than Affymetrix and NimbleGen	(Wolber et al., 2006)
Solexa sequencing	Whole-genome methylation mapping Analysis of imprinted loci	Quantitative Does not require hybridization Concurrent genotype information	New technology Requires purchase of an expensive instrument	(Barski et al., 2007; Bentley, 2006)

NimbleGen also offers dye-labeling and array hybridization as a service, which could be useful to researchers who do not have access to a microarray facility.

Single-nucleotide polymorphism arrays

Single-nucleotide polymorphism (SNP) arrays have probes that selectively bind to specific polymorphic sequences, thus providing genotype information based on relative hybridization to the polymorphic probes. Using SNP arrays for DNA methylation analysis allows the genotyping of methylated DNA that has been isolated from polymorphic individuals (Hellman and Chess, 2007; Yuan et al., 2006). A recent study used Affymetrix SNP arrays to distinguish methylation of the active and inactive X chromosomes, and found that transcribed regions of genes were preferentially methylated on the active X (Hellman and Chess, 2007). This approach should be generally useful for analyzing DNA methylation that is associated with allele-specific processes, such as genomic imprinting and X inactivation.

High-throughput sequencing

The newest and most promising methodology for genome-scale analysis of DNA methylation is high-throughput sequencing. A number of high-throughput sequencing technologies exist, most of which are still in development (Bentley, 2006; Braslavsky et al., 2003; Levene et al., 2003; Margulies et al., 2005; Meyers et al., 2004; Shendure et al., 2005; Vercoutere et al., 2003). The aim of each approach is to produce a very large amount of sequence information, more rapidly and at a lower cost than conventional Sanger sequencing, and without the need for cloning.

High-throughput sequencing can be employed as an alternative to analyzing DNA methylation with oligonucleotide arrays. Instead of labeling and hybridizing the test and control samples, as in array experiments, they can be sequenced directly. The frequency of a given sequence will be represented by its abundance in the sample. With enough sequences, information density comparable to microarray data can be achieved. Direct sequencing offers a number of advantages. Counts of sequence reads provide a quantitative measure of methylation abundance, rather than the relative measure that array-based methods provide. The sample does not need to be amplified, except as part of the sequencing strategy. Single-molecule sequencing methods, although still in development, would eliminate the need for amplification entirely (Braslavsky et al., 2003; Levene et al., 2003; Vercoutere et al., 2003). Biases that affect hybridization, such as sequence composition, are generally not an issue in this approach. There is also no need to represent the entire genome on an array to identify the methylated fraction. As with SNP arrays, sequencing provides allele-specific information in polymorphic individuals, but the SNPs do not need to be known in advance – a potentially major advantage. Finally, any microarray is only as good as the quality of the genome sequence. Similarly, because short reads produced by high-throughput sequencing are very

DEVELOPMENT

challenging to assemble de novo, a high-quality reference sequence is required. But one advantage of high-throughput sequencing is that the data can be re-analyzed following improvements in the reference sequence.

The methylation detection technique best suited to high-throughput sequencing depends on the genome to be analyzed. For a smaller, less repetitive genome like *Arabidopsis*, the direct sequencing of bisulfite-converted DNA is a good option. However, the analysis of data generated by this approach presents a special challenge because of C to T conversion. The benefit of this approach is that it provides the best possible resolution: quantitative information about the methylation status of every cytosine. Restriction enzyme- and affinity-based methods are also suitable for the analysis of smaller genomes (see Fig. 4). Such data are easier to analyze, but do not provide the resolution of bisulfite analysis. For larger genomes with a high repeat content, such as the human and mouse genomes, direct bisulfite sequencing would be more challenging, but has already been performed on a small scale (Meissner et al., 2005). Affinity-based purification of methylated DNA would also be challenging, because most of the genome is methylated. The best approach might be to enrich unmethylated DNA, either by affinity purification or by utilizing restriction enzymes (Fig. 4).

Two high-throughput sequencing platforms are currently commercially available: a high-throughput pyrosequencing approach developed by 454 Life Sciences (Margulies et al., 2005), and a fluorescent nucleotide-based system developed by Solexa (now Illumina) (Bentley, 2006). The 454 system can produce 400,000 reads of over 100 bases per run. The Solexa system can produce 40 million reads of about 25-35 bases. For DNA methylation analysis, the number of reads is more important than the length, because most ~30-mers can be unambiguously aligned to a reference genome sequence (Bentley, 2006). Therefore, the Solexa system is the best currently available for this type of analysis, and has already been successfully used to analyze the distribution of a number of post-translational histone modifications in the human genome (Barski et al., 2007).

Conclusions

After decades of work, DNA methylation research is entering a new phase. The ability to analyze methylation patterns of whole genomes will enable us, for the first time, to obtain the most basic type of information about this modification – its location within the genome. This, in turn, should enable the elucidation of how DNA methylation influences chromatin function, and the role it plays in development and disease. The choice of which approach is best to analyze DNA methylation ultimately depends on one's biological question, model organism, budget, and to some extent adventurousness (see Fig. 4, Table 1). Microarray technology is transitioning from an esoteric tool used primarily to measure mRNA levels to a general approach that might soon be as common as Southern blotting and PCR. Exciting developments in high-throughput sequencing might make this the next technology of choice, replacing microarrays within a few years. Further down the road, single-molecule sequencing technologies promise to make a reality the sensitive analysis of tiny quantities of a sample that is free of amplification artifacts. The transition of genomics from the province of highly specialized laboratories and consortia to a standard tool of molecular biology promises to revolutionize every field in which, as in DNA methylation research, genomic information is of use.

References

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837.

Bentley, D. R. (2006). Whole-genome re-sequencing. *Curr. Opin. Genet. Dev.* **16**, 545-552.

Bibikova, M., Chudin, E., Wu, B., Zhou, L., Garcia, E. W., Liu, Y., Shin, S., Plaia, T. W., Auerbach, J. M., Arking, D. E. et al. (2006a). Human embryonic stem cells have a unique epigenetic signature. *Genome Res.* **16**, 1075-1083.

Bibikova, M., Lin, Z., Zhou, L., Chudin, E., Garcia, E. W., Wu, B., Doucet, D., Thomas, N. J., Wang, Y., Vollmer, E. et al. (2006b). High-throughput DNA methylation profiling using universal bead arrays. *Genome Res.* **16**, 383-393.

Bird, A. P. and Southern, E. M. (1978). Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from Xenopus laevis. *J. Mol. Biol.* **118**, 27-47.

Bird, A., Taggart, M., Frommer, M., Miller, O. J. and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* **40**, 91-99.

Braslavsky, I., Hebert, B., Kartalov, E. and Quake, S. R. (2003). Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci. USA* **100**, 3960-3964.

Brena, R. M., Huang, T. H. and Plass, C. (2006). Quantitative assessment of DNA methylation: potential applications for disease diagnosis, classification, and prognosis in clinical settings. *J. Mol. Med.* **84**, 365-377.

Callinan, P. A. and Feinberg, A. P. (2006). The emerging science of epigenomics. *Hum. Mol. Genet.* **15** Spec No 1, R95-R101.

Chan, S. W., Henderson, I. R. and Jacobsen, S. E. (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nat. Rev. Genet.* **6**, 351-360.

Clark, S. J., Harrison, J., Paul, C. L. and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* **22**, 2990-2997.

Clark, S. J., Statham, A., Stirzaker, C., Molloy, P. L. and Frommer, M. (2006). DNA methylation: bisulphite modification and analysis. *Nat. Protoc.* **1**, 2353-2364.

Cross, S. H., Charlton, J. A., Nan, X. and Bird, A. P. (1994). Purification of CpG islands using a methylated DNA binding column. *Nat. Genet.* **6**, 236-244.

Dalma-Weiszhausz, D. D., Warrington, J., Tanimoto, E. Y. and Miyada, C. G. (2006). The affymetrix GeneChip platform: an overview. *Methods Enzymol.* **410**, 3-28.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V. K., Attwood, J., Burger, M., Burton, J., Cox, T. V., Davies, R., Down, T. A. et al. (2006). DNA methylation profiling of human chromosomes **6**, 20 and 22. *Nat. Genet.* **38**, 1378-1385.

Ehrich, M., Nelson, M. R., Stanssens, P., Zabeau, M., Liloglou, T., Xinarianos, G., Cantor, C. R., Field, J. K. and van den Boom, D. (2005). Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc. Natl. Acad. Sci. USA* **102**, 15785-15790.

Ehrich, M., Field, J. K., Liloglou, T., Xinarianos, G., Oeth, P., Nelson, M. R., Cantor, C. R. and van den Boom, D. (2006). Cytosine methylation profiles as a molecular marker in non-small cell lung cancer. *Cancer Res.* **66**, 10911-10918.

Fan, J. B., Gunderson, K. L., Bibikova, M., Yeakley, J. M., Chen, J., Wickham Garcia, E., Lebruska, L. L., Laurent, M., Shen, R. and Barker, D. (2006). Illumina universal bead arrays. *Methods Enzymol.* **410**, 57-73.

Freitag, M. and Selker, E. U. (2005). Controlling DNA methylation: many roads to one modification. *Curr. Opin. Genet. Dev.* **15**, 191-199.

Gehring, M. and Henikoff, S. (2007). DNA methylation dynamics in plant genomes. *Biochim. Biophys. Acta* **1769**, 276-286.

Goll, M. G. and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481-514.

Gowher, H., Leismann, O. and Jeltsch, A. (2000). DNA of Drosophila melanogaster contains 5-methylcytosine. *EMBO J.* **19**, 6918-6923.

Heard, E. and Disteche, C. M. (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.* **20**, 1848-1867.

Hellman, A. and Chess, A. (2007). Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141-1143.

Jacobsen, S. E., Sakai, H., Finnegan, E. J., Cao, X. and Meyerowitz, E. M. (2000). Ectopic hypermethylation of flower-specific genes in Arabidopsis. *Curr. Biol.* **10**, 179-186.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R. A., Niveleau, A., Cedar, H. et al. (2006). Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat. Genet.* **38**, 149-153.

Khulan, B., Thompson, R. F., Ye, K., Fazzari, M. J., Suzuki, M., Stasiek, E., Figueroa, M. E., Glass, J. L., Chen, Q., Montagna, C. et al. (2006). Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res.* **16**, 1046-1055.

Klose, R. J. and Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31**, 89-97.

Korch, C. and Hagblom, P. (1986). In-vivo-modified gonococcal plasmid pJD1. A model system for analysis of restriction enzyme sensitivity to DNA modifications. *Eur. J. Biochem.* **161**, 519-524.

DEVELOPMENT

Kreil, D. P., Russell, R. R. and Russell, S. (2006). Microarray oligonucleotide probes. *Methods Enzymol.* **410**, 73-98.
Laird, P. W. (2003). The power and the promise of DNA methylation markers. *Nat. Rev. Cancer* **3**, 253-266.
Levene, M. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G. and Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**, 682-686.
Lieb, J. D., Beck, S., Bulyk, M. L., Farnham, P., Hattori, N., Henikoff, S., Liu, X. S., Okumura, K., Shiota, K., Ushijima, T. et al. (2006). Applying whole-genome studies of epigenetic regulation to study human disease. *CytoGenet. Genome Res.* **114**, 1-15.
Lindsay, S. and Bird, A. P. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* **327**, 336-338.
Lippman, Z., Gendrel, A. V., Black, M., Vaughn, M. W., Dedhia, N., McCombie, W. R., Lavine, K., Mittal, V., May, B., Kasschau, K. D. et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471-476.
Lippman, Z., Gendrel, A. V., Colot, V. and Martienssen, R. (2005). Profiling DNA methylation patterns using genomic tiling microarrays. *Nat. Methods* **2**, 219-224.
Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z. et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.
Meissner, A. and Jaenisch, R. (2006). Mammalian nuclear transfer. *Dev. Dyn.* **235**, 2460-2469.
Meissner, A., Gnirke, A., Bell, G. W., Ramsahoye, B., Lander, E. S. and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868-5877.
Meyers, B. C., Vu, T. H., Tej, S. S., Ghazal, H., Matvienko, M., Agrawal, V., Ning, J. and Haudenschild, C. D. (2004). Analysis of the transcriptional complexity of Arabidopsis thaliana by massively parallel signature sequencing. *Nat. Biotechnol.* **22**, 1006-1011.
Mito, Y., Henikoff, J. G. and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* **37**, 1090-1097.
Mito, Y., Henikoff, J. G. and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. *Science* **315**, 1408-1411.
Nuwaysir, E. F., Huang, W., Albert, T. J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J. P., Ballin, J. et al. (2002). Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.* **12**, 1749-1755.
Penterman, J., Zilberman, D., Huh, J. H., Ballinger, T., Henikoff, S. and Fischer, R. L. (2007). DNA demethylation in the Arabidopsis genome. *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.0701861104.
Rai, K., Nadauld, L. D., Chidester, S., Manos, E. J., James, S. R., Karpf, A. R., Cairns, B. R. and Jones, D. A. (2006). Zebra fish Dnmt1 and Suv39h1 regulate organ-specific terminal differentiation during development. *Mol. Cell. Biol.* **26**, 7077-7085.
Reynaud, C., Bruno, C., Boullanger, P., Grange, J., Barbesti, S. and Niveleau, A. (1992). Monitoring of urinary excretion of modified nucleosides in cancer patients using a set of six monoclonal antibodies. *Cancer Lett.* **61**, 255-262.
Richards, E. J. (2006). Inherited epigenetic variation – revisiting soft inheritance. *Nat. Rev. Genet.* **7**, 395-401.
Rollins, R. A., Haghighi, F., Edwards, J. R., Das, R., Zhang, M. Q., Ju, J. and Bestor, T. H. (2006). Large-scale structure of genomic methylation patterns. *Genome Res.* **16**, 157-163.
Schatz, P., Dietrich, D. and Schuster, M. (2004). Rapid analysis of CpG methylation patterns using RNase T1 cleavage and MALDI-TOF. *Nucleic Acids Res.* **32**, e167.
Schatz, P., Distler, J., Berlin, K. and Schuster, M. (2006). Novel method for high throughput DNA methylation marker evaluation using PNA-probe library hybridization and MALDI-TOF detection. *Nucleic Acids Res.* **34**, e59.
Schumacher, A., Kapranov, P., Kaminsky, Z., Flanagan, J., Assadzadeh, A., Yau, P., Virtanen, C., Winegarden, N., Cheng, J., Gingeras, T. et al. (2006). Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res.* **34**, 528-542.
Selker, E. U., Tountas, N. A., Cross, S. H., Margolin, B. S., Murphy, J. G., Bird, A. P. and Freitag, M. (2003). The methylated component of the Neurospora crassa genome. *Nature* **422**, 893-897.
Shendure, J., Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., Wang, M. D., Zhang, K., Mitra, R. D. and Church, G. M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728-1732.
Soppe, W. J., Jacobsen, S. E., Alonso-Blanco, C., Jackson, J. P., Kakutani, T., Koornneef, M. and Peeters, A. J. (2000). The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6**, 791-802.
Sutherland, E., Coe, L. and Raleigh, E. A. (1992). McrBC: a multisubunit GTP-dependent restriction endonuclease. *J. Mol. Biol.* **225**, 327-348.
Tompa, R., McCallum, C. M., Delrow, J., Henikoff, J. G., van Steensel, B. and Henikoff, S. (2002). Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr. Biol.* **12**, 65-68.
Tost, J. and Gut, I. G. (2006). Analysis of gene-specific DNA methylation patterns by pyrosequencing(r) technology. *Methods Mol. Biol.* **373**, 89-102.
Tost, J., Schatz, P., Schuster, M., Berlin, K. and Gut, I. G. (2003). Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. *Nucleic Acids Res.* **31**, e50.
Ushijima, T. (2005). Detection and interpretation of altered methylation patterns in cancer cells. *Nat. Rev. Cancer* **5**, 223-231.
Vercoutere, W. A., Winters-Hilt, S., DeGuzman, V. S., Deamer, D., Ridino, S. E., Rodgers, J. T., Olsen, H. E., Marziali, A. and Akeson, M. (2003). Discrimination among individual Watson-Crick base pairs at the termini of single DNA hairpin molecules. *Nucleic Acids Res.* **31**, 1311-1318.
Waalwijk, C. and Flavell, R. A. (1978). MspI, an isoschizomer of HpaII which cleaves both unmethylated and methylated hpaII sites. *Nucleic Acids Res.* **5**, 3231-3236.
Walsh, C. P. and Bestor, T. H. (1999). Cytosine methylation and mammalian development. *Genes Dev.* **13**, 26-34.
Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L. and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* **37**, 853-862.
Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M. and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457-466.
Wolber, P. K., Collins, P. J., Lucas, A. B., De Witte, A. and Shannon, K. W. (2006). The Agilent in situ-synthesized microarray platform. *Methods Enzymol.* **410**, 28-57.
Xiao, W., Custard, K. D., Brown, R. C., Lemmon, B. E., Harada, J. J., Goldberg, R. B. and Fischer, R. L. (2006). DNA methylation is critical for Arabidopsis embryogenesis and seed viability. *Plant Cell* **18**, 805-814.
Yuan, E., Haghighi, F., White, S., Costa, R., McMinn, J., Chun, K., Minden, M. and Tycko, B. (2006). A single nucleotide polymorphism chip-based method for combined genetic and epigenetic profiling: validation in decitabine therapy and tumor/normal comparisons. *Cancer Res.* **66**, 3443-3451.
Zhang, X. and Jacobsen, S. E. (2006). Genetic analyses of DNA methyltransferases in Arabidopsis thaliana. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 439-447.
Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S. W., Chen, H., Henderson, I. R., Shinn, P., Pellegrini, M., Jacobsen, S. E. et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. *Cell* **126**, 1189-1201.
Zhou, Y., Bui, T., Auckland, L. D. and Williams, C. G. (2002). Undermethylated DNA as a source of microsatellites from a conifer genome. *Genome* **45**, 91-99.
Zilberman, D., Gehring, M., Tran, R. K., Ballinger, T. and Henikoff, S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* **39**, 61-69.