FORUM REVIEW ARTICLE

Deciphering the Epigenetic Code: An Overview of DNA Methylation Analysis Methods

Muhammad Umer^{1,2} and Zdenko Herceg¹

Abstract

Significance: Methylation of cytosine in DNA is linked with gene regulation, and this has profound implications in development, normal biology, and disease conditions in many eukaryotic organisms. A wide range of methods and approaches exist for its identification, quantification, and mapping within the genome. While the earliest approaches were nonspecific and were at best useful for quantification of total methylated cytosines in the chunk of DNA, this field has seen considerable progress and development over the past decades. **Recent Advances:** Methods for DNA methylation analysis differ in their coverage and sensitivity, and the method of choice depends on the intended application and desired level of information. Potential results include global methyl cytosine content, degree of methylation at specific loci, or genome-wide methylation maps. Introduction of more advanced approaches to DNA methylation analysis, such as microarray platforms and massively parallel sequencing, has brought us closer to unveiling the whole methylome. **Critical Issues:** Sensitive quantification of DNA methylation from degraded and minute quantities of DNA and high-throughput DNA methylation mapping of single cells still remain a challenge. **Future Directions:** Developments in DNA sequencing technologies as well as the methods for identification and mapping of 5-hydroxymethylcytosine are expected to augment our current understanding of epigenomics. Here we present an overview of methodologies available for DNA methylation analysis with special focus on recent developments in genome-wide and high-throughput methods. While the application focus relates to cancer research, the methods are equally relevant to broader issues of epigenetics and redox science in this special forum. *Antioxid. Redox Signal.* 18, 1972–1986.

Introduction

OVER THE PAST DECADE, research in epigenetics has been making it into mainstream research. This is due to the fact that epigenetic mechanisms have emerged as key events involved in the regulation of critical biological processes. Epigenetic mechanisms have been proposed to function as an interface between the genome and environment; therefore, epigenetic deregulation is likely to be involved in the etiology of human diseases associated with environmental exposures (22, 25). However, the high variability of epigenetic states between different cell types within the same organism as well as epigenetic fluctuation in time, even within a single cell, represents an important challenge to a complete understanding of the normal epigenetic landscape and its dynamic variability.

DNA methylation is the most extensively studied epigenetic mechanism, and it plays multiple roles in key cellular processes, including regulation of gene expression, embryonic development, genomic imprinting, and chromosome stability (Fig. 1A). Although methylation of other nucleotides does exist, at least for this article, DNA methylation refers to the attachment of a methyl group to the 5-carbon (C5) position of cytosine, mostly in the context of so-called CpG dinucleotide. This process is mediated by two main categories of DNA methyltransferases (DNMTs): *de novo* DNA methyltransferases (DNMT3A and DNMT3B) and the maintenance methyltransferase (DNMT1) (Fig. 1B). DNA methylation changes, like other epigenetic changes, are in principle reversible, which makes them attractive targets for the epigenetic therapy (Fig. 1C).

A plethora of experimental studies demonstrate that deregulation of DNA methylation is intimately linked with

¹Epigenetics Group, International Agency for Research on Cancer (IARC), Lyon, France.

²Drug Discovery & Structural Biology Group, National Institute for Biotechnology & Genetic Engineering (NIBGE), Faisalabad, Pakistan.

FIG. 1. DNA methylation/demethylation machinery and potential for novel drug targets. (A) Role of DNA methylation in various cellular processes. (B) DNA methylation machinery. (C) DNA methylation is a reversible phenomenon. DNMT, DNA methyltransferase. (To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/ars.)

many human diseases, most notably cancer. In tumor cells, DNA methylation changes are found in two forms: gene promoter-associated (CpG island [CGI]-specific) hypermethylation and concurrent global loss of 5-methyl-cytosine (global hypomethylation) (Fig. 2).

A plethora of experimental studies demonstrate that epigenetic deregulation is intimately linked with many human disease, most notably cancer. The ubiquity and early appearance of epigenome alterations in human malignancies make them attractive targets for biomarker discovery, therapeutic intervention, and prevention (37, 48). In recent years, we have witnessed an emergence of powerful technologies in epigenomics that allow a high-throughput detection of epigenetic changes with a genome-wide scope. (Fig. 3) These remarkable advances, notably those linked to the development and application of massively parallel sequencing technologies, have tremendously accelerated epigenomic research and opened up new perspectives. In this review, we summarize methodologies for DNA methylation analysis and recent developments in genome-wide and high-throughput methods that are increasingly available and discuss their potential for use in cancer research. A comparative summary of methods discussed in the article is given in Table 1. For more comprehensive overviews of methodologies for epigenetic/epigenomic analysis, readers are directed to these excellent recent reviews (21, 35).

Preparing DNA for Methylation Analysis

Owing to inability of standard molecular biology techniques such as hybridization and sequencing in differentiating cytosine from its methylated counterpart, and erasure of methylation information during PCR amplification, sample DNA has to undergo certain pretreatments. These pretreatments are methylation dependent and fall under three

categories: (i) restriction endonuclease (RE) treatment, (ii) affinity enrichment, and (iii) bisulfite treatment.

RE treatment

REs are very important tools in molecular biology research. Activity of a few restriction enzymes is dependent on the methylation status of particular nucleotides in their recognition sequence. Some enzymes are inhibited (methylation sensitive), while some are only active if their recognition site is methylated (methylation dependent) at a particular base. Methylation-sensitive and insensitive isoschizomer pairs have long been handy tools in methylation research. *HpaII* (sensitive)/*MspI* (insensitive) and *SmaI* (sensitive)/*XmaI* (insensitive) pairs are among the most widely used. *McrBC* is one of the most commonly used methylation-dependent REs. (cut only if cytosines in recognition sequence are methylated). Its unique recognition sequence [$R^mC(N)_{55-103}R^mC$] helps indirect enrichment of unmethylated sections of genome.

Affinity enrichment

Affinity enrichment of the methylated DNA fraction exploiting 5-methylcytosine (5mC)-specific antibodies or in other case proteins that have particular affinity for methylated CpGs (methyl-binding domain [MBD]-containing proteins) provides a powerful unbiased alternative overcoming constraint of recognition site dependence related to RE-based strategies. Methylated DNA immunoprecipitation (MeDIP) utilizes 5mC-specific antibodies to immunoprecipitate methylated DNA (58). Among MBD family proteins, MeCP2 was the first one to be used (12). Multimerized methyl DNA-binding domains of another MBD protein, MBD1, have also been widely used (31). Methylated CGI recovery assay (MIRA) is another MBD protein-based approach that utilizes

FIG. 2. Genomic distribution of DNA methylation. **A** (i) & (ii): In a normal cell, most promoter sequences are protected from methylation (i), while oncogene promoters and repetitive elements are heavily methylated helping keep oncogenes silent and maintain genomic stability (ii). **A** (iii) & (iv): Aberrant DNA methylation in tumor cells causes promoter hypermethylation-mediated silencing of genes (iii). Loss of DNA methylation at previously hypermethylated repetitive elements causes genomic instability (iv). **(B)** Methylation status of individual cytosines may differ from others in close vicinity. In addition to CpG Islands, flanking shores and shelves also play important role in gene regulation (26). Empty stars (☆) indicate unmethylated cytosines. Filled stars (★) indicate methylated cytosines.

a high-affinity protein complex made up of short isoform of MBD2 (MBD2B) and its binding partner MBD3L1 (47).

Bisulfite treatment

Although it was long known that sodium bisulfite treatment of DNA deaminates unmethylated cytosines into uracil at a rate much faster than the methylated ones, it was not until the description of the method by Frommer *et al.* in 1992 that bisulfite treatment was introduced into methylation research (24). By using optimal conditions, all unmethylated cytosines are converted to uracils, while 5mC remains unconverted. Subsequent PCR amplification will replace all unmethylated cytosines (uracil) with thymine, and this way an epigenetic methyl mark is converted into a genetic difference.

Global Estimation of 5mC Content

Global methylation content of DNA can provide basic information about the disease process and progress and can be

useful in various clinical settings as biomarker as well drug screening (32). Global content of 5mC can be quantified by high-performance capillary electrophoresis (HPCE) or high-precision liquid chromatography (HPLC)-based separation of individual nucleosides (generated by enzymatic hydrolysis of DNA) coupled with electrospray ionization/mass spectrometry. Both HPCE and HPLC can separate unmethylated cytosines from the methylated ones (2).

Based on one of the most common isoschizomer pairs used in DNA methylation analysis, *HpaII*/*MspI* Karimi *et al.* reported a luminometric method for global DNA methylation quantification (luminometric methylation assay [LUMA]). Using *EcoRI* as an internal control, the DNA is digested in two separate reactions (*MspI* + *EcoRI* and *HpaII* + *EcoRI*). *EcoRI* leaves a 5' overhang—AATT—while *HpaII*/*MspI* leaves a 5' overhang—GC. Using a pyrosequencing platform, these overhangs are filled during the sequential addition of nucleotides. A failure in digestion by *HpaII* owing to methylation at the recognition site will result in an absence or relative

FIG. 3. Summary of DNA methylation analysis methods. 5mC, 5-methylcytosine; BS, bisulfite; DMH, differential methylation hybridization; HPCE-ESI/MS, high-performance capillary electrophoresis–electron spray ionization/mass spectrometry; LUMA, luminometric methylation assay; MALDI-TOF, matrix-assisted laser desorption ionization–time of flight; MSP, methylation specific PCR; NGS, next-generation sequencing; RLGS, restriction landmark genomic scanning.

reduction (in comparison to *MspI*-digested aliquot) of peaks during G and C nucleotide dispensations. Using mathematical calculations, the degree of methylation at all *HpaII/MspI* sites across the genome is quantified and is represented as a% methylation (32). Other enzyme combinations can also be used.

Repetitive elements are spread across the genome and are normally heavily methylated. The degree of methylation at repetitive elements such as long interspersed element-1 (LINE-1) is directly proportional to global methylation content (60). Bisulfite pyrosequencing of repetitive elements can provide a global picture of DNA methylation levels and has been widely used.

Locus-Specific DNA Methylation Analysis

Numerous methods have been reported for locus-specific analysis. While the earlier ones relied exclusively on restriction enzymes, application of bisulfite conversion has revolutionized the field. Methylation at one or more CpG sites within a particular locus can be determined either qualitatively (presence or absence) or quantitatively. DNA samples are usually derived from a heterogeneous population of cells, in which individual cells may vary vastly in their DNA methylation patterns. Hence, most of the methods aimed at quantitative measurement of DNA methylation determine the average methylation level across many DNA molecules. The results are presented as percent methylation—percentage of DNA molecules, in a given sample, methylated at the specific cytosine position under investigation (see more details in the following paragraphs and also Fig. 4). There is a long list of locus-specific methods available; hence, only a few will be discussed here keeping in mind the extent and ease of their usage.

Methylation-specific PCR (MSP) is one of the most widely used methods in DNA methylation studies. Using primers that can discriminate between methylated (M primer pair) or

unmethylated (U primer pair) target region after bisulfite treatment, DNA is PCR amplified. One primer in both M and U primer pairs necessarily contains a CpG site near its 3' end. This CpG site is the one under investigation, and both M and U primers contain the same site. A forward primer in M pair having the C nucleotide in its sequence for the CpG position under investigation will fail to amplify the region if that particular cytosine is unmethylated (hence, converted to uracil during bisulfite reaction) and vice versa. Success or failure in amplification can qualitatively determine the methylation status of the target site (26). Although rapid and easy to use, MSP suffers from various disadvantages such as reliance on gel electrophoresis and the fact that only a very few CpG sites can be analyzed using a given primer pair. The method reported originally by Herman *et al.* (26) was at best qualitative. However using TaqMan® technology, Eads *et al.* reported a quantitative MSP-based method for DNA methylation analysis (17). In this method, named as MethyLight, bisulfite-converted DNA is amplified and detected by methylation-state-specific primers and TaqMan® probes in a real-time PCR. As the two strands of DNA no longer remain complementary after bisulfite conversion, primers and probes are targeted for either of the resulting strands. Initial template quantity can be measured by traditional real-time PCR calculations. Incorporation of various quality controls for bisulfite conversion and recovery of DNA after bisulfite treatment have improved the quantitative reliability of this method. MethyLight has many advantages over MSP and other locus-specific DNA methylation analysis methods. It avoids gel electrophoresis, restriction enzyme digestion, radiolabeled dNTPs, and hybridization probes, yet there are a few shortcomings as well to this method, the biggest one being PCR bias. As mentioned above, individual DNA molecules originating from a mixed population of cells can vastly differ from each other in their methylation status. Such a variable

TABLE 1. SUMMARY AND COMPARISON OF DNA METHYLATION ANALYSIS METHODS

S. No	Method	Abbreviation	Advantages	Disadvantages	References
<i>Locus-specific methods</i>					
1.	Methylation-specific PCR	MSP	Cost effective. Rapid. Easy to carry out	False positive, only very few CpG sites can be analyzed for a given primer pair	(26)
2.	MethylLight	—	Avoids gel electrophoresis. Does not require restriction enzyme digestion.	PCR Bias	(17)
3.	Combined bisulfite restriction analysis	COBRA	Cost effective	Limited number of sites that can be exploited for this purpose	(7)
4.	Bisulfite (Sanger) sequencing	BS	The methylation status of individual CpG sites can be analyzed with great precision	Cloning may be required	(24)
5.	Bisulfite pyrosequencing	—	Small quantities of DNA required. Compatible with degraded DNA obtained from archived samples.	Relatively short fragments can be analyzed. Robust pyrosequencing assays cannot be easily designed for all regions of interest	(54)
6.	Base-specific cleavage coupled to mass spectrometry	—	Quantitative determination of DNA methylation at individual CpG sites with high sensitivity.	High cost associated with required equipment	(19)
7.	HeavyMethyl	—	Highly sensitive, particularly useful for samples with very low percentage of methylated sequences	Many oligonucleotides are required	(15)
8.	Methylation-sensitive single-nucleotide primer extension	MS-SnuPE	Requires small amount of DNA. Multiplexing is possible.	Requires radioactive labeling	(34)
9.	Methylation-sensitive melting curve analysis	MS-MCA	Less laborious. Avoids contamination for being closed tube method.	Interpretation of results can be difficult if too many CpG sites are present in amplicon	(34)
10.	Methylation-sensitive high-resolution melting	MS-HRM	Overcomes PCR bias. Less laborious. Avoids contamination for being closed tube method.	Interpretation of results can be difficult if too many CpG sites are present in amplicon	(34)
<i>Estimation of global methyl-cytosine content</i>					
11.	Luminometric methylation assay	LUMA	Robust. Simple experimental setup. High throughput	Limited by the number of enzyme recognition sites	(32)
<i>Genome-scale Approaches</i>					
<i>Earlier methods</i>					
12.	Restriction landmark genomic scanning	RLGS	Can scan thousands of sites in one reaction	Large quantities of DNA are required	(11)
13.	Methylation-specific arbitrarily primed PCR	MS-AP-PCR	Simple experimental setup	Labor intensive. Reliance on gel electrophoresis	(36)
14.	Amplification of intermethylated sites	AIMS	Simple experimental setup	Labor intensive. Reliance on gel electrophoresis	(23)
15.	Methylation-sensitive representational differential analysis	MS-RDA	Enrichment of heavily methylated repetitive sequences by use of MSREs	Complex procedures. Not much efficient with fragments <1Kb	(55)
<i>Microarray based approaches</i>					
<i>Restriction enzyme treatment</i>					
16.	Differential methylation hybridization	DMH	Use of <i>MseI</i> precutting depletes the sample from CpG-poor sequences	Coverage limited by the number of RE sites	(29, 35)

(continued)

TABLE 1. CONTINUED

S. No	Method	Abbreviation	Advantages	Disadvantages	References
17.	Comprehensive high-throughput arrays for relative methylation	CHARM	Increased sensitivity to densely methylated regions	Moderate resolution	(30)
18.	<i>HpaII</i> tiny-fragment enrichment by ligation mediated PCR coupled with microarray	HELP Assay ^a	Use of <i>HpaII</i> tiny fragments helps analyze non-CpG island regions as well	More sensitive for CpG-poor sequences as compared to CpG-rich sequences	(30, 33)
19.	Methyl CpG island amplification (MCA) coupled with microarray	MCAM	Targeting methylated CpG islands before amplification increases specificity	Low resolution as compared to other RE-based methods that use a pool of REs	(20, 35)
<i>Affinity enrichment</i>					
20.	Methylated DNA immunoprecipitation	MeDIP-Chip ^a	Efficient for CGI and repetitive sequences	Require substantial quantities of high quality DNA. Bias toward low CpG density regions	(30, 43, 58)
21.	Methylated CpG island recovery assay	MIRA-Chip ^a	Efficient for CGI and repetitive sequences	Require substantial quantities of high-quality DNA	(47)
<i>Bisulfite treatment</i>					
22.	GoldenGate ®	—	Compatible with degraded DNA obtained from FFPE samples. High throughput	Very limited number of CpG sites can be analyzed	(4)
23.	Illumina Infinium HumanMethylation27 BeadChip	—	Fairly high throughput. Can analyze around 27,000 CpG sites located in the promoter regions of <i>CCDS</i> genes	WGA step might introduce bisulfite PCR bias	(5)
24.	Illumina Infinium HumanMethylation450 BeadChip	—	Includes CpG island shores, 99% RefSeq genes, and 96% CpG islands	WGA step might introduce bisulfite PCR bias	(3)
<i>Sequencing-based approaches</i>					
<i>Bisulfite treatment</i>					
25.	BS-Seq	—	Provides unbiased single-nucleotide resolution methylation maps	Cost intensive and low sample throughput	(10)
26.	MethylC-Seq	—	Provides unbiased single-nucleotide resolution methylation maps	Cost intensive and low sample throughput	(39, 40)
27.	Reduced representational bisulfite sequencing ^b	RRBS	Reduces sequence redundancy and amount of sequencing required	Bias toward high-CpG-density regions	(42)
<i>Affinity Enrichment</i>					
28.	MethylCap-seq	—	Step-wise elution helps stratify DNA into fractions of variable methyl-CpG density	Cannot distinguish CpG and non-CpG methylation	(8)
29.	MBD-isolated genome sequencing	MIGS	Alleviates sporadically methylated CpGs in low-CpG-density regions, which might have an uncertain biological relevance	Cannot distinguish CpG and non-CpG methylation. Bias toward high-CpG-density regions	(49)

^aHELP, MeDIP and MIRA all have been used with downstream sequencing as well and except for the general superiority of sequencing over microarray, advantages and disadvantages of all these enrichment methods remain almost the same.

^bRRBS employs both restriction digestion and bisulfite conversion.

FFPE, formalin-fixed paraffin-embedded; Res, restriction endonucleases; WGA, whole-genome amplification.

FIG. 4. Bisulfite pyrosequencing. During a pyrosequencing reaction, sequencing primer annealed to template DNA molecule is extended by sequential addition of nucleotides A, T, C, and G, one type of nucleotide at a time. A pyrophosphate molecule is released as a result of successful nucleotide incorporation, which is then utilized by sulfurylase to generate a single ATP molecule. The released ATP is used to generate a light signal by luciferase, which is then recorded by a camera and presented as peaks in the output called pyrogram as shown in figure. The height of peak is directly proportional to intensity of light generated, which in turn is proportional to the number of nucleotides added in one cycle (e.g., if a single T is present in sequence, the peak height will be lowest while a stretch of three T nucleotides will result in a taller peak. Compare nucleotide positions 20 and 30 in both *upper* and *lower* panels in figure). If a nucleotide dispensation order is designed in such a way that for every cytosine position under investigation, T and C nucleotides are added in succession (see highlighted areas in figure), it can reveal methylation levels in a quantitative way in given DNA sample. For variable positions, the proportion of alleles (methylated cytosines *vs.* unmethylated ones—C/T) can be deduced by relative peak heights.

population will result in DNA molecules differing widely in their cytosine content after bisulfite treatment; highly methylated molecules will be C-rich, whereas unmethylated ones will be T-rich. These two populations can sometimes amplify with varying efficiencies, a phenomenon termed as PCR bias. Warnecke *et al.* reported preferential amplification of T-rich unmethylated sequences (57). This PCR bias can potentially affect the accurate quantitative estimation of DNA methylation.

Combined bisulfite conversion restriction analysis (COBRA) is based on the principle of loss or retention of a restriction enzyme site after bisulfite treatment, depending on the methylation status of targeted cytosine. Alternatively, creation of a new restriction site is also possible (7). Both of these facts can be utilized to analyze DNA methylation in a target region whose sequence is already known. Bisulfite-treated DNA is PCR amplified using primers flanking the target site and is subsequently digested with a restriction enzyme. Combined with an electrophoresis through microfluidic chips, such as Agilent™ 2100 Bioanalyzer, a procedure named as Bio-COBRA, accurate assessment of all the resulting restriction fragments can provide a quantitative measurement of the methylation status at the target region (7). Although quite useful and robust, the biggest disadvantage of this method is limited number of restriction sites that can be used.

Bisulfite sequencing of DNA using Sanger chemistry has also been used for locus-specific methylation analysis.

Originally introduced by Frommer *et al.* (24), this method is based on PCR amplification of bisulfite-converted DNA using specific primers, followed by cloning. To determine an average picture of methylation across millions of DNA molecules, a few clones are randomly selected and sequenced. The massively parallel revolution in sequencing has also benefited locus-specific approaches. More than 100 PCR products from different tissues were analyzed on 454 sequencing platform (Roche/454 Life Sciences, Branford, CT) in a single run (51). For each PCR product, on the average, >1600 sequences were generated, far larger than the clones usually sequenced (around 20 for each PCR product). This way more accurate measurement of average methylation levels in the given population of DNA molecules can be carried out (51). High throughput of this method can prove to be very useful in candidate region approaches, as many samples can be analyzed simultaneously in one single run that too for potentially more than one target regions.

Pyrosequencing is sequencing by synthesis approach and has been widely used for DNA methylation analysis. Pyrosequencing offers a highly reliable, quantitative, and high-throughput method for analysis of DNA methylation at multiple CpG sites with built-in internal control for completeness of bisulfite treatment. As bisulfite treatment converts unmethylated cytosines into uracils (which will be converted to thymine upon subsequent PCR amplification)

leaving methylated ones unchanged, the methylation difference between cytosines is converted into a C/T genetic polymorphism and can be quantified likewise (54). Although bisulfite pyrosequencing is one of the most widely used methods for quantitative determination of methylation, it is limited by a few drawbacks. Thermal instability of enzymes used in pyrosequencing reactions, particularly luciferase, requires the reaction to be carried out at 28°C. Therefore, optimal amplicon size to be subsequently used for pyrosequencing reaction is around 300 bp or less to avoid secondary structures (13). As bisulfite conversion results in low-complexity DNA molecules (A, T, and G nucleotides, except very few methylated cytosines), designing optimal primer sets for every region of interest is a difficult task (52). For more details, see Figure 4.

In yet another high-throughput quantitative approach, bisulfite-treated DNA is first amplified with specific primers. The reverse primer is tagged at the 5' end with T7 promoter sequence to facilitate *in vitro* transcription by phage RNA polymerase in the next step. Endonuclease RNase A, which cuts after every C and U in an RNA molecule, is used to generate short fragments. However, for only C-specific or U-specific cleavage, two separate *in vitro* transcription reactions are run. In an U-specific cleavage reaction, dCTP is used instead of CTP. This blocks cleavage after C and RNase A only cut after U. Similarly, a separate C-specific cleavage reaction is set up. This way, a complex mixture of short oligonucleotides of varying lengths is generated. Methylation-dependent C/T polymorphism in bisulfite-converted DNA is reflected as G/A in transcribed RNA molecules and results in a 16-Da mass difference for each CpG site in cleavage products, which is then analyzed by matrix-assisted laser desorption ionization–time-of-flight (MALDI-TOF) spectrometry (19).

HeavyMethyl method relies on nonextendable oligonucleotide blockers specific for bisulfite-converted unmethylated DNA. Primers flanking the region of interest are used to amplify target DNA fragment. The blockers anneal to a region overlapping with amplification primer-binding site, blocking primer binding to unmethylated DNA and inhibiting amplification. This results in selective amplification of methylated DNA only. In real-time PCR, amplification is monitored by a methylation-specific probe. This method is particularly useful for samples where the percentage of methylation might be very low (methylated sequences have been quantified in a background of 8000-fold excess of unmethylated DNA) (15).

Other important single-locus methods include single-nucleotide primer extension on bisulfite-converted DNA template with primer terminating immediately 5' of cytosine to be assayed (methylation-sensitive single-nucleotide primer extension [MS-SnuPE]) and methods based on differential melting properties of amplicons generated from bisulfite-converted methylated and unmethylated DNA (methylation-sensitive melting curve analysis [MS-MCA] and methylation-sensitive high-resolution melting [MS-HRM]) [see Ref. (34) and references therein].

Genome-Scale Approaches for DNA Methylation Analysis

Technological advances in the past two decades have geared up research in the fields of genomics and molecular biology ushering into an era of rapid discovery, -omics-scale approaches, and high-throughput analysis. Although locus-

specific and global analyses of DNA methylation have contributed important insights into biology and wide ranging role of DNA methylation, a comprehensive understanding requires methodologies that provide higher coverage, ideally up to single-base resolution spreading over whole genome.

Earlier approaches

Among the earliest forays into broad DNA methylation profiling, restriction landmark genomic scanning proved to be the most comprehensive and has been widely used, though labor intensive. Two-dimensional electrophoretic separation of methylation-sensitive restriction enzyme (MSRE)-cut genomic DNA, with cut ends labeled with radionucleotides, produces a radiograph of thousands of spots (~2000), essentially lacking methylated sites, which failed to be cut and hence radiolabeled. Differences in methylation between two samples are inferred from the differences in patterns of these spots. This method is CGI biased, as mostly those enzymes are used as landmarks that have recognition sites more prevalent in CGIs such as *NotI* and *AscI* (11).

Two similar PCR-based methods, namely methylation-specific arbitrarily primed PCR (MS-AP-PCR), and amplification of intermethylated sites (AIMS) have also been widely used. Both methods rely on methylation-sensitive isoschizomer digestion followed by PCR amplification, which produces a distinct fingerprint when electrophoresed. However, they differ in the REs and primers used. MS-AP-PCR uses the *HpaII*/*MspI* pair for digestion followed by amplification with arbitrary primers under low-stringency conditions (36). On the other hand, AIMS use the *SmaI*/*XmaI* pair. Both enzymes share the same recognition sequence (CCCGGG). The DNA is first treated with methylation-sensitive *SmaI*, which cuts between third C and G leaving a blunt end, thus eliminating unmethylated sites. Treatment with methylation-insensitive *XmaI* cleaves between first and second C of methylated-hence-uncut sites, leaving a four base overhang—CCGG, for adaptor ligation. Adaptor-ligated fragments are PCR amplified and resolved to produce a fingerprint of anonymous DNA bands representing methylome of the cell (23).

Methyl CpG Island Amplification (MCA) is an approach similar to AIMS, and has been used in combination with downstream dot-blot or representational difference analysis (RDA) (53). Methylation-sensitive RDA (MS-RDA) employs MSREs (*e.g.*, *HpaII*) to enrich unmethylated fraction, followed by PCR amplification and RDA (55)

Microarray-based DNA methylation analysis

Introduction of microarray technologies opened unprecedented horizons in methylome research. In contrast to hitherto discussed methods, thousands of regions of interest can be analyzed simultaneously. Various platforms have been used for studying DNA methylome differing in their resolution and regions targeted. Ranging from CGI or promoter region-specific platforms to oligonucleotide-tiling arrays virtually covering whole genome with high resolution, various arrays have been custom designed or are commercially available.

The earliest approaches to this end coupled restriction digestion with downstream hybridization to microarrays. Differential methylation hybridization (DMH) was among the first of such kind (29). Fragmented with a frequent cutter

MseI, linker-ligated DNA is either digested with an MSRE such as *Bst*UI or left untreated. This is useful to identify methylation spots in a particular sample. Alternatively, using a combination of MSREs (*Bst*UI, *Hpa*II, and *Hha*I) to increase genomic representation, two samples can be compared for differential methylation (59). Using linker-specific primers, only uncut fragments, methylated or with no recognition site, will be amplified. Radioisotope (^{32} P) labeling of DNA fractions (*Bst*UI treated and untreated) in the former case (29) or differential fluorescent labeling in later case (59) allows quantification of methylation upon hybridization of sample to microarray. Based on same principle, but using a methylation-dependent enzyme McrBC, Nouzova *et al.* studied methylation changes during leukemia cell differentiation (44). In this case, only unmethylated fragments will remain intact to be amplified, and methylation status of a particular locus can be studied in the same manner as discussed above. Use of McrBC increases sensitivity to densely methylated regions. Further improvements were made to McrBC-based microarray analysis by proposing a novel array design and optimized data processing—comprehensive high-throughput arrays for relative methylation (CHARM) (30).

*Hpa*II tiny-fragment enrichment by ligation-mediated PCR (HELP) (33) is another restriction enzyme-based strategy. Using *Hpa*II and its isoschizomer *Msp*I, two separate aliquots of DNA are fragmented. This way, the methylated fraction of DNA is eliminated from an *Hpa*II-digested sample. Fragments are ligated to adaptors and amplified by ligation-mediated PCR. Both fractions are labeled with different fluorescent dyes and hybridized to an oligonucleotide array. Absence or relative weakness of signal in the *Hpa*II fraction as compared to the one treated with *Msp*I will show the methylation status of a particular position. HELP assay was further improved by using dual-adaptor approach to amplify smaller fragments (50–200 bp) (45). MCA (discussed above) was later combined with downstream CGI microarray hybridization and named as MCAM (20). Differential labeling of two samples and subsequent array hybridization will yield comparative picture of DNA methylation profile.

Only a very small portion of methylome can be studied using restriction enzyme-based approaches. Due to this limitation, affinity enrichment of DNA has emerged as an effective alternative. DNA can be precipitated by either of the methods discussed above and has been widely used in combination with microarrays. As a general approach, DNA is first sheared randomly (either by sonication or a frequent cutter enzyme), and a portion of DNA is set aside to be used later as reference (input control). The resulting enriched fraction and input control are differentially labeled and hybridized to a custom-made or commercially available array platform. Using a custom array, Weber *et al.* developed MeDIP-chip (58). MIRA has also been used in variety of studies; for example, when combined with a whole-genome tiling array, a 100-bp resolution methylation profile human B cells was generated (47).

Bisulfite treatment of DNA results in reduced sequence complexity and increased redundancy, thus reducing hybridization specificity (35). Therefore, coupling bisulfite treatment with array hybridization has not been much successful. However, Illumina GoldenGate® BeadArray™ and Infinium® are well suited for this purpose and have been widely used. Principle of both platforms is shown in Figure 5.

GoldenGate® Human Cancer Panel I is a typical example of this assay, and up to 1536 CpG sites spanning 807 genes can be studied using this platform; however, custom-made arrays can also be used alternatively (4). An example of results obtained from GoldenGate® platform is shown in Figure 6 [adapted from Hernandez-Vargas *et al.* (27)]. The first version of Illumina Infinium Methylation assay called Infinium HumanMethylation27 BeadChip can investigate 27,578 CpG sites spread over the proximal promoter regions of 14,475 consensus coding sequences, including 110 miRNA promoters in the human genome (5). Improved version of this chip, called Infinium HumanMethylation450 BeadChip, can be used to investigate >485,000 CpG sites per sample. This version of chip has included more than 99% of RefSeq genes with multiple CpG sites per gene spread across the promoter, 5' untranslated region (UTR), first exon, gene body, and 3' UTR. In addition to 96% CGIs, the chip also includes various CpG sites corresponding to CGI shores and flanking regions (3). Both the chips can be used to analyze up to 12 samples per chip simultaneously.

Sequencing-based approaches

Sequencing a whole methylome has not been possible until very recently. Although the use of DNA-sequencing platforms to identify and quantify the methylation of individual cytosines is as old as application of bisulfite treatment for DNA methylation analysis (24), the methodology reported by Frommer *et al.* was at best useful for a very limited number of targeted regions even with the use of brute-force Sanger sequencing (18). Massively parallel revolution in sequencing has shifted the paradigm of genome-wide DNA methylation analysis. Whole methylome map of a particular cell type can now be generated in a matter of 3 to 5 days (38) at costs much lower than previously possible. As compared to array hybridization, sequencing-based analysis provides more detailed information with less DNA input and does not rely on laborious efforts needed to design an appropriate array. In addition, complexity of methylation patterns in DNA samples originating from heterogeneous population of cells and existence of cytosine methylation outside traditionally analyzed CpG-rich regions (30) demand an unbiased approach that can quantitatively determine the level of methylation ideally at single-nucleotide resolution throughout the genome.

The pioneering efforts at generating a whole-genome single base-pair resolution methylome map of eukaryotic organisms were made by two independent groups in the year 2008. By using two different shotgun bisulfite high-throughput sequencing protocols named BS-seq (10) and MethylC-seq (39), both groups generated comprehensive cytosine methylation maps of *Arabidopsis thaliana* genome. In the BS-seq approach, genomic DNA from *A. thaliana* aerial tissues was used. An interesting finding of this study was identification of a pattern of periodicity between sites of methylation, 167 nucleotides, which roughly correlates with internucleosome linker length in plants, highlighting the functional correlation of methylation machinery with nucleosomal positioning (10). Using a slightly different approach named MethylC-seq, Lister *et al.* sequenced DNA from *A. thaliana* flower buds, mapping 79% of total cytosines (16× average coverage) to the reference genome (39). MethylC-seq was also later applied to generate the first ever single-base resolution map of human methylome in

FIG. 5. Illumina methylation analysis platforms. (A) On a GoldenGate® platform, two pairs of oligos, allele-specific oligo (ASO) and locus-specific oligo (LSO), each ASO-LSO pair specific for methylated or unmethylated target CpG, are used. Each ASO has two parts: a sequence complementary to target and a priming site (P1 or P2). LSO consists of three parts: a stretch of sequence complementary to target, a priming site (P3) and an address sequence which identifies its genomic location and is complementary to capture probes on BeadArray. Hybridization of Pooled oligos to bisulfite converted fragmented immobilized DNA results in specific hybridization. ASO is then extended and ligated to corresponding LSO. During the following PCR amplification, methylated and unmethylated fractions are differentially labeled with fluorescent dyes using differentially labeled P1 and P2 primers. Hybridization to bead array results in annealing of address sequence in LSO with a unique capture probe on array. Resulting fluorescence intensities can be compared to quantitatively measure the methylation status at a particular locus (4) (see also Figure 6). (B) Infinium® platform is based on similar principle, but a little different design. An Illumina BeadArray is made up of 3-micrometer silica beads capable of self-assembling in microwells on either of two substrates: fiber optic bundles or planar silica slides. On an Illumina methylation assay chip, two types of beads are present: (M), for methylated template and (U) for unmethylated. Each bead is covered with hundreds of thousands of copies of a specific 50-mer oligonucleotide. Bisulfite-converted DNA is subjected to whole-genome amplification, fragmented, and hybridized to chip. This is followed by single base extension with hapten-labeled ddNTPs. The single base extension is only possible when correct base pairing is achieved. After complex immunohistochemical assays differentiate the two bead types and produce type-specific signals, relative intensities of which are used for quantitative estimation of methylation status of corresponding CpG site (5).

embryonic stem cells and fetal fibroblasts (40). Remarkably, this study identified that nearly one quarter of cytosine methylation in embryonic stem cells is in non-CG context, emphasizing the need of unbiased, genome-wide approach (Fig. 7).

RE treatment and affinity enrichment methods have been adapted to downstream massive sequencing, with additional advantage of reduction in target DNA complexity and

amount of sequencing. Improving original HELP assay by using two sets of adaptors to amplify <200-bp fragments during the LM-PCR step, and coupling HELP output with NGS, Oda *et al.* could analyze 98.5% CGIs in human genome. Using MspI digestion as control, methylated spots are identified by their absence in the *Hpa*II-cut fraction's sequence reads (45). Using an essentially similar approach named Methyl-seq (9), important differences in methylation patterns

FIG. 6. An example of GoldenGate® output. The fluorescent signals from array are read using a fluorescent reader. Using specialized software such as BeadStudio, signal values (AVG-Beta) corresponding to the ratio of the signal from methylated allele to the sum of the fluorescent signals of both methylated and unmethylated alleles are calculated. Using heat map feature of BeadStudio, differential methylation patterns between two samples (e.g., tumor and surrounding) can be easily detected, as shown in figure. Figure is adapted from Hernandez-Vargas *et al.* (27) (To see this illustration in color, the reader is referred to the web version of this article at www.liebertpub.com/ars.)

FIG. 7. Shotgun bisulfite sequencing on high-throughput platform (Illumina GA). (A) BS-Seq: Sonicated and end-repaired DNA is first ligated to a set of double-stranded adaptors containing *DpnI* sites (methylated A). Adaptor-ligated fragments are then bisulfite treated and PCR amplified using primers specific for bisulfite-converted adaptors. After PCR, products are digested with *DpnI*, and after generating, 3' single-A extensions are ligated to double-stranded sequencing adaptors (Illumina). Size-selected (120–170 bp) adaptor-ligated fragments are subjected to a second round of PCR amplification using adaptor-specific primers and subsequently subjected to sequencing. (B) MethylC-Seq: Genomic DNA is sonicated to generate 50–500-bp fragments, end repaired, and ligated to methylated adaptors (Illumina). Adaptor-ligated DNA fragments are size selected (120–170 bp) and subjected to two successive bisulfite treatments followed by purification. After enriching DNA by 18 cycles of PCR using adaptor-specific primers, samples are sequenced by Illumina GA.

FIG. 8. Affinity enrichment coupled to high-throughput sequencing. These workflows are based on protocols reported by (42) and (46) and may differ from other reports. A portion of DNA is kept aside (input control), which is used to compare relative enrichment of methylated fraction in precipitated samples. qPCR validation using primers for regions whose degree of methylation is already known is used to confirm success of affinity enrichment. If no such information is available, spiked in DNA standards with known methylation level can also be used in qPCR validation step. MeDIP, methylated DNA immunoprecipitation; MIRA, methylated CpG island recovery assay.

between hESCs, their *in vitro* differentiated derivatives, and human tissues have been identified. Another widely used method named as reduced representation bisulfite sequencing (RRBS) couples RE representation followed by BS sequencing on a massively parallel platform (42). *MspI* digestion before BS conversion allows reducing redundancy by selecting a CpG-rich genomic subset (42). Another RE-based approach dubbed as methyl-sensitive cut counting (MSCC) (1) was used in parallel analysis of human B lymphocyte methylome along with padlock probe-captured DNA (discussed later).

Affinity-enrichment-based methods have been adapted to downstream analysis by massive parallel sequencing. MIRA-seq (46) MeDIP-seq (16), MethylCap-seq (using MeCP2) (8), MBD-isolated genome sequencing (49) follows more or less similar protocols (except affinity enrichment itself). As an example, two of the reported workflows are summarized in Figure 8 (46, 50). Although similar in approach, enrichment methods target different compartments of the genome. While MeDIP captures methylated regions with low CpG density, MBD favors high-CpG-density regions. Moreover, inclusion of multiple elution steps with increasing salt concentrations in MBD capture protocols can enrich moderately methylated regions, making it more useful as compared to MeDIP (43).

As affinity enrichment methods target methylated DNA, the unmethylated state of DNA is inferred from its absence in reads. A confidence in this inference is highly dependent on the sequencing depth. To overcome this potential issue, Maunakea *et al.* (41) took on a unique approach to elucidate methylome of the human frontal cortex gray matter. They used two complementary methods, MeDIP (used to enrich methylated fraction of genome) and MSRE (MRE) digestion (used to enrich unmethylated fraction), followed by high-throughput sequencing to achieve a greater genome coverage. For MRE-seq, DNA is first digested in three separate reactions with REs *HpaII*, *AciI*, and *Hin6I*, combined in equal quantities and size selected (100–300 bp). Sequencing DNA from both MeDIP and MRE fractions highlighted the role of intragenic

and intergenic methylation in gene regulation, pointing out to the complexity of cytosine methylation code (41).

Other important sequence selection strategies used before sequencing include simultaneous capture and amplification of bisulfite-converted targeted regions using padlock probes (14) and capture of specific sequences by array hybridization (28).

Raw sequencing reads from each type of high-throughput platform need to undergo dedicated and complex bioinformatics analysis pipelines, which differ from each other according to the platform used and the particular type of experiment and protocol. Such bioinformatics tools are beyond the scope of this article. As a general approach, enzyme- and affinity-based sequencing methods determine the relative abundance of different genomic regions in enriched fraction by counting the number of reads that uniquely map to the reference genome as compared to input control. On the other hand, bisulfite sequencing extracts information directly from the sequence (35).

All the sequencing approaches discussed above have their own advantages and shortcomings. Shotgun bisulfite sequencing, though still a gold standard because of its genome coverage, is cost and effort intensive. This makes this approach unfeasible for studies involving large number of samples, for example, cancer epigenomic studies. Sequence selection strategies though useful are invariably prone to particular biases. While RE-based strategies are limited by the number and distribution of enzyme recognition sites, affinity enrichment methods cannot yield information on individual CpG dinucleotides (35). Therefore, while the whole-genome approach is useful to generate reference methylome maps, sequence selection strategies can yield useful information about most relevant regions. Relative merits and demerits of each method have been excellently reviewed elsewhere (35).

Future Challenges

Unlike genetic code, which is the same in every single cell of an organism, epigenetic code shows wide-ranging variability

across different cell types and also in the same cells at different developmental stages and under the influence of various environmental stimuli. This plasticity of epigenetic code poses a significant challenge in terms of effort, resources, and technology. In our opinion, it would be necessary in the long run to develop robust methods capable of analyzing methylome of single cell in a heterogeneous population of cells. From the standpoint of translational applications of DNA methylome research, such as in clinical settings as diagnostic biomarkers, methods capable of analyzing methylation patterns accurately in minute quantities of DNA need to be developed. DNA from archived formalin-fixed paraffin-embedded samples collected as a part of epidemiological studies is a special challenge, owing to its degraded state. Although a few adaptations to previously established protocols can help [for example (56)], this area is in need of further development.

Innovation

Recent discoveries of 5-hydroxymethylcytosine (5hmC) occurrence being more widespread in mammalian cells than previously thought and the role it might play in demethylation have opened up a whole new dimension to epigenetic layer of coding. As most of the methods currently in use for 5mC detection cannot differentiate between 5mC and 5hmC, future developments in this area need to be focused upon. Affinity purification and downstream sequencing with 5hmC-specific antibodies, similar to that for 5mC, cannot determine precise location of 5hmC. Recently, two excellent methods have been reported for sequencing 5hmC up to single-nucleotide resolution (6, 61).

Acknowledgments

We apologize to authors whose relevant publications were not cited due to space limitation. The work of the IARC Epigenetics Group is supported by the grants from the National Cancer Institute (NIH), United States; l'Association pour la Recherche sur le Cancer (ARC), France; la Ligue Nationale Contre le Cancer, France; the Swiss Bridge Award; and the Bill and Melinda Gates Foundation (to Z.H.).

References

1. Ball MP, Li JB, Gao Y, Lee J-H, LeProust EM, Park I-H, Xie B, Daley GQ, and Church GM. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotech* 27: 361–368, 2009.
2. Berdasco Ma, Fraga MF, Esteller M, and Tost J. Quantification of Global DNA Methylation by Capillary Electrophoresis and Mass Spectrometry. In: *Methods in Molecular Biology*, edited by Walker JM. Clifton, NJ: Humana Press, 2009, pp. 23–34.
3. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, Delano D, Zhang L, Schroth GP, Gunderson KL, Fan J-B, and Shen R. High density DNA methylation array with single CpG site resolution. *Genomics* 98: 288–295, 2011.
4. Bibikova M, Fan J-B, and Tost J. GoldenGate® Assay for DNA Methylation Profiling. In: *Methods in Molecular Biology*, edited by Walke JM. Clifton, NJ: Humana Press, 2009, pp. 149–163.
5. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, and Gunderson KL. Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics* 1: 177–200, 2009.
6. Booth MJ, Branco MR, Ficiz G, Oxley D, Krueger F, Reik W, and Balasubramanian S. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336: 934–937, 2012.
7. Brena RM, Plass C, and Tost J. Bio-COBRA: absolute quantification of DNA methylation in electrofluidics chips. In: *Methods in Molecular Biology*, edited by Walker JM. Clifton, NJ: Humana Press, 2009, pp. 257–269.
8. Brinkman AB, Simmer F, Ma K, Kaan A, Zhu J, and Stunnenberg HG. Whole-genome DNA methylation profiling using MethylCap-seq. *Methods* 52: 232–236, 2010.
9. Brunner AL, Johnson DS, Kim SW, Valouev A, Reddy TE, Neff NF, Anton E, Medina C, Nguyen L, Chiao E, Oyulu CB, Schroth GP, Absher DM, Baker JC, and Myers RM. Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res* 19: 1044–1056, 2009.
10. Cokus SJ, Feng S, Zhang X, Chen Z, Merriman B, Haubenschild CD, Pradhan S, Nelson SF, Pellegrini M, and Jacobsen SE. Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452: 215–219, 2008.
11. Costello JF, Hong C, Plass C, Smiraglia DJ, and Tost J. Restriction landmark genomic scanning: analysis of CpG islands in genomes by 2D gel electrophoresis. In: *Methods in Molecular Biology*, edited by Walker JM. Clifton, NJ: Humana Press, 2009, pp. 131–148.
12. Cross SH, Charlton JA, Nan X, and Bird AP. Purification of CpG islands using a methylated DNA binding column. *Nat Genet* 6: 236–244, 1994.
13. Dejeux E, El abdalaoui H, Gut IG, and Tost J. Identification and quantification of differentially methylated loci by the pyrosequencing technology. In: *Methods in Molecular Biology*, edited by Walker JM. Clifton, NJ: Humana Press, 2009, pp. 189–205.
14. Deng J, Shoemaker R, Xie B, Gore A, LeProust EM, Antosiewicz-Bourget J, Egli D, Maherali N, Park I-H, Yu J, Daley GQ, Eggan K, Hochedlinger K, Thomson J, Wang W, Gao Y, and Zhang K. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotech* 27: 353–360, 2009.
15. Distler J. Quantification of methylated DNA by Heavy-Methyl duplex PCR. In: *Methods in Molecular Biology*, edited by Walker JM. Clifton, NJ: Humana Press, 2009, pp. 339–346.
16. Down TA, Rakyan VK, Turner DJ, Flicek P, Li H, Kulesha E, Graf S, Johnson N, Herrero J, Tomazou EM, Thorne NP, Backdahl L, Herberth M, Howe KL, Jackson DK, Miretti MM, Marioni JC, Birney E, Hubbard TJP, Durbin R, Tavare S, and Beck S. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotech* 26: 779–785, 2008.
17. Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D, Danenberg PV, and Laird PW. MethyLight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res* 28: E32, 2000.
18. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, Burger M, Burton J, Cox TV, Davies R, Down TA, Haefliger C, Horton R, Howe K, Jackson DK, Kunde J, Koenig C, Liddle J, Niblett D, Otto T, Pettett R, Seemann S, Thompson C, West T, Rogers J, Olek A, Berlin K, and Beck S. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet* 38: 1378–1385, 2006.

19. Ehrlich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, Cantor CR, Field JK, and van den Boom D. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. *Proc Natl Acad Sci U S A* 102: 15785–15790, 2005.
20. Estecio MR, Yan PS, Ibrahim AE, Tellez CS, Shen L, Huang TH, and Issa JP. High-throughput methylation profiling by MCA coupled to CpG island microarray. *Genome Res* 17: 1529–1536, 2007.
21. Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 8: 286–298, 2007.
22. Feil R and Fraga MF. Epigenetics and the environment: emerging patterns and implications. *Nat Rev Genet* 13: 97–109, 2012.
23. Frigola J, Ribas M, Risques RA, and Peinado MA. Methylation profiling of cancer cells by amplification of inter-methylated sites (AIMS). *Nucleic Acids Res* 30: e28, 2002.
24. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, and Paul CL. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* 89: 1827–1831, 1992.
25. Herceg Z and Vaissiere T. Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics* 6: 804–819, 2011.
26. Herman JG, Graff JR, Myohanen S, Nelkin BD, and Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* 93: 9821–9826, 1996.
27. Hernandez-Vargas H, Lambert M-P, Le Calvez-Kelm F, Gouysse Gr, McKay-Chopin S, Tavtigian SV, Scazecz J-Y, and Herceg Z. Hepatocellular carcinoma displays distinct dna methylation signatures with potential as clinical predictors. *PLoS One* 5: e9749, 2010.
28. Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, Zhang MQ, Ye K, Bhattacharjee A, Brizuela L, McCombie WR, Wigler M, Hannon GJ, and Hicks JB. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res* 19: 1593–1605, 2009.
29. Huang TH, Perry MR, and Laux DE. Methylation profiling of CpG islands in human breast cancer cells. *Hum Mol Genet* 8: 459–470, 1999.
30. Irizarry RA, Ladd-Acosta C, Carvalho B, Wu H, Brandenburg SA, Jeddellah JA, Wen B, and Feinberg AP. Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res* 18: 780–790, 2008.
31. Jorgensen HF, Adie K, Chaubert P, and Bird AP. Engineering a high-affinity methyl-CpG-binding protein. *Nucleic Acids Res* 34: e96, 2006.
32. Karimi M, Johansson S, and Ekstrom TJ. Using LUMA: a Luminometric-based assay for global DNA-methylation. *Epigenetics* 1: 45–48, 2006.
33. Khulan B, Thompson RF, Ye K, Fazzari MJ, Suzuki M, Stasiak E, Figueroa ME, Glass JL, Chen Q, Montagna C, Hatchwell E, Selzer RR, Richmond TA, Green RD, Melnick A, and Grealley JM. Comparative isoschizomer profiling of cytosine methylation: the HELP assay. *Genome Res* 16: 1046–1055, 2006.
34. Kristensen LS and Hansen LL. PCR-based methods for detecting single-locus DNA methylation biomarkers in cancer diagnostics, prognostics, and response to treatment. *Clin Chem* 55: 1471–1483, 2009.
35. Laird PW. Principles and challenges of genomewide DNA methylation analysis. *Nat Rev Genet* 11: 191–203, 2010.
36. Liang G, Gonzalgo ML, Salem C, and Jones PA. Identification of DNA methylation differences during tumorigenesis by methylation-sensitive arbitrarily primed polymerase chain reaction. *Methods* 27: 150–155, 2002.
37. Lima SC, Hernandez-Vargas H, and Herceg Z. Epigenetic signatures in cancer: Implications for the control of cancer in the clinic. *Curr Opin Mol Ther* 12: 316–324, 2010.
38. Lister R and Ecker JR. Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res* 19: 959–966, 2009.
39. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, and Ecker JR. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133: 523–536, 2008.
40. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, and Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462: 315–322, 2009.
41. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJM, Haussler D, Marra MA, Hirst M, Wang T, and Costello JF. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466: 253–257, 2010.
42. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, and Lander ES. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454: 766–770, 2008.
43. Nair SS, Coolen MW, Stirzaker C, Song JZ, Statham AL, Strbenac D, Robinson MD, and Clark SJ. Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. *Epigenetics* 6: 34–44, 2011.
44. Nouzova M, Holtan N, Oshiro MM, Isett RB, Munoz-Rodriguez JL, List AF, Narro ML, Miller SJ, Merchant NC, and Futscher BW. 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.
45. Oda M, Glass JL, Thompson RF, Mo Y, Olivier EN, Figueroa ME, Selzer RR, Richmond TA, Zhang X, Dannenberg L, Green RD, Melnick A, Hatchwell E, Bouhassira EE, Verma A, Suzuki M, and Grealley JM. 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.
46. Park JH, Park J, Choi JK, Lyu J, Bae MG, Lee YG, Bae JB, Park DY, Yang HK, Kim TY, and Kim YJ. Identification of DNA methylation changes associated with human gastric cancer. *BMC Med Genomics* 4: 82, 2011.
47. Rauch TA, Wu X, Zhong X, Riggs AD, and Pfeifer GP. A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci U S A* 106: 671–678, 2009.
48. Rodriguez-Paredes M and Esteller M. Cancer epigenetics reaches mainstream oncology. *Nat Med* 17: 330–339, 2011.
49. Serre D, Lee BH, and Ting AH. MBD-isolated Genome Sequencing provides a high-throughput and comprehensive

- survey of DNA methylation in the human genome. *Nucleic Acids Res* 38: 391–399, 2010.
50. Taiwo O, Wilson GA, Morris T, Seisenberger S, Reik W, Pearce D, Beck S, and Butcher LM. Methylome analysis using MeDIP-seq with low DNA concentrations. *Nat Protoc* 7: 617–636, 2012.
  51. Taylor KH, Kramer RS, Davis JW, Guo J, Duff DJ, Xu D, Caldwell CW, and Shi H. Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. *Cancer Res* 67: 8511–8518, 2007.
  52. Tost J, El abdalaoui H, and Gut IG. Serial pyrosequencing for quantitative DNA methylation analysis. *Biotechniques* 40: 721–722, 724, 726, 2006.
  53. Toyota M, Ho C, Ahuja N, Jair KW, Li Q, Ohe-Toyota M, Baylin SB, and Issa JP. Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res* 59: 2307–2312, 1999.
  54. Uhlmann K, Brinckmann A, Toliat MR, Ritter H, and Nurnberg P. Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* 23: 4072–4079, 2002.
  55. Ushijima T, Morimura K, Hosoya Y, Okonogi H, Tatematsu M, Sugimura T, and Nagao M. Establishment of methylation-sensitive-representational difference analysis and isolation of hypo- and hypermethylated genomic fragments in mouse liver tumors. *Proc Natl Acad Sci U S A* 94: 2284–2289, 1997.
  56. Vaissiere T, Cuenin C, Paliwal A, Vineis P, Hoek G, Krzyzanowski M, Airolidi L, Dunning A, Garte S, Hainaut P, Malaveille C, Overvad K, Clavel-Chapelon F, Linseisen J, Boeing H, Trichopoulou A, Trichopoulos D, Kaladidi A, Palli D, Krogh V, Tumino R, Panico S, Bueno-De-Mesquita HB, Peeters PH, Kumle M, Gonzalez CA, Martinez C, Dorronsoro M, Barricarte A, Navarro C, Quiros JR, Berglund G, Janzon L, Jarvholm B, Day NE, Key TJ, Saracci R, Kaaks R, Riboli E, and Herceg Z. Quantitative analysis of DNA methylation after whole bisulfite amplification of a minute amount of DNA from body fluids. *Epigenetics* 4: 221–230, 2009.
  57. Warnecke PM, Stirzaker C, Melki JR, Millar DS, Paul CL, and Clark SJ. Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic Acids Res* 25: 4422–4426, 1997.
  58. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, and Schubeler D. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37: 853–862, 2005.
  59. Yan PS, Chen CM, Shi H, Rahmatpanah F, Wei SH, and Huang TH. Applications of CpG island microarrays for high-throughput analysis of DNA methylation. *J Nutr* 132: 2430S–2434S, 2002.
  60. Yang AS, Estecio MR, Doshi K, Kondo Y, Tajara EH, and Issa JP. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res* 32: e38, 2004.
  61. Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, Min JH, Jin P, Ren B, and He C. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149: 1368–1380, 2012.

Address correspondence to:

Dr. Zdenko Herceg

Epigenetics Group

International Agency for Research on Cancer (IARC)

150 Cours A. Thomas

Lyon 69008

France

E-mail: hercegz@iarc.fr

Date of first submission to ARS Central, September 3, 2012; date of final revised submission, October 23, 2012; date of acceptance, November 4, 2012.

Abbreviations Used

5hmC = 5-hydroxymethylcytosine
 5mC = 5-methylcytosine
 AIMS = amplification of intermethylated sites
 BS = bisulfite
 CGI = CpG island
 CHARM = comprehensive high-throughput arrays for relative methylation
 COBRA = combined bisulfite conversion restriction analysis
 DMH = differential methylation hybridization
 DNMT = DNA methyltransferase
 HELP = *HpaII* tiny-fragment enrichment by ligation-mediated PCR
 HPCE = high-performance capillary electrophoresis
 HPLC = high-precision liquid chromatography
 LUMA = luminometric methylation assay
 MALDI-TOF = matrix-assisted laser desorption ionization–time of flight
 MBD = methyl-binding domain
 MCA = methyl CpG island amplification
 MeDIP = methylated DNA immunoprecipitation
 MiGS = MBD-isolated genome sequencing
 MIRA = methylated CpG island recovery assay
 MS-AP-PCR = methylation-specific arbitrarily primed PCR
 MS-HRM = methylation-sensitive high-resolution melting
 MS-MCA = methylation-sensitive melting curve analysis
 MSP = methylation-specific PCR
 MS-RDA = methylation-sensitive representational difference analysis
 MSREs = methylation sensitive restriction enzymes
 MS-SnuPE = methylation-sensitive single-nucleotide primer extension
 RDA = representational difference analysis
 REs = restriction endonucleases
 RLGS = restriction landmark genomic scanning
 RRBS = reduced representation bisulfite sequencing
 UTR = untranslated region