

Open Access Article

The authors, the publisher, and the right holders grant the right to use, reproduce, and disseminate the work in digital form to all users.



Technology in Cancer Research and Treatment

ISSN 1533-0346

Volume 9, Number 2, April 2010

©Adenine Press (2010)

Profiling DNA Methylomes from Microarray to Genome-Scale Sequencing

www.tcrt.org

DNA cytosine methylation is a central epigenetic modification which plays critical roles in cellular processes including genome regulation, development and disease. Here, we review current and emerging microarray and next-generation sequencing based technologies that enhance our knowledge of DNA methylation profiling. Each methodology has limitations and their unique applications, and combinations of several modalities may help build the entire methylome. With advances on next-generation sequencing technologies, it is now possible to globally map the DNA cytosine methylation at single-base resolution, providing new insights into the regulation and dynamics of DNA methylation in genomes.

Key words: DNA methylation; Microarray; Next-generation sequencing.

Introduction

Tumorigenesis is caused by not only genetic but also epigenetic changes. Genetic modification changes the DNA sequences and results in alternation of the structure and function of a gene's product. Epigenetic modification can be categorized into two types: modifications at the DNA level or the protein (histone) level. Changes on histone tails include but not limit to methylation, acetylation and phosphorylation. DNA methylation at the carbon-5 position of cytosine pyrimidine ring is catalyzed by DNA methyltransferases and can be heritable without changing DNA sequence. DNA methylation is the most stable of all epigenetic modifications and was first discovered in 1948 (1). Subsequently, CpG islands (CGIs), corresponding to genomic regions with a high frequency of CpG sites (2) were uncovered. The "p" in CpG refers to the phosphodiester bond between the cytosine and the guanine. DNA methylation is associated with transcriptional silencing of imprinted genes in normal cells and with inactivation of female X-chromosomes (3, 4). Furthermore, hypermethylation of CpG islands in the promoter region of tumor suppressor genes is linked to cancer development (5, 6).

Over the past decade, many approaches have been developed for DNA methylation analysis. Typically, these analyses can be divided into two categories, typing and profiling technologies. Typing technologies are used when few loci are assayed

Yi-Wen Huang, Ph.D.*
Tim H.-M. Huang, Ph.D.
Li-Shu Wang, Ph.D.*

The Ohio State University
Comprehensive Cancer Center,
Columbus, OH 43210, USA

*Corresponding Author:
Email: Yi-wen.huang@osumc.edu
Li-shu.wang@osumc.edu

Abbreviations: CGIs: CpG islands; CHARM: Comprehensive high-throughput arrays for relative methylation; DMH: Differential methylation hybridization; HELP: *HpaII* tiny fragment enrichment by ligation-mediated PCR; MBD: Methyl-CpG binding domain; MCAM: Methylated CpG island amplification microarray; 5-mC: 5-methyl cytosine; MeDIP: Methylated DNA immunoprecipitation; MIRA: Methylated-CpG island recovery assay; MMAS: Microarray-based methylation assessment of single samples; MSCC: Methyl-sensitive cut counting; MSO: Methylation-specific oligonucleotides; PMAD: Promoter-associated methylated DNA amplification DNA chip; RRBS: Reduced representation bisulfite sequencing; NGS: Next-generation sequencing.

in multiple samples. Currently, a variety of technologies are used for typing applications, including polymerase chain reaction (PCR)-, restriction- and mass spectrometry-based methodologies (7). Readers are referred to recent review articles (7-9) as this type of assay will not be described here. This review will focus on the current and emerging technologies on DNA methylation profiling. In particular, we will discuss the strength and weaknesses associated with microarray and next generation sequencing for profiling DNA methylation (Table I).

Microarray-Based DNA Methylation Profiling

Microarray-based DNA methylation profiling is a relatively new field. Its three major techniques are endonuclease restriction-, bisulfite conversion- and affinity-based microarray analyses (Table I).

Endonuclease Restriction-Based Microarray Analysis

Differential methylation hybridization (DMH) (10) was the first genome wide DNA methylation profiling assay. It is used to identify hypermethylated loci in normal tissues, in cancer samples or in cancer cell lines by using CpG island microarrays. The basis of DMH is the application of site-specificity and methylation-dependent restriction enzymes. Genomic DNA is fragmented by restriction enzymes such as *MseI*, a 4-base cutter, or by sonication into smaller fragments. The resultant materials are ligated to synthetic linkers and their methylation status is interrogated by methylation-sensitive endonucleases, *BstUI* and/or *HpaII*. *BstUI* is the enzyme of choice as most CpG islands contain *BstUI* cut sites. The methylated fragments, which are protected from restriction, will be amplified by linker-dependent PCR. In contrast, the unmethylated fragments are digested by the

Table I
Summary of microarray and next-generation sequencing assays in profiling DNA methylome.

	Restriction enzyme	Bisulfite modification	DNA precipitation	Advantages	Limitations	Ref.
Microarray-based methylation profiling						
Endonuclease restriction-based microarray analysis						
CHARM	<i>McrBC</i>	-	-	Identify hyper- or hypomethylated CpG sites	Moderate resolution; limit to restriction enzymes digested sites	22, 23
DMH	<i>BstUI/HpaII</i>	-	-			10, 11
HELP	<i>HpaII/MspI</i>	-	-			16, 17
MCAM	<i>SmaI/XmaI</i>	-	-			12
McrBC	<i>McrBC</i>	-	-			19, 20
MMASS	<i>McrBC</i> /cocktail	-	-			21
PMAD	<i>HpaII/MspI</i>	-	-			15
Bisulfite conversion-based microarray analysis						
MSO	-	Yes	-	Cost efficiency	Moderate resolution	24
Bead array	-	Yes	-			28, 29
Affinity-based microarray analysis						
MeDIP	-	-	Yes	Efficient assays to CGIs and repetitive sequences	Less sensitive to CpG poor sites Depending on MBD binding ability	30
MIRA	-	-	Yes			32-34
Sequencing-based methylation profiling						
BS-seq	-	Yes	-	High resolution	Higher cost at this stage	39
MethylC-Seq	-	Yes	-			44, 45
RRBS	<i>MspI</i>	Yes	-	Each sequencing-read contains at least one CpG site	Limit to restriction enzymes digested sites	41
HELP-seq	<i>HpaII/MspI</i>	-	-			17
Padlock probe	Could be	Yes	-	Quantify DNA methylation in a smaller genome fraction with high resolution at this stage	Limit to probes designing	42, 43
MeDIP-seq	-	-	-	Moderate resolution	Less sensitive to CpG poor sites	31

endonucleases and will not be amplified. The PCR amplicons from normal and tumor specimens are labeled with fluorescent dyes, pooled, and co-hybridized to a microarray chip spotted with CpG islands probes. After hybridization, chips are washed and scanned with a laser beam. The intensities of fluorescent dyes reflect the methylation status within each locus in a tumor relative to its normal counterpart. This assay was used to identify methylated genes in breast, ovarian and colon cancers (11). The limitation of this assay is a result of the reliance on *Bst*UI/*Hpa*II to interrogate methylation status, which restricts the analysis of a fraction of CpG sites in the genomes.

In 1999, methylated CpG island amplification (MCA) (12) was developed with novel restriction enzymes and PCR-based assay for DNA methylation analysis. Later, in 2007, MCA was combined with CpG island microarray and named MCAM (13), for methylated CpG island amplification microarray. MCAM was reported as a suitable method to uncover methylated genes and to profile methylation changes in clinical samples in a high-throughput fashion (13). In principle, genomic DNA is first digested with *Sma*I (a methylation sensitive endonuclease, CCC/GGG) to generate blunt end fragments and to eliminate unmethylated sites. Next, the cleaved DNA is further digested with *Xma*I (a methylation impaired endonuclease, C/CCGGG) to create sticky ends and leaves CCGG overhangs in methylated sites. Methylated fragments are ligated with adaptors and then amplified by linker ligation-mediated PCR. The amplicons from a tumor and a normal control are labeled with different fluorescent dyes and co-hybridized to a CpG island microarray chip. This technique simultaneously reduces complexity and increases specificity by targeting methylated CpG islands before amplification (13).

In 2002, Hatada and colleagues adapted the MCA method and developed a similar assay called methylation amplification DNA chip (MAD) (14). MAD also uses *Sma*I and *Xma*I endonucleases because about 70–80% of CpG islands contains at least two closely spaced *Sma*I sites (<1 kb). Later, in 2006, MAD was modified to become the promoter-associated methylated DNA amplification DNA chip (PMAD) assay (15) which incorporated *Hpa*II and *Msp*I endonucleases as most of CpG islands contain their recognition sequence, “CCGG”. *Msp*I is the methylation insensitive isoschizomer of *Hpa*II. For each sample, methylated *Hpa*II-resistant DNA fragments and *Msp*I-cleaved (including both unmethylated and methylated) DNA fragments are amplified and labeled with Cy3 and Cy5, respectively. Then both fragments are co-hybridized to a microarray containing the promoters of cancer-related genes. Signals from *Hpa*II-resistant (methylated) DNA (Cy3) are normalized to signals from *Msp*I-cleaved (unmethylated and methylated) DNA fragments (Cy5). Normalized signals from tumors are compared to signals from

normal control. This assay eliminates the potential for false discoveries by applying the methylation insensitive isoschizomer of *Hpa*II (*i.e.*, *Msp*I) in a control reaction. It was successfully used in the identification of genes methylated in lung cancer (15).

The *Hpa*II tiny fragment enrichment by ligation-mediated PCR (HELP) assay, reported in 2006, applied a similar restriction fractionation approach (*Hpa*II or *Msp*I) as PMAD and used *Msp*I digested fragments as an internal control (16). The HELP assay was reported as a robust and quantitative measurement of cytosine methylation in the mouse genome, and it identified novel tissue-specific differentially methylated regions. More recently, in 2009, HELP (17) underwent additional modifications to provide a new representation of the genome with a dual adapter approach and to expand to smaller fragments (50–200 bp) compared with the larger amplicons (200–2,000 bp) in the original protocol. The digested DNA fragments are labeled with fluorescent dyes and co-hybridized to oligonucleotide microarray. The modified protocol significantly enhances the resolution of this assay. It was able to identify more than 1.32 million loci in the human genome and represented 98.5% of CpG islands and 91.1% of refSeq promoters (17), especially in the most CG-rich regions. HELP has also been applied to massively parallel sequencing applications and clinical specimens with limited amount of DNA amounts (17).

A similar approach to HELP-seq, termed Methyl-seq, was adapted to investigate DNA methylation alternations from specimens representing various stages during human development (18). Although this assay requires *Hpa*II sites and sequencing reads which are biased toward CpG islands, more than half of the regions identified by this method were outside annotated reference genes. This assay has effectively identified 65% of all annotated CpG islands and the DNA methylation status of more than 90,000 regions in the human genome. The method was reported as a sensitive, highly specific with very low background, reproducible, simple to execute assay, and relatively inexpensive when sequenced on NGS instruments (18). This assay, however, is also plagued with limitations. For instance, DNA fragments contain a specific subset of *Hpa*II cleavage sites (those that appear within 35–75 bp of each other in the human genome), which limit the accessible number of CpG islands. Methyl-seq in current status only defines sequencing reads to binary calls of “methylated” and “unmethylated, which might be a problem in measuring methylation quantitatively in certain regions. Because, an “unmethylated” region is defined within an *Hpa*II fragment in the next-generation sequencing and reads as methylation lost, this assay limits the ability of Methyl-seq to detect the quantitative methylation state of any individual CpG site(s) which might be partial methylation or imprinted loci. Finally, Methyl-seq is not effective in detecting regions with low CpG

density which might include only one single *HpaII* cut site, as it relies mainly on small *HpaII* fragments with multiple *HpaII* cutting sites. Despite all the limitations, Methyl-seq is still a valuable assay for detecting the methylation status of CpG dinucleotides in the human genome as well as any other vertebrate genome (18).

As opposed to methylation sensitive endonucleases in DMH and MCA assays, which digest unmethylated loci and generate an enrichment of methylated DNA fragments, *McrBC* (19, 20) was selected because it prefers methylated DNA as a substrate which results in selecting the unmethylated fractions of DNA from a genotype of interest. *McrBC*, an endonuclease, recognizes 2 closely spaced methylated cytosines (55 bp to 3 kb) in the context of (G/A) ^mC (^m means methylated). It has a broad capacity to digest densely methylated regions of DNA, such as abnormally methylated CpG islands in cancer cells, and methylated repetitive sequence elements in most cells and tissues. Briefly, the size of genomic DNA is reduced by nebulization or *MseI* digestion. The restricted samples are ligated with linker primers and then divided into two fractions, and one treated with and the other without *McrBC* digestion. The *McrBC* digested DNA and an untreated sample are size-fractionated, differentially labeled, and co-hybridized to a CpG island microarray. The ratio of their hybridization intensities thus provides a measure of DNA methylation. *McrBC* is preferred to other methylation-sensitive enzymes such as *HpaII*, because it does not require a highly specific sequence motif and therefore cuts more frequently. This methodology not only assays both hypermethylated and hypomethylated loci but also extracts other genomic information, such as copy number differences. One other advantage of this assay is that it does not require prior methylation information from a reference genome to serve as a control and therefore can be applied widely to clinical specimens without a matched control (19, 20).

In 2006, microarray-based methylation assessment of single samples (MMASS) (21) was developed to directly compare methylated and unmethylated sequences within a sample, and to improve the detection sensitivity of differential methylation within a single hybridization. Briefly, genomic DNA is first digested with *MseI*, followed by adaptor ligation and then divided into two equal fractions. One fraction is treated with *McrBC* to enrich unmethylated sequences. The other fraction is optimized with a combination of methylation-sensitive enzymes (*AciI*, *HinPII*, *HpyCH4IV* and *HpaII*) to enrich methylated sequences. The restricted materials from both fractions are amplified with linker-mediated PCR, which corresponded to unmethylated and methylated regions, respectively. The amplicons are labeled, mixed and hybridized to a CpG island array. MMASS utilizes bioinformatic tools to provide detailed

annotation of all probes on a publicly available CpG island array and uses this information to develop and validate as a high-throughput method (21).

In comparison to MMASS, CHARM (comprehensive high-throughput arrays for relative methylation) (22), developed in 2008, provided an improvement on statistical procedures and array design algorithm beyond that of *McrBC*-based assays. CHARM also uses *McrBC* to digest genomic DNA and detects hypermethylated CpG sites in CpG island core and “shore” regions (stretches of ~2 kb bordering CpG islands). The identification of hypermethylated CpG island shores is important as they are associated with gene expression silencing (23).

Bisulfite conversion-based microarray analysis

The gold standard for investigating DNA methylation is bisulfite sequencing. Bisulfite treatment converts unmethylated cytosines from CpG sites to uracils, while methylated cytosines are protected from conversion. After PCR amplification, uracils are read as thymidines, but methylated cytosines remain unchanged. This process alters an epigenetic difference into a quantifiable genetic difference. In a methylation-specific oligonucleotides microarray experiment, PCR amplicons generated after bisulfite conversion of genomic DNA function as probes to hybridize targets corresponding to methylated and unmethylated regions in genes of interest (24). The quantitative differences in hybridization, which are assessed by the fluorescent intensity, indicate the methylation status of a particular locus. This approach has been applied to various human cancers (24-27), such as breast, non-Hodgkin's lymphoma and colorectal.

The first commercial bead array-based platform was Illumina's GoldenGate assay and is still widely used. The GoldenGate Methylation Cancer Panel I covers 1,505 CpG sites selected from 807 genes (28). In brief, bisulfite-treated genomic DNA is immobilized on beads. A pooled query oligonucleotides containing two allele-specific (ASO) and two locus-specific oligonucleotides (LSO) are annealed to the genomic DNA under a controlled hybridization program, and then washed to remove excess or mis-hybridized oligonucleotides. Hybridized oligonucleotides are then extended and ligated to generate amplifiable templates. A PCR reaction is performed with fluorescently labeled universal PCR primers. The extent of methylation at a given CpG site is determined by comparing the proportion of signal from methylated and unmethylated alleles in the DNA sample. This assay was recently applied to compare the matched formalin-fixed, paraffin-embedded and frozen surgical specimens (29). Results from this study showed the complete preservation of the cancer methylome among differently archived lymph nodes tissues. More recently,

Illumina has developed Infinium, a system that employs the same principle but with a higher resolution BeadChip than GoldenGate. HumanMethylation27 is an assay of Infinium used to profile DNA methylation status at 27,578 CpG sites (spanning more than 14,000 genes) per sample at single-nucleotide resolution while analyzing 12 samples on a single chip. This high throughput technology features low sample input and low cost.

Affinity-Based Microarray Analysis

Due to the limitations of enzyme recognition site(s) within CpG sites and the reduction of sequence complexity following bisulfite conversion, a third technology has been developed to enrich the methylated (or unmethylated) fraction of the genome, *i.e.*, methylated DNA immunoprecipitation (MeDIP) or affinity chromatography over an MBD (methyl-binding domain). MeDIP-chip (30) was developed based on immunoprecipitating the methylated fraction of a genomic DNA with a monoclonal antibody against methylated cytosine and then hybridizing the immunoprecipitated fraction against the input or total fraction on a microarray. As technologies evolve, the next generation sequencing becomes preferred platform to analyze the methylated fraction from MeDIP, known as MeDIP-seq (31), and will be discussed later in this review. Batman, a new algorithm, was developed to measure the quantitative DNA methylation of MeDIP-seq. However, this assay requires the antibody to recognize single strand DNA, which is sometimes difficult to achieve in CpG-poor DNA regions.

An alternative approach to obtaining enriched methylated DNA is to exploit the high binding affinity of MBD (methyl-binding domain). Unlike MeDIP which relies on a specific monoclonal antibody against 5-mC (5-methyl cytosine), methylated CpG island recovery assay (MIRA) (32-34) utilizes the specific binding capacity of the MBD2/MBD3L1 complex to double-strand methylated DNA sequences which is not sequence dependent except for methylated CpGs (33). MIRA was demonstrated as a specific and sensitive assay, sufficient to detect as low as one methylated CpG site in an *in vitro* analysis (32). MIRA-enriched fractions could be coupled with a whole-genome tiling microarray providing a 100-bp resolution of methylation profiling (34). It is worth noting that the MIRA approach has now been adapted and commercialized by several companies including Life Technologies (Invitrogen). Despite the development of all the approaches described in this review, no single method can profile the methylome in great detail. A combination of several methodologies is needed to comprehensively interrogate DNA methylation by microarray. The current rapid development in the NGS will overtake array-based methylome analysis.

Microarray Platforms

Microarray platforms can be grouped based upon characteristics such as the nature of the probe, the solid-surface support used, and the specific method for probe addressing and/or target detection (35). The microarray signal intensity relies on probes. In general, shorter probe lengths reduce the errors introduced during probe synthesis and often have greater specificity. While longer probes have higher melting temperatures, greater mismatch tolerance, and increased sensitivity, they also result in decreased specificity. A longer probe length and greater probe number per target region could reduce random signal variation, referred to as noise (35-36). A detailed discussion of the complexities of probe design and probe-specific signal interpretation can be found in a recent review (36). There are three types of microarrays used to profile DNA methylation: printed and *in situ* synthesized oligonucleotide microarrays and high-density bead arrays.

Printed microarrays are relatively simple, inexpensive and flexible as it is easy to modify the probes on the array to reflect the latest annotation, novel targets and spliced variations (35). The medium on which the probes are arrayed is glass slides. It is economical, nonporous, stable to various ranges of hybridization temperatures and wash stringing. Glass also produces minimal background fluorescence and allows for efficient kinetics during hybridization. The printed microarrays are usually lower in density (~10,000 to 30,000 features) when compared to other platforms (35).

Most commercial microarrays, including Affymetrix, Agilent and Roche NimbleGen, (Table II) are *in situ* synthesized oligonucleotide microarrays. This format is not conducive to user-defined development due to the complex nature of chemical synthesis and the expense involved in production (35). The major advantages of this manufacturing system include reproducibility of production probes, and the standardization of reagents, instrumentation, and data analysis. Other advantages include controls, such as reference probes for intensity normalization, internal standards of known concentrations, and probes arranged in a checkerboard pattern that are homologous to an internal control provided in the hybridization mix (35).

Affymetrix GeneChips are the most widely used platform and its probes are short (20-25 bp), synthesized using semiconductor-based photochemical synthesis. Their advantages include multiple probes per target to improve sensitivity and a probe set to increase specificity. The probe set includes one perfect-match probe and one mismatch probe which contains a 1-bp difference in the middle position of the probe. In contrast to Affymetrix microarrays which are limited to one color label for hybridizing, both Agilent and Roche NimbleGen platforms are used with multicolor labeling. The Agilent system

Table II
Summary of current microarray platforms to profile DNA methylation.

		Probe Length (bp)	Probe Spacing (bp)	Density	Total probe number	Array number
Affymetrix	Promoter	25	35	4.60×10^6	4.6 million	1
	Tiling	25	35	3.25×10^6	Up to 45 million	Up to 14
Agilent	Promoter	60	100	244×10^3	237,220	1
	Tiling	60	100	244×10^3	14.5 million	Up to 6
Roche NimbleGen	Promoter	50-75	100	385×10^3 or 2.1×10^6	Various	Up to 2
	Tiling	50-75	100	385×10^3 or 2.1×10^6	Various	Up to 10

uses longer oligonucleotide probes (60 bp) and employs five-ink (4 bases plus catalyst) inkjet technology for probe production. Each Agilent microarray can contain up to 244,000 probe features in format. The Roche NimbleGen is the third *in situ* synthesized oligonucleotide microarray. NimbleGen arrays also contain long probes (50 to 100 bp) and use maskless photo-mediated synthesis chemistry to generate probes. Maskless array synthesizer technology uses programmable micromirrors to create digital masks that reflect the desired pattern of DNA sequences onto arrays (35). Each NimbleGen microarray can contain up to 2.1 million features per slide.

The third platform, BeadArrays (Illumina), is based on 3- μ m silica beads that randomly self assemble in microwells onto either fiber optic bundles or planar silica slide substrates. Each bead is covered with hundreds of thousands of copies of a specific oligonucleotide that captures specific sequences. BeadArrays can support up to 10^5 to 10^6 features and have built-in redundancy. This redundancy is crucial as an experimental control to compare data among arrays since each manufactured microarray may not be identical. Another unique advantage is that the bead pattern alternations provide a reflector to identify spatial bias (35).

Sequencing-Based DNA Methylation Profiling

The biggest advantage of next-generation sequencing (NGS) technologies is that a single run can provide higher coverage at relatively low cost when compared with the above described assays for genome-wide DNA methylation. NGS technology is in an exponential development stage and detailed discussion can be found in recent publications (37, 38) as only the platforms and characteristics will be discussed in this review.

Today, sequencing of every CpG in the genome with high accuracy is already technically achievable by using whole-genome shotgun bisulfite sequencing (named BS-Seq) (39) and MethyIC-Seq (43, 44). Although it is still prohibitively expensive, the rate of cost reduction should continue on its exponential trajectory for the near future. BS-Seq combines bisulfite treatment of genomic DNA with Solexa sequencing

technology and measures cytosine methylation within specific sequence contexts. BS-Seq provides much more accurate promoter methylation results in comparison with that of microarray approaches. It can also be used to detect repetitive sequences that are difficult to study using microarrays because of the difficulty in designing probes for these regions (39).

Taking advantage of the next generation sequencing technology in time-effort and saving cost, targeted loci can be selected for high-throughput parallel bisulfite sequencing at a single cytosine resolution. Taylor *et al.*, (40) directly sequenced 122 bisulfite PCR products from human lymphocytes, lymphomas and leukemia, spanning 25 genes in a single sequencing run without subcloning. This approach was able to detect more than a thousand individual sequences for each PCR amplicons. This is far more than the number of clones that can be analyzed by traditional bisulfite sequencing cloning (10~20 clones). Another approach to study selected loci is the RRBS method (reduced representation bisulfite sequencing) (41) in which 90% of CpG islands within the mouse genome were include in *MspI* digested fractions. DNA fragments are then subjected to bisulfite conversion and high-throughput sequencing. The advantage of RRBS is the assurance that each sequencing read contains at least one CpG as sequencing fragments are bracketed by *MspI* sites. This method covers nearly 1 million distinct CpG dinucleotides utilizing short 36 bp reads generated by the Illumina genome analyzer platform.

More recently, padlock (molecular inversion) probes were developed to capture targeted-sequences in bisulfite-converted DNA (42, 43). Thousands of padlock probes are generated and synthesized by computational programs based on targeted DNA sequences. The targeted CpG sites within bisulfite converted DNA are in the middle of circular padlock probe. The CpG methylation levels are captured in a single ligation-amplification reaction which amplifies many multiplexed padlock probes in a single tube. The captured targeted CpG sites in padlock loops are then subjected to Illumina sequencing. Padlock probes can achieve 90-99% targeted-specificity after pooling ~10,000 probes in selected chromosomes (42, 43). Ball *et al.*, (43) also applied padlock

probes using a second method, called methyl-sensitive cut counting (MSCC) using the methylation sensitive enzyme, *HpaII*. MSCC profiles all genomic locations within *HpaII* cut sites that totals to ~1.4 million unique sites in the human genome. Using this strategy, the authors found a positive correlation between hypermethylation in gene bodies and transcription activation, consistent with biological observation (43).

Another relatively high resolution strategy is MeDIP-seq (31) was used to study the whole-genome DNA methylation profile (DNA methylome) in the mammalian genome. MeDIP depends on the efficiency of immunoprecipitation as well as the density and configuration of methylated CpG sites as described in previous section. MeDIP fragments are subjected to Illumina Genome Analyzer sequencer. Down and colleagues (31) developed a cross-platform algorithm, Batman, to estimate absolute methylation level utilizing MeDIP-chip or MeDIP-seq. They found that MeDIP-seq provides coverage of 90% of all CpG sites within CpG islands, promoters and other regulatory sequences, exons and introns, and 60% coverage of all CpGs in the human genome.

A recent assay, MethylC-Seq, is a bisulfite-based and single-base-resolution technology and has been used to map methylated cytosines in genomes of *Arabidopsis* (44) and of mammals (45). This approach starts with fragmented genomic DNA treated with sodium bisulfite to convert cytosine, but not methyl-cytosine, to uracil, and is then sequenced by Illumina Genome Analyzer. In the human genome (45), MethylC-Seq interrogated at least one sequence read for over 86% of both strands of the 3.08 Gb human reference sequence which accounted for 94% of the cytosines. This method provides, for the first time, the information that methylation in non-CG contexts is enriched in gene bodies and depleted in protein binding sites and enhancers, which is usually overlooked by other methodologies. It also produces a novel observation that patterns of abundant methylation in the CG and non-CG sites of gene bodies might have different correlations with gene expression.

Next-Generation Sequencing Technologies

Next-generation sequencing (NGS) is referred to as a new technology arising after the automate Sanger method. These

technologies can generate hundreds of millions of sequences of short DNA fragments in a single run. Different experiments can be performed by varying strategies including template preparation, sequencing, imaging, genome alignment and assembly steps. In the following, we will briefly discuss the three main next-generation sequencing platforms and their characteristics are summarized in Table III. A detailed discussion including the chemistry of synthesizing sequences for the platforms can be found in a recent review (46). We will also mention the strengths, weaknesses and challenges of NGS technologies.

Roche/454 Genome Sequencer FLX system was the first NGS platform on the market and is based on pyrophosphate detection. DNA templates are ligated with specific adapters which immobilize one DNA fragment onto one bead. These fragments are amplified by emulsion PCR which consists of water droplets containing one bead and PCR reagents immersed in oil. The amplification is necessary to obtain sufficient light signal intensity for reliable detection in the sequencing-by-synthesis reaction steps. The sequencing signals are collected through the fluorescence generated from the luciferin substrate during sequencing reaction. The system can generate more than one million individual reads per run at lengths up to 500 bases. With continuous new chemistry developments, read lengths are expected to extend to 1000 bases in 2010 according to Roche website.

The Illumina/Solexa GAIIX platform dominates the NGS current market. The key features of this platform are cyclical reversible termination for the four nucleotides each labeled with a fluorescent dye, and a mutant DNA polymerase that incorporates a single nucleotide at a time. DNA fragments ligated with adapters at both ends are immobilized on a solid support and create a 'bridge' structure by hybridizing with its free end to the complementary adapter on the surface of the support. PCR amplification is performed using the adapters on the surface that act as primers to obtain sufficient light signal intensity. The PCR products, named clusters, are created and subjected to real-time sequencing. The sequencing reagents contain primers, four reversible terminator nucleotides and the DNA polymerase. After incorporation into the DNA strand, the terminator nucleotide is detected and identified via its fluorescent dye by the CCD camera. The terminator group at the-end of the base and the fluorescent

Table III
Summary of next-generation sequencing technologies.

	Template preparation	Synthesis method	Read length (bases)	Run time	Gb per run
Roche/454 GS FLX Titanium	Emulsion PCR	Pyrosequencing	Average 400	10 hr	0.4
Illumina/Solexa Genome Analyzer IIX	Bridge PCR	Reversible terminators	Up to 100	4*, 9.5** days	17*, 33**
Life/SOLiD 3	Emulsion PCR	Ligation	Up to 50	7*, 14** days	30*, 60**

* Single-end read sequencing; ** Paired-end read Sequencing.

dye are then removed from the base and the synthesis cycle is repeated. Currently, the sequence read length is about 100 nucleotides. However, longer reads will reduce accuracy because of signal decay and de-phasing.

The Life Technologies/Support Oligonucleotide Ligation Detection (SOLiD) system is based on sequencing by ligation technique. The system generates sequences on the fragmented DNA templates by ligating a pool of fluorescently labeled probes which contain random oligonucleotide combinations. Each cycle of hybridization and ligation is preceded by cleavage of the 3' end of ligated probes and the addition of the next fluorescent probe. A detailed description on the sequencing chemistry can be found in a recent review (46). A total of 400 million sequence tags are produced per run and the length of each read can reach up to 50 bases.

NGS offers many advantages over microarray-based assays. First, NGS provides higher base pair resolution, with the exception of tiling arrays which still require a large number of probes to reach its high resolution. Second, NGS has relatively fewer artifacts, such as noise in the form of cross-hybridization, generated by the hybridization step on microarrays. Third, the genome coverage is not limited in NGS by the repertoire of probe sequences fixed on the array. This is particularly important for the analysis of repetitive regions of the genome, which are typically masked out on microarrays. Finally, NGS has a larger dynamic range and provides high-coverage thereby increasing the confidence of the resultant data.

The main weaknesses of NGS are its current cost and availability. The overall cost of NGS including machine depreciation and reagent expense is much higher than those needed in microarray. However, the profiling of whole genome in high resolution is still cheaper in NGS experiments than those in microarray approaches. With the improvements of sequencing chemistry and institutional support for the procurement of sequencing platforms, NGS will become the main choice for genome-wide profiling experiments.

The main challenges of NGS technologies are data management and analysis. NGS generates an immense amount of data that can reach terabytes per machine run in raw image files. This makes data storage a challenge even for facilities with considerable expertise in the management of genomic data (47). Genomic alignment and assembly require bioinformatics skills. New and improved algorithms are needed to identify genome enrichment. Downstream data analyses are time ensuring and expensive. For example, motif finding is the most common analysis for a transcription factor binding on DNA. The process of computing statistical significance in NGS experiments is rather different from that of microarray. It is impossible to discuss issues related to NGS in detail, but

taken together, NGS offers higher resolution and well defined data than array-based experiments for profiling the whole-genome (46, 47). With cost reduction and increase in instrument, as well as improvements in bioinformatics skill sets, data analysis pipelines and statistical analyses, NGS technologies will be highly suitable for genome-wide profiling DNA methylation in large-scale projects.

Conclusions

Epigenetic study has made remarkable progress in the field of cancer research. Utilizing genome-wide approaches, such as microarray and next-generation sequencing, profiling of DNA methylation has come of age; methylome analysis is now widely applied in exploring the functional genome in health and disease. As next-generation sequencing technology improves, new platforms will provide greater sequencing length and depth at lower cost. Targeting selective representations or sampling whole genome, bisulfite sequencing methods can be used for analyzing methylation status at single base resolution. These high throughput analyses can be utilized to identify methylation signatures from cancers and normal specimens that serve as molecular biomarkers for early detection and/or diagnosis. These technologies also enhance the knowledge of disease progress thereby forming a rational basis to epigenetic therapies. In conclusion, the human methylome is coming into its golden age with the aid of next-generation sequencing technologies.

Acknowledgments

The authors thank supports from the NIH grants, U54 CA113001 and R01 CA148818, and Pearly Yan and Claire Seguin for critical reading of the manuscript.

References

- Hotchkiss, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J Biol Chem* 175, 315-332 (1948).
- Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* 321, 209-213 (1986).
- Li, E., Beard, C., Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* 366, 362-365 (1993).
- Riggs, A. D. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* 14, 9-25 (1975).
- Kass, S. U., Pruss, D., Wolffe, A. P. How does DNA methylation repress transcription? *Trends Genet* 13, 444-449 (1997).
- Baylin, S. B., Herman, J. G. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet* 16, 168-174 (2000).
- Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 8, 286-298 (2007).
- Chang, J. W., Huang, T. H., Wang, Y. C. Emerging methods for analysis of the cancer methylome. *Pharmacogenomics* 9, 1869-1878 (2008).
- Ammerpohl, O., Martín-Subero, J. I., Richter, J., Vater, I., Siebert, R. Hunting for the 5th base: techniques for analyzing DNA methylation. *Biochim Biophys Acta* 1790, 847-862 (2009).

10. Huang, T. H., Perry, M. R., Laux, D. E. Methylation profiling of CpG islands in human breast cancer cells. *Hum Mol Genet* 8, 459-470 (1999).
11. Deatherage, D. E., Potter, D., Yan, P. S., Huang, T. H., Lin, S. Methylation analysis by microarray. *Methods Mol Biol* 556, 117-139 (2009).
12. Toyota, M., Ho, C., Ahuja, N., et al. Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res* 59, 2307-2312 (1999).
13. Estecio, M. R., Yan, P. S., Ibrahim, A. E., et al. High-throughput methylation profiling by MCA coupled to CpG island microarray. *Genome Res* 17, 1529-1536 (2007).
14. Hatada, I., Kato, A., Morita, S., et al. A microarray-based method for detecting methylated loci. *J Hum Genet* 47, 448-451 (2002).
15. Fukasawa, M., Kimura, M., Morita, S., et al. Microarray analysis of promoter methylation in lung cancers. *J Hum Genet* 51, 368-374 (2006).
16. Khulan, B., Thompson, R. F., Ye, K., et al. Comparative isoschizomer profiling of cytosine methylation: The HELP assay. *Genome Res* 16, 1046-1055 (2006).
17. Oda, M., Glass, J. L., Thompson, R. F., et al. High-resolution genome-wide cytosine methylation profiling with simultaneous copy number analysis and optimization for limited cell numbers. *Nucl Acids Res* 37, 3829-3839 (2009).
18. Brunner, A. L., Johnson, D. S., Kim, S. W., et al. Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res* 19, 1044-1056 (2009).
19. Lippman, Z., Gendrel, A-V., Black, M., et al. Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430, 471-476 (2004).
20. Nouzova, M., Holtan, N., Oshiro, M. M., et al. Epigenomic changes during leukemia cell differentiation: analysis of histone acetylation and cytosine methylation using CpG island microarrays. *J Pharmacol Exp Ther* 311, 968-981 (2004).
21. Ibrahim, A. E., Thorne, N. P., Baird, K., et al. MMAS: an optimized array-based method for assessing CpG island methylation. *Nucl Acids Res* 34, e136 (2006).
22. Irizarry, R. A., Ladd-Acosta, C., Carvalho, B., et al. Comprehensive high-throughput arrays for relative methylation (CHARM). *Genome Res* 18, 780-790 (2008).
23. Irizarry, R. A., Ladd-Acosta, C., Wen, B., et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* 41, 178-186 (2009).
24. Gitan, R. S., Shi, H., Chen, C. M., Yan, P. S., Huang, T. H. Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res* 12, 158-164 (2002).
25. Adorjan, P., Distler, J., Lipscher, E., et al. Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucl Acids Res* 30, e21 (2002).
26. Kimura, N., Nagasaka, T., Murakami, J., et al. Methylation profiles of genes utilizing newly developed CpG island methylation microarray on colorectal cancer patients. *Nucl Acids Res* 33, e46 (2005).
27. Aihara, T., Takatsukaa, Y., Itohb, K., et al. Phase II study of concurrent administration of doxorubicin and docetaxel as first-line chemotherapy for metastatic breast cancer. *Oncology* 64, 124-130 (2003).
28. Bibikova, M., Lin, Z., Zhou, L., et al. High-throughput DNA methylation profiling using universal bead arrays. *Genome Res* 16, 383-393 (2006).
29. Killian, J. K., Bilke, S., Davis, S., et al. Large-scale profiling of archival lymph nodes reveals pervasive remodeling of the follicular lymphoma methylome. *Cancer Res* 69, 758-764 (2009).
30. Weber, M., Davies, J. J., Wittig, D., et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 37, 853-862 (2005).
31. Down, T. A., Rakyen, V. K., Turner, D. J., et al. A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis. *Nat Biotech* 26, 779-785 (2008).
32. Rauch, T., Pfeifer, G. P. Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Lab Invest* 85, 1172-1180 (2005).
33. Rauch, T., Li, H., Wu, X., Pfeifer, G. P. MIRA-assisted microarray analysis, a new technology for the determination of DNA methylation patterns, identifies frequent methylation of homeodomain-containing genes in lung cancer cells. *Cancer Res* 66, 7939-7947 (2006).
34. Rauch, T. A., Wu, X., Zhong, X., Riggs, A. D., Pfeifer, G. P. A human B cell methylome at 100-base pair resolution. *Proc Natl Acad Sci USA* 106, 671-678 (2009).
35. Miller, M. B., Tang, Y. W. Basic concepts of microarrays and potential applications in clinical microbiology. *Clin Microbiol Rev* 22, 611-633 (2009).
36. Kreil, D.P., Russell, R. R., Russell, S. Microarray oligonucleotide probes. *Methods Enzymol* 410, 73-98 (2006).
37. Ansorge, W. J. Next-generation DNA sequencing techniques. *N Biotechnol* 25, 195-203 (2009).
38. Lister, R., Ecker, J. R. Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res* 19, 959-966 (2009).
39. Cokus, S. J., Feng, S., Zhang, X., et al. Shotgun bisulfite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215-219 (2008).
40. Taylor, K. H., Kramer, R. S., Davis, J. W., et al. Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. *Cancer Res* 67, 8511-8518 (2007).
41. Meissner, A., Mikkelsen, T. S., Gu, H., et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766-770 (2008).
42. Deng, J., Shoemaker, R., Xie, B., et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat Biotech* 27, 353-360 (2009).
43. Ball, M. P., Li, J. B., Gao, Y., et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotech* 27, 361-368 (2009).
44. Lister, R., O'Malley, R. C., Tonti-Filippini, J., et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133, 523-536 (2008).
45. Lister, R., Pelizzola, M., Dowen, R. H., et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315-322 (2009).
46. Metzker, M. L. Sequencing technologies-the next generation. *Nat Rev Genet* 11, 31-46 (2010).
47. Park, P. J. ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 10, 669-680 (2009).

Received: October 21, 2009; Revised: January 1, 2010;

Accepted: January 27, 2010

