2008

Development of Oligonucleotide Microarray for High Throughput Dna Methylation Analysis

Xiaopeng Li
Cleveland State University

DEVELOPMENT OF OLIGONUCLEOTIDE MICROARRAY FOR

HIGH THROUGHPUT DNA METHYLATION ANALYSIS

XIAOPENG LI

Bachelor of Chemistry

Zhengzhou University, China

July, 2004

Submitted in partial fulfillment of requirements for the degree

DOCTOR OF PHILOSOPHY IN CLINICAL/BIOANALYTICAL CHEMISTRY

at the

CLEVELAND STATE UNIVERSITY

December, 2008

This dissertation has been approved for

the Department of Chemistry and for

the College of Graduate Studies of

Clevelnd State University

By

_____ Date_____
Dr. Baochuan Guo, Dept of Chemistry/Cleveland State University
Major Advisor

_____ Date_____
Dr. Aimin Zhou, Dept of Chemistry/Cleveland State University
Advisory committee Member

_____ Date_____
Dr. Anton Komar, Dept of BGES/Cleveland State University
Advisory committee Member

_____ Date_____
Dr. John Turner, Dept of Chemistry/Cleveland State University
Advisory committee Member

_____ Date_____
Dr. Xue-Long Sun, Dept of Chemistry/Cleveland State University
Advisory committee Member

ACKNOWLEDGEMENTS

DEVELOPMENT OF OLIGONUCLEOTIDE MICROARRAY FOR HIGH THROUGHPUT DNA METHYLATION ANALYSIS

XIAOPENG LI

ABSTRACT

DNA methylation is a key event regulating gene expression. DNA methylation analysis plays a pivotal role in unlocking association of epigenetic events with cancer. However, simultaneous evaluation of the methylation status of multiple genes is still a technical challenge. Microarray is a promising approach for high-throughput analysis of the methylation status at numerous CpG sites within multiple genes of interest. In this dissertation study, we conducted a systematic study to examine the use of microarray methods for methylation analysis.

First, a robust universal microarray was established with more flexible in design and content, and potential cost saving over commercial arrays. In order to produce high quality microarray data, we optimized the attachment chemistry for the modified oligonucleotides, searched for the good combination of fluorescent dyes, and

hybridization conditions. To improve the specificity of the microarray, we conducted a study to experimentally search for a set of highly discriminative tag Sequences.

Second, SBE-TAGs microarray was successfully adapted from the SNP detection for methylation analysis of multiple genes. SBE-TAGs microarray performed quite well in multiplex methylation analysis of cell lines if a standard calibration curve method was used. 10 CpG sites of 9 tumor suppressor genes (MGMT, GATA4, HLTF, SOCS1, p16, RASSF2, CHFR, TPEF, and Reprimo) were selected for this study.

Third, a novel method called CHZMA (Competing-Hybridization-Zipcode-MicroArray) was developed for methylation analysis of tumor tissue samples, which is based on two steps of hybridization to achieve the specific detection of methylation on microarray. On the basis of analysis of seven genes (MGMT, GATA4, HLTF, SOCS1, RASSF2, ER, 3-OST-2), we found that the CHZMA assay can robustly detect methylation of multiple genes in the samples containing as low as 10% of methylated DNA. With the strict control group test and statistical analysis, CHZMA can be a good high-throughput method in place of MSP for methylation analysis of tumor tissue samples.

These studies provide reliable and robust tools for methylation analysis, and could be used for cancer prognosis and diagnosis in the future.

TABLE CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BAC	Bacterial Artificial Chromosome
BiMP	Bifulfite Methylation Profiling
CGH	Comparative Genomic Hybridization
ESI-MS	Electrospray Mass Spectrometry
MBD	5-Methylcytosine Binding Domain
MeDIP	Methylated DNA Immunoprecipitation
HELP	HpaII Tiny Fragment Enrichment by Ligation-Mediated PCR
HPCE	High Performance Capillary Electrophoresis
HPLC	High Performance Liquid Chromatography
MALDI-TOF	Matrix Assisted Laser Desorption/Ionization Time-of-Flight
m^5C	5-methylcytosine
m^5dC	5- Methyl-2'-Deoxycytidine
MSP	Methylation-Specific PCR
MSO	Methylation-Specific Oligo
PNA	Peptide Nucleic Acid
PCR	Polymerase Chain Reaction
RLGS	Restriction Landmark Genomic Scanning
RP-HPLC	Reverse-Phase HPLC
SAM	S-Adenosyl-L-Methionine
SBE	Single Base Extension
TLC	Thin-Layer Chromatography

CHAPTER I

CURRENT DNA METHYLATION ANALYSIS METHODOLOGY

1.1 Introduction

1.1.1 DNA Methylation: the Major Form of Epigenetic Change in Genome

It is now clear that genetic and epigenetic are the two forms of information included in genome. Genetic information provides the blueprint to manufacture all of the proteins, while epigenetic information provides the instruction to use these proteins (Robertson 2001). In the mammalian cell, methylation is a major form of epigenetic change, which is modified by DNA methyltransferases (DNMTs) adding a methyl group on the 5-position of cytosine within the CpG dinucleotide site (Doerfler 1983). Approximately, 70-80% of CpG sites in the mammalian genome contain methylated cytosines (Antequera and Bird 1993; Bird 1995). Methylated cytosines are widely spread throughout the genome, predominantly in repetitive genomic regions, including satellite DNA and parasitic elements (Yoder et al.1997). Unmethylated CpG sites are primarily

confined to DNA regions with high relative densities of CpG, called CpG islands, ranging from 0.5 to 5 kb and occurring on average every 100 kb (Bird 1986; Gardiner-Garden and Formmer 1987). Recently, it is estimated at least 29, 000 CpG islands in the human genome (Lander et al. 2001; Venter et al. 2001), are distributed in a non-random pattern, particularly in the promoters and the first exon regions of protein coding genes (Bird 1986).

DNA methylation is a crucial epigenetic alteration, involved in embryonic development, transcription, chromatin structure, X chromosome inactivation, genomic imprinting and chromosome stability. In the normal cell, most CpG islands remaining unmethylated are associated with transcriptionally active genes, and some CpG islands were found normally methylated in imprinted genes (Li et al. 1993; Paulsen and Ferguson-Smith 2001), X chromosome inactivation genes (Panning and Jaenisch 1998), and parasitic DNA suppression genes (Walsh et al. 1998). Whereas DNA methylation is a crucial mechanism to regulate the genetic information, it also adds an additional burden, such as abnormal methylation was found in human carcinogenesis.

1.1.2 DNA Methylation and Cancer

As the two major forms of methylation, genome-wide hypomethylation and region-specific hypermethylation have been broadly found in carcinogenesis. In human carcinogenesis, the hypermethylation in promoters will repress transcription via the inhibiting the binding of specific transcription factors directly, or via local changes in histone modification and chromatin structure by methyl-CpG-binding proteins indirectly (Robertson 2005). In contrast, the biological significance of genome-wide hypomethylation in the repetitive regions is less understood. However, global

measurements of DNA hypomethylation remain a valuable tool for understanding the molecular pathology of human cancer, and for monitoring therapeutic responses to potential compounds in human clinical trials. It is clear that DNA hypomethylation can foster chromosome instability (Ehrlich 2002; Eden et al. 2003), increase the mobility of transposable elements (Walsh et al. 1998), and induce tumorigenesis in different mouse models (Eden et al. 2003; Gaudet et al. 2003; Yamada et al. 2005; Jones and Baylin 2007). In addition, the correlation between hypomethylation and increased gene expression was reported for many oncogenes (Feinberg and Vogelstein 1983; Hanada et al. 1993).

In cancer, promoter hypermethylation is associated with gene silencing, and together with point mutations and deletions, serves as one of the most common mechanisms for the loss of function of tumor suppressor genes (Baylin et al. 1998; Costello and Plass 2001; Jones and Baylin 2002; Robertson 2001; Robertson2005). Because of the importance of hypermethylation in carcinogenesis, a growing number of studies have provided fundamental insights into cancer development. In virtually every type of human neoplasm, promoter methylation is found to be associated with inappropriate transcriptional silencing of genes (Costello et al. 2000; Toyooka et al. 2001; Rashid et al. 2001; Martens et al. 2005; Sato et al. 2005; Kusano et al. 2006). As DNA methylation is more readily reversible than genetic events, DNA hypermethylation in promoter is emerging as a potential diagnostic and prognostic tool (Karpf and Jones 2002).

1.1.3 DNA Methylation Analysis

DNA methylation research can be approached from a wide range of techniques. Each technique has its own peculiarities to offer a different perspective for occurrence

3

and localization of methylation in the genome. In this chapter, a summary of the most commonly used techniques will be provided for methylation research, in particularly for hypermethylation analysis. A general outline of the principles, advantages, and disadvantages of these techniques will be discussed in each section. The current methods can be broadly classed into two major approaches: global methylation analysis (genome-wide); and site-specific methylation status determination (Oakeley 1999; Fraga and Esteller 2002; Laird 2003). In the following, all of the methods will be grouped into these two categories (Fig. 1). In global methylation analysis, high-performance separation (HPLC and HPCE), enzymatic reaction, chemical reaction, and immunology-based methods will be discussed. In site-specific methylation analysis, the methods are divided into two subcategories, bisulfite and non-bisulfite treatment. For the non-bisulfite methods, we focused on the application of methylation-sensitive and methylation-insensitive restriction endonucleases. For the bisulfite treatment methods, a series of methods including sequencing, methylation-specific PCR (MSP), quantitative MSP (QMSP), single base extension, ligation, fragment analysis, and microarray et al., will be discussed.

1.2 Global Methylation Analysis

1.2.1 Reverse-Phase HPLC (RP-HPLC)

Global genomic DNA methylation status can be determined by reverse-phase HPLC using chemical or enzymatic hydrolysis of DNA. In the chemical hydrolysis, DNA is incubated in formic acid (Eick et al. 1983) or hydrofluoric acid (Catania et al. 1987). After the hydrolysis, liberated cytosine, thymine, guanine, adenine, and 5-methylcytosine
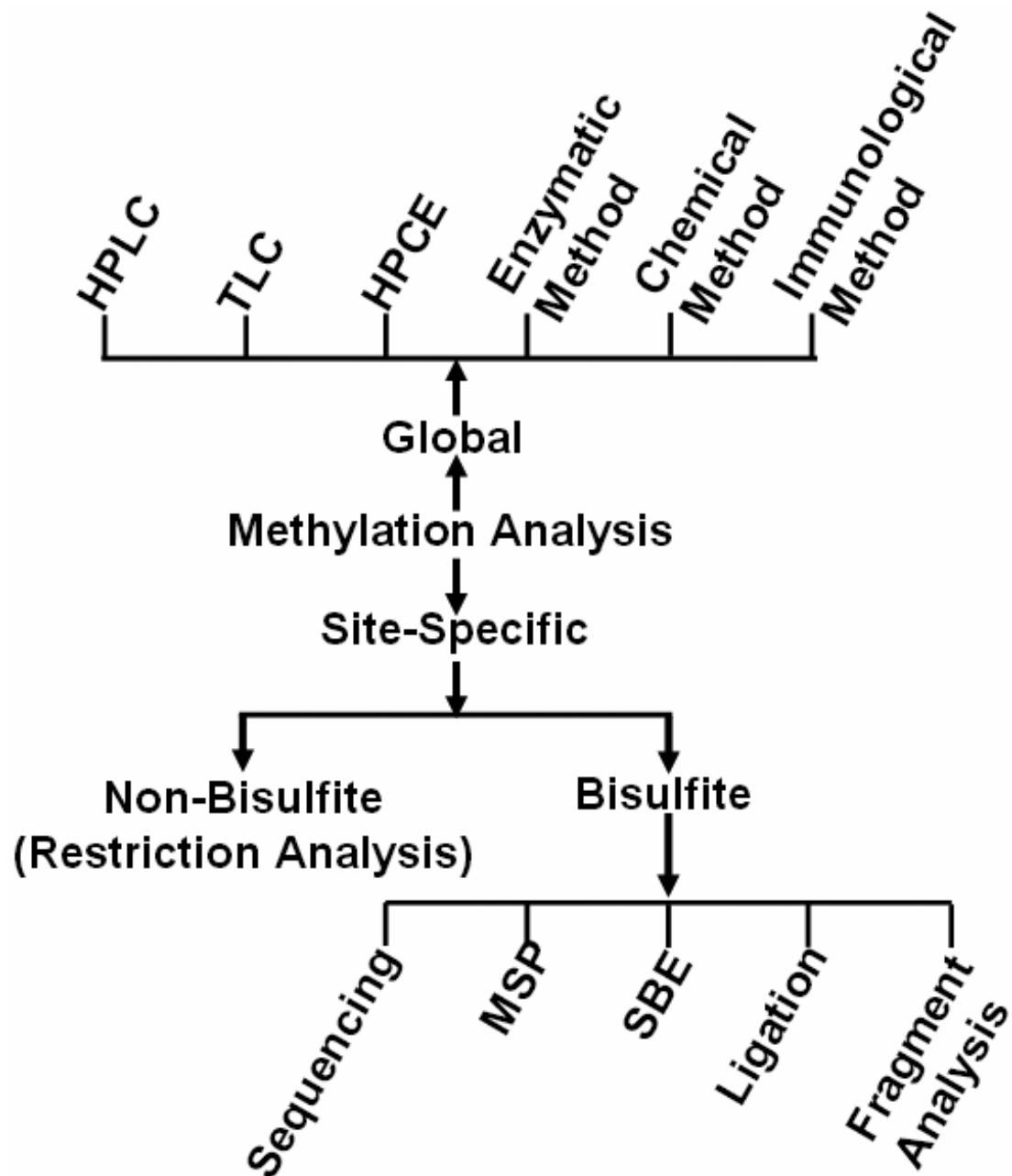
Fig. 1 The methods of DNA methylation analysis. The existing methylation analysis methods are assorted into global and site-specific analysis. The site-specific analysis includes two categories: non-bisulfite and bisulfite analysis. MSP: methylation-specific PCR; SBE: single base extension.

are subsequently separated and determined by reverse-phase HPLC. Since the deamination of cytosine and 5-methylcytosine occurs during the hydrolysis procedure using formic acid, quantitation of 5-methylcytosine in DNA is not as accurate as hydrofluoric acid is used in hydrolysis.

As an alternative, enzymatic hydrolysis presents more power in quantitative methylation analysis. The DNA sample is the hydrolyzed to produce 2'-deoxymononucleosides, using DNase I, Nuclease P1, or snake venom phosphodiesterase, followed by alkaline phosphatase treatment (Kuo et al. 1980; Ehrlich 1982; Gomes and Chang 1983). The liberated 2'-deoxymononucleosides are separated and detected by standard HPLC-UV system, which can detect 2–10% of the 5-methylcytosine (m^5C) (Kuo et al. 1980; Wagner and Capesius 1981; Gama-Sosa et al. 1983). HPLC operating with fluorescence detection can provide a 10-fold higher sensitivity, but has drawbacks with tedious labeling procedures (Sonoki et al. 1998; Wirtz et al. 2004).

Recently, with the utilizing of HPLC-MS system, the sensitivity of methylation detection has been greatly improved. Babinger and co-workers used RP-HPLC combined with electrospray mass spectrometry (ESI-MS) to analyze *Volvox* DNA. As little as 1.1% of 5-methylcytosine was detected in total deoxycytidine residues of genome (Babinger et al. 2001). Friso also detected about 1.7-1.9% of 5-methylcytosine in total deoxycytidine with <1μg DNA sample, using HPLC-ESI system (Friso et al. 2002). However, RNA contamination of the DNA sample could cause the quantification to fail, as ESI-MS was incapable of distinguishing 5-methyl-2'-deoxycytidine from 5-methyl-2'-cytidine (Havlis and Trgbusek 2002). Alternatively, Sony et al. have developed a highly specific and

sensitive assay with tandem mass spectrometry (Song et al. 2005). With this HPLC-ESI-MS/MS system, methylation levels ranging from 0.05 to 10% can be detected in 4 ng of DNA. This detection sensitivity should permit the use of this method for applications having limiting amounts of DNA in clinical samples.

1.2.2 Thin-Layer Chromatography (TLC)

The TLC was initially reported for separation and determination of 5-methyldeoxycytosine (m^5dC) in 1981 (Wagner and Capesius 1981). In this assay, the restriction enzyme *MspI* cuts the target site CCGG regardless of whether cytosine is methylated or not. Bestor and Schmitt groups labeled the internal cytosine with [^{32}P]ATP and polynucleotide kinase, then hydrolyzed the nucleotides with nuclease P1 to individual nucleotides, which is separated on 1-D or 2-D cellulose thin-layer chromatography plates (Bestor et al. 1984; Schmitt et al. 1997). In Leonard and Gowher studies, the nucleotides were digested to individual nucleotides, which are separated by RP-HPLC to isolate both m^5dC and dC. Then, the individual m^5dC and dC were labeled with [^{32}P]ATP and polynucleotide kinase. Finally, the 1-D or 2-D TLC was used to identify the methylation status (Leonard et al. 1993; Gowher et al. 2000). This sensitive assay could detect the presence of m^5C in genomic DNA at the level of 1 in 1000-2000 cytosine residues (Leonard et al. 1993). One of the shortcomings of this method is that digesting the nucleotides to produce individual m^5dC and dC prior to the labeling makes the dC much easier to label than m^5dC. This will reduce the detection limit of m^5dC in genomic sample. In addition, noisy signals from A, G, and T were observed on the TLC plates (Oakeley 1999).

7

1.2.3 High Performance Capillary Electrophoresis (HPCE)

Two major modifications of capillary electrophoresis: micellar-electrokinetic capillary chromatography and capillary zone electrophoresis, are frequently used in global methylation analysis. Fraga et al. has reported a new open-tube capillary electrophoresis system to separate the acidic/enzymatic hydrolyzed genomic DNA for methylation analysis. The dC and m^{5}dC were successfully separated in SDS micellar-electrokinetic system with UV detection (Fraga et al. 2000). In contrast to HPLC, HPCE does not need specific running buffer, and is faster than HPLC with 10min/sample. The limitation of detection is one m^{5}dC in 200 dC, using a 1μg DNA sample. Sandoval Guerrero et al. optimized this system to more efficient peaks, a flatter base line, and shorter analysis time with employing a fused-silica capillary (75 mm inner diameter) of 44.5 cm effective length, 20 mM carbonate buffer (pH 9.6) plus 80 mM sodium dodecyl sulfate (Sandoval Guerrero et al. 2005). To overcome the limitation of the UV detection at 280 nm, Schmitz's group developed a laser-induced fluorescence detection system to improve the sensitivity (Lyko et al. 2004; Wirtz et al. 2004). In the assay, the hydrolyzed mono-nucleotides, including dC and m^{5}dC, were modified with fluorescent dye (Bodipy FLEDA), then separated in micellar-electrokinetic chromatography and detected with the laser-induced fluorescence system. After optimization, the method can determine the methylation level in 100-ng DNA samples (10 folds better than UV) with a standard deviation of less than 5%, suitable for clinical sample testing.

Recently, a new capillary zone electrophoresis method applying short-end injection technique was reported (Sotgia et al. 2008). By injecting the sample at the capillary end, closest to the detector window, the migration distance of analytes is

significantly reduced to10.2 cm. Thus, dC and m^{5}dC after acid hydrolysis were separated with a good resolution in less than 1.5 min, which is much shorter than normal injection with 6-10min. The lowest amount of DNA required to evaluate methylation is about 0.5 μg.

1.2.4 Enzymatic Method

In vivo, *SssI* methyltransferase (CpG methylase) catalyses the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to C5 of cytosine within CpG dinucleotide sequences in genomic DNA shortly after DNA replication (Baylin 1997). Under this mechanism, *SssI* methyltransferase based enzymatic methods were developed for methylaiton analysis (Adams et al. 1991; Balaghi et al. 1993; Wu et al. 1993; Belinsky et al. 1995; Schmitt et al. 1997). In this assay, [Methyl-^{3}H]SAM is used as a substrate to incorporate into all of the unmethylated cytosines in CpG sites of genomic DNA under the catalyses of *SssI* methyltransferase. After that, the labeled DNA is collected on filter paper, washed, and counted in a scintillation counter. The amount of incorporated radioactive Tritium is proportional to the unmethylated level of the DNA sample. Due to the crude or semipurified preparations of *SssI* methyltransferase, RNA and protein remaining in genomic DNA may possibly be labeled with [Methyl-^{3}H]SAM in the presence of other methyltransferases to yield artificial results. Furthermore, the nonspecific absorption of radioactivity in the precipitates can also yield error results (Chiang et al. 1996). Fiala et al. developed a modified *SssI* methyltransferase assay for hypomethylation analysis in tissue samples, using HPLC with radioflow detection after enzymatic hydrolysis, enhancing specificity and reliability (Fiala et al. 1998). The problem with the *SssI* methyltransferase assay is the handling and disposal of Tritium.

9

1.2.5 Chemical Method

Oakeley et al. developed a method to label any m^5dC in genomic sequence with sodium bisulfite and chloracetaldehyde reaction (Oakeley et al. 1999). First, the genomic DNA was subjected into bisulfite treatment, which coverted all of dC into dU, while having no effect on m^5dC (Frommer et al. 1992). Second, using sulphuric acid, bisulfite treated DNA was depurinated to eliminate 99% of dA, which was precipitated with silver nitrate. The supernatant was incubated with chloracetaldehyde to form ethenocytosine (derivative of m^5dC), and residual ethenoadenine (derivative of dA). Both of these two derivatives are intensely fluorescent and with two different major peaks. Only dA, dC, and m^5dC can react with chloracetaldehyde. Hence, the bisulfite conversion and dA removal are critical to the assay. After the fluorescence scanning, the methylation signal was determined by removing the contamination of ethenoadenine. As the chloracetaldehyde is quite toxic, special care must be taken in the labeling and detoxification.

Recently, Okamoto group reported a m^5dC-selective reaction with potassium osmate, functional bipyridine, and potassium hexacyanoferrate (III) (Okamoto et al. 2006; Tanaka et al. 2007 a, b). This osmate reaction distinguished m^5dC and dC with high selectivity. All of the study of this new method is focused in site-specific methylation detection with a hybridization probe. But this reaction has great potential for global methylation analysis, if combined with other techniques. Chemical based analysis, as a good alternative mean, can detect any m^5dC in the genomic DNA.

1.2.6 Immunological Method

Using monoclonal antibodies directed against m5C is another approach for global DNA methylation analysis. Podestà et al. described the preparation and basic properties of monoclonal IgG1 (Podestà et al. 1993), whose specificity was validated by using methylated and unmethylated oligonucleotides (Oakeley et al.1997) In this assay, a DNA sample is immobilized on the DEAE membrane, then incubated in monoclonal IgG1, followed by incubating in the secondary antibody of FITC-labeled goat anti-mouse IgG for fluorescence scanning (Oakeley et al.1997; Zluvova et al. 2001). The methylation level is proportional to the intensity of the fluorescent signal on the membrane. The amount of DNA loaded on membrane is measured by ethidium bromide fluorescence staining after the fluorescent scanning.

The outstanding feature of immunology assay is to visualize the chromosomal methylation patterns by fluorescence microscopy on a cell-by-cell basis. The DNA in chromosomes can be labeled only when m^5dC is not base-paired. In the early research, UV light, alkali, or heat treatment was used for long time denaturation of chromosome (Miller et al. 1974; Barbin et al. 1994; Montpellier et al. 1994; Rougier et al. 1998). After the denaturation, the mono-antibodies combined with fluorescence dye were incorporated into chromosome. Due to the disruption of chromosome structure in the denaturation, gentle conditions were used in denaturation with a short UV light exposure (three hours) at a low temperature (4^oC) (Bensaada et al. 1998). This method can be used to determine the methylated status of chromosomes, and to map not only the structural (banding) but also the functional (methylation status) properties of the different chromosome domains in cells.

1.3 Restriction Analysis for Site-Specific Methylation Detection

Methylation-sensitive digestion combined with insensitive restriction endonucleases provide a simple tool for methylation analysis of large quantities DNA. As the most common isoschizomers, *Hpa*II and *Msp*I are used to cleave the DNA at CCGG target, where *Hpa*II is not able to cut the sequence when m^{5}dC is present in the CpG site (Cedar et al. 1979; Singer et al. 1979; Bird 1980). The DNA samples are digested by *Hpa*II and *Msp*I, respectively. Then, the different size of the digests will be revealed on Southern blot by hybridizing with radio isotope labeled probe. If the *Hpa*II and *Msp*I digests have the same size hybridized bands, the target site covered by probe is unmethylated. In the contrast, if the hybridized bands of *Hpa*II are bigger than *Msp*I, the target site is methylated. To improve the sensitivity, PCR was used to amplify the digests (Singer-Sam et al. 1990). The PCR amplification region must contain single CCGG or more. If the CCGG is methylated, only DNA digested with *Hpa*II can be amplified, and no PCR amplification is from *Msp*I digests. If the CCGG is unmethylated, both of *Hpa*II and *Msp*I digests can not amplify with the cleaved templates. For PCR detection, less than 10ng DNA is required for digestion, whereas 10μg is required for Southern blot detection. Furthermore, 1 m^{5}dC in 1000 dC is able to be detected (Kutueva et al. 1996). Although this method is straightforward, some problems arise from the limitation of restriction sites, false-positives caused by partial digest, and the high molecular weight DNA requirement.

Recently, Methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) was reported to detect 40 sequences simultaneously, using only 20ng of DNA (Nygren et al. 2005). In the assay, the methylated and unmethylated DNA samples

were denatured and hybridized with one pair of ligation probes with universal sequence, followed by simultaneously ligated and digested with ligase and methylation sensitive endonucleases. The ligated unmethylated sequence was cleaved by methylation sensitive endonucleases, and could not be amplified by universal primers. The ligated methylated sequence remained intact as a template for PCR amplification. The PCR products with different sizes were measured by fragment analysis.

Restriction landmark genomic scanning (RLGS) is another restriction enzyme based method, which can evaluate over 2000 loci simultaneously with two-dimensional gel electrophoresis of digested genomic DNA (Hatada et al. 1991). In the methylation analysis, DNA is digested with methylation-sensitive enzyme, like *NotI* (GCGGCCGC) or *AscI* (GGCGCGCC), and directly labeled with a radioactive isotope. Methylated sites are not digested and are therefore not labeled, thus they do not contribute to the two-dimensional pattern of RLGS fragments. The DNA samples require the high-quality, high-molecular-weight to prevent nonspecific labeling of degraded fragments. After the first restriction digest with methylation-sensitive enzyme, the DNA is subjected to the second restriction to produce short strands for the first-dimension electrophoresis with 0.8% agarose gel. Following that, DNA in agarose is restriction digested at 3rd time to produce shorter strands for second-dimension electrophoresis in a large 5% acrylamide gel. Finally, the digested sites from the unmethylation sequence display with corresponding spots with the autoradiography. With the control profiles from normal group, they can determine the absent spot, which results from the failure of methylation-sensitive digest and represents the methylation site in the sample (Shibata et al. 1995; Akama et al. 1997; Costello et al. 2000; Matsuyama et al. 2003). Although

RLGS is highly reproducible, the assignment of specific genomic sequences to RLGS spots is still a challenge. Plass and co-workers set up the standard data system, called Master profile, for each locus assignment (Smiraglia and Plass 2002; Rusha and Plass 2002). With the completion of the human and mouse genome sequences, a virtual RLGS profile based on the bioinformatics approach became available for RLGS fragment identification. In the virtual RLGS profile each restriction fragment's migration in both dimensions could be predicted (Rouillard et al. 2001; Smiraglia et al. 2007). As much, RLGS has become a versatile tool in hypomethylation and hypermethylation analysis for normal tissues, primary tumors, and cancer cell lines in various organisms.

1.4 Bisulfite-Based Methods for Site-Specific Methylation Detection

1.4.1 Bisulfite Conversion

The analysis of DNA methylation was revolutionized by sodium bisulfite conversion, which permits the indentification of any CpG sites in the genome, rather than the specific sequences targeted by restriction endonucleases. The different conversion rates of cytosine and 5-methylcytosine deaminated by sodium bisulfite to yield uracil and thymine were described in 1970s (Hayatsu et al. 1970; Shapiro et al. 1970; Hayatsu and Shiragami 1979). This converstion did not get much attention until Frommer et al. used it to distinguish the dC and m^5dC. The breakthrough in this area is that bisulfite converts all of dCs to uracils whereas the m^5dC residues remain intact. Then, the treated DNA is subjected to PCR amplification to yield fragments wherein Ts replace Us (former Cs), and Cs replace the methylated Cs (Frommer et al. 1992). Based on this conversion, DNA methylation research can be approached from a wide range of techniques, including

14

sequencing, methylation-specific PCR, combined bisulfite restriction assays, microarray, and others (Fig. 2).

In the bisulfite treatment, the incomplete conversion of dCs will cause false-positive results. To prevent the non-converted dCs, special attention should be paid for DNA denaturation, desulfonation, bisulfite sodium concentration, and reaction conditions, whereas, 84-96% of the DNA is degraded under this condition (Grunau et al. 2001). More significant losses are found during the subsequent reaction and purification for short strand DNA, thus more than 99% of DNA can be degraded (Munson et al. 2007; Tanaka and Okamoto 2007).

As the bisulfite conversion is susceptible to processing errors, several attempts have been made to simplify the treatment. Raizis et al. reported that using high concentration (5M) of sodium bisulfite can reduce the degradation with 4h incubation (Raizis et al. 1995). Olek et al. performed the bisulfite treatment and subsequent PCR steps with low melting point agarose beads, in which the denatured DNA will be kept in single strand form to avoid strand annealing (Olek et al. 1996). The single strand DNA improves the reaction efficiency, and reduces the reaction time to 4h. On the other hand, the subsequent reaction and purification of bisulfite converted DNA within agarose beads will be gentle to reduce the DNA strand cleavage. With these two modifications, 500 pg of starting DNA can provide reliable PCR products. Many research works focused on clinical specimens rely on this agarose beads based method to improve the recovery of bisulfite treated DNA, in particularly in tiny tissue and serum specimens, which often provide less than 50ng DNA (Kitazawa et al. 2000; Kim et al. 2005; Peng et al. 2006). To reduce the matrix binding loss in the desulfonation and purification, centrifugal filtration
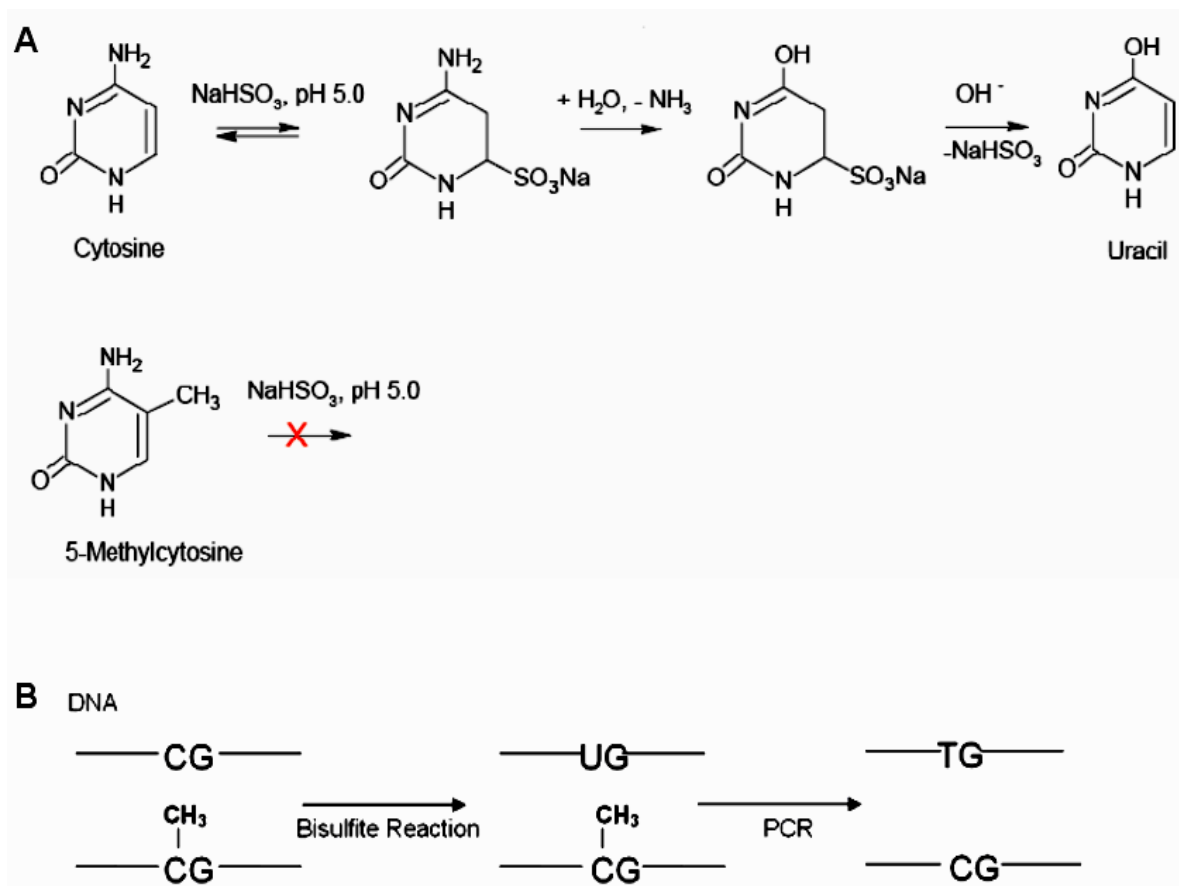
Fig. 2 Bisulfite conversion. (A) In the bisulfite treatment, only normal cytosine is converted to uracil, and 5-methylcytosine is intact. (B) The treated DNA is amplified in PCR to generate fragments wherein Ts replace Us (former Cs), and Cs replace the methylated Cs. These sequence variation can be analyzed using conventional SNP or mutation analysis methods.

was used to provide a simplified, high-recovery procedure that lends itself to increased throughput and can be applied to 300 pg starting DNA (Boyd and Zon 2004). Recently, the molecular diagnostic researchers target the tumor-derived short strand DNA in samples such as blood (Cottrell and Laird 2003; Lofton-Day et al. 2008), urine (Westra et al. 2006; Yu et al. 2007), sputum or saliva (Konno et al. 2004) and stool (Chen et al. 2005; Zhang et al. 2007; Wang and Tang 2008). However, the degradation of the tiny DNA in bisulfite treatment is still a big challenge to improve the sensitivity and specificity.

1.4.2 Sequencing

Bisulfite sequencing genomic DNA is the most common method in methylation analysis. The PCR primers are designed without CpG sites contained in their sequence to ensure unbiased amplification of both methylated or unmethylated DNA. After the PCR amplifies, the PCR product can be subjected for the direct sequencing, or subcloning followed by sequencing the subclones (Frommer et al. 1992). The direct sequencing is fast, simple, and may better reveal the overall methylation status of the DNA sample. The human epigenome project in Europe used direct sequencing of bisulfite PCR products (Rakyan et al. 2004; Eckhardt et al. 2006). In this method, the methylation status of any given CpG site is the average of all amplicons generated during PCR. Sequencing the subcloned PCR products is time consuming, but can deliver the methylation patterns of individual molecules. This provides a better understanding of complex pattern of methylation such as monoallelic methylation and incomplete allele methylation. Because of the DNA methylation patterns with the ability to reflect different functions of the DNA methylation system (Colot and Rossignol 1999), sequencing of subclones has been used to study cancer related methylated genes such as APC (Esteller et al. 2000), Rb (Stirzaker

et al. 1997), TMS (Levine et al. 2003), and RASSF1A (Peters et al. 2007).

As alternative sequencing to conventional chain termination sequencing (Sanger et al. 1977), pyrosequencing holds great promise for both confirmatory sequencing and de novo sequencing (Ronaghi et al. 1996; Ronaghi et al. 1998). Pyrosequencing is a sequencing-by-synthesis technique based on the detection of released pyrophosphate (PPi) during DNA synthesis. With ATP sulfurylase, the released PPi is subsequently converted to ATP, which provides the energy to luciferase to oxidize luciferin and generate light. The intensity of the light is proportional to the number of incorporated nucleotides and the sequence is determined by added nucleotide in the cycling reaction.

However, this technique has not been used for genome sequencing due to the short read length about 1-300 nucleotides (Mashayekhi and Ronaghi 2007). The interesting thing is that this limitation makes pyrosequencing suitable for bisulfite treated DNA with short strands (Uhlmann et al. 2002). Moreover, since pyrosequencing works better on small PCR products (100-150 bp), it provides good resolution for clinical or archival sample with partial degraded DNA, which is a big problem with other techniques (Colella et al. 2003; Tost et al. 2003; Shaw et al. 2006). Recently, a novel parallel pyrosequencing in microfabricated high-density picolitre reactors was developed by 454 sequencing (Margulies et al. 2005). With this new platform, around 300,000 DNA templates with 100 bp length can be simultaneously sequenced in 5.5 h with an accuracy of 99.6%. This technology provides high-throughput and ultradeep methylation analysis for multiple tumor types, and multiple genes. Taylor et al. described the methylation pattern analysis of 294,631 DNA fragments with an average read length of 131bp from 25 gene-related CpG rich regions in more than 40 primary cells (Taylor et al. 2007).

Korshunova et al. reported the methylation patterns of more than 700,000 DNA fragments derived from breast tissues and circulating DNA in more than 50 individuals (Korshunova et al. 2008).

1.4.3 Methylation-Specific PCR

Methylation-specific PCR (MSP) is the most common used method for methylation analysis, because of its high sensitivity and simplicity (Herman et al. 1996). The differences of methylated and unmethylated sequences derived from bisulfite conversion are the basis of MSP, in particularly in CpG islands. MSP reveals the methylation status of the CpG sites within the two short sequences targeted by MSP primers. Hence, the primer design is crucial to the discrimination power of MSP, where methylation-specific and unmethylated DNA-specific primers cover the same CpG sites and yield different PCR products, respectively.

To avoid the labor-intensive gel electrophoresis after the MSP, Lo et al. and Eads et al. adopted the fluorescence-based real-time PCR (TaqMan) for high-throughput quantitative MSP (QMSP), which was highly sensitive and can detect one methylated allele in 10,000 copies of unmethylated alleles (Lo et al. 1999; Eads et al. 2000). In this QMSP, the methylation level was calculated by dividing the methylated signal with the sum of methylated signal and unmethylated signal. Sidransky and co-workers simplified this QMSP using the unique internal reference gene such as MyOD1 (Jerónimo et al. 2001), and ACTB (Harden et al. 2003; Hoque et al. 2006), which do not have CpG sites in their sequence. The ratio of signal from methylated amplification to internal reference was used to measure the relative methylation level of the target gene. In this simplified QMSP system, only the primers for methylated allele were contained, but no primers for

unmethylated allele.

However, MSP is prone to false-positive because of the low discrimination of primers, especially in incompletely converted sequences in the bisulfite- treated DNA (Aggerholm et al. 2000; Rand et al. 2002). To avoid the false-positive results caused by non-specific amplification, sensitive melting analysis after real time (SMART)-MSP was reported for sensitive DNA methylation detection based on probe-free real-time PCR (Kristensen et al. 2008). In this assay, the false-positive results due to incomplete bisulfite conversion or false priming can be identified by high-resolution melting.

1.4.4 Single Base Extension

Gonzalgo et al. first used the single base extension (SBE) for quantitative methylation analysis of individual CpG sites using bisulfite-PCR products (Gonzalgo and Jones 1997). SBE is then performed with oligonucleotides designed to hybridize immediately upstream of the CpG site being interrogated. The product can be detected by the adding radioactive ddC/ddT (ddG/ddA) on polyacrylamide gels. Several new detection methods have been developed to improve the performance. Based on the different masses and hydrophobicities of incorporated ddNTP, the methylated and unmethylated CpG sites can be discriminated and quantified by ion pair reverse phase (IP RP) HPLC (El-Maarri et al. 2002). Using this method, several CpG sites contained in different SBE products can be measured simultaneously with high reproducibility. Dye labeled ddNTPs are reported for the SBE reaction, followed by capillary electrophoresis with ABI Prism® 3100 Genetic Analyzer (Applied Biosystems, CA) (Hong et al. 2005; Kaminsky et al. 2005). In this method, the first discrimination is to use methylated-specific and unmethylated- specific primers with different length but

20

covering the same CpG sites to anneal the template. The second discrimination is the SBE reaction with different dye labeled ddNTPs. After the electrophoresis, the peaks of methylated and unmethylated primers with extension will show at different positions.

Since MALDI-TOF mass spectrometry combined with primer extension has been shown to be a powerful tool in SNP analysis, Schatz et al. described the combination of MALDI-TOF and primer extension for methylation analysis (Schatz et al. 2003). In the assay, the extension with α-S-ddNTPs results in a product with a phosphorothioate backbone. After SBE, the part of the primer containing regular phosphate bonds is digested off with 5'-phosphodiesterase, while the first phosphorothioate group from the 5'-end inhibits the digestion of the product. Finally the phosphorothioate functions are measured by MALDI-TOF.

In chapter IV, we will present our study with SBE based microarray for high-throughput analysis of multiple genes, and multiple CpG sites.

1.4.5 Ligation

Recently, with bisulfite treatment, two ligation-based microarray methods were developed for methylation analysis (Bibikova et al. 2006; Cheng et al. 2006). In the first method, Barany and coworkers developed a PCR/LDR method for methylation analysis (Cheng et al. 2006). Briefly, it first utilizes multiplexed PCR to amplify multiple target DNA sequences, followed by ligation chain reactions, where methylated-specific and unmethylated-specific probes modified with two different dyes (Cy3/Cy5). The ligation products are then analyzed using microarray to determine the methylation status at each target CpG site via the different fluorescence signals.

In the second method, Fan and his team from Illumina apply a genotyping

system to methylation analysis (Bibikova et al. 2006). Fan's approach first utilizes extension and ligation to produce both methylated and unmethylated alleles, followed by multiplexed PCR with fluoresce labeled universal primers to amplify the sequences containing the target CpG site. The methylation status at each CpG site is analyzed using microarray. In Fan's method, in order to improve the specificity, a gap (1-20bp) was designed between the ligation probes. This gap can be filled by extension of DNA polymerase. Ligation-based microarray improves specificity, allowing for methylation analysis of cancer cell lines or tumor tissues that contain high abundance of tumor cells. However, it is still not a robust method for methylation analysis of tumor tissues that contain a minority of tumor cells. This is because the specificity of annealing a probe to a given sequence varies greatly from one sequence to another (Dahl and Guldberg 2007).

1.4.6 Fragment Analysis

Fragment analysis as the most straightforward method was developed by Boyd et al. from ABI (Applied Biosystems, CA) (Boyd et al. 2006). In this assay, amplicons with different size bisulfite-converted DNA are analyzed immediately after multiplex PCR with fluorescence labeled primers. In gel-capillary electrophoresis, methylated amplicon migrates faster than the unmethylated amplicon. The PCR products of methylated allele and unmethylated allele can be separated due to the accumulation of several cytosines or thymines in the amplicons, respectively. Furthermore, this assay can be used for quantitative analysis the overall methylation level in the selected area. To improve the resolution of methylated and unmethylated peaks due to presence of only a few CpG sites, the modified dCTP is incorporated in the PCR (Boyd et al. 2007). This modified dCTP incorporated in PCR products allows for the detection of single C/T

22

difference derived from single CpG after bisulfite conversion. The only problem with this method is the methylation status can not be assigned to every CpG site in PCR product with multiple CpG sites. But this direct analysis of bisulfite-converted PCR products without any additional sample processing provides a robust tool for high-throughput quantitative methylation analysis, in particular for the hypermethylation analysis in CpG islands.

1.5 Conclusion

DNA methylation is a crucial epigenetic change and is involved in regulating many cellular processes, particularly in relation with gene silencing in disease. Consistent with these important roles, a growing number of methylation detection techniques have been developed to provide fundamental insights into DNA methylation. In this short review, some representative techniques have been introduced to address global and site-specific methylation analysis. Each technique has its own peculiarities to provide a unique approach for a specific problem. In some cases, DNA methylation analysis will combine several methods together to get reliable results. Microarray as an alternative has been widely used in methylaiton detection. Since the microarray application in methylation is the major study of this project, the different platforms of microarray related to methylation analysis will be described in detail in chapter II.

CHAPTER II

HIGH-THROUGHPUT DNA METHYLATION PROFILING USING

MICROARRAY

2.1 Introduction

DNA microarrays are miniature arrays of DNA fragments attached to solid supports, enabling a simultaneous analysis of a large amount of genetic information. Two major forms of microarray were developed by fixing a target DNA fragment or detecting oligonucleotide probes. The first form of microarray was evolved from Southern blotting, whereby oligonucleotide probes are attached to a substrate and then hybridized with the target DNA fragments (Southern et al. 1992). In the second form, Drmanac and co-workers immobilized target DNA on the solid support first, then hybridized with oligonucleotide probes (Drmanac et al. 1989; Drmanac et al. 1993). At the beginning, the microarray was used for quantitative monitoring of gene expression, where thousands of genes were evaluated simultaneously (Schena et al. 1995). Then, different platforms of microarray were developed for mutation and SNP analysis, including direct hybridization

(Chee et al. 2005; Kinoshita 2007), and ligation (Gunderson et al. 1998; Gerry et al. 1999; Broude et al. 2001; Hashimoto et al. 2005).   Inspired by the success of microarray on mutation and SNP analysis, many investigators adapted these microarray platforms to methylation analysis. In general, two major steps were involved in methylation research. The first step is to identify new methylation markers using gene expression (Suzuki et al. 2002; Sato et al. 2003). The second step is to determine which genes among the candidate makers discovered in the first step are silenced by methylation and how they are associated with cancer or other diseases with microarray. Microarray-based methylation profiling techniques have been categorized into three types: bisulfite conversion based method, restriction enzyme cleavage based method, and immunology based method. In this chapter, the several major platforms of microarray will be reviewed for methylation profiling.

2.2 Bisulfite Conversion Based Microarray

2.2.1 Methylation-Specific Oligonucleotides Microarray

In the methylation-specific oligonucleotides (MSO) microarray, methylated and unmethylated probes are designed from target DNA sequences with CpG sites, and immobilized on a solid support. This pair of methylated and unmethylated probes was hybridized with bisulfite PCR products via perfect match to evaluate the methylation status of a DNA sample (Adorjâan et al. 2002; Balog et al. 2002; Gitan et al. 2002). This direct hybridization based microarray potentially allows for rapid screening of the methylation status at numerous CpG sites within one or multiple genes of interest. After the bisulfite treatment and PCR, all Cs (not 5mCs) are replaced by Ts. This conversion

creates a nearly C-less sequence of mostly 3-base-DNA having predominantly A, G, and T (~50%). The reduction in sequence complexity facilitates the non-specific hybridization between probes and targets (Tusnady et al. 2005). The investigator developed strict criteria for probe selection, but the cross-hybridization of imperfect-match between probes and targets was still observed. Mund and co-workers tried 876 probes for the methylation analysis of P16 gene, and only 22 high discriminative probes were found to be good enough for the analysis at last (Mund et al. 2005). The challenge is to select suitable sequences as probes with similar melting temperatures. Furthermore, to improve the specificity of hybridization, two or more CpG sites are included in one probe, which may lack the sensitivity to distinguish partially methylated sequences from fully methylated or unmethylated DNA. These limitations make it impossible to use MSO microarray for high-throughput methylation analysis. Although additional work was done to improve the specificity of hybridization by designing better probes (Kimura et al. 2005; Piotrowski et al. 2006), direct hybridization-based microarray is still not sufficiently robust for methylation analysis.

The use of peptide nucleic acid (PNA) hybridization probes has been claimed to enhance the hybridization stability and sensitivity, but with this modification, the base composition and sequence influence the stability of the duplex (Griffin et al. 1997; Song et al. 2005). The high cost ($600/probe) is also another limitation for using PNA in large panels of microarray.

2.2.2 Ligation Based Microarray

Recently, two ligation-based microarray methods were developed for

26

methylation analysis (Bibikova et al. 2006; Cheng et al. 2006). Both of these two approaches use bisulfite conversion to create the C/T mutation sites for methylation research. The only difference between these two methods is PCR first or ligation first. In the first method, Barany and coworkers used traditional mutation and SNP detection method, PCR/Ligation, for multiplexed profiling of CpG island methylation status (Cheng et al. 2006). Briefly, it first utilizes two-run multiplexed PCR to amplify multiple target DNA sequences, followed by multiplex ligation reactions with two discriminating and one common ligation primers. The discriminating primers contain 5' fluorescent label (Cy5/Cy3) and 3' discriminating nucleotides corresponding to methylation or unmethylation CpG sites. The common ligation primers contain distinct zip-code sequences, which are designed with similar thermodynamic properties and high specificity. These unique zip-code sequences allow for the hybridization to be performed at the same temperature under more stringent conditions to reduce the non-specific interaction (Gerry et al. 1999; Fan et al. 2000; Favis et al.2000). Finally, the ligation products are analyzed using the zip-code universal microarray to determine the methylation status at each target CpG site. Using this assay, 75 CpG sites from 15 tumor suppressor genes are evaluated simultaneously. Furthermore, as the zip-code microarray can be easily expanded, additional CpG sites and genes can be added to the panel.

In the second method (Bibikova et al. 2006), the investigators from Illumina apply their GoldenGate genotyping system to methylation analysis (Fan et al. 2000). This approach first utilizes ligation on a bisulfite-converted template to produce both methylated and unmethylated alleles, followed by multiplexed PCR to amplify the ligated sequences containing the target CpG site. The methylation status at each CpG site is

analyzed using a zip-code microarray as well. In this assay, a gap of 1-20 bases was designed between one pair of ligation primers to help sequence selection and to improve the specificity of primer-template annealing. The elongation with polymerase and ligation with ligase takes place simultaneously. Another advantage of this assay is the improvement of multiplex PCR efficiency. As the pairs of ligation primers have identical universal sequences, the multiplex PCR can be performed with only one pair of primers having universal sequences. This is different from the PCR/Ligation assay, where a number of primers are used in one reaction for amplification and the efficiency of each amplification varies greatly. In PCR amplification based assays, the PCR products are critical to the microarray analysis. Compared with direct hybridization based microarray, ligation based microarray improves specificity, allowing for methylation analysis of cancer cell lines or tumor tissues that contain a high abundance of tumor cells.

2.2.3 Bisulfite Based Genome-Wide Microarray

Due to DNA degradation in bisulfite treatment, PCR products have not been used for genome-wide methylation profiling. Recently, a novel whole-genome methylation profiling method, called BiMP (Bisulfite Methylation Profiling), was developed to amplify 100ng of bisulfite-converted DNA (Reinders et al. 2008). Unlike the standard whole-genome amplification method, BiMP reduces the length of the random primers from 6mers to 4mers to favor the priming of smaller DNA fragments, and to reduce the amplification bias for bisulfite-converted DNA. All of the random primers are labeled with the same sequences, which can be used as universal primers for the PCR amplification. As a result, improved uniformity of amplification of whole-genome was achieved. Finally, the PCR products are hybridized with standard

high-density oligo tiling microarray from Affymetrix. In the future, the detection can be performed on custom arrays with methylation-specific oligos as the design in MSO microarray. This new strategy should improve the sensitivity and specificity of BiMP.

2.3 Restriction Enzyme Based Method

Recently, many variations of restriction enzyme based methods were developed for methylation analysis, in particular for genome-wide methylation profiling (Tompa et al. 2002; Lippman et al. 2004; Lippman et al. 2005; Hatada et al. 2006 ; Khulan et al. 2006; Schumacher et al. 2006; Yuan et al. 2006 ; Shann et al. 2008). Via the cleavage of either methylation-sensitive or methylation-insensitive restriction endonucleases, either methylated and unmethylated alleles can be enriched for large-scale analysis. All of the restriction enzyme based microarray can be assorted into three categories: digested and non-digested samples; digesting with methylation-sensitive and methylation-insensitive enzymes; digesting mutant and wild-type samples (Zilberman and Henikoff et al. 2007).

2.3.1 Digestion and Non-Digestion

As the most common approach, two portions of the same DNA sample are treated with or without restriction enzyme, followed by comparison of these two groups using microarray. Lippman et al. used the restriction enzyme *McrBC*, which prefers methylation DNA as a substrate, to selectively exclude the methylated fractions in whole sequence (Lippman et al. 2004; Lippman et al. 2005; Ordway et al. 2006). Then the enzyme treated DNA and non-treated DNA are gel-purified to keep the fragments bigger than 1kb, which are labeled with Cy5/Cy3, respectively. Labeled two portions are mixed together, and detected on one tiling microarray, which contains contiguous stretches of

chromosomes without bias to evaluate the methylation status of whole sequence simultaneously at high resolution. If the signal intensity from a non-treated sequence is higher than the treated sequence, this sequence targeted region is methylated. All methylated sequences are excluded by *McrBC*, whereas the non-treated sample retains all methylated DNA. In contrast, the unmethylated sequences can be excluded by *TspRI-HpaII-ExoIII* microarray assay, where DNA is digested with *TspRI* and *HpaII*, followed with *ExoIII* (Shann et al. 2008). Correspondingly, another portion of DNA is not treated with *HpaII* as control. These two portions of fragments are labeled with different dyes after the digestion for the microarray detection.

To improve the sensitivity, Yuan et al. used Affymetrix SNP chip combing with methylation-sensitive restriction enzyme *HpaII* for methylation profiling (Yuan et al. 2006). In the Affymetrix technology, PCR amplification with ligation adaptors reduces the required amount of DNA sample. In the assay, the sequences without methylation are cleaved with *HpaII*, and can not be amplified by PCR and can not be observed by the final array detection. The parallel DNA sample with HpaII digestion can serve as a control. With this improvement, only 600ng genomic DNA is required on 50K chips, instead more than 10μg needed for non-PCR amplification approach (Lippman et al. 2005). Again, with PCR amplification, the DNA derived from formalin-fixed, and paraffin-fixed breast cancer tissue has been interrogated (Melnikov et al. 2008). This assay avoids degradation of the DNA sequence in bisulfite treatment for a limited clinical sample.

2.3.2 Methylation-Sensitive and Methylation-Insensitive Enzymes Digestion

As we described in Chapter 1, the methylation-sensitive restriction enzyme

HpaII, and methylation-insensitive isoschizomer *MspI* are most two common enzymes used in methylation analysis. Based on the digestion of these two enzymes, Khulan et al. developed a powerful method called HELP (*HpaII* tiny fragment Enrichment by Ligation-mediated PCR) for methylation profiling (Khulan et al. 2006). The HELP assay is based on the comparison of DNA samples digested by *HpaII* and *MspI* using microarray. Two portions of the DNA sample with high molecular weight are digested with *HpaII* and *MspI*, and ligated with an oligonucleotide pair on the site cleaved by restriction enzyme, respectively. Primed by these ligated oligonucleotides, the digested DNA can be amplified and labeled with different fluorescent dyes. Finally, the *HpaII* and *MspI* representations are hybridized on high density microarray. The ratio between these two representations can reveal the methylation status. The *MspI* representation provides a robust internal control to compare with *HpaII* representation, and detects the copy number of cancer cells, in which amplification and deletions are common. Furthermore, the HELP assay can be used on the other commercial microarray slides with the suitable combination of restriction enzyme isoschizomers for high-throughput genome-wide methylation profiling. Since more than 80% of all CpG islands are not related to regulation of gene expression, Hatada et al. developed more specific Microarray-based Integrated Analysis of Methylation by Isoschizomers (MIAMI) for promoter methylation profiling (Hatada et al. 2006). In this assay, lung cancer cell lines and normal lung cells are subjected to *HpaII* digestion, adaptor-ligation/PCR, followed by MspI digestion, and 2nd PCR and labeling. With microarray detection, the *HpaII* resistance was calculated for both the cancer cell and normal cell lines. Moreover, another portion of lung cancer cell and normal lung cells are subjected to *MspI* digestion, followed by using the same

procedure as the *HpaII* digestion. With the microarray detection, the *MspI* resistance was calculated as well. After analysis of *HpaII* resistance and *MspI* resistance, threshold criteria was made to judge the methylation status of the microarray spots to the corresponding sequence in the promoter. Under this criteria, new epigenetic mutations can be detected in cancer cells.

Instead of using single enzyme for DNA digestion, methylation-sensitive restriction enzyme cocktails were used to improve the resolution of detection (Schumacher et al. 2006). To enrich the unmethylated fraction, several combinations of methylation-sensitive enzymes, *HpaII*, *Hin6I*, *AciI* and *HpyCh4IV* were used to digest the genomic DNA, interrogating more than 41% CpG sites, followed by adaptor-ligation. *McrBC* was used to remove all of the methylated Cytosine in CpG sites. The cleaved methylated sequences can not be amplified or labeled. To enrich the methylated fraction, methylation-insensitive enzyme *TasI* or *Csp6I* was used for the digestion. After the adaptor-ligation, methylation-sensitive cocktails were used to remove the unmethylated CpG sites. Both of these two fractions were detected by microarray. This assay provides a useful tool for both hypomethylation and hypermethylation detection.

2.3.3 Digestion of Mutant and Wild-Type Samples

Comparison between the digestions of mutant and wild-type samples is the most common method in microarray. Yan et al. digested the tissue and wild-type control samples with *MseI*, followed by adaptor-ligation (Yan et al. 2001). Thereafter, the ligation DNA was digested with methylation-sensitive enzymes *BstUI* and *HpaII*. The tissue and wild-type control samples were labeled with different fluorescent dyes after the PCR amplification, and cohybridized on microarray, which is derived from a CpG island

library. Since the *MseI* fragments are not abundant in CpG islands, it was found that both non-CpG and CpG sequences were amplified. To improve the sensitivity and reduce the amplification of non-CpG sequences, a new adaptor-ligation method was used for methylation-specific amplification (Hatada et al. 2002). In this approach, the cancer cell lines and normal control samples were digested with methylation-sensitive enzyme *SmaI* firstly, followed by methylation-insensitive enzyme *XmaI* cleavage. Then, the adaptor-ligation takes place on *XmaI* cleaved sites. Hence, only methylated sequences could be amplified and detected by microarray. In addition, higher percentages of DNA fragments derived from hypermethylated loci were selected from the CpG island library.

For the sample containing large amounts of DNA, Tompa et al. used *MspI* to digest the mutant sample and wild-type control, followed by size fraction to get fragments less than 2.5kb. The fractions of mutant and wild-type control without PCR amplification were labeled with different dyes for the direct microarray detection (Tompa et al. 2002). Recently, two other methylation-sensitive enzymes, *HpyCH4IV* and *HpaII,* were used for methylation profiling based on the same principle (Tran et al. 2005).

2.4 Immunology Based Method

The restriction enzyme based microarray requires high-molecular-weight DNA and is limited by the distribution of the restriction sites of the chosen endonucleases. For example, the *HpaII* sites only comprise 8% of the CpG sites in the human genome. To overcome these limitations, 5-Methylcytosine antibody or 5-Methylcytosine binding domain (MBD), immunoprecipitation and affinity chromatography followed by tiling microarray detection have been used for methylation profiling (Weber et al. 2005; Keshet et al. 2006; Zhang et al. 2006; Hayashi et al. 2007; Penterman et al. 2007; Zilberman et al.

2007). In the tiling microarray, nucleic acid probes can either overlap, lay end-to-end, or be spaced in the target genomic region. There can be classified into two major classes of tiling microarray on the basis of the fabrication of tiling microarray: oligonucleotide tiling microarray; PCR products or bacterial artificial chromosome (BAC) microarray. In the oligonucleotide tiling microarray, relatively short oligonucleotide probes (25-85bp) can be directly synthesized on the surface of the solid support by photolithography (Fodor et al. 1991; Pease et al. 1994), ink-jet device (Hughes et al. 2001; Wolber et al. 2006), or programmable mirrors techniques (Singh-Gasson et al. 1999). This high-density microarray can provide 6.6 million spots in $<2cm^2$. Alternatively, the synthesized oligonucleotide probes can be printed in low-density on the solid support by mechanical printer. In this low-density printing array, 10,000-40,000 spots can be spotted in single chip (1 inch X 3 inch). In the PCR products or BAC tiling microarray, the PCR products or BAC clones are printed on microarray array.

2.4.1 PCR Products and BAC Based Tiling Microarray

BAC tiling microarray has been used in comparative genomic hybridization (CGH) for identification of chromosomal imbalances and variation in DNA copy-number (Albertson et al. 2000; Snijders et al. 2001; Ishkanian et al. 2004). This quantitative and high-throughput technique is now attracting wide-spread interest, especially among methylation researchers. Weber et al. developed methylated DNA immunoprecipitation (MeDIP) combined with BAC tiling microarray for whole-genome as well as promoter-specific methylation analysis, where methylated DNA was captured by 5-methylcytosine antibody, and hybridized to microarray to achieve a more unbiased analysis (Weber et al. 2005). In this assay, sheared genomic DNA between 300-1,000 bp

was incubated with the antibody against 5-methylcytosine. Then the captured methylated DNA fragments were isolated by immunoprecipitation using Dynabeads with antibody. The methylated DNA and input DNA (total DNA) were labeled with different fluorescent dyes for microarray hybridization. In the data analysis, the signal ratio of methylated DNA and input DNA could reveal the methylation status. In the BAC microarray fabrication, 32,433 BAC clones with 80Kb resolution in genome-wide, and 12,192 BAC clones derived from CpG island library were prepared for the slides printing. In this assay, the resolution and coverage of CpG sites are mainly determined on the microarray BAC fragments distribution, unlike the restriction sites of endonuclease.

As an alternative, PCR products can be used for tiling microarray printing (Rinn et al. 2003; Odom et al. 2004). Combined with MeDIP, they were also used in genome-wide methylation analysis (Keshet et al. 2006). In the microarray fabrication, 13,000 individual PCR amplifications derived from promoters were printed on array slides one by one. As in the case of the BAC tiling array, the input DNA and methylated DNA from immunoprecipitation were labeled with different dyes for microarray detection. After stringent analysis, the MeDIP based method can recognize the methylation region with the CpG density of larger than 2% and is ideally suitable for CpG island methylation detection. One has used this assay for colon cancer cell lines Caco-2 methylation analysis and a set of 135 gene promoters were found methylated, 127 of which were contained in CpG islands.

The problem with BAC and PCR tiling microarrays is that they are labor intensive. For example, in the tiling microarray fabrication for human chromosome 22, more than 20 000 PCR reactions must be designed to achieve a 1-Kb resolution (Rinn et

al. 2003). If the entire human genome tiling array was made in this way, more than two million PCR reactions are required (Royce et al. 2005). Hence, researchers generally adopted oligonucleotide based tiling microarray for methylation research.

2.4.2 Oligonucleotide Based Tiling Microarray

Tiling microarray can be classified into two groups on the basis of the length of oligonucleotides: short oligonucleotides (25-mer) from Affymetrix and long oligonucleotides (45-mer to 85-mer) from NimbleGen and Agilent. In the Affymetrix platform, each perfect-match oligonucleotide probe, alone with a single mismatch probe in the central base (position 13 of 25), is synthesized on glass slides using the photolithographic technology. The mismatch probe will provide a measurement of nonspecific target DNA binding to the perfect match. These short oligonucleotides improve the specificity, but suffer from decreased sensitivity, probe variability, and noise. To avoid these drawbacks, several stringent statistical methods have been developed for the data analysis, including the two-state hidden Markov model based on probe-level t-statistics (Carroll et al. 2005; Ji and Wong et al. 2005; Li et al. 2005), non-parametric Wilcoxon rank sum test (Cawley et al. 2004; Bernstein et al. 2005), and non-parametric Kolmogorov-Smirnov test (Zhang et al. 2006). In the long oligonucleotide microarray platform, only perfect match probes to genomic sequence are built on the microarray by photolithographic technology (NimbleGen), or ink-jet device (Agilent). These longer probes result in reducing the microarray density, but provide a better balance of specificity, sensitivity, and non-specific binding than the Affymetrix shorter platform (Kreil et al. 2006). Furthermore, the data derived from these long oligonucleotide platforms requires less statistical analysis.

36

With the Affymetrix platform, Zhang et al. reported the first comprehensive, high-resolution DNA methylation mapping of the Arabidopsis thaliana genome (Zhang et al. 2006). This whole genome tiling array contains ~6.4 million of 25-mer oligonucleotide probes for each 35 bp genomic region, which covers ~97% of one strand from five chromosomes (~120Mb). In the sample preparation, methylated DNA and unmethylated DNA were fractionated by MeDIP using 5-methylcytosine antibody, or by affinity chromatography using 5-Methylcytosine binding domain (MBD). Then, the fractions were subject to whole genome amplification and labeling with different dyes for microarray detection. The methylation patterns revealed in this method are consistent with previously individual loci analysis, and provide an invaluable reference for research in the individual loci from the genome in the future (Schöb and Grossniklaus 2006). Hiroshi et al. developed the tiling array covering ~30 Mb of human genome (1%), using Affymetrix platform (Hayashi et al. 2007). Combined with MeDIP, this unbiased, scalable, and detailed mapping technique was used for human colorectal cancer cell methylaiton profiling. As the tiling microarray for the whole human genome is being developed by Affymetrix, we can expect this will lead to a significant impact on human genome-wide methylation research.

The group led by Henikoff successfully adopted the NimbleGen microarray from gene expression to plant methylation research (Mito et al. 2005; Mito et al. 2007; Penterman et al. 2007; Zilberman et al. 2007). Due to the large amount of DNA isolated from plant, the fractions of input and methylated DNA derived from MeDIP were labeled with different dyes and cohybridized on the tiling array without PCR. Using raw DNA can result in good results without the complicated statistical manipulation. Moreover, the

custom tiling arrays built by NimbleGen allows for the more flexible experimental design.

Again, in human methylation research, this flexible tiling array system has shown great power recently. Weber et al. migrated from BAC tiling microarray to NimbleGen oligonucleotide microarray to study the impact of methylation on gene expression in human genome (Weber et al. 2007). In this study, ~16,000 high-confidence promoters in human primary somatic and germline cells and each with 15 oligonucleotides were analyzed using MeDIP combined with tiling microarray. By mapping DNA methylation within the major human promoters, researchers that show the CpG island promoters with high ratio of CpG sites are mostly unmethylated, whereas the CpG-poor promoters are mostly hypermethylated in somatic cells. This comprehensive promoter methylation map provides intriguing insights into genome evolution, cellular differentiation, and their relation to tumorigenesis (Zilberman 2007). Later, Rauch et al. described another application of NimbleGen and Agilent platforms to high-resolution profiling of hypermethylation and hypomethylation in lung cancer (Rauch et al. 2007; Rauch et al. 2008). Briefly, methylated fractions were enriched by MBD proteins from lung squamous cell carcinomas and normal tissues extracted DNA, respectively. The fractions were cleaved by *MseI* digestion, followed by adaptor-ligation. After PCR amplification, these two fractions were labeled with different dyes for microarray cohybridization.

2.5 Conclusion

In this chapter, major microarray platforms can be grouped into three groups: bisulfite conversion, restriction enzyme, and immunology. Due to the significant losses

and cleavage of template DNA in bisulfite conversion, bisulfite based microarray is appropriate for the analysis for individual loci, specific genes, or specific sites. On the other hand, the conversion creates a nearly C-less sequence of mostly 3-base-DNA having predominantly A, G, and T. The reduction in sequence complexity facilitates the non-specific binding between bisulfite-treated template and probe (primer). It is difficult to maintain the specificity and sensitivity in genome-wide analysis. As alternatives, restriction enzyme and immunology based methods are capable for genome-wide methylation analysis, combined with high-resolution tiling microarray. The high-molecular-weight requirement and a limited number of recognizing sites for chosen endonucleases are the two drawbacks of restriction enzymes based microarray. To avoid these two limitations, 5-methylcytosine and MBD proteins based methods are developed for methylation analysis in genome-wide and promoters. All of the microarray platforms for methylation research are very young technologies at present. Undoubtedly, with development and refinement, these platforms hold great promises to unlock the relationship between cancer and methylation, as well as to provide valuable prognostic and diagnostic tools.

CHAPTER III

MICROARRAY FABRICATION

This chapter discusses how we fabricate home-made microarray. The commercial oligonucleotide microarrays manufactured by Affymetrix, Illumina, NimbleGen, and Agilent have facilitated research in gene expression, mutation detection, and genotyping. However, the cost and complexity of these platforms make them less adaptable by many researchers, who prefer printed microarray, which is more flexible in design and content, and potential cost saving over commercial arrays. Due to the technical variability in printed microarray, significant differences exist in microarray fabrication and post manipulation. In order to produce high quality microarray data, the following parameters have been extensively studied: solid supports selection, oligonucleotide modification, printing procedure optimization, fluorescent dyes selection, hybridization procedure optimization, and highly discriminative zip-code sequences selection.

3.1 Solid Support Selection and Spotting Optimization

3.1.1 Solid Support Selection

To develop reliable oligonucleotide microarrays, the immobilization of probes on solid support is one of the most essential and important subjects. A number of noncovalent and covalent methods can be used to immobilize oligonucleotides on solid support, which, in general, has two major forms: nylon membrane and glass slide. Due to the high fluorescent background, chemiluminescence is the most common detection method for nylon membrane. As the two-color or four-color fluorescent detection system is broadly used in microarray, glass slides with different functional surfaces are more suitable for fluorescence-based microarray. Glass as a support material has many advantages over other support: its non-porous structure requires low reaction volumes; it is durable to chemicals and high temperatures; low-background is very well suited for the fluorescence-based detection. The glass slides are commonly modified with several functional groups such as carboxyl (Lindroos et al. 2001), thiol (Rogers et al. 1999), poly-L-lysine (Schena et al. 1995), amine (Pack et al. 2007), aldehyde (Dawson et al. 2005), and epoxy (Wrobel et al. 2003). In the noncovalent immobilization, the DNA can be fixed on the glass surface by acid-base interaction, hydrogen bonding, and electrostatic interactions. In contrast, the terminal covalent attachment allows the entire oligonucleotide to be available for hybridization and to withstand the high temperature and salt concentration often required during the stringent washing conditions in subsequent steps of microarray processing. Hence, to yield reproducible results, covalent immobilization was used in our platform development. Direct covalent coupling is most

41

commonly achieved using amino-terminated oligonucleotides and either aldehyde or epoxy modification glass. For these reasons, this study focused on the immobilization of amino-terminated oligonucleotides to aldehyde and epoxy modified glass slides.

First, epoxy or aldehyde slides (CEL Associates, Pearland, TX) were printed, blocked, and washed under the same conditions. Then, mini-sequencing products were hybridized with the immobilized oligonucleotides under the same condition. In Fig. 3, it is seen that the background of epoxy is much stronger than aldehyde slides, although there was no significant difference between the strength of signals. The reason for the high background is the different property of these two modifications. The aldehyde surface is hydrophilic, whereas the epoxy surface is hydrophobic. The fluorescent dyes, modified from ddNTP or dNTP, mostly are hydrophobic. After the hybridization, it is difficult to clean the surface. Hence, we preferred aldehyde modification to epoxy modification. The aldehyde slides made by different companies such as Erie Scientific, CEL, Corning, and TeleChem were tested. Finally, CEL aldehyde slides were chosen in this thesis work for their high binding capacity, low background, and cost-saving features.

3.1.2 Optimization of Spotting, Immobilization, and Wash Conditions

Several recent studies that investigated immobilization of short NH_2–oligonucleotides on aldehyde–glass employed a number of vastly different protocols for spotting and processing of microarray slides (Lindroos et al. 2001; Booth et al. 2003; Al-Khaldi et al. 2004; Hou et al. 2004; Deyong 2005). In these protocols, different immobilized solutions are used, such as DMSO, SSC, alkaline phosphate,
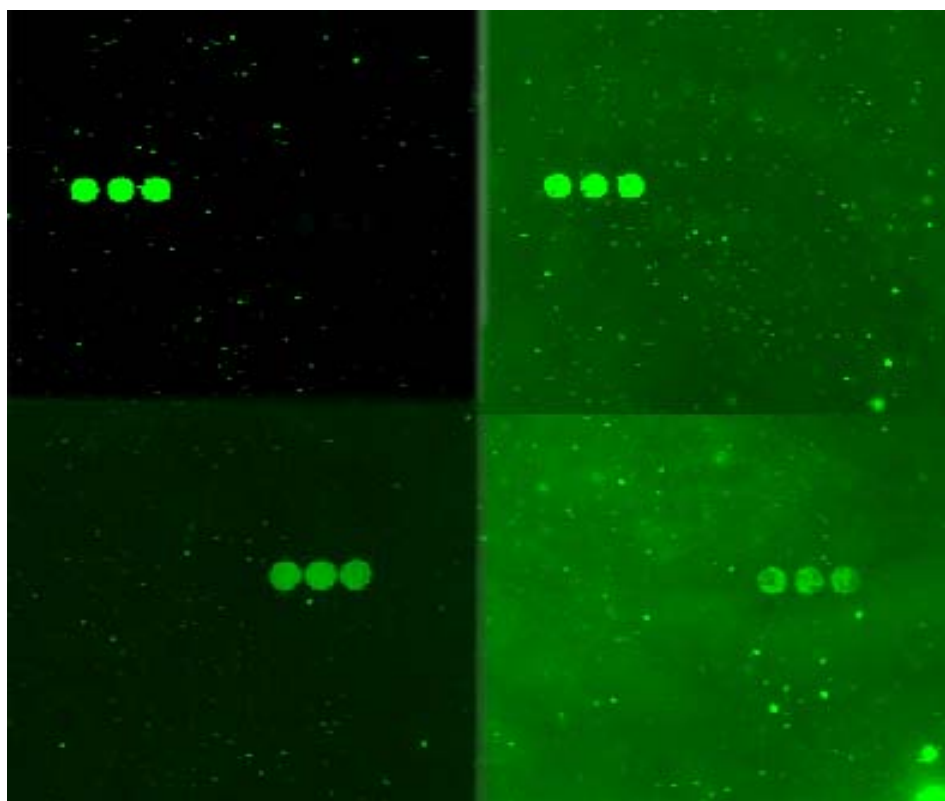
A   **B**

Fig. 3 Modified glass selection. (A)Aldehyde modification glass support and (B) epoxy modification glass support were immobilized, blocked, and washed under the same condition and hybridized with the same SBE products.

sodium carbonate, and some commercial solutions. Furthermore, different temperature (from room temperature to 75 °C), humidity (from <30% to 100%), and time (from 15min to 48h) were used in the printing. These examples demonstrate a clear need to investigate a number of variables in oligonucleotide microarray fabrication and processing procedures, as it is impossible to compare the results of similar experiments conducted under such vastly different conditions.

Laura R. Kuck and co-worker tested and compared different types of spotting buffer and chose 3XSSC/0.05% SDS/0.001% Chaps as the spotting buffer, which was also used in our studies (Dawson et al. 2005). After printing, glass slides were kept for 24 hours at room temperature in the hood (humidity <30%), a sealed container with saturated NaCl solution, and a sealed container with water respectively. All of these slides were blocked, washed, and hybridized under the same condition. We found that under the low humidity the printing spots were not uniform. And the spots formed 'donut' circles without signals in the center. When we increased the humidity using saturated NaCl solution and pure water, the donut disappeared gradually (Fig. 4 A). Hence, saturated humidity must be kept in the post-spotting process to improve the reaction between oligonucleotides and the modified surface. To optimize the post-spotting temperature, we studied the slide performance at room temperature, 37°C, and 55 °C for 12 hours, and 80 °C for 1 hour, respectively. We didn't find any significant difference except for the higher temperature inducing high fluorescent backgrounds. The reason for this could be that after the incubation at high temperature, the salt and oligonucleotides changed the surface of spot area, which was displayed as a fluorescent signal under high sensitivity scanning. We did the blank test at room temperature and 80 °C, respectively. The later condition showed
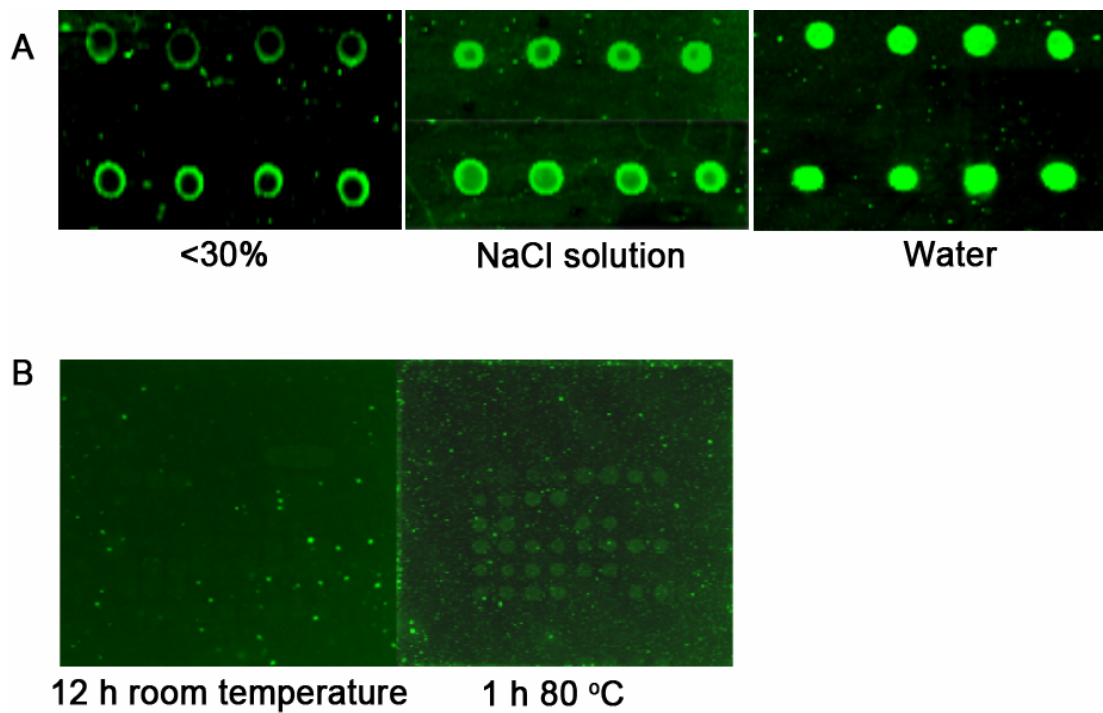
Fig. 4 Post-printing humidity and temperature selection. (A) Humidity Test: less than 30% humidity at room temperature (left); the humidity of saturated NaCl solution at room temperature (middle); the humidity of water at room temperature (right). (B) Blank Test: printed slides were left at room temperature for 12 h (left); printed slides were baked at 80 °C for 1 h (right).

high background and produce the false-positive signal in the test (Fig. 4 B). Hence, we performed the post-spotting at room temperature for 12-48 hours. In the blocking and washing step, the excess of aldehyde-reactive groups were deactivated by keeping the slides for 25 min at room temperature in Pronto™ pre-soak solution (Corning, NY), followed by washing with buffer I and II of Pronto™ Universal Microarray Validation kit (Corning, NY).

3.2 Oligonucleotide Modification Optimization

In this study, we attached an amino group to all of the oligonucleotides. The 3' terminal or 5' terminal could be chosen as modification site. To reduce steric interference from the solid surface that affects the hybridization between probes and targets, the immobilized oligonucleotides must be lifted up from the surface. When the oligonucleotides were synthesized, C3, C5, C6 or C12 was linked to amino to increase the space (Gerry et al. 1999; Gitan et al. 2002; Moreno et al.2006). C6 and C12 linkers are the most common in the microarray. In order to increase the space, a spacer consisting of a poly thymine chain up to 15 bases in length was added in the synthesis (Lindroos et al. 2003; Hultin et al. 2005; Pourmand et al. 2007). In our study, C3, C6, C12 linkers and 15T spacer were tested. We chose one oligonucleotide, which was synthesized with different carbon linkers to hybridize with the same PCR product. The strongest signal was C3 amino linker, and the weakest signal was C12 amino linker (Fig. 5). Five different oligonucleotides were tested by this way, and the same result was seen. The reason for this gradual change is probably the hydrophobic property of C12 and the hydrophilic property of aldehyde surface. Another reason is the solubility of C12 linker in the spotting buffer. The C12 linked to amino makes it difficult to dissolve amino in the

46
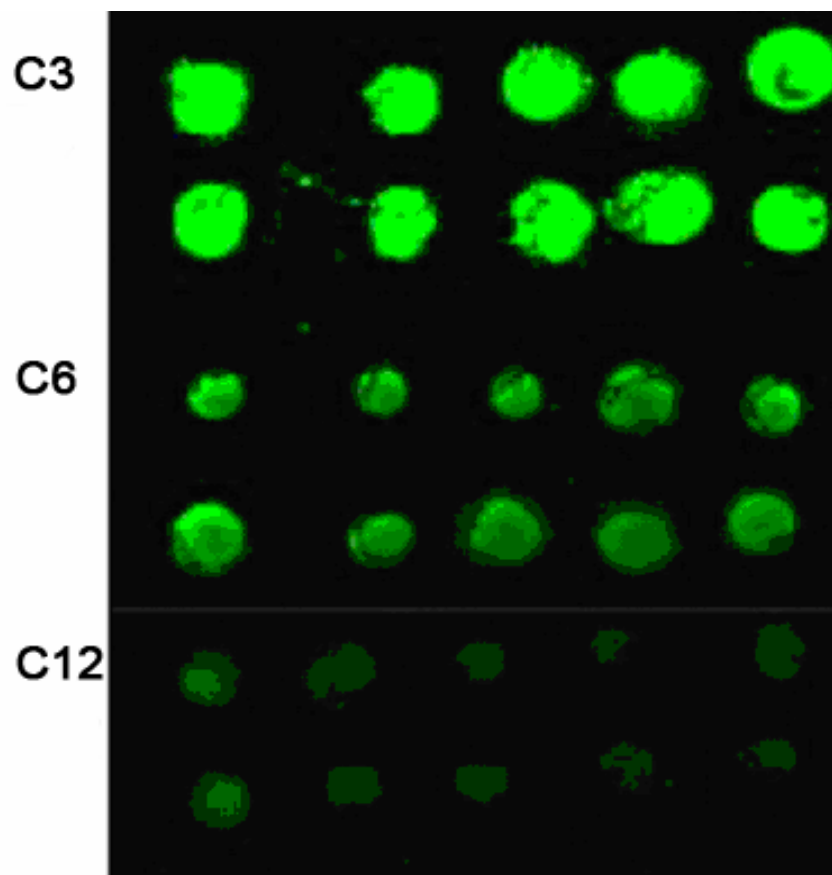
Fig. 5 Carbon linker selection. C3, C6, and C12 linkers comparison: the same oligonucleotide with different carbon linkers were synthesized in the amino modification. Each oligonucleotide was printed 10 times on the same aldehyde slide, then hybridized with the same PCR product.

spotting buffer. As a result, the amino and aldehyde reactive group on the slides were not fully contacted. This decreased the immobilized efficiency of an amino modified oligonucleotide on aldehyde slides.

15T spacer was also tested with 6 oligonucleotides, three of which included 15T spacer and three of did not. No significant difference was found by us. Again, in our SBE-TAGs and CHZMA platforms, the hybridization efficiency was not influenced by 15T spacers. Hence, in the universal zip-code microarray, oligonucleotides were synthesized with a 3C linker and without a 15T spacer.

3.3 Fluorescent Dyes Selection

As an alternative to radioisotopic labeling, fluorescent labeling enables the signal detection with a high resolution in microarray. On the other hand, the combination of multiple fluorescent dyes used in a single reaction became available for the sequence variation changes detection such as SNP and methylation. Two kinds of labeling, direct and indirect, are used in the microarray system. In the direct labeling, fluorescence modified oligonucleotides, ddNTPs, and dNTPs can be incorporated into DNA strands by enzymatic reactions, such as PCR, ligation, SBE (Gitan et al. 2002; Lindroos et al. 2002; Bibikova et al. 2006; Cheng et al. 2006). In the indirect labeling, biotin or aminoallyl modified oligonucleotides, ddNTPs, and dNTPs can be incorporated into DNA strands by enzymatic reaction, followed by conjugating with fluorescent dyes using chemical or enzymatic reactions (Hoen et al. 2003; Zhang et al. 2006). Generally, four-color fluorescent labeling is used in the SNP and mutation detection and two-color and single-color fluorescent labeling are used in the methylation detection.

Due to the fact that different dyes have considerable differences in their photostabilities and fluorescence efficiencies, fluorescein, Texas-red, Cy3, and Cy5 were tested by us on a laser-induced fluorescence system, Typhoon 9410, which has four lasers including: 457nm, 488nm, 532nm, and 632nm. All of these four dyes are equally separated on the major fluorescent area from 400nm to 750nm with minimal overlap (Fig. 6). Using single-base extension, we tested several combinations of these four dyes for methylation analysis. The combinations are listed in Table I. After the comparison, Combination 3 is determined to be the best candidate for our methylation analysis. As an addition, combination 5 is another good candidate except for the low fluorescent efficiency of fluorescein.

The problems of Combination 1 are that fluorescein has low excitation efficiency (83000), and Texas-red does not have a matched laser for excitation. In Typhoon 9410, Texas-red is excited by a 532nm laser, which is far from the optimum absorption (583nm) of Texas-red. With the same problem of Texas-Red, Combination 4 is not a good candidate either.

We next compared Combination 2 with Combination 3, where unmethylated PCR product was used as template for SBE in the presence of Cy5 ddATP and Cy3 ddGTP (Combination 2) or Cy5 ddGTP and Cy3 ddATP (Combination 3). In the SBE reaction, only Cy5 ddATP is incorporated into extended products with the unmethylated template in the case of Combination 2. But as shown in Fig. 7, both Cy3 and Cy5 channels show signals after we scanned array slide with 583nm and 632nm lasers, respectively. The reason for this is that Cy5 ddATP can be excited by both 583nm and 632nm lasers, but the signal from 583nm laser is corresponding to the methylation

49
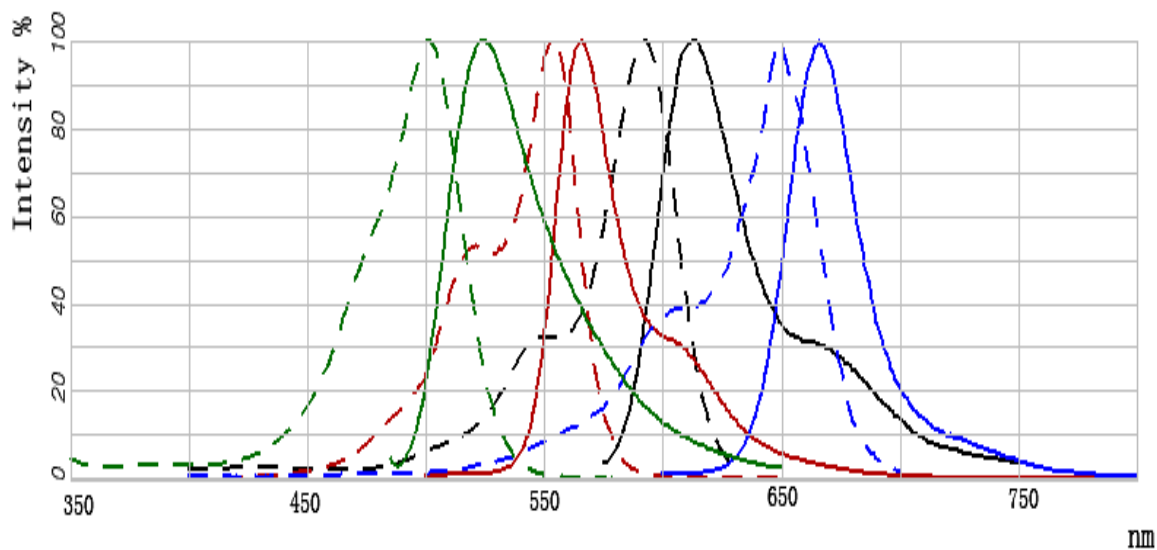
Fig. 6 Fluorescence spectra of selected dyes. Absorption and emission of fluorescence Spectra: broken line is for the absorption peak and solid line is for the emission peak; Green, red, black, and blue are corresponding to fluorescein, Texas-red, Cy3, and Cy5, respectively.

Table I. Different combinations of ddATP and ddGTP were studied in the SBE reaction.

	Combination 1	Combination 2	Combination 3	Combination 4	Combination 5
ddATP	Texas-red	Cy5	Cy3	Texas-red	Cy5
ddGTP	Fluorescein	Cy3	Cy5	Cy5	Fluorescein
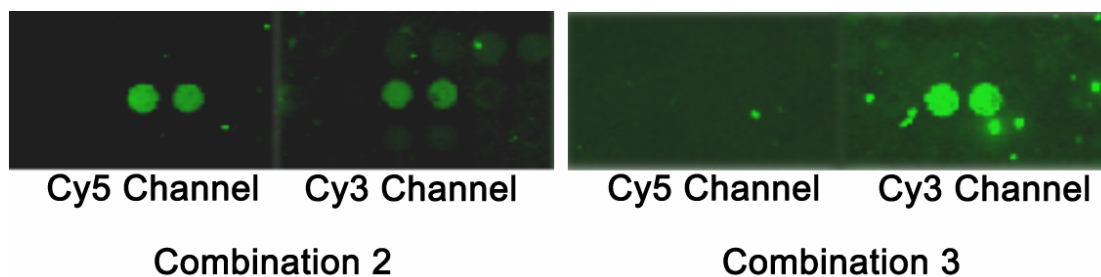
Fig. 7 ddATP and ddGTP fluorescent labeling selection. Combination 2 and 3 were used for the SBE reaction using unmethylated PCR product as the template. After hybridization, microarray slides were scanned in Cy5 (632nm) and Cy3 (583nm) channels, respectively.

detection. This can lead to false-positives for the methylation evaluation. Hence, Combination 2 can not be selected for the SBE reaction. In contrast, using Combination 3 for the SBE reaction with unmethylated template, only Cy3 channel has signal. No signal was found in Cy5 channel because Cy3 ddATP can not be excited by the 632nm laser.

3.4 Hybridization Optimization

Hybridization between single-stranded probes and target molecules occur through hydrogen bonds formation between the bases of complementary nucleic acid sequences. Sequence composition, target and probe length, hybridization temperature, secondary structure, degree of homology, salt concentration, pH, and a number of other factors influence the hybridization efficiency and the strength of the double helix structure. LifterSlip system (Erie scientific, NH) and ProPlate™ Multi-Array Slide System (Grace Bio labs, OR) were compared for the hybridization. The LifterSlip system uses 15~23μL hybridization volume and 3 areas in a slide. The ProPlate™ Multi-Array Slide System uses 40~200μL hybridization volume and 16 areas in a slide. Under the same condition, High background was detected in the LifterSlip system (Fig. 8 A). After 8~12 hours hybridization, unknown substances bond to on the surface of the array slide. Hence, we used ProPlate™ Multi-Array Slide System for the hybridization of our home-made microarray.

We next studied the hybridization buffers. SSC series buffer, SlideHyb buffer (Ambion, TX), Pronto™ short oligo hybridization buffer (Corning, NY), and PCR buffer (GE healthcare, NJ) were tested. High background and poor signal were detected in the first two hybridization buffers. Low background and high discrimination were seen by

A
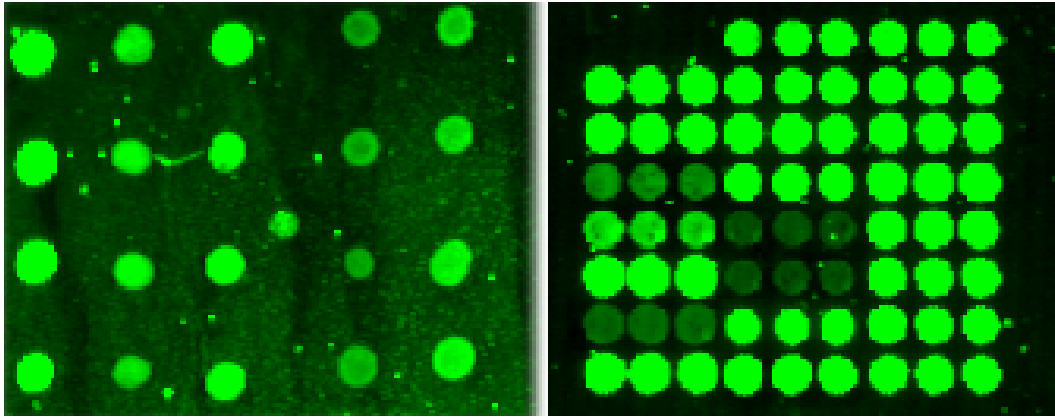
B
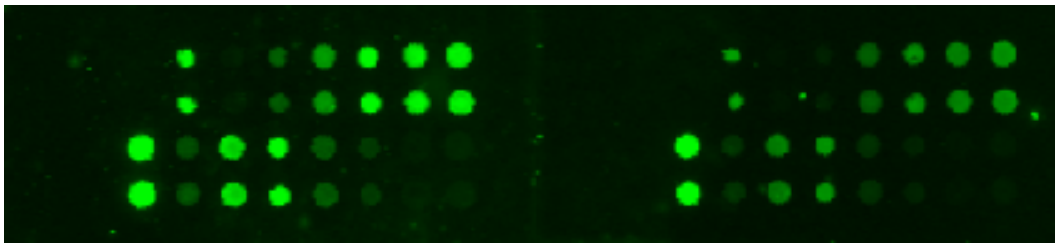
Fig. 8 Microarray hybridization condition optimization. (A)LifterSlip hybridization system and ProPlate™ Multi-Array Slide System comparison: the left is LifterSlip system with higher background and the right is ProPlate™ Multi-Array Slide System. (B) Coring hybridization buffer (right) and GE healthcare PCR buffer (left) comparison.

using Coring buffer. Low background and strong signal were detected by using PCR buffer. As seen from Fig. 8 B, the signal from the PCR buffer is stronger than that from Coring buffer, but with more crosslink and nonspecific binding between different sequences. Hence we selected Pronto™ short oligo hybridization buffer as our hybridization buffer.

In the universal zip-code microarray, all of the zip-code sequences immobilized on the slides are 23 bases, with the melting point of 62~66°C. In order to obtain the optimal hybridization temperature to achieve the sufficient specificity, we tested the hybridization from 38 ^{o}C to 63 ^{o}C. Finally, 54 ^{o}C was selected for this group of zip-code probes.

3.5 Highly Discriminative Zip-Code Sequences Selection

A highly discriminative and specific sequence is the critical factor for the microarray. To design the zip-code sequences, all possible 20~24 mers were subject to a computational screen to generate a set of sequences with similar GC content, thermodynamic properties, and minimal cross-hybridization (Gerry et al. 1999; Shoemaker et al. 1996; Giaever et al. 1999; Winzeler et al. 1999; Favis et al. 2000). However, individual oligonucleotides still differ in their performance. It is not possible currently to predict all of the performance parameters of particular probes in microarray experiments simply by theoretical calculations (Pozhitkov et al. 2006). In our low density tag microarray, we must ensure that the methylation status of each gene or each specific region. As a result, stronger and more specific interactions between the zip-code sequences and complement zip-code sequences are required.

The strategy to select highly discriminative sequences is illustrated in Fig. 9. In the selection, all of the candidate zip-code oligonucleotides with 3C amino linker at 3' terminal (MWG-Biotech Inc., AL) were printed on aldehyde modified microarray slides (CEL, TX) with XactII Compact Microarrayer (LabNEXT Inc., IL). In the SBE reaction, a chemically synthesized single-strand oligonucleotide was used as the template for all of the candidate sequences. All SBE primers are designed with the identical gene-specific sequence for the chemically synthesized template and different candidate zip-code sequences. All of the SBE for different sequences were performed in individual tubes and the hybridization took place in a different subarray area. Typically, the selection criteria is that the signal from a perfect match spot is at least 10 folds higher than the signal from a mismatch spot. We were able to select 32 discriminative and specific zip-code sequences from 70 candidates for the fabrication of a universal tag microarray. The sequences of 32 zip-code are listed in the Appendices.

3.6 Conclusion

In this chapter, we describe the study of the fabrication of our home-made microarray. It generates the standard procedure for our future research. First, the highly discriminative tag probes (MWG-Biotech Inc., AL) are synthesized with 3C amino modification at 3' terminal without poly-T spacer. Second, the probes are dissolved in the printing buffer of 3XSSC/0.05% SDS/0.001% Chaps, and printed on aldehyde modified microarray slides (CEL, TX) with XactII Compact Microarrayer (LabNEXT Inc., IL). After spotting, the printed slides were incubated in a container with the saturated humidity of water at room temperature for 12-48 h. Then, the printed slides are placed in
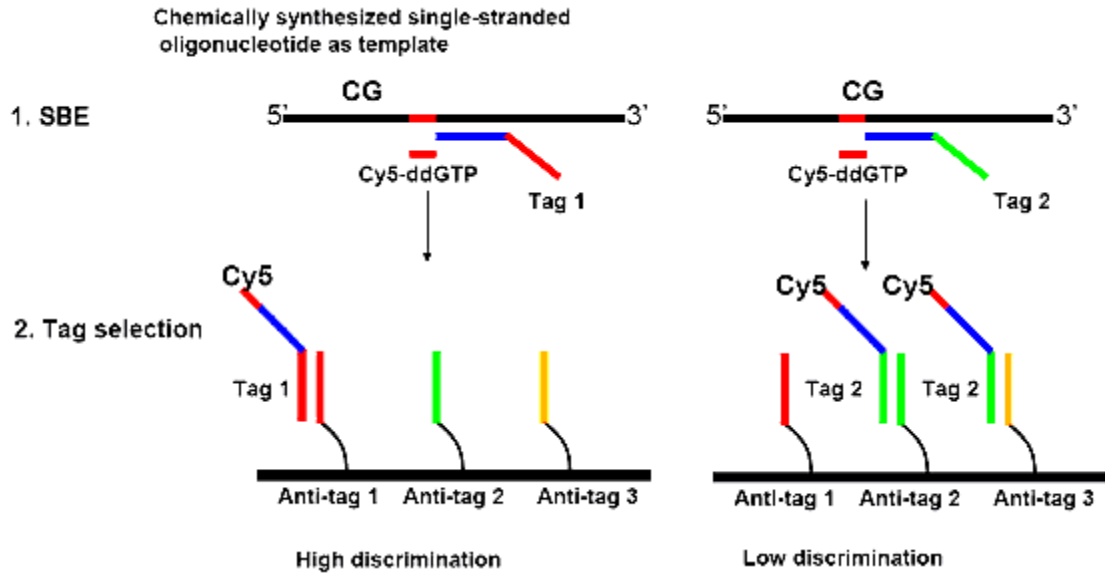
Fig. 9 Schematic outline for highly discriminative tag selection. A chemically synthesized single-strand oligonucleotide is used as the template for SBE reaction of tag selection. The SBE primers are designed with the identical gene-specific sequence (highlighted in blue) and different tag sequences including: tag 1, tag 2, and tag 3. In the SBE, all tag primers are added to different tubes for the cyclic reaction separately. Correspondingly, the hybridization is performed in different subarray area for each tag primer separately. For example, if the SBE primer with tag 1 (highlighted in red) is detected at the position of anti-tag 1 on the microarray, tag 1 is thus highly discriminative. Contrarily, in the tag 2 selection, if two positions of anti-tag 2 and anti-tag 3 are all detected with positive signal, then tag 2 has low discrimination.

the desiccator for future use. Before the hybridization, the excess of aldehyde-reactive groups are deactivated by placing the slides in Pronto™ pre-soak solution (Corning, NY) for 25 min at room temperature, followed by washing with buffer I and II in Pronto™ Universal Microarray Validation kit (Corning, NY). Third, in the two-color microarray system, Cy5 and Cy3 can be used for the labeling of methylation and unmethylation signal, respectively. Fourth, the hybridization of microarray slides can be processed on ProPlate™ Multi-Array Slide System (Grace Bio labs, OR) using Pronto™ short oligo hybridization buffer (Corning, NY) at 54 ^{o}C for 3-12 h. Fifth, the array slides are rinsed with SSC buffer and ddH$_2$O, and scanned in a laser-induced fluorescence system, Typhoon 9410. And the data is analyzed by ImageQuant or ArrayVision.

CHAPTER IV

MULTIPLEX METHYLATION ANALYSIS BY SBE-TAGS

MICROARRAY

4.1 Abstract

DNA methylation analysis plays a pivotal role in unlocking association of epigenetic events with cancer. However, simultaneous evaluation of the methylation status of multiple genes is still a technical challenge. In this work, we describe the use of SBE-TAGs microarray for the simultaneous determination of methylation status of multiple genes. 10 CpG sites of 9 tumor suppressor genes (MGMT, GATA4, HLTF, SOCS1, p16, RASSF2, CHFR, TPEF, and Reprimo) were selected for this study. We found that SBE-TAGs microarray performed quite well in multiplex methylation analysis if a standard calibration curve method was used. In addition, SBE-TAGs is robust and reproducible. Hence, this work suggests that SBE-TAGs microarray can be a high-throughput method for multiplex methylation analysis.

4.2 Introduction

DNA methylation is a key event regulating gene expression. In vertebrates, the majority of 5'-methylcytosine occur at CpG sites within the promoter, often resulting in transcriptional inactivation of genes (Huang et al. 1999; Jones and Laird 1999; Toyota et al. 1999; Bird 2002). The hypermethylation of the promoter regions of some genes plays a pivotal role in the governance of normal and disease development (Baylin et al. 1998). It was found that in cancer, promoter hypermethylation is one of the most common mechanisms for the loss of function of tumor suppressor genes. In virtually every type of human neoplasm, promoter methylation is found to be associated with the inappropriate transcriptional silencing of genes (Baylin and Herman 2000; Warnecke and Bestor 2000). Because of the enormous potential of epigenetic markers in both cancer screening and treatment, considerate effort was made to search for methylated genes that can be novel therapeutic targets or screening biomarkers (Hayatsu et al. 1970; Frommer et al. 1992; Herman et al. 1996; Sadri et al. 1996; Gonzalgo and Jones 1997; Xiong and Laird 1997; Eads et al. 2000; Cottrell and Laird 2003; Tost et al. 2003; Zeschnigk et al. 2004).

Most of existing methods for methylation analysis are based on bisulfite treatment that converts cytosine to uracil whereas the methylated cytosine residues are unaltered. The treated DNA is then PCR amplified to yield fragments wherein Ts replace Us (former Cs), and Cs replace the methylated Cs. Thereafter, the methylation status of the target DNA is examined using conventional DNA analysis methods. Currently, methylation-specific PCR (MSP) is the method of choice for methylation analysis (Herman et al. 1996; Cottrell and Laird 2003), which utilizes methylation-specific primers to discriminate methylated DNA from unmethylated DNA in PCR. MSP is

sensitive and can detect one methylated allele in a large background of unmethylated alleles. Its real time PCR variation can also quantify the abundance of methylated DNA (Zeschnigk et al. 2004). However, its problem is low throughput. In general, one MSP reaction reveals methylation of a gene at a time.

Microarray is a method for multiplex analysis. Attempt has been made to use microarray for high-throughput methylation analysis (Gitan et al. 2002; Kimura et al. 2005; Bibikova et al. 2006; Cheng et al. 2006; Piotrowski et al. 2006). The conventional application of microarray to methylation analysis involved direct hybridization, utilizing oligonucleotide probes designed to target either the methylated or unmethylated alleles within target sequences (Gitan et al. 2002). However, its poor specificity makes it impossible for it to be used in routine methylation analysis. It is not totally surprised to see that it is a challenge to apply microarray to methylation analysis. This is because bisulfite treatment eliminates all Cs in unmethylated DNA and in non-CpG sites of methylated DNA and thus creates a nearly all C-less sequence of most 3-base-DNA having predominantly A, G, and T. This reduction in sequences, relative to normal 4-base-DNA, makes hybridization less specific. Because of this problem, each of the conventional genetic analysis methods must be carefully examined to determine if it is suitable for methylation analysis.

Recently, we conducted a systematic study to examine the use of microarray methods for methylation analysis and one of the methods that we tested was SBE-TAGs microarray, which was originally developed for SNP analysis (Pastinen et al. 1997; Hirschhorn et al. 2000; Lindroos et al. 2002; Lovmar et al. 2003; Favis et al. 2004; Fan et al. 2006). We found that SBE-TAGs microarray performed quite well in methylation

analysis and could reveal the methylation status of the individual CpG sites, thus the methylation status of the corresponding genes. Importantly, SBE-TAGs microarray allows for the simultaneous determination of the methylation status of multiple genes. In addition, SBE-TAGs microarray is robust and reproducible. Clearly, this study established that SBE-TAGs microarray is an effective microarray method for multiplex methylation analysis.

4.3 Materials and Methods

4.3.1 Samples and DNA Isolation

Unmethylated control genomic DNA and methylated control genomic DNA were purchased with CpGenome^{TM} DNA modification Kit (Millipore, Billerica, MA). The mixture samples containing both methylated and unmethylated DNA were created by spiking methylated DNA control genomic DNA into unmethylated control genomic DNA at a given ratio. The colorectal carcinoma cell lines HCT116 and HTB38 were cultured under standard conditions and genomic DNA were extracted from the cells using QIAamp DNA mini kits (QIAGEN, Valencia, CA) according to the manufacturer's protocol.

4.3.2 Bisulfite Treatment

The CpGenome^{TM} DNA modification Kit (Millipore, Billerica, MA) was used for bisulfite treatment in this study. Bisufite treatment was performed using the protocol suggested by the manufacturer. The only exception was the desalting and desulfonation step. Briefly, after bisulfite treatment for 15 h at 50°C, a 50K Microcon centrifuge column (Millipore, Billerica, MA) was used in desalting, followed by washing the column 3

times with 500μL ddH$_2$O on Eppendorf Centrifuge 5415C (Eppendorf North America, Westbury, NY) at 9000RPM . In the desulfonation step, we transferred 500μL of 0.1MNaOH solution to the column, and then spun it until the solution went through at 3000RPM. The column was then washed twice with 500μL of ddH$_2$O and 1 time TE buffer. Finally, the modified DNA was recovered from the column. The converted DNA was used immediately or stored at -20°C for further analysis.

4.3.3 Multiplex PCR

The first PCR contained 2μL of bisulfite-converted DNA, 300-400nM of each primer, 1× AmpliTaq Gold PCR buffer, 400 μM of each dNTP, 4 mM MgCl$_2$, and 2.5 U AmpliTaq Gold polymerase (Applied Biosystems, Foster City, CA). The first run of PCR was performed at 95°C for 10min, then 42 cycles of 95°C for 30 sec, 58°C for 45 sec, 72°C for 45 sec and a final extension at 72°C for 5 min on iCycler Thermal Cycler (Bio-Rad Laboratories, Hercules, CA). The second run of PCR contained 0.5μL of template DNA produced from the first PCR, 400 nM universal primer (T3 AATTAACCCTCACTAAAGGG; T7 GTAATACGACTCACTATAGG), 1× AmpliTaq Gold PCR buffer, 400 μM of each dNTP, 4 mM MgCl$_2$, and 2.5 U AmpliTaq Gold polymerase. The second PCR were carried out at 95°C for 10min, then 35 cycles of 95°C for 30 sec, 57°C for 30 sec, 72°C for 30 sec and a final extension at 72°C for 5 min.

4.3.4 Multiplex SBE

The final PCR product (4μL) was first treated with 1U/μL Exonuclease I and 0.1U/μL shrimp alkaline phosphatase (USB Corporation, Cleveland, OH) at 37°C for 60 min, followed by inactivation of the enzyme at 95°C for 10min. The SBE reaction was

performed in a multiplex manner, containing all SBE primers (10nM each), 0.1μM of Cy3-ddATP, 0.1μM of Cy5-ddGTP (PerkinElmer,Waltham, MA), 0.05U/μL of ThermoSequenase, 1× SBE buffer (GE healthcare, Piscataway, NJ). The cyclic reactions were performed for 94°C 5min, followed by 60 cycles of 94°C and 55°C for 30 sec each.

4.3.5 Array Fabrication

Microarray used was home-made. Briefly, synthetic oligonucleotides with 3C amino linker at 3' terminal (MWG-Biotech, Piscataway, NJ) were dissolved in 3XSSC, 0.05% SDS, and 0.001% Chaps (Sigma-Aldrich, St. Louis, MO) spotting solution, as this buffer could improve immobilization of oligonucleotides on an aldehyde modified slide (CEL Associates, Pearland, TX). The oligonucleotides were spotted on the slides using XactII Compact Microarrayer (LabNEXT Inc., Glenview, IL). After printing, the slides were placed in a box at a 75%~100% relative humidity for 12~72 h at room temperature. Then, the excess of aldehyde-reactive groups were deactivated by incubating the slides with Pronto! Pre-Soak solution (Corning, NY) for 25 min., followed by washing the slides with dH$_2$O several times.

4.3.6 Hybridization

After SBE, 1μL of the SBE product, 1μL of 10mg/ml salmon sperm DNA, 8μL of 10×PCR buffer (GE healthcare, Piscataway, NJ) that was used as hybridization buffer, and 70μL of dH$_2$O are mixed together, followed by denaturing the mixture solution at 95°C for 5 min. Then, the solution was added to the ProPlate™ Multi-Array Slide System (Grace Bio-Labs, Bend, OR) and incubated with the slide at 56°C for 3~5 h. After hybridization, the slides were rinsed with 2×SSC, 1×SSC, 0.5×SSC, and ddH$_2$O,

64

respectively. Finally, the slides were spun dry immediately for 5 min at 200×g before signal detection.

4.3.7 Signal Detection and Data Analysis

Fluorescence signal from each spot of microarray was detected with Typhoon 9410 (GE healthcare, Piscataway, NJ) using the excitation laser: Green (532 nm) and Red (633 nm). The images acquired by the scanner were analyzed with the software ArrayVision 8.0 (GE healthcare, Piscataway, NJ). The signal intensity from each spot was corrected by subtracting the average background from the mean average intensity for each spot. The data generated from ArrayVision was exported to Microsoft Excel for further analysis. The methylation level of a target CpG site was calculated by dividing the methylation signal intensity (Cy5 signal) by the sum of the signal of both methylation (Cy5) and unmethylation (Cy3): Percentage of Methylation = Signal $_{Cy5}$/(Signal$_{Cy3}$+Signal$_{Cy5}$).

4.3.8 MALDI-TOF Based Mini-Sequencing

MALDI-TOF based mini-sequencing was used to verify the result of SBE-TAGs. Briefly, mini-sequencing was first carried out in the presence 0.2μM of ddA, ddT, and ddC, and 1μM of dG. The concentration of each mini-sequencing primer is 200nM. Other mini-sequencing conditions were similar to those used in SBE. Thereafter, the MALDI sample was prepared as described previously (Sun et al. 2000). The sample was then analyzed using MALDI-TOF (Bruker Daltonics Inc., Billerica, MA). The negative ion mode was used to collect all spectra.

4.4 Results and Discussion

4.4.1 Assay Design

SBE-TAGs microarray was originally developed for multiplex SNP genotyping. In this work, we modified it for multiplex methylation analysis. Fig. 10 displays the schematic representation of multiplex methylation analysis with SBE-TAGs microarray. Briefly, a sample is first subjected to bisulfite treatment that converts cytosine to uracil whereas the methylated cytosine residues are unaltered. The treated DNA is then amplified by two-runs of multiplex PCR to yield DNA fragments wherein Ts replace Us (former Cs), and Cs replace the methylated Cs. Thereafter, the methylation status of each target CpG site is determined using SBE-TAGs microarray, in which single base extension (SBE) is performed involving extension of a SBE primer located adjacent to the C position of a target CpG site using DNA polymerase in the presence of the fluorescence labeled Cy3-ddATP and Cy5-ddGTP, resulting in extension of the methylated allele by Cy5-ddGTP and extension of the unmethylated allele by Cy3-ddATP.

The detection of SBE products is achieved with a tag array. Briefly, a unique sequence tag is added to the 5' end of each SBE primer, allowing for specific hybridization with one probe spotted on the microarray. Each of the spotted probes is complementary only to one sequence tag added to the SBE primers. As a result, all SBE reactions can be distinguished and detected by a single microarray.

10 CpG sites in 9 different genes were studied to examine the performance of SBE-TAG microarray. In this study, we use only one SBE primer to target one CpG site in each gene, except for p16 in which two SBE primers were used to target two different
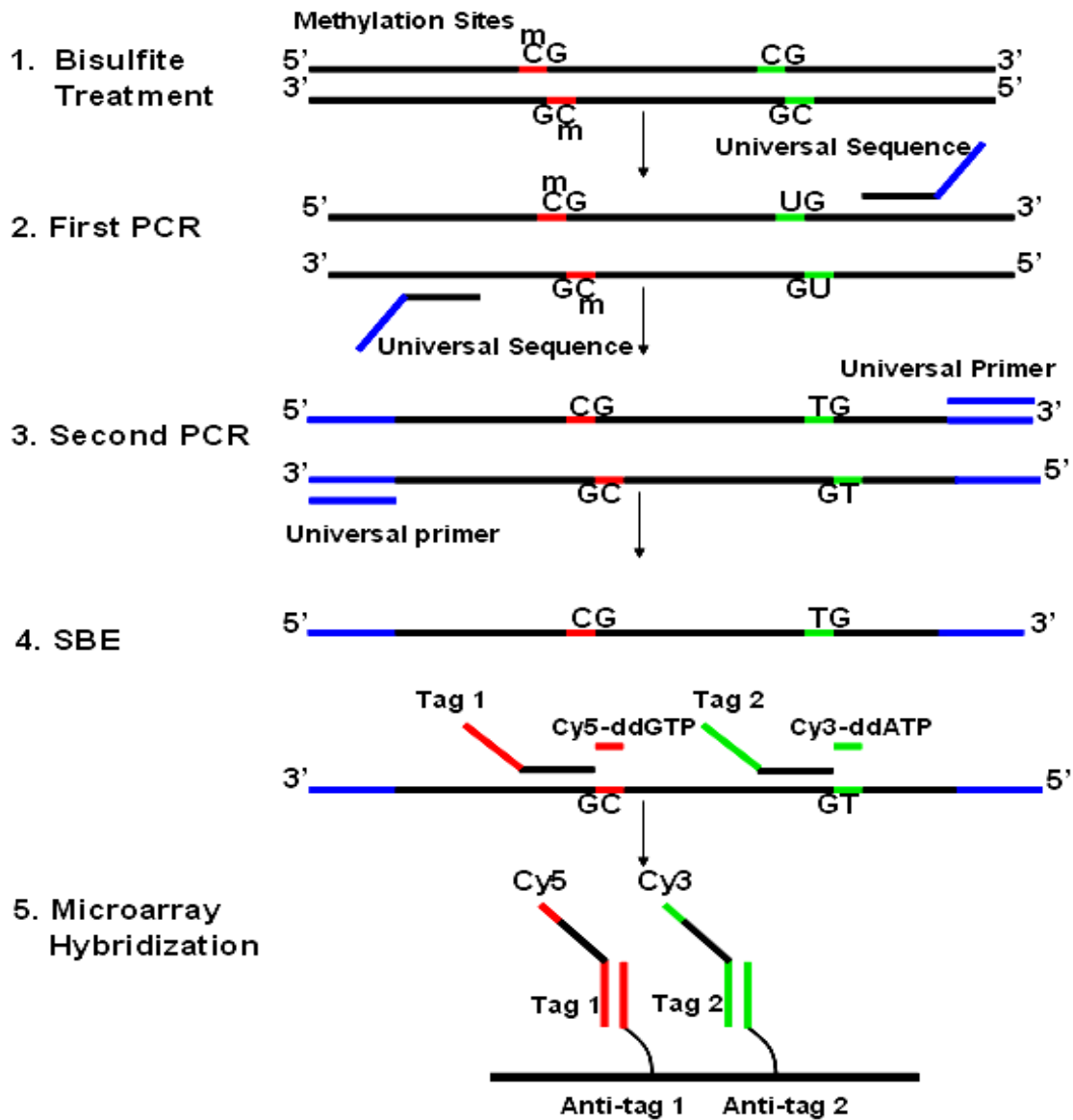
Fig. 10 Schematic representation of SBE-TAGs microarray for multiplex methylation analysis. Two hypothetical CpG sites 1 and 2 are designed as methylated and unmethylated, respectively. After bisulfite conversion, all of modified DNA template is amplified by using the primers containing the gene-specific 3' portion and an upstream universal sequence (highlighted in blue) in the first PCR. Introduction of this universal sequence allows for the amplification of all target sequences in the second PCR by using the same universal primers. In the multiplex SBE reactions, the SBE primers, one for each target CpG site, are extended in the presence of Cy5-ddGTP and Cy3-ddATP, where Cy5-ddGTP is added to the SBE primers when the allele is methylated, while Cy3-ddATP is added to the SBE primers if the allele is unmethylated, respectively. Each SBE primer is also tagged with one specific sequence tail that is fully complementary to the sequence of a detection probe spotted on microarray. The labeled SBE products are analyzed by microarray. The methylation status of a target CpG site is determined on the basis of the signals of Cy5 and Cy3 emitted from its corresponding spots on microarray.

CpG sites.

4.4.2 Multiplex PCR Amplification and SBE

Multiplexed PCR is essential to multiplex methylation analysis, but multiplexed PCR of the bisulfite treated DNA is a challenge. Bisulfite treatment eliminates all Cs in unmethylated DNA and in non-CpG sites of methylated DNA and thus creates a nearly all C-less sequence of most 3-base-DNA having predominantly A, G, and T. This reduction in sequences, relative to normal 4-base-DNA, makes multiplexed PCR of bisulfite treated DNA difficulty. A two-run PCR strategy is used in this work. The first PCR primers have both a target-specific part and universal tail part. The target-specific part allows for amplification of a specific sequence, while the tail part introduces the universal sequences for the second run of PCR that employs universal primers complementary to the tail parts introduced by the first run of PCR. The T7 and T3 sequences were used as the universal tail of the forward and reverse primers for the second PCR, respectively. The target-specific part of the first PCR primers was designed to tolerate internal CpG sites to allow its hybridization to both methylated and unmethylated sequences. Because the first run of PCR is intended to produce sufficient amounts of methylated DNA molecules for the enrichment step (note that amplification is mainly achieved in the second run of PCR in this two-staged multiplexing PCR strategy), its amplification is not requested to be as efficient as conventional PCR. As a result, this reduces the difficulty of designing the PCR primers. In general, the target-specific part of the primers was designed to target the regions containing a number of cytosines, but without CpG sites. When the CpG site cannot be avoided in some cases, the PCR primers were designed to contain both C and T or both G and A in the C position of the CpG sites. Table II lists 9

68

Table II List of the Sequences of the Primers Used in the First PCR.

Primer Set	Forward Primer	Reverse Primer	Size (bp)
MGMT	T7-TTT**Y**GGATATGTTGGGATAGTT	T3-AAAC**R**ACCCAAACACTCACCAAATC	109
GATA4	T7-GGTTTGTTTTATTTTTAGGGCGGAGG	T3-CATCCCGAACTACCCTACTAAATCCC	119
HLTF	T7-GTTGGTTATTAGAGGGTGGGGCGGA	T3-AACCGAAAACTCCATACTACTCCCC	124
SOCS1	T7-GAGTT**Y**GTGGGTATTTTTTTGGTG	T3-ACTACCATCCAAATAAAAAC**R**ACC	135
p16	T7-TGGAGTTTT**Y**GGTTGATTGGTTGGTTA	T3-CTACAAACCCTCTACCCACCTAAATC	144
RASSF2	T7-GTTTGTTTAGTTTGGTTGGATTTGAGTT	T3-CTTACCTTCACCCCGAACCAACTACATC	150
CHFR	T7-ATTAAGAGCGGTAGTTAAAGCGGAGT	T3-TCCCTTCTACCCCAACATAACATAAC	156
Reprimo	T7-TTTTGGGAAATTTTTACGGTTTAGGGAT	T3-TCGCTACTATTAACCAAAAACAAACCC	225
TPEF	T7-T**Y**GTGTTTTGTTGTTGTTGTTTTTTTGA TATTTT	T3- ATAAACTACCCGCACACTCATATACCAT TC	198

Y is the combination of C and T.
R is the combination of G and A

.

pairs of primers used in this work, where Y and R indicates that the primer contains either both C and T or both G and A at that given base.

In the same way, SBE was carried out in a multiplex manner. In other words, all 10 SBE reactions were performed in the same reaction tube. Prior to multiplex SBE, all of the SBE primers listed in table III were tested individually to check their performance and specificity. It is noted that the SBE primers were also designed to have both C and T or both G and A in the C position of the CpG sites in the case where there are CpG sites located within the sequences covered by the SBE primer sequences.

4.4.3 Tag Sequence Selection

Selection of highly discriminative sequences that can be used as the tags is critically important to SBE-TAGs microarray. During the course of this work, we found that many of the tag sequences reported in the literatures could still interact with each other, creating a high background that reduces the effectiveness of SBE-TAGs microarray. Hence, we first conducted a study to experimentally search for a set of good sequence tags. This is achieved by using a SBE method. Briefly, a synthesized oligo was used as the template for SBE. Each of the SBE primers consists of an identical sequence complementary to the sequence of the template and a unique sequence tag selected from 70 candidate sequences (Favis et al. 2004). After the SBE reaction, all of the products with fluorescence labeling were hybridized on microarray slides with candidate sequences individually. We examined each of the 70 sequences to determine whether it has cross-hybridization with other sequences. With the selection criteria of signal/noise > 10 folds, we were able to select 32 highly discriminative sequences from the 70 sequences.

Table III. List of the Sequences of the SBE and Mini-Sequencing Primers.

Primer Set	Microarray SBE primer	MALDI-TOF Mini-Sequencing primer
MGMT	TC**R**AAAC**R**CAAAAC**R**TTCTAAAAAC	GCAAAACGTTCTAAAAAC
GATA4	CCAAAC**R**C**R**ACCCCACC**R**AAACTACC	GACCCCACCGAAACTACC
HLTF	C**R**CTACCTACTCTCCCCCTCTCC	CCTACTCTCCCCCTCTCC
p16_1	CACCTCCTCTACCC**R**ACCCC	CCTCCTCTACCCGACCCC
p16_2	CCCCC**R**CCTCCAACAAC**R**CCC	CCGCCTCCAACAACGCCC
SOCS1	AAAAAACAATTCC**R**CTAAC**R**ACTATC	ATTCCGCTAACGACTATC
RASSF2	ACTTCAACTAC**R**CCTC**R**ACCAAC	AACTACGCCTCGACCAAC
CHFR	CCTTCTACCCCAACATAACATAAC	ACCCCAACATAACATAAC
TPEF	CC**R**CTCCAAAAC**R**CCCCCTACC	TCCAAAACGCCCCCTACC
Reprimo	C**R**AAAACTCCTC**R**CTCTACTTTC	ACTCCTCGCTCTACTTTC

R is the combination of G and A.
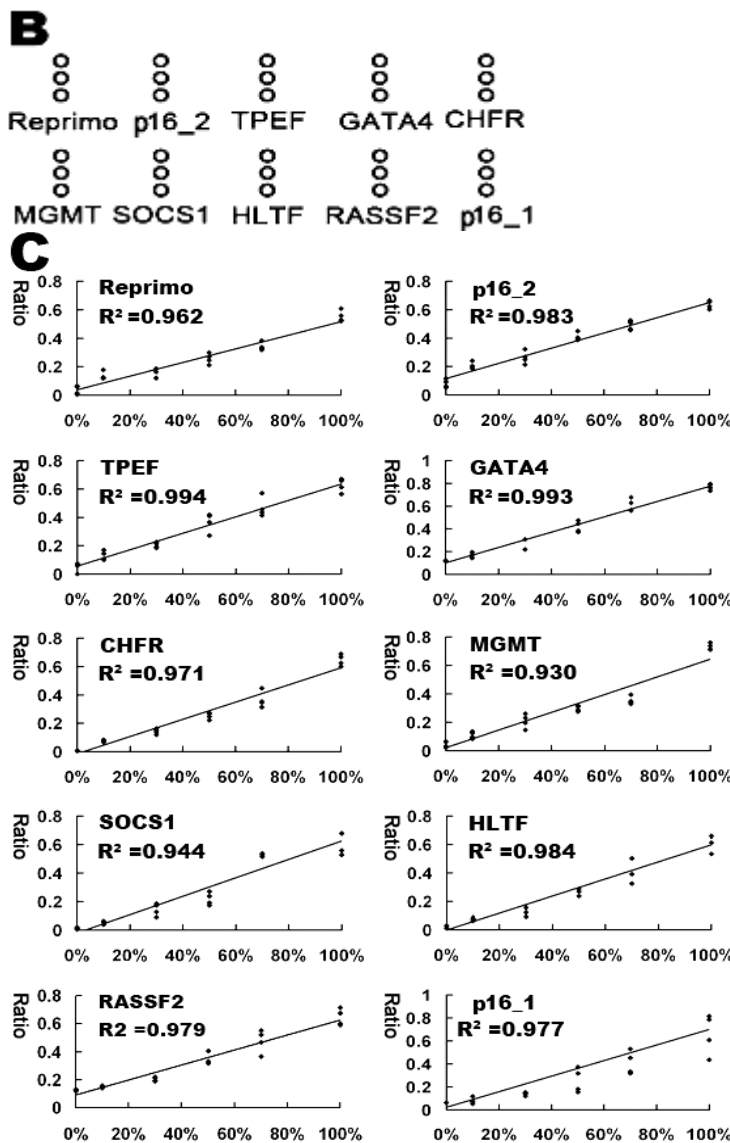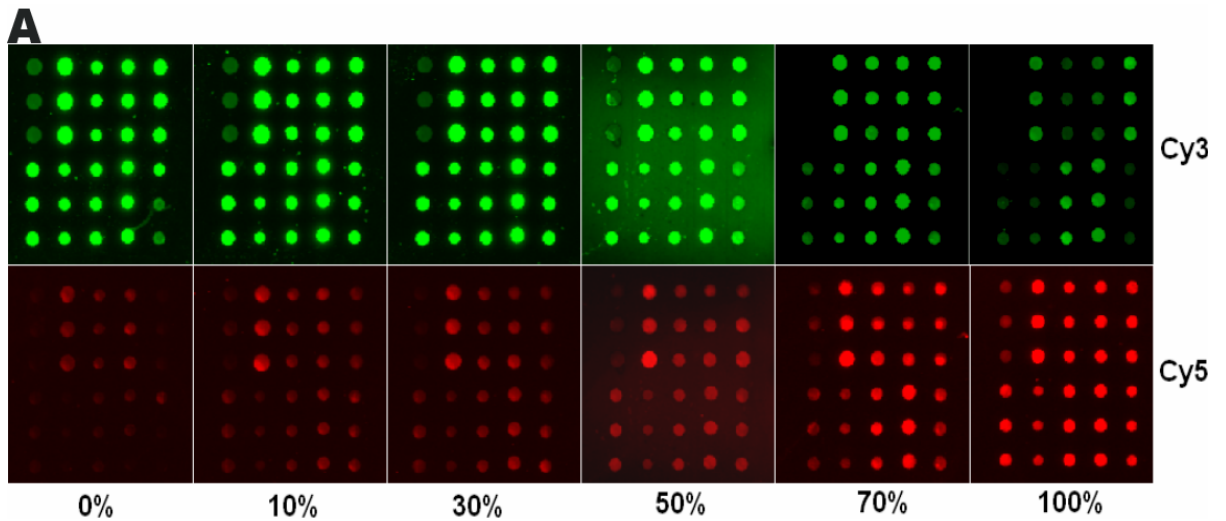
Fig. 11 Analysis of Control DNA with SBE-TAGs Microarray. (A) Typical array spectra obtained from the samples containing 0%, 10%, 30%, 50%, 70%, and 100% of methylated DNA. The green color shows the signal of Cy3 (the signal of the unmethylated allele) and the red color is the signal of Cy5 (the signal of the methylated allele). The intensity of the two channels reflects the difference in the degree of methylation; (B) The printing pattern of the 10 probes spotted on microarray, in which three spots were printed for each probe and one probe detects methylation in one CpG site. Other than gene of p16 in which two CpG sites are targeted, only one CpG site is targeted in 8 other genes; (C) The standard calibration curves (plot of M/(M+U) vs. the percentage of methylated DNA) for all 10 targeted CpG sites.

4.4.4 Analysis of Control DNA with SBE-TAGs Microarray

We first studied the DNA samples containing control genomic DNA to generate criteria for determining the methylation status of each target CpG. Each CpG site in methylated control DNA is methylated, while each CpG site is unmethylated in unmethylated control DNA. The samples containing 0%, 10%, 30%, 50%, 70%, and 100% of methylated DNA were used in this study, respectively, where the samples containing both methylated and unmethylated DNA were created by spiking methylated control genomic DNA into unmethylated control genomic DNA.

The array spectra obtained from each sample were showed in Fig. 11A. Each array was printed with the fixed printing pattern shown in Fig. 11B and each probe was spotted on three positions to generate the average signal. It is seen that the Cy3 (unmethylated) signal (the top panel spectra of Fig. 11A) decreased while the Cy5 (methylated) signal (the bottom panel spectra of Fig. 11A) increases from left to right. This is consistent with the fact that the concentration of methylated DNA increased from left to right in the tested samples. It is also seen that the background signal still appears in the arrays even after we used highly discriminated tag probes. For example, the array spots for the CpG sites of p16_2, TPEF, and GATA4 yield the signal in the Cy5 channel when the sample containing no methylated DNA was used. The array spots corresponding to the CpG sites of CHFR, RASSF2, HLTF, and p16_2 yields the signal in the Cy3 channel when the sample containing 100% of methylated DNA was used. One possible source for this background signal is cross-hybridization between the extension part of a SBE primer with the probes targeting other SBE reactions.

Nevertheless, we found that the SBE-TAG microarray can still distinguish

methylated DNA from unmethylated DNA even in the presence of this background if we utilize the following quantitative approach. Briefly, the average signal of a methylated (M) allele or its unmethylated (U) allele of a target CpG site was first derived from three spots. The signals were then used to calculate the ratio of M/(M+U) or Cy5/(Cy5+Cy3), where the M/(M+U) ratio represents the methylation level or the percentage of the methylated allele in a CpG site. The ratio of M/(M+U) for each target CpG site was plotted vs. the percentage of methylated DNA to establish the standard calibration curves (Fig. 11C).

It was seen from Fig. 11C that for each target CpG site, there is a good linear relationship between M/(M+U) and the percentage of methylated DNA, and that M/(M+U) increased proportionally with the percentage of methylated DNA in the samples. We repeated this calibration at least four times and found that SBE-TAGS was robust and reproducible. This result shows that SBE-TAGs microarray in conjunction with the standard calibration approach can be used for multiplex methylation analysis.

4.4.5 Analysis of Methylation of Cell Line DNA with SBE-TAGs Microarray

Next, we utilize the calibration curves developed above to analyze the methylation status of DNA in two colorectal cancer cell lines of HCT116 and HTB38. HCT116 has been subjected to the comprehensive methylation analysis (Liang et al. 2000; Akiyama et al. 2003; Kondo et al. 2003; Akinoet al. 2005; Brandes et al. 2005; Mund et al. 2005; Takahashi et al. 2005; Takahashi et al. 2006), allowing us to examine the accuracy of the SBE-TAGs assay by comparing our result with the published data. HTB38 was rarely studied for its methylation and this work can yield insights into methylation of HTB38.

Fig. 12 shows the results of four separate studies of HCT116 and HTB38. The
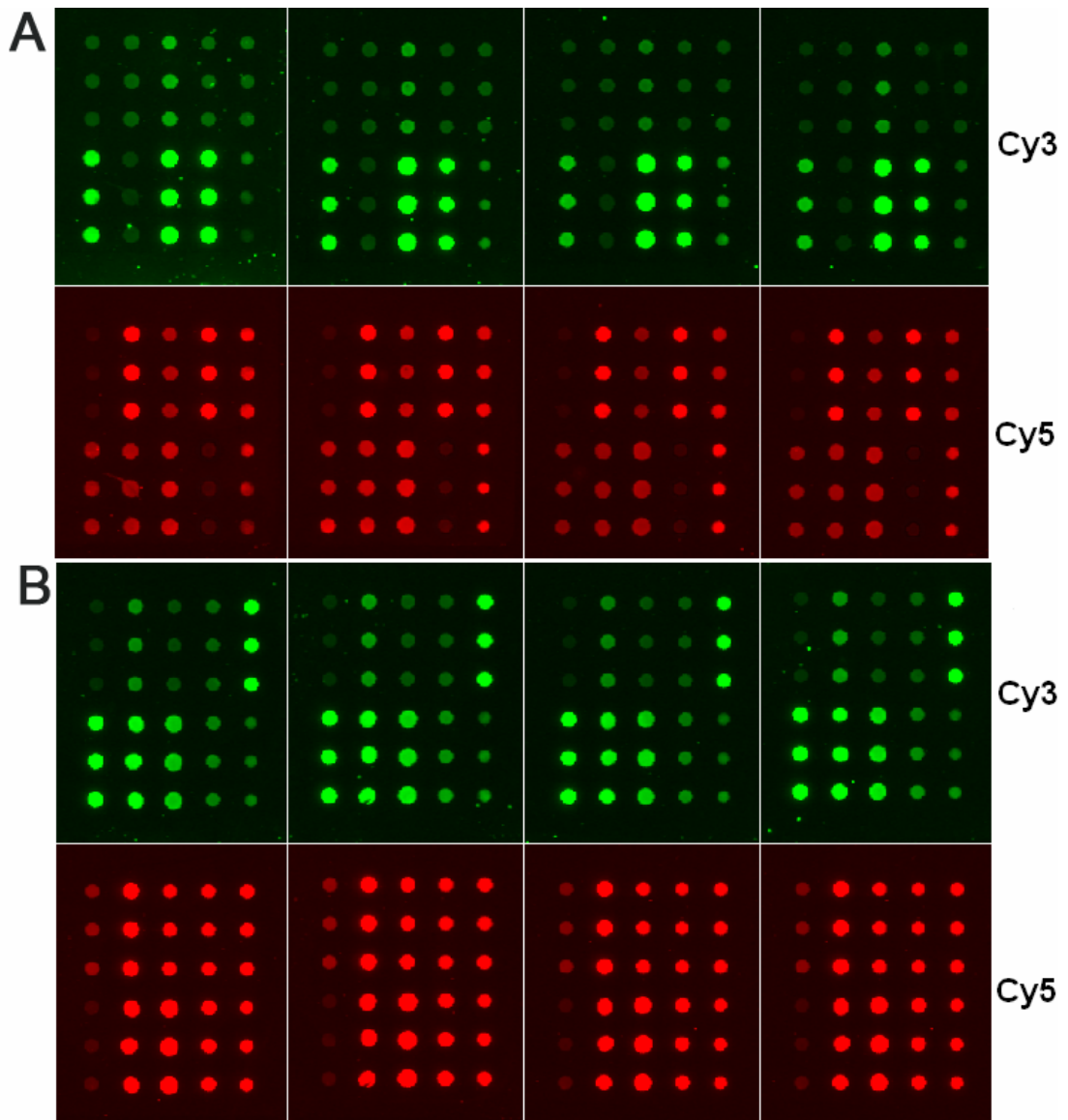
Fig. 12 Using SBE-TAGs microarray for cell lines analysis. Array spectra obtained from analysis of cell lines of HCT116 (A) and HTB38 (B). Each cell line sample was independently studied four times.

signal of a probe in each array was averaged from 3 spots to determine the ratio of M/(M+U). The mean values of M/(M+U) were determined from four separate arrays and are listed in Table IV. It is seen that the SD (Standard Deviations) of the mean of M/(M+U) for all target CpG sites are between 2.4E-03 and 1.6E-05, establishing that the SBE-TAGs assay is reproducible. We also used the standard curves derived from Fig 11C to determine the methylation status of the 10 target CpG sites and found that in HCT116, 8 of 9 genes and 9 of 10 CpG sites were methylated with the exception of RASSF2. This finding is in good agreement with the reported methylation level in HCT116. For example, published studies shown that RASSF2 is unmethylated while, other eight genes are methylated in HCT116 (Liang et al. 2000; Akiyama et al. 2003; Kondo et al. 2003; Akinoet al. 2005; Brandes et al. 2005; Mund et al. 2005; Takahashi et al. 2005; Takahashi et al. 2006). In HTB38, 8 of 9 genes (9 of 10 CpG sites) were found methylated, while MGMT was unmethylated. The main difference between HCT116 and HTB38 is that RASSF2 is unmethylated in HCT116, while MGMT is unmethylated in HTB38.

4.4.6 Methylation Status Validation by MALDI-TOF based Mini-sequencing

We also carried out a study of using MALDI-TOF based mini-sequencing (Sun et al. 2000) to determine the methylation status of the 10 target CpG sites in HCT116 and HTB38. MALDI-TOF can directly determine the molecular weight of mini-sequencing products without hybridization, eliminating any potential effects of cross-hybridization on methylation analysis. Hence, the MALDI-TOF based analysis can provide additional insights into the accuracy of SBE-TAGs microarray. Each mini-sequencing primer used is the same to the extension part of the corresponding SBE primer. In other words, mini-sequencing primers target the same CpG sites, but without the tag tail used in the

Table IV. Results of Analyzing 10 CpG Sites of 9 Genes in HCT116 and HTB38.

	HCT116		HTB38	
	Mean of M/(M+U)	SD	Mean of M/(M+U)	SD
p16_1	0.788	9.80E-05	0.913	1.40E-04
RASSF2	0.123	1.60E-04	0.893	5.60E-05
HLTF	0.215	2.90E-04	0.775	2.60E-04
SOCS1	0.893	1.60E-04	0.587	9.20E-04
MGMT	0.334	1.20E-03	0.028	1.60E-05
CHFR	0.839	3.80E-04	0.45	1.90E-03
GATA4	0.943	1.30E-04	0.877	2.60E-04
p16_2	0.549	2.40E-03	0.919	3.60E-05
TPEF	0.935	8.60E-05	0.888	1.70E-04
Reprimo	0.303	1.50E-03	0.879	4.40E-04
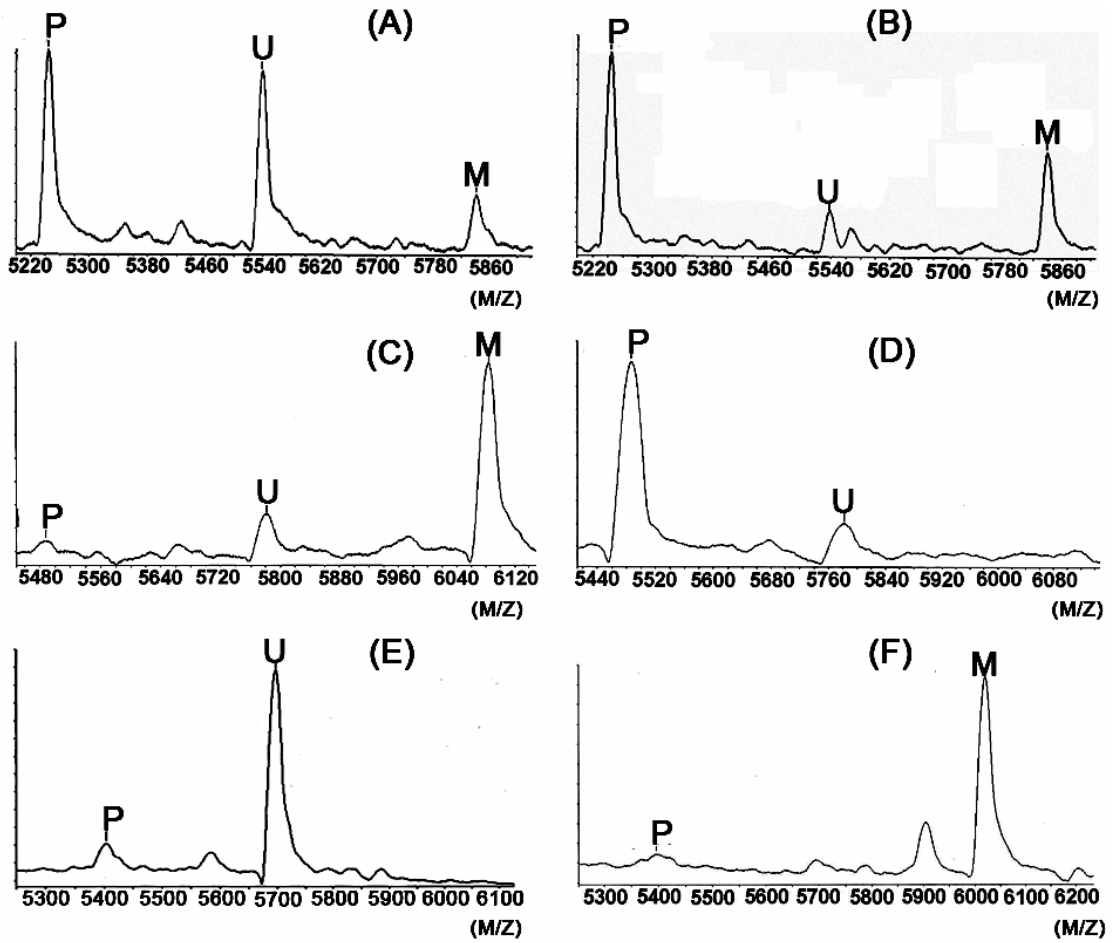
Fig. 13 Typical MALDI-TOF spectra obtained from analysis of HCT116 and HTB38. Peak labeled by P, U, and M correspond to the primer peak, the peak extended from the unmethylated allele, and the peak extended from the methylated allele, respectively. (A) Analysis of HLTF in HCT116; (B) Analysis of HLTF in HTB38; (C) Analysis of MGMT in HCT116; (D) Analysis of MGMT in HTB38; (E) Analysis of RASSF2 in HCT116; (F) Analysis of RASSF2 in HTB38.

78

SBE primer. In addition, each CpG site was individually studied by MALDI-TOF. The result of MALDI-TOF based mini-sequencing was consistent with the result of SBE-TAGs. Several representative mass spectra obtained from MALDI-TOF were showed in Fig. 13. Figs. 13A and 13B shows the result of analyzing the targeted CpG site of HLTF in HCT116 and HTB38, respectively, where the M peak (methylated) was seen in both spectra, indicating that HLTF is methylated in both cell lines. Figs. 13C and 13D shows the results of analyzing the CpG site of MGMT in both HCT116 and HTB38, respectively, where only Fig. 13C has the observable M peak, indicating that MGMT is methylated in HCT116, but unmethylated in HTB38. Figs. 13E and 13F shows the results of analyzing the CpG site of RASSF2 in both HCT116 and HTB38, respectively, where only Fig. 4F has the observable M peak, indicating that MGMT is methylated in HTB38, but unmethylated in HCT116. Clearly, this MALDI-TOF mini-sequencing study further confirms that SBE-TAGs microarray is reliable and can accurately determine the methylation status of many genes in a multiplex manner.

In conclusion, the SBE-TAGs microarray assay described in this report can accurately determine the methylation status of a number of CpG sites in a multiplexing manner. In addition, this SBE-TAGs microarray is robust and reproducible. Clearly, this work establishes that SBE-TAGs microarray is an effective method for multiplex methylation analysis.

CHAPTER V

CHZMA: A NOVEL MICROARRAY METHOD FOR MULTIPLEX

METHYLATION ANALYSIS OF TUMOR TISSUE SAMPLES

5.1 Abstract

Multiplex analysis of methylation in tumor tissue samples has been a challenge. This report describes a novel approach called Competing-Hybridization-Zipcode-MicroArray (CHZMA) for multiplex methylation analysis of tumor tissue samples, which is based on two steps of hybridization to achieve the specific detection of methylation on microarray. On the basis of analysis of seven genes (MGMT, GATA4, HLTF, SOCS1, RASSF2, ER, 3-OST-2), we found that the CHZMA assay can robustly detect methylation in the samples containing as low as 10% of methylated DNA. In addition, the CHZMA assay is validated with breast tumor tissues samples. This work suggests that CHZMA can be a new method in place of MSP for methylation analysis of tumor tissue samples.

5.2 Introduction

Because of the enormous potential of epigenetic markers in both cancer screening and treatment, considerable effort has been made to search for methylated genes that can be novel therapeutic targets or screening biomarkers (Mori et al. 2006; Liu et al. 2007; Schuebel et al. 2007). The commonly used approach to identify methylated genes in cancer consists of two steps. The first is to identify the genes that are underexpressed in cancer through gene expression of cell lines. The second step is to determine which genes among the candidate genes discovered in the first step are silenced by methylation and how they are associated with cancer. This is achieved by analysis of a number of non-tumor tissue and tumor tissue samples.

Most of existing methods for methylation analysis are based on bisulfite treatment that converts cytosine to uracil whereas the methylated cytosine residues are unaltered. The treated DNA is then PCR amplified wherein Ts replace Us (former Cs), and Cs replace the methylated Cs. Thereafter, methylation analysis is performed using conventional genetic analysis methods. Although other methods are available (Toyota et al. 1999), the most commonly used methylation analysis method has been methylation-specific PCR (MSP) (Herman et al. 1996) or its real-time PCR variation (Eads et al. 2000) because of its high sensitivity and simplicity. However, the problem of MSP is its low throughput. In general, one MSP reaction reveals methylation of a gene at a time. This weakness has prevented one from performing large-scale methylation analysis.

Microarray allows for parallel evaluation of the methylation status at numerous CpG sites within multiple genes of interest. The early application of microarray to

methylation analysis involved direct hybridization, utilizing oligonucleotide probes to target either the methylated or unmethylated alleles within target sequences (Gitan et al. 2002). However, it is not sufficiently sensitive and robust for analysis of tissues containing low abundance of methylated DNA. Although additional work was done to improve direct hybridization by designing better probes (Kimura et al. 2005; Piotrowski et al. 2006), it is still not robust for methylation analysis. Ligation-based microarray methods were also developed for methylation analysis (Bibikova et al. 2006; Cheng et al. 2006), but they are also not sufficiently robust for methylation analysis of tissue samples. This is due in part to the fact the specificity of annealing a ligation probe to a given sequence can vary greatly from one sequence to another (Bibikova et al. 2006, Dahl et al. 2007). Furthermore, bisulfite treatment eliminates all Cs in unmethylated DNA and in non-CpG sites of methylated DNA and thus creates a nearly all C-less sequence of most 3-base-DNA having predominantly A, G, and T. This reduction in sequences, relative to normal 4-base-DNA, makes hybridization less specific.

In this report, we describe a novel method called CHZMA (Competing-Hybridization- Zipcode-MicroArray) for methylation analysis. Unlike other microarray methods, CHZMA utilizes two steps of hybridization to improve its specificity. Based on the principle of CHZMA, we have developed a multiplex assay to simultaneously survey methylation of seven genes in the presence of a 10 folds more excess of unmethylated DNA, indicating that CHZMA is sufficiently sensitive to detect methylated DNA in tumor tissue samples.

5.3 Materials and Methods

5.3.1 Sample Collection and DNA Isolation

Unmethylated control genomic DNA and methylated control genomic DNA were purchased from Millipore within CpGenome[TM] DNA modification Kit (Millipore). Samples containing both methylated and unmethylated control DNA were created by spiking methylated control DNA into unmethylated control DNA. Tissue DNA was extracted from breast tissues samples using a QIAamp DNA mini kit (QIAGEN) according to the manufacture's protocol.

5.3.2 Sodium Bisulfite Treatment

The CpGenome[TM] DNA modification Kit (Millipore, Billerica, MA) was used for bisulfite treatment in this study. Bisufite treatment was performed using the protocol suggested by the manufacturer. The only exception was the desalting and desulfonation step. Briefly, after bisulfite treatment for 15 h at 50°C, a 50K Microcon centrifuge column (Millipore, Billerica, MA) was used in desalting, followed by washing the column 3 times with 500μL ddH$_2$O on Eppendorf Centrifuge 5415C (Eppendorf North America, Westbury, NY) at 9000RPM. In the desulfonation step, we transferred 500μL of 0.1MNaOH solution to the column, and then spun it until the solution went through at 3000RPM. The column was then washed twice with 500μL of ddH$_2$O and 1 time TE buffer. Finally, the modified DNA was recovered from the column. The converted DNA was used immediately or stored at -20°C for further analysis.

5.3.3 Multiplex PCR

The first PCR reaction contained 2μL of bisulfite-converted DNA, 300-400nM of each primer (Table V), 1× AmpliTaq Gold PCR buffer, 400 μM of each dNTP, 4 mM MgCl$_2$, and 2.5U of AmpliTaq Gold polymerase (Applied Biosystems, Foster City, CA), respectively. The first run of PCR was performed at 95°C for 10min, then 42 cycles of 95°C for 30 sec, 58°C for 45 sec, 72°C for 45 sec and a final extension at 72°C for 5 min on iCycler Thermal Cycler (Bio-Rad Laboratories, Hercules, CA). The second run PCR reaction contained 0.5μL of the template DNA produced from the first PCR, 400 nM of each universal primer (T3 AATTAACCCTCACTAAAGGG; T7 GTAATACGACTCACTA TAGG), 1× AmpliTaq Gold PCR buffer, 400 μM of each dNTP, 4 mM MgCl$_2$, and 2.5 U of AmpliTaq Gold polymerase, respectively. The second PCR were carried out at 95°C for 10min, then 35 cycles of 95°C for 30 sec, 57°C for 30 sec, 72°C for 30 sec and a final extension at 72°C for 5 min.

5.3.4 Microarray Array Fabrication

Microarray used was home-made. Briefly, synthetic oligonucleotides (MWG-Biotech, Piscataway, NJ) were dissolved in 3xSSC, 0.05% SDS, and 0.001% Chaps (Sigma-Aldrich, St. Louis, MO) spotting solution, as this buffer could improve immobilization of oligonucleotides on an aldehyde modified slide (CEL Associates, Pearland, TX). The oligonucleotides were spotted on the slides using XactII Compact Microarrayer (LabNEXT Inc., Glenview, IL). After printing, the slides were placed in a box at a 75%~100% relative humidity for 12~72 h. Then, the excess of aldehyde-reactive groups were deactivated by incubating the slides with Prontol Pre-Soak solution

Table V. PCR Primer Sequences.

Primer Set	Forward Primer	Reverse Primer	Size (bp)
MGMT	T7-TTTYGGATATGTTGGGATAGTT	T3-AAACRACCCAAACACTCACCAAATC	109
GATA4	T7-GGTTTGTTTTATTTTTAGGGCGGAGG	T3-CATCCCGAACTACCCTACTAAATCCC	119
HLTF	T7-GTTGGTTATTAGAGGGTGGGGCGGA	T3-AACCGAAAACTCCATACTACTCCCC	124
SOCS1	T7-GAGTTYGTGGGTATTTTTTTGGTG	T3-ACTACCATCCAAATAAAAACRACC	135
RASSF2	T7-GTTTGTTTAGTTTGGTTGGATTTGAGTT	T3-CTTACCTTCACCCCGAACCAACTACATC	150
ER	T7-GAGGTGTATTTGGATAGTAGTAAGTT	T3-AAAAAAACCCCCCAAACCGTTAAAACC	174
3-OST-2	T7-TTTTGGTTAGTAGTTTTCGGAGAAGA	T3-ATAAACCATAACTCCATAAACCCGC	168
ACTB	T7-TGGTGATGGAGGAGGTTTAGTAAGT	T3-AACCAATAAAACCTACTCCTCCCTTAA	133

Y is the combination of C and T.
R is the combination of G and A.

(Corning, NY) for 25 min., followed by washing the slides with dH$_2$O several times.

5.3.5 Competing Hybridization

A mixture of seven M probes (1μM of each), seven U probes (40-50μM of each), and an ACTB internal reference probe (1μM) was incubated with the PCR products for hybridization. Thehybridization was carried out on a PCR cycler by first denaturing DNA at 95 $^{\circ}$C for 5min, followed by stepwise reducing the temperature at 0.1 $^{\circ}$C/sec to 72 $^{\circ}$C, then slowly reducing the temperature from 72 $^{\circ}$C to 50 $^{\circ}$C within 1.5h. Finally the hybridization temperature was maintained at 50 $^{\circ}$C for 30min. The M probes and competing probes are listed in table VI.

5.3.6 Probe Isolation

6μL of streptavidin-coated magnetic beads (Dynabeads® M-280) were incubated with the above hybridization solution at the room temperature for one hour, allowing for bead capturing of the hybrids formed by the probes and PCR products. After washing, the beads were resuspended in 60μL of the microarray hybridization buffer, followed by heating the solution at 95 $^{\circ}$C for 5min to release the captured probes.

5.3.7 Microarray Hybridization and Data Analysis

The captured probes were spotted on the ProPlate™ Multi-Array Slide System (Grace Bio-Labs, Inc.). The slide was incubated at 37 $^{\circ}$C for 12h. After hybridization, the slide was rinsed one time each with 2×SSC, 1×SSC, 0.5×SSC, and ddH$_2$O. Finally, the slide was spun dried immediately for 5 min at 200×g. The fluorescence signal emitted from an array slide was detected with a Typhoon 9410 (GE healthcare). The images

Table VI. Sequences of the Probes Used in Competing Hybridization.

Probe Set	Methylation-Specific Detection Probe	Competing Probe
MGMT	**Cy5-Zip-Code 1**-AAAACCTACGAACGTCGAAA	AAAACCTACAAACATCAAAA
GATA4	**Cy5-Zip-Code 2**-CGCCCGAAACGCCTCCGAACGC	CACCCAAAACACCTCCAAACAC
HLTF	**Cy5-Zip-Code 3**-CGCAACCGCCGAACGCACGC	CACAACCACCAAACACACAC
SOCS1	**Cy5-Zip-Code 4**-CACGCGAATACTCGTAAATCCC	CACACAAATACTCATAAATCCC
RASSF2	**Cy5-Zip-Code 5**-CGCCTCGACCAACGATCGACGA	CACCTCAACCAACAATCAACAA
ER	**Cy5-Zip-Code 6**-GCGTTAACGACGACCGCGACG	ACATTAACAACAACCACAACA
3-OST-2	**Cy5-Zip Code 7**-ACGACCACGCGAATCGAACGTT	ACAACCACACAAATCAAACATT
ACTB	**Cy5-Zip-Code 8**-ACCACCACCCAACACACAATAAC	

acquired were then analyzed with Software ArrayVision 8.0 (GE healthcare). The signal intensity from each spot was corrected by subtracting the average background from the mean average intensity of that spot. The data generated from ArrayVision was exported in the Excel format for further analysis. The methylation status of a gene of interest was determined by the ratio of the signal of its corresponding M probe to the signal of the internal reference gene probe.

5.3.8 Statistical Analysis

The optimal cutoff values for the assignment of the methylation status were determined by the receiver-operator characteristic (ROC) curve analysis using MedCalc Software (Mariakerke, Belbium). This is achieved by maximizing both sensitivity and specificity (Enokida et al. 2005; Hoque et al.2006).

5.3.9 MSP Analysis

Each MSP reaction contained 1~10ng of bisulfite-converted DNA, 400 nM of each primer, 1×AmpliTaq Gold PCR buffer, 400 μM of each dNTP, 4 mM $MgCl_2$, and 1.5 units of AmpliTaq Gold polymerase (Applied Biosystems). PCR products were examined on 4% agarose gel.

5.4 Results and Discussion

5.4.1 Principle of CHZMA

Fig. 14 displays the principle of CHZMA. In general, a CHZMA assay consists of five steps. Firstly, a DNA sample is subject to bisulfite treatment that converts cytosine to uracil whereas the methylated cytosine residues are resistant to this conversion.
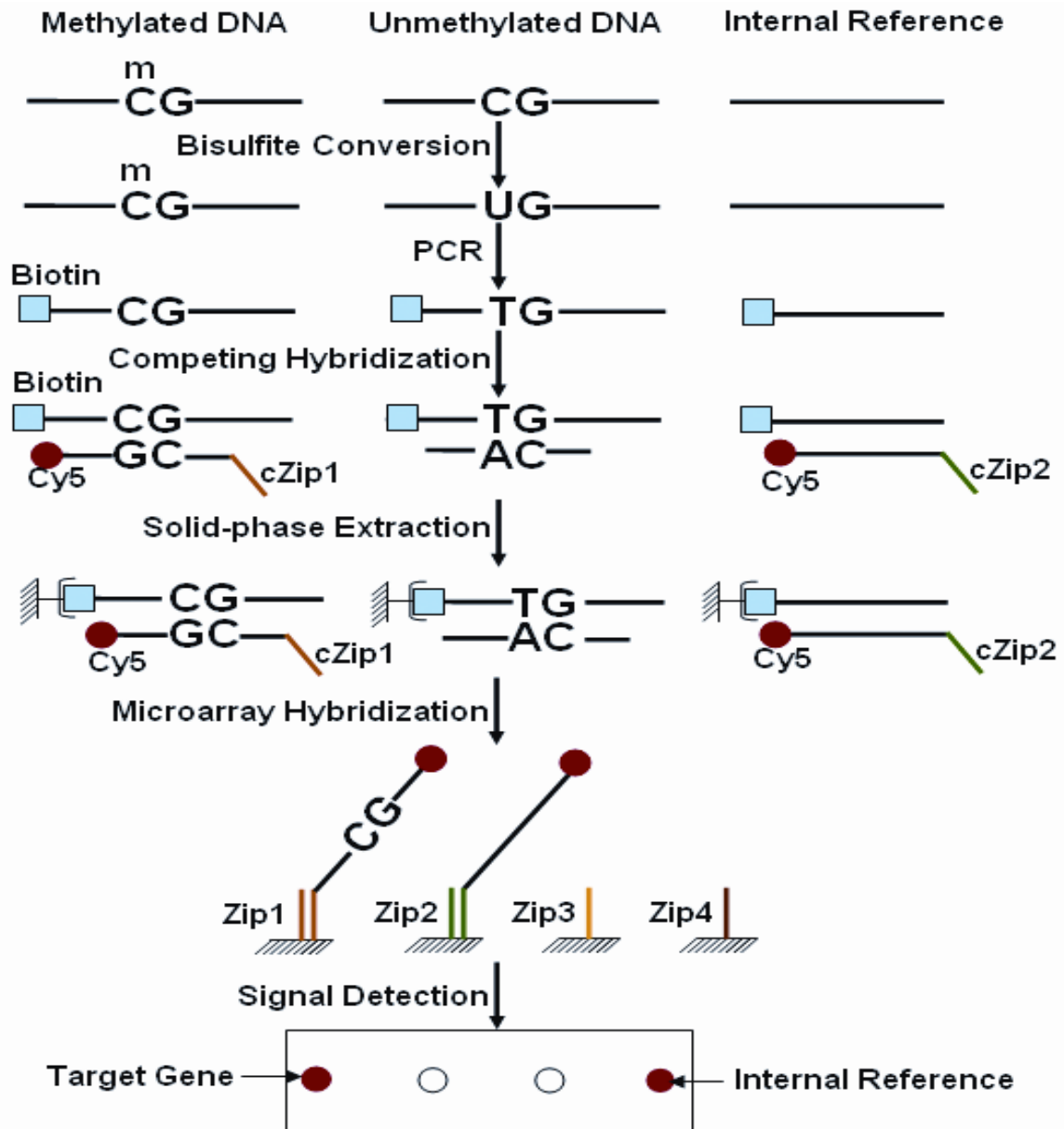
Fig. 14 Schematic representation of the principle of CHZMA. Two hypothetical CpG sites 1 and 2 are designed as methylated and unmethylated, respectively. After bisulfite conversion, the treated DNA is subject to PCR amplification using a gene-specific 3' portion and an upstream universal sequence (highlighted in blue) in the first PCR. This universal sequence is used as a PCR primer to amplify all of target genes in the second PCR. all of the PCR products were labeled with biotin. In the competing hybridization, only methylated probe and internal reference probe with zip-code sequences were labeled with Cy5. To improve the hybridization specificity, excessive amount of the competing probes without zip-code and Cy5 modification were added in the competing system. After the following solid-phase extraction, only methylation probes and internal reference could hybridized on microarry slides for the signal detection. The relative methylated level could be present by the ratio of target gene and internal reference.

89

Secondly, the treated DNA is subject to multiplex PCR amplification by using a two-run PCR, in which all target sequences are first amplified by the first run of PCR. Each of the primers used in the first PCR has a target-specific part and a universal tail part. The target-specific part amplifies one of the target genes, while the tail part introduces a universal sequence for the second run of PCR. In this work, the T7 and T3 sequences are used as the universal tail of the forward and reverse primers, respectively. Then, the products of the first PCR are further amplified using the T7 and T3 universal primers. A biotin tag is also added to the PCR products during the course of the second PCR.

Thirdly, PCR products are incubated with a set of the methylation-specific detection probes (or M probe) and their respective competing probes (U probe). Each of the M probes has two parts: a methylation-specific sequence complementary to a methylated sequence of interest, and a zip-code tag complementary to a zip-code probe spotted on microarray. In addition, the M probes are labeled by a Cy5 dye. Each of the U probes only contains a sequence complementary to an unmethylated sequence. In this step, a competing hybridization scheme is utilized to improve the hybridization specificity, in which DNA fragments bind to either the M or U probes. The role that the U probes play herein is to inhibit non-specific hybridization between unmethylated sequences and the M probes. It was observed that the presence of a greatly more excess of the U probes could substantially improve the specificity

of hybridization. As a result of this hybridization, the M probes can preferentially bind to the methylated DNA fragments.

Fourthly, the hybrids formed in Step 3 are captured by solid-phase extraction

90

involving the use of streptavidin-coated magnetic bead, as each PCR product contains a biotin tag. The beads are washed, leading to removal of the un-reacted U and M probes. Then, the beads are heated to release the probes captured. The result of step 4 is to isolate the M probes that preferentially bind to the methylated DNA fragments. The quantity of an M probe captured can reflect the quantity of the methylated allele in a target sequence.

Finally, the methylation status of each target sequence is analyzed using microarray. This is done by incubating the probes isolated in Steps 3 and 4 with microarray pre-spotted with a variety of zip-code probes. As mentioned above, the second part of an M probe is a zip-code tag that is complementary to a zip-code probe spotted on microarray. Each zip-code tag has a unique sequence containing all four nucleotides of A, G, T, C, minimizing cross-hybridization. It is noted that the U probes do not contain the zip-code tag, thus not being captured and detected by microarray. Similar to quantitative MSP (Harden et al. 2003), CHZMA also utilizes an unmethylated gene, ACTB, as internal reference to measure the methylation status of a gene. In other words, the methylation status of a gene is determined on the basis of the signal ratio of its corresponding M probe to the zip-code probe for detecting ACTB.

5.4.2 Specificity and Sensitivity of CHZMA

We first studied the specificity and sensitivity of CHZMA using the samples created by spiking methylated control DNA into unmethylated control DNA. The methylated status at each CpG site is known in control DNA. Seven genes were selected for this study and they were MGMT, GATA4, HLTF, SOCS1, RASSF2, ER, 3-OST-2, respectively. Only one methylation-specific detection probe was used to determine the methylation status of one target gene. The samples containing 0%, 3%, 5%, 10%, 50%,

and 100% of methylated DNA were studied, respectively. We found that the specificity of hybridization involved in Step 3 was poor in the absence of the U probes, but the presence of an excessive amount of the U probes greatly improved the specificity. Fig. 15A shows the result when the same quantity of the M and U probes were used, while Fig. 15B shows the result when a 40-folds more excess of the U probes were present. It was seen that the presence of a 40-50 folds more excess of the U probes led to the best discrimination. Hence, a 40 fold more excess of the U probes were used in the remaining of this study.

Experimentally, we first determine the background signal (R_b), i.e. the signal ratio of a M probe to the internal reference gene probe when no methylated DNA is present. Then, we measured the same signal ratio (R) in the samples containing 3%, 5%, 10%, 25%, and 50% of methylated DNA, respectively. Table VII lists the result of the statistical analysis of the CHZMA assay when it was used to analyze the samples containing 0% or 10% of methylated DNA, respectively. Each of the samples was analyzed 25 times. Then, we determined the average signal ratios of R and R_b for each gene. It was seen that in each case, the R value determined for the samples containing 10% of mehtylated DNA is about 3 times or more larger than R_b. Statistical analysis also suggests that the difference between R and R_b is statistically significant with the p value of less than 1.9×10^{-5} in all seven genes, suggesting that CHMZA is still robust to detect all methylated genes if the samples contain only 10% of methylated DNA. It was noted that in some cases, CHZMA could detect as little as 3% of methylated DNA.

Receiver operating characteristics (ROC) curve (plotting sensitivity vs. 100-specificity) is a useful method to evaluate the performance of a diagnostic test in
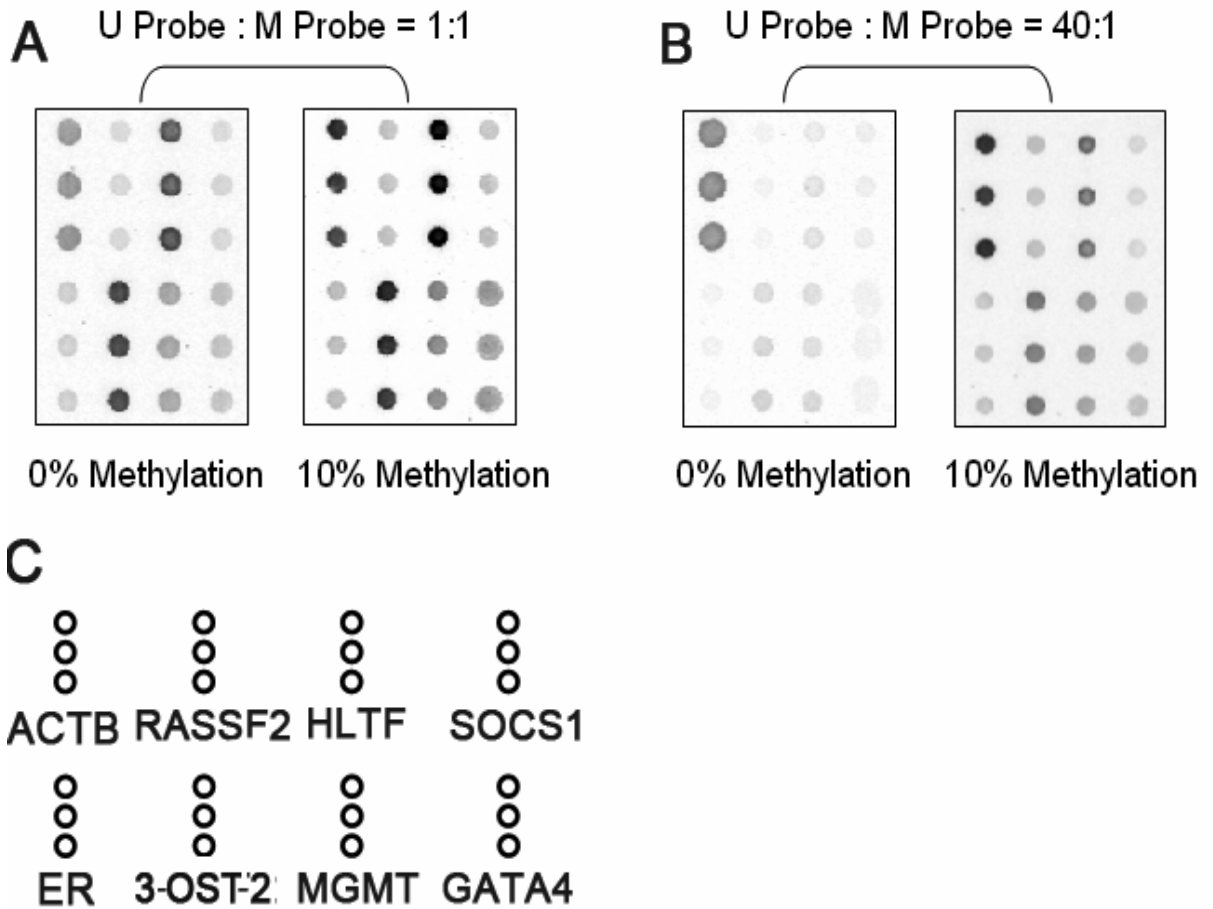
Fig. 15 Effect of amounts of the U probes on the specificity of hybridization. (A) The molar quantity of an M probe used in hybridization is the same to that of its corresponding U probe; (B) a 40 folds more excess of the U probe than the M probe was used in hybridization; (C) The fixed printing pattern of 8 zip-code probes spotted on a assay slide. Seven probes are used to detect seven genes (one for each gene) and the eighth probe is for the detection of the internal reference gene, ATCB.

Table VII. Statistical Analysis of the Performance of CHZMA.

	R_b	R	P-value	100% Specificity		100% Sensitivity	
				Cutoff Value[a]	Sensitivity %	Cutoff Value[b]	Specificity %
MGMT	0.248	0.631	1.5×10^{-13}	0.382	88	0.364	88
GATA4	0.141	0.716	7.8×10^{-10}	0.249	96	0.226	88
HLTF	0.253	1.775	2.0×10^{-15}	0.526	100	0.526	100
SOCS1	0.070	0.370	1.1×10^{-07}	0.144	100	0.144	100
RASSF2	0.075	0.425	1.9×10^{-05}	0.142	92	0.123	96
ER	0.082	0.295	3.1×10^{-23}	0.121	100	0.121	100
3-OST-2	0.490	1.886	1.7×10^{-20}	0.856	100	0.856	100

Cutoff Value[a] is determined at 100% specificity.
Cutoff Value[b] is determined at 100% sensitivity.
R_b is the ratio of unmethylated control DNA, and R is the ratio of 10% methylated control DNA.
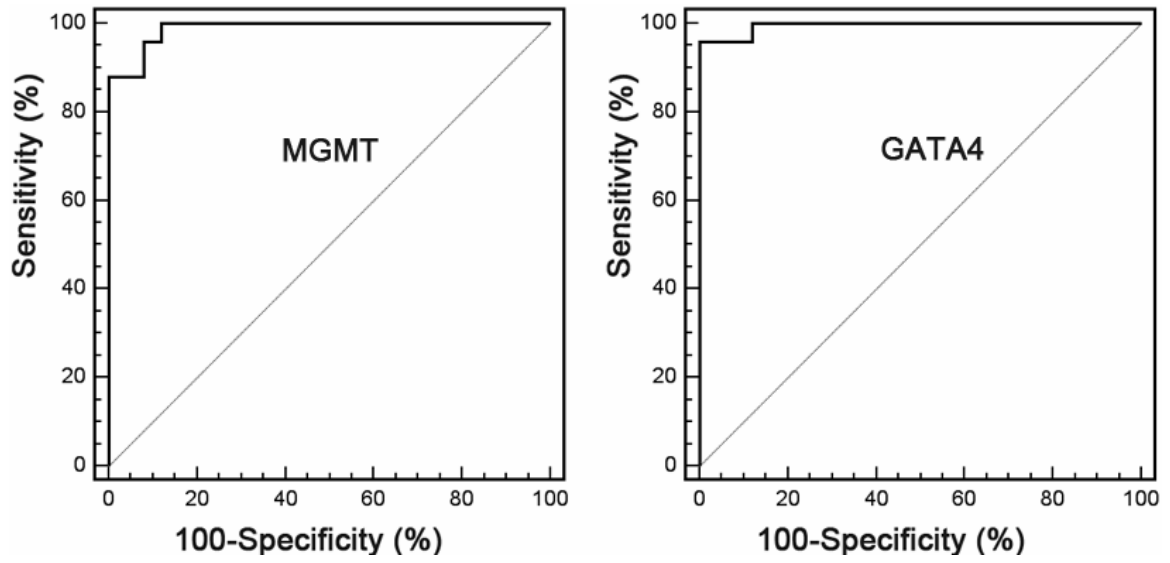
Fig. 16 ROC curve. Representative receiver-operator characteristic (ROC) curves for analysis of MGMT (left) and GATA4 (right).

Hoque et al. 2006), in which the sensitivity and specificity of an assay can be determined on the basis of the cutoff value. Fig. 16 shows the representative ROC curves for analysis of MGMT and GATA4, respectively. The curves were obtained for the samples containing 10% of methylated DNA.

In this study, the optimum cutoff values, maximizing the sum of sensitivity and specificity, were derived. Sensitivity is defined as the percentage of the samples that are correctly identified as methylated in the samples containing the same abundance of methylated DNA. Specificity is defined as the percentage of the samples that are correctly called as unmethylated in the samples containing only unmehtylated DNA. Table 3 lists the cutoff values when the samples contain 10% of methylated DNA at 100% of sensitivity and 100% of specificity, respectively. As seen from Table 3, with a specificity of 100%, CHZMA has a sensitivity of 88% (MGMT), 92% (RASSF2), 96% (GATA4), and 100% (HLTF, SOCS1, ER, and 3-OST-2), respectively. With a sensitivity of 100%, CHZMA has a specificity of 88% (MGMT), 96% (RASSF2), 88% (GATA4), and 100% (HLTF, SOCS1, ER, and 3-OST-2), respectively.

5.4.3 Analysis of Breast Cancer Tissues

The CHZMA assay developed in this work can robustly reveal methylation in the samples containing as little as 10% of methylation DNA, allowing for methylation analysis of tumor tissues. Hence, we next apply this assay to analysis of 5 normal breast tissues and 11 breast cancer tissues. The DNA extracted from the tissue samples was subject the CHZMA analysis in the same manner as outlined above. Briefly, after the microarray data analysis, the methylation status of each gene in each sample was classified using the cutoff values (100% of specificity) listed in Table 3. In other words,

96

Table VIII. Methylated Primer Sets for Methylation-Specific PCR.

Primer set	Forward Primer	Reverse Primer	Size (bp)	Anneal Temperature
MGMT	CGTTTCGACGTTCGTAGGTTTTCGC	AACCACTCGAAACTACCACCGTCC	121	60℃
GATA4	GGTTTTGCGTTCGGAGGCGTTTC	ATAAACGACTCCGTCGACTCCCG	120	66℃
HLTF	GATCGCGTGCGTTCGGCGGTT	GACCCCGAACCGCGACCGTAA	132	64℃
SOCS1	CGCGTGTATTTTTAGGTCGGTCGT	GCGAAACTCTTCCCGACGAATAAA	117	59.5℃
RASSF2	TTCGTCGATCGTTGGTCGAGGC	CGCGCCTTACCTTCACCCCGA	124	64℃
ER	TCGCGGTCGTCGTTAACGCGTA	AACGAACTCGAAAACACGCTATTAAAT	133	60℃
3-OST-2	TAACGTTCGATTCGCGTGGTCGT	AACTCCATAAACCCGCGCCGAAA	124	60℃

for a given gene, this gene is called methylated if its measured R value is larger than the corresponding cutoff value, otherwise this gene is called unmethylated. Fig. 17A displays the result of analyzing the 16 tissue samples. The red color indicates that the measure R value for a given gene is larger than its correposning cutoff value, thus reveal that this gene is methylated in a sample of interest. The green color indicates that a gene is unmethylated in a tissue sample of interest. 90.9% (10 of 11) of cancer samples were found methylated in at least one target gene, while only one normal tissue of S5 was methylated in ER, 3-OST-2, and GATA4. It has reported that aging can also cause methyaiton in ER, 3-OST-2, and GATA4 (Issa et al. 1997; Takahashi et al. 2006). Hence, it is not totally surprised to observe methylated ER in a normal breast tissue.

To validate the result of CHAZM, we also used MSP to determine the methylation status of the same seven genes in two tissue samples of S8 and S9, in which, one of the two primers used in a MSP reaction was designed to target the same sequence covered by an M probe, allowing for the direct comparison of MSN with CHZMA. In addition, only MSP reactions for detecting methylated DNA were performed and ach gene was individually studied by MSP. The MSP result was shown in Fig. 17B. It is seen that the MSP result is concordant with the methylaiton status determined by CHAMZ, except for the case of 3-OST-2 in the sample of S9, where MSP revealed that 3-OST-2 was methylation, while 3-OST-2 was called unmethylated by CHZMA. This discrepancy may be due to the difference in the detection sensitivity. MSP is highly sensitive and can detect one methylated allele in 1,000 copies of unmethylated alleles (Eads et al. 2000), while CHZMA is robust to detect 10% or higher abundance of methylated DNA. During the course of this work, we found that the methylation level can vary greatly from gene to

A

	RASSF2	HLTF	SOCS1	ER	3-OST-2	MGMT	GATA4
S1	0.031	0.235	0.138	0.112	0.480	0.285	0.192
S2	0.015	0.202	0.051	0.026	0.534	0.293	0.141
S3	0.022	0.208	0.041	0.095	0.672	0.205	0.059
S4	0.011	0.082	0.067	0.020	0.456	0.195	0.080
S5	0.036	0.391	0.106	0.253	1.517	0.326	0.255
S6	0.053	0.371	0.073	0.189	1.485	0.427	0.181
S7	0.066	0.267	0.143	0.138	1.033	0.252	0.139
S8	0.041	0.960	0.230	0.397	3.761	0.900	0.661
S9	0.028	0.230	0.114	0.180	0.727	0.157	0.096
S10	0.054	0.326	0.093	0.209	1.134	0.201	0.158
S11	0.053	0.469	0.281	0.240	1.847	0.560	0.286
S12	0.044	0.301	0.082	0.198	1.103	0.396	0.319
S13	0.152	0.532	0.196	0.416	1.726	0.409	0.309
S14	0.060	0.291	0.078	0.086	0.695	0.340	0.284
S15	0.088	0.218	0.103	0.137	1.317	0.327	0.175
S16	0.131	0.281	0.030	0.095	0.557	0.184	0.135

B
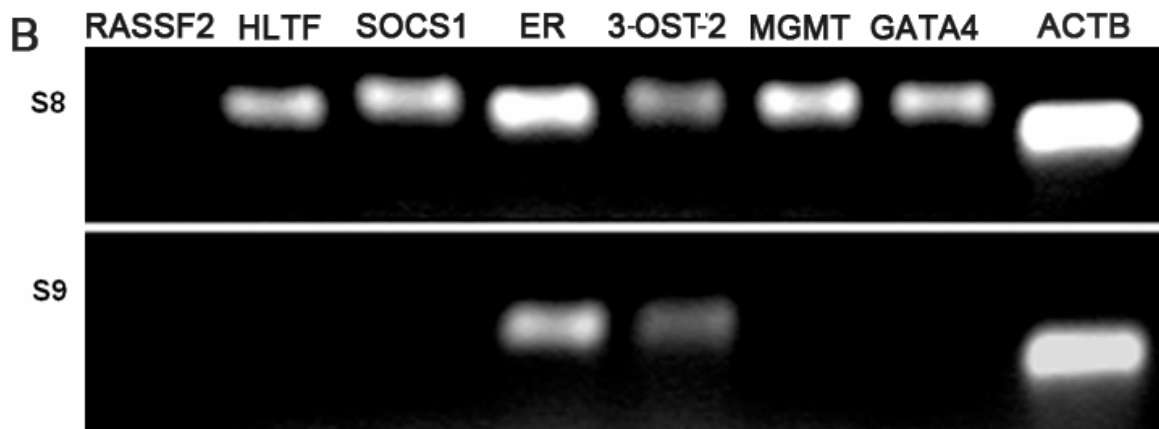
Fig. 17 Methylation analysis of tissue samples. (A) The result of analysis of 5 normal breast tissues (S1-S5) and 11 breast cancer tissues (S6-S16) with the CHZMA assay, in which the measured R value for each target gene in a given tissue samples is listed. The red color indicates that the measure R value for a gene is larger than the its correposning cutoff values, while the green color indicates that the R value for a gene is smaller than the corresponding cutoff value. (B) MSP analysis of Samples S8 and S9. A clear band suggests that the corresponding gene is methylated in tissue. Otherwise the gene is unmethylated.

99

gene even in the tissue sample. The abundance of the methylated gene of 3-OST-2 may simply be lower than 10% in Sample of S9 and thus cannot be detected by CHZMA.

Clearly, this study demonstrates that CHZMA is a reliable method for methylation analysis, which can robustly detect as low as 10% of methylated DNA. In principle, the number of target sequences that can be simultaneously amplified by PCR limits the throughput of a CHZMA assay. During the course of this study, we found that the PCR strategy used in this work can simultaneously amplify more than 10 different sequences. Therefore, one CHZMA assay should be able to detect methylation in at least 10 genes and this throughput is 10 folds higher than that of MSP. Considering the fact that CHZMA is sufficiently sensitive to detect 10% of methylated DNA, we believe that CHZMA can be a good high-throughput method in place of MSP for methylation analysis of tumor tissue samples.

Recently, a novel random whole-genome amplification method called BiMP was developed to amplify bisulfite-treated DNA (Reinders et al. 2008). Unlike the standard whole-genome method, BiMP reduces the length of the random primers from 6mers to 4mers to favor the priming of smaller DNA fragments. As a result, improved uniformity of amplification of whole-genome was achieved. A combination of BiMP with CHZMA may eventually offer a technology that allows for genome-wide, high-resolution DNA methylation profiling using bisulfite-treated DNA.

BIBLIOGRAPHY

Adams RLP, Rinaldi A, and Seiwright C. (1991). Microassay for DNA methyltransferase. *J. Biochem. Biophys. Meth.* **22**:19–22.

Adorjâan P, Distler J, Lipscher E, Model F, Mèuller J, Pelet C, Braun A, Florl AR, Gèutig D, Grabs G, et al. (2002). Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res.* **30**:e21.

Aggerholm A and Hokland P. (2000). DAP-kinase CpG island methylation in acute myeloid leukemia: methodology versus biology? *Blood* **95**:2997–2999.

Akama TO, Okazaki Y, Ito M, Okuizumi H, Konno H, Muramatsu M, Plass C, Held WA, and Hayashizaki Y. (1997). Restriction landmark genomic scanning (RLGS-M)-based genome-wide scanning of mouse liver tumors for alterations in DNA methylation status, *Cancer Res.* **57**:3294–3299.

Akino K, Toyota M, Suzuki H, Mita H, Sasaki Y, Ohe-Toyota M, Issa JP, Hinoda Y, Imai K, and Tokino T. (2005). The ras effector RASSF2 is a novel tumor-suppressor gene in human colorectal cancer. *Gastroenterology* **29**:156-169.

Akiyama Y, Watkins N, Suzuki H, Jair KW, Engeland VM, Esteller M, Sakai H, Ren CY, Herman JG, and Baylin SB, et al. (2003). GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. *Mol. Cell Biol.*, **23**:8429-8439.

Albertson DG, Ylstra B, Segraves R, Collins C, Dairkee SH, Kowbel D, Kuo WL, Gray JW, and Pinkel D. (2000). Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nat. Genet.* **25**:144-146.

Al-Khaldi SF, Myers KM, Rasooly A, and Chizhikov V. (2004). Genotyping of Clostridium perfringens toxins using multiple Oligonucleotide microarray hybridization. *Mol. Cell. Probes* **18**: 359–367.

Antequera F and Bird AP. (1993). Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci.* **90**:11995.

Babinger P, Kobl I, Mages W, and Schmitt R. (2001). A link between DNA methylation and epigenetic silencing in transgenic Volvox carteri. *Nucleic Acids Res.* **29**:1261-1271.

Balaghi M and Wagner C. (1993). DNA methylation in folate deficiency: use of CpG methylase. *Biochem. Biophys. Res. Commun.* **193**:1184–1190.

Balog RP, de Souza YE, Tang HM, DeMasellis GM, Gao B, Avila A, Gaban DJ, Mittelman D, Minna JD, Luebke KJ, et al. (2002). Parallel assessment of CpG methylation by two-color hybridization with oligonucleotide arrays. *Anal. Biochem.* **309**:301–310.

Barbin A, Montpellier C, Kokalj-Vokac N, Gibaud A, Niveleau A, Malfoy B, Dutrillaux B, and Bourgeois CA. (1994). New sites of methylcytosine-rich DNA detected on metaphase chromosomes. *Hum. Genet.* **94**:684-692.

Baylin SB. (1997). Tying it all together: Epigenetics, genetics, cell cycle and cancer. *Science* **277**:1948–1949.

Baylin SB, Herman JG, Graff JR, Vertino PM, and Issa JP. (1998). Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv. Cancer Res.* **72**:141-196.

Baylin SB and Herman JG. (2000). DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet.* **16**:168–174.

Belinsky SA, Nikula KJ, Baylin SB, and Issa JP. (1995). A microassay for measuring cytosine DNA methyltransferase activity during tumor progression. *Toxicol. Lett.* **83**:335–340.

Bensaada M, Kiefer H, Tachdjian G, Lapierre JM, Cacheux V, Niveleau A, and Métézeau P. (1998). Altered patterns of DNA methylation on chromosomes from leukemia cell lines: identification of 5-methylcytosines by indirect immunodetection. *Cancer Genet. Cytogenet.* **103**:101-109.

Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ 3rd, Gingeras TR, et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**:169-181.

Bestor TH, Hellewell SB, and Ingram VM. (1984). Differentiation of two mouse cell lines is associated with hypomethylation of their genomes. *Mol. Cell Biol.* **4**:1800–1806.

Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, Doucet D, Thomas NJ, Wang Y, Fan J-B., et al. (2006). High-Throughput DNA methylation profiling using universal bead arrays. *Genome Res*. **16:**383-393.

Bird AP. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* **8**:1499–1504.

Bird AP. (1986). CpG-rich islands and the function of DNA methylation. *Nature* **321**:209-213.

Bird AP. (1995). Gene number, noise reduction and biological complexity. *Trends Genet.* 11:94.

Bird AP. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev*. **16:** 6–21.

Boyd VL and Zon G. (2004). Bisulfite conversion of genomic DNA for methylation analysis: protocol simplification with higher recovery applicable to limited samples and increased throughput. *Anal. Biochem.* **326**:278-280.

Boyd VL, Moody KI, Karger AE, Livak KJ, Zon G, and Burns JW. (2006). Methylation-dependent fragment separation: direct detection of DNA methylation by capillary electrophoresis of PCR products from bisulfite-converted genomic DNA. *Anal. Biochem.* **354**:266-273.

Boyd VL, Moody KI, Karger AE, Livak KJ, Zon G, and Burns JW. (2007). Methylation-dependent fragment separation: novel analysis of 5-methyl cytosine by capillary electrophoresis of amplified DNA using PCR incorporation of chemically modified dCTP. *Nucleosides Nucleotides Nucleic Acids* **26**:629-634.

Booth SA, Drebot MA, Martin IE, and Ng L-K. (2003). Design of Oligonucleotide arrays to detect point mutations: molecular typing of antibiotic resistant strains of Neisseria gonorrhoeae and hantavirus infected deer mice. *Mol. Cell. Probes* **17**:77–84.

Brandes JC, Engeland VM, Wouters KA, Weijenberg MP, and Herman JG. (2005). CHFR promoter hypermethylation in colon cancer correlates with the microsatellite instability phenotype. *Carcinogenesis* **26**:1152-1156.

Broude NE, Woodward K, Cavallo R, Cantor CR, and Englert D. (2001). DNA microarrays with stem-loop DNA probes: preparation and applications. *Nucleic Acids Res.* **29**:e92.

Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR, et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**:33-43.

Catania J, Keenan BC, Margison GP, and Fairweather DS. (1987). Determination of 5-methylcytosine by acid hydrolysis of DNA with hydrofluoric acid. *Anal. Biochem.* **167**:347-351.

Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* **116**:499-509.

Cedar H, Solage A, Glaser G, and Razin A. (1979). Direct detection of methylated cytosine in DNA by use of the restriction enzyme MspI. *Nucleic Acids Res.* **6**:2125-2132.

Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, and Fodor SP. (1996). Accessing genetic information with high-density DNA arrays. *Science* **274**:610−614.

Chen WD, Han ZJ, Skoletsky J, Olson J, Sah J, Myeroff L, Platzer P, Lu S, Dawson D, Markowitz SD, et al. (2005). Detection in fecal DNA of colon cancer-specific methylation of the nonexpressed vimentin gene. *J. Natl. Cancer Inst.* **97**:1124-1132.

Cheng Y, Shawber C, Notterman D, Paty P, and Barany F. (2006). Multiplexed profiling of candidate genes for CpG island methylation status using a flexible

PCR/LDR/Universal Array assay. *Genome Res*. **16**:282-289.

Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K, and McCann PP. (1996). S-Adenosylmethionine and methylation. *FASEB J*. **10**:471–480.

Colella S, Shen L, Baggerly KA, Issa JP, and Krahe R. (2003). Sensitive and quantitative universal Pyrosequencing methylation analysis of CpG sites. *Biotechniques* **35**:146-150.

Colot V, Maloisel IL, and Rossignol J-L. (1996). Interchromosomal transfer of epigenetic states in Ascobolus: Transfer of DNA methylation is mechanistically related to homologous recombination. *Cell* **86**:855–864.

Costello JF, Fruhwald MC, Smiraglia DJ, Rush LJ, Robertson GP, Gao X, Wright FA, Feramisco JD, Peltomaki P, Lang JC, et al. (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat. Genet.* **24**:132-138.

Costello JF and Plass C. (2001). Methylation matters. *J. Med. Genet.* **38**:285-303.

Cottrell SE and Laird PW. (2003). Sensitive detection of DNA methylation. *Ann. NY Acad. Sci.* **983**:120-130.

Dahl C and Guldberg PA. (2007). A Ligation assay for multiplex analysis of CpG methylation using bisulfite-treated DNA. *Nucleic Acids Res.* **35**:e144.

Dawson ED, Reppert AE, Rowlen KL, and Kuck LR. (2005). Spotting optimization for oligo microarrays on aldehyde-glass. *Anal. Biochem.* **341**:352-360.

Deyong Z, Willingmann P, Heinze C, Adam G, Pfunder M, Frey B, and Frey JE. (2005). Differentiation of Cucumber mosaic virus isolates by hybridization to Oligonucleotides in a microarray format, *J. Virol. Methods* **123**:101–108.

Doerfler W. (1983). DNA methylation and gene activity. *Annu. Rev. Biochem.* 2: 93-124.

Eden A, Gaudet F, Waghmare A. and Jaenisch R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300**:455.

Drmanac R, Labat I, Brukner I, and Crkvenjakov R. (1989). Sequencing of megabase plus DNA by hybridization: theory of the method. *Genomics* **4**:114-128.

Drmanac R, Drmanac S, Strezoska Z, Paunesku T, Labat I, Zeremski M, Snoddy J, Funkhouser WK, Koop B, Hood L, et al. (1993). DNA sequence determination by hybridization: a strategy for efficient large-scale sequencing. *Science* **260**:1649-1652.

Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, Shibata D, Danenberg PD, and Laird PW. (2000). Methylight: a high-throughput assay to measure DNA methylation. *Nucleic Acids Res*. **32:**e10.

Eckhardt F, Lewin J, Cortese R, et al. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat. Genet.* **38**:1378–1385.

Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, and Gehrke CW. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res.* **10**:2709-2721.

Ehrlich M. (2002). DNA methylation in cancer: too much, but also too little. *Oncogene* **21**:5400–5413.

Eick D, Fritz HJ, and Doerfler W. (1983). Quantitative determination of 5-methylcytosine in DNA by reverse-phase high-performance liquid chromatography. *Anal. Biochem.* **135**:165-171.

El-Maarri O, Herbiniaux U, Walter J, and Oldenburg J. (2002). A rapid, quantitative,

non-radioactive bisulfite-SNuPE-IP RP HPLC assay for methylation analysis at specific CpG sites. *Nucleic Acids Res.* **30**:e25

Enokida H, Shiina H, Urakami S, Igawa M, Ogishima T, Li LC, Kawahara M, Nakagawa M, Kane CJ, Dahiya R, et al. (2005). Multigene methylation analysis for detection and staging of prostate cancer. *Clin. Cancer Res.* **11**:6582-6588.

Erdogan F, Kirchner R, Mann W, Ropers HH, and Nuber UA. (2001). Detection of mitochondrial single nucleotide polymorphisms using a primer elongation reaction on oligonucleotide microarrays. *Nucleic Acids Res.* **29**:e36.

Esteller M, Sparks A, Toyota M, Sanchez-Cespedes M, Capella G, Peinado MA, Gonzalez S, Tarafa G, et al. (2000). Analysis of *adenomatous polyposis coli* promoter hypermethylation in human cancer. *Cancer Res.* **60**:4366-4371.

Fan J-B, Chen X, Halushka MK, Berno A, Huang X, Ryder T, Lipshutz RJ, Lockhart DJ, and Chakravarti A. (2000). Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res.* **10**:853-860.

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

Favis R, Day JP, Gerry NP, Phelan C, Narod S, and Barany F. (2000). Universal DNA array detection of small insertions and deletions in BRCA1 and BRCA2. *Nat. Biotechnol.* **18**:561-564.

FavisR, Huang J, Gerry NP, Culliford A, Paty P, Soussi T, and Barany F. (2004). Harmonized microarray/mutation scanning analysis of TP53 mutations in undissected colorectal tumors. *Hum. Mutat.* **24**:63-75.

Feinberg AP and Vogelstein B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301:**89–92.

Fiala ES, Staretz ME, Pandya GA, El-Bayoumy K, and Hamilton SR. (1998). Inhibition of DNA cytosine methyltransferase by chemopreventive selenium compounds, determined by an improved assay for DNA cytosine methyltransferase and DNA cytosine methylation. *Carcinogenesis* **19**:597-604.

Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, and Solas D. (1991). Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**:767–773.

Fraga MF and Esteller M. (2002). DNA Methylation: A Profile of Methods and Applications. *BioTechniques,* **33**:632-649.

Fraga MF, Rodríguez R, Cañal MJ. (2000). Rapid quantification of DNA methylation by high performance capillary electrophoresis. *Electrophoresis* **21**:2990-2994.

Friso S, Choi SW, Dolnikowski GG, and Selhub J. (2002). A method to assess genomic DNA methylation using high-performance liquid chromatography/electrospray ionization mass spectrometry. *Anal. Chem.* **74**:4526-4531.

Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, and Paul CL. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci.* **89**:1827-1831.

Gama-Sosa, MA, Midgett RM, Slagel VA, Githens S, Kuo KC, Gehrke CW, and Ehrlich M. (1983). Tissue-specific differences in DNA methylation in various mammals. *Biochim. Biophys. Acta* **740**:212-219.

Gardiner-Garden M and Frommer M. (1987). CpG islands in vertebrate genomes. *J. Mol.*

Biol. **196**:261-282.

Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J, Gray JW, Leonhardt H, and Jaenisch R. (2003). Induction of tumors in mice by genomic hypomethylation. *Science* **300**:489–492.

Gerry NP, Witowski NE, Day J, Hammer RP, Barany G, and Barany F. (1999). Universal DNA microarray method for multiplex detection of low abundance point mutations. *J. Mol. Biol.* **292**:251-262.

Gitan RS, Shi H, Chen CM, Yan PS, and Huang TH. (2002). Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res.* **12**:158-164.

Gomes JD and Chang CJ. (1983). Reverse-phase high-performance liquid chromatography of chemically modified DNA. *Anal. Biochem.* **129**:387–391.

Gonzalgo ML and Jones PA. (1997). Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE) *Nucleic Acids Res.* **25**:2529–2531.

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

Griffin TJ, Tang W, and Smith LM. (1997). Genetic analysis by peptide nucleic acid affinity ALDI-TOF mass spectrometry. *Nat. Biotechnol.* **15**:1368-1372.

Grunau C, Clark SJ, and Rosenthal A. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.* **29**:e65.

Gunderson KL, Huang XC, Morris MS, Lipshutz RJ, Lockhart DJ, and Chee MS. (1998). Mutation detection by ligation to complete n-mer DNA arrays. *Genome Res.*

8:1142-1153.

Hanada M, Delia D, Aiello A, Stadtmauer E, and Reed JC. (1993). bcl-2 gene hypomethylation and high-level expression in B-cell chronic lymphocytic leukemia. *Blood* **82**:1820–1828.

Hacia JG, Brody LC, Chee MS, Fodor SP, and Collins FS. (1996). Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. *Nat. Genet.* **14**:441-447.

Hacia JG, Makalowski W, Edgemon K, Erdos MR, Robbins CM, Fodor SP, Brody LC, and Collins FS. (1998). Evolutionary sequence comparisons using high-density oligonucleotide arrays. *Nat. Genet.* **18**:155-158.

Hacia JG, Fan J-B, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, et al. (1999). Determination of ancestral alleles for human singlenucleotide polymorphisms using high-density oligonucleotide arrays. *Nat. Genet.* **22**:164-167.

Halushka MK, Fan J-B, Bentley K, Hsie L, Shen N, Weder A, Cooper R, Lipshutz R, and Chakravarti A. (1999). Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. *Nat. Genet.* **22**:239-247.

Harden SV, Guo Z, Epstein JI, and Sidransky D. (2003). Quantitative GSTP1 methylation clearly distinguishes benign prostatic tissue and limited prostate adenocarcinoma. *J. Urol.* **169**:1138-1142.

Hashimoto M, Hupert ML, Murphy MC, Soper SA, Cheng YW, and Barany F. (2005). Ligase detection reaction/hybridization assays using three-dimensional microfluidic networks for the detection of low-abundant DNA point mutations.

Anal. Chem. **77**:3243-3255.

Hatada I, Hayashizaki Y, Hirotsune S, Komatsubara H, and Mukai T. (1991). A genomic scanning method for higher organisms using restriction sites as landmarks. *Proc. Natl. Acad. Sci.* **88**:9523–9527.

Hatada I, Kato A, Morita S, Obata Y, Nagaoka K, Sakurada A, Sato M, Horii A, Tsujimoto A, and Matsubara K. (2002). A microarray-based method for detecting methylated loci. *J. Hum. Genet.* **47**:448-451.

Hatada I, Fukasawa M, Kimura M, Morita S, Yamada K, Yoshikawa T, Yamanaka S, Endo C, Sakurada A, Sato M, et al. (2006). Genome-wide profiling of promoter methylation in human. *Oncogene* **25**:3059-3064.

Havlis J and Trgbusek M. (2002). 5-Methylcytosine as a marker for the monitoring of DNA methylation. *J. Chromatogr. B* **781**:373-392.

Hayashi H, Nagae G, Tsutsumi S, Kaneshiro K, Kozaki T, Kaneda A, Sugisaki H, and Aburatani H. (2007). High-resolution mapping of DNA methylation in human genome using oligonucleotide tiling array. *Hum. Genet.* **120**:701-711.

Hayatsu H, Wataya Y, Kai K, and Iida S. (1970) Reaction of sodium bisulfite with uracil, cytosine, and their derivatives. *Biochemistry* **9**:2858–2865.

Hayatsu H and Shiragami M. (1979) Reaction of bisulfite with the 5-hydroxymethyl group in pyrimidines and in phage DNAs. *Biochemistry* **18**:632–637.

Head SR, Rogers YH, Parikh K, Lan G, Anderson S, Goelet P, and Boyce-Jacino MT. (1997). Nested genetic bit analysis (N-GBA) for mutation detection in the p53 tumors uppressor gene. *Nucleic Acids Res.* **25**:5065-5071.

Herman JG, Graff JR, Myohanen S, Nelkin BD, and Baylin SB. (1996).

Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci.* **93:**9821-9826.

Hirschhorn JN, Sklar P, Lindblad-Toh K, Lim YM, Ruiz-Gutierrez M, Bolk S, Langhorst B, Schaffner S, Winchester E, and Lander ES. (2000). SBE-TAGS: an array-based method for efficient single-nucleotide polymorphism genotyping. *Proc. Natl. Acad. Sci.* **97**:12164-12169.

Hoen PA, de Kort F, van Ommen GJ, and den Dunnen JT. (2003). Fluorescent labelling of cRNA for microarray applications. *Nucleic Acids Res.* **31**:e20.

Hoque MO, Begum S, Topaloglu O, Chatterjee A, Rosenbaum E, Van Criekinge W, Westra WH, Schoenberg M, Zahurak M, Sidransky DJ, et al. (2006). Quantitation of promoter methylation of multiple genes in urine DNA and bladder cancer detection. *Natl. Cancer Inst.* **98**:996-1004.

Hong KM, Yang SH, Guo M, Herman JG, and Jen J. (2005). Semiautomatic detection of DNA methylation at CpG islands. *Biotechniques* **38**:354-358.

Hou P, Ji M, Li S, He N, and Lu Z. (2004). High-throughput method for detecting DNA methylation, *J. Biochem. Biophys. Methods* **60**:139–150.

Hu G, Wang H-Y, Greenawalt DM, Azaro MA, Luo M, Tereshchenko IV, Cui X, Yang Q, Gao R, Shen L, et al. (2006). AccuTyping: new algorithms for automated analysis of data from high-throughput genotyping with oligonucleotide microarrays. *Nucleic Acids Res.* **34**:e116.

Huang TH, Perry MR, and Laux DE. (1999). Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet.* **8**:459–470.

Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM,

Ziman M, Schelter JM, Meyer MR, et al. (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer, *Nat. Biotechnol.***19**:342– 347.

Hultin E, Kaller M, Ahmadian A, and Lundeberg J. (2005). Competitive enzymatic reaction to control allele-specific extensions. *Nucleic Acids Res.* **33**:e48.

Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, et al. (2004). A tiling resolution DNA microarray with complete coverage of the human genome. *Nat. Genet.* **36**:299-303.

Issa JP, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, and Baylin SB. (1997). Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat. Genet.* **7:**536-540.

Jerónimo C, Usadel H, Henrique R, Oliveira J, Lopes C, Nelson WG, and Sidransky D. (2001). Quantitation of GSTP1 methylation in non-neoplastic prostatic tissue and organ-confined prostate adenocarcinoma. *J. Natl. Cancer. Inst.* **93**:1747-1752.

Ji H and Wong WH. (2005). TileMap: create chromosomal map of tiling array hybridizations. *Bioinformatics* **21**:3629-3636.

Jones PA and Laird PW. (1999). Cancer epigenetics comes of age. *Nat. Genet.*, **21**:163–167.

Jones PA and Baylin SB. (2002). The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3**:415-28.

Jones PA and Baylin SB. (2007). The epigenomics of cancer. *Cell* **128:**683–692.

Kaminsky ZA, Assadzadeh A, Flanagan J, and Petronis A. (2005). Single nucleotide extension technology for quantitative site-specific evaluation of metC/C in GC-rich regions. *Nucleic Acids Res.* **330**:e95.

Karpf AR and Jones DA. (2002). Reactivating the expression of methylation silenced genes in human cancer. *Oncogene* **21**:5496-5503.

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

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

Kim JY, Tavaré S, and Shibata D. (2005).Counting human somatic cell replications: methylation mirrors endometrial stem cell divisions. *Proc. Natl. Acad. Sci.* **102**:17739-17744.

Kimura N, Nagasaka T, Murakami J, Sasamoto H, Murakami M, Tanaka N, and Matsubara N. (2005). Methylation profiles of genes utilizing newly developed CpG island methylation microarray on colorectal cancer patients. *Nucleic Acids Res.* **33:**e46.

Kinoshita K, Fujimoto K, T Yakabe, Saito S, Hamaguchi Y, Kikuchi T, Nonaka K, S. Murata, Masuda D, Takada W, et al. (2007). Multiple primer extension by DNA polymerase on a novel plastic DNA array coated with a biocompatible polymer. *Nucleic Acids Res.* **35**:e3.

Kitazawa S, Kitazawa R, and Maeda S. (2000). Identification of methylated cytosine

from archival formalin-fixed paraffin-embedded specimens. *Lab. Invest.* **80**:275–276.

Kondo Y, Shen L, and Issa JP. (2003). Critical role of histone methylation in tumor suppressorgene silencing in colorectal cancer. *Mol. Cell Biol.* **23**:206–215.

Konno S, Morishita Y, Fukasawa M, Shu Y, Wang D, Tanaka R, Minami Y, Iijima T, and Noguchi M. (2004). Anthracotic index and DNA methylation status of sputum contents can be used for identifying the population at risk of lung carcinoma. *Cancer* **102**:348-354.

Korshunova Y, Maloney RK, Lakey N, Citek RW, Bacher B, Budiman A, Ordway JM, McCombie WR, Leon J, McPherson JD, et al. (2008). Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA. *Genome Res.* **18**:19-29.

Kreil DP, Russell RR, and Russell S. (2006). Microarray oligonucleotide-probes. *Methods Enzymol.* **410**:73-98.

Kristensen LS, Mikeska T, Krypuy M, and Dobrovic A. (2008). Sensitive Melting Analysis after Real Time-Methylation Specific PCR (SMART-MSP): high-throughput and probe-free quantitative DNA methylation detection. *Nucleic Acids Res.* **36**:e42.

Kuo KC, McCune RA, Gehrke CW, Midgett R, and Ehrlich M. (1980). Quantitative reversed-phase high performance liquid chromatographic determination of major and modified deoxyribonucleosides in DNA. *Nucleic Acids Res.* **8**:4763–4776.

Kusano M, Toyota M, Suzuki H, Akino K, Aoki F, Fujita M, Hosokawa M, Shinomura Y,

Imai K, and Tokino T. (2006). Genetic, epigenetic, and clinicopathologic features of gastric carcinomas with the CpG island methylator phenotype and an association with Epstein-Barr virus. *Cancer* **106:** 1467-1479.

Kutueva, LI, Ashapkin VV, and Vanyushin BF. (1996). The methylation pattern of a cytosine DNA-methyltransferase gene in *Arabidopsis thaliana* plants. Biochem. *Mol. Biol. Int.* **40**:347-353.

Laird PW. (2003). The power and the promise of DNA methylation markers. *Nat. Rev. Cancer* **3**:253-266.

Lander ES, Linton LM, Birren B, et al. (2001). Initial sequencing and analysis of the human genome. *Nature* **409**:860–921.

Leonard SA, Wong SC, and Nyce JW. (1993). Quantitation of 5-methylcytosine by onedimensional high-performance thin-layer chromatography.*J. Chromatogr.* **645**:189-192.

Levine JJ, Stimson-Crider KM, and Vertino PM. (2003). Effects of methylation on expression of TMS1/ASC in human breast cancer cells. *Oncogene* **22**:3475-3488.

Li E, Beard C, and Jaenisch R. (1993). Role for DNA methylation in genomic imprinting. *Nature* **366**:362-365.

Li W, Meyer CA, and Liu XS. (2005). A hidden Markov model for analyzing ChIP-chip experiments on genome tiling arrays and its application to p53 binding sequences. *Bioinformatics* Suppl **1**:274-282.

Liang G, Robertson KD, Talmadge C, Sumegi J, and Jones PA. (2000). The gene for a novel transmembrane protein containing epidermal growth factor and follistatin

domains is frequently hypermethylated in human tumor cells. *Cancer Res.* **60**:4907–4912.

Lindroos K, Liljedahl U, Raitio M, and Syvänen AC. (2001). Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries. *Nucleic Acids Res.* **29**:e69.

Lindroos K, Sigurdsson S, Johansson K, Ronnblom L, and Syvänen AC. (2002). Multiplex SNP genotyping in pooled DNA samples by a four-colour microarray system. *Nucleic Acids Res*. **30**:e70.

Lindroos K, Liljedahl U, and Syvänen AC. (2003). Genotyping SNPs by minisequencing primer extension using oligonucleotide microarrays. *Methods Mol. Biol.* **212**:149-165.

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**:471-476.

Lippman Z, Gendrel AV, Colot V, and Martienssen R. (2005). Profiling DNA methylation patterns using genomic tiling microarrays. *Nat. Methods* **2**:219-224.

Liu JW, Kim MS, Nagpal J, Yamashita K, Poeta L, Chang X, Lee J, Park HL, Jeronimo C, Sidransky D, et al. (2007). Quantitative hypermethylation of NMDAR2B in human gastric cancer. *Int. J. Cancer* **121:**1994-2000.

Lo YM, Wong IH, Zhang J, Tein M, Ng MH, and Hjelm NM. (1999). Quantitative analysis of aberrant p16 methylation using real-time quantitative methylation-specific polymerase chain reaction. *Cancer Res.* **59**:3899–3903.

Lofton-Day C, Model F, Devos T, Tetzner R, Distler J, Schuster M, Song X, Lesche R,

Liebenberg V, Ebert M, et al. (2008). DNA methylation biomarkers for blood-based colorectal cancer screening. *Clin. Chem.* **54**:414-423.

Lovmar L, Fredriksson M, Liljedahl U, Sigurdsson S, and Syvänen AC. (2003). Quantitative evaluation by minisequencing and microarrays reveals accurate multiplexed SNP genotyping of whole genome amplified DNA. *Nucleic Acids Res*. **31**:e129.

Lyko F, Stach D, Brenner A, Stilgenbauer S, Dohner H, Wirtz M, Wiessler M, and Schmitz OJ. (2004). Quantitative analysis of DNA methylation in chronic lymphocytic leukemia patients. *Electrophoresis* **25**:1530-1535.

Margulies M, Egholm M, Altman WE, et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**:376–380.

Maris JM, Hii G, Gelfand CA, Varde S, White PS, Rappaport E, Surrey S, and Fortina P. (2005). Region-specific detection of neuroblastoma loss of heterozygosity at multiple loci simultaneously using a SNP-based tag-array platform. *Genome Res.* **15**:1168-1176.

Martens JW, Nimmrich I, Koenig T, Look MP, Harbeck N, Model F, Kluth A, Bolt-de Vries J, Sieuwerts AM, Portengen H, et al. (2005). Association of DNA methylation of phosphoserine aminotransferase with response to endocrine therapy in patients with recurrent breast cancer. *Cancer Res.* **65:**4101-4117.

Mashayekhi F and Ronaghi M. (2007). Analysis of read length limiting factors in Pyrosequencing chemistry. *Anal. Biochem.* **363**:275-287.

Matsuyama T, Kimura MT, Koike K, Abe T, Nakano T, Asami T, Ebisuzaki T, Held WA, Yoshida S, and Nagase H. (2003). Global methylation screening in the

Arabidopsis thaliana and Mus musculus genome: applications of virtual image restriction landmark genomic scanning (Vi-RLGS). *Nucleic Acids Res.* **31**:4490-4496.

Melnikov AA, Scholtens DM, Wiley EL, Khan SA, and Levenson VV. (2008). Array-based multiplex analysis of DNA methylation in breast cancer tissues. *J. Mol. Diagn.*10:93-101.

Miller OJ, Schnedl W, Allen J, and Erlanger BF (1974). 5-Methylcytosine localisation in mammalian constitutive heterochromatin. *Nature* **251**:236–237.

Mito Y, Henikoff JG, and Henikoff S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. *Nat. Genet.* **37**:1090-1097.

Mito Y, Henikoff JG, and Henikoff S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. *Science* **315**:1408-1411.

Montpellier C, Bourgeois CA, Kokallj-Vokac N, Muleris M, Niveleau A, Reynaud C, Gibaud A, Malfoy B, and Dutrillaux B. (1994). Detection of methylcytosine-rich heterochromatin on banded chromosomes. Application to cells with various of DNA methylation. *Cancer Genet. Cytogenet.* **78**:87–93.

Moreno V, Gemignani F, Landi S, Gioia-Patricola L, Chabrier A, Blanco I, González S, Guino E, Capellà G, and Canzian F. (2006). Polymorphisms in genes of nucleotide and base excision repair: risk and prognosis of colorectal cancer. *Clin. Cancer Res.* **12**:2101-2108.

Mori Y, Cai K, Cheng Y, Wang S, Paun B, Hamilton JP, Jin Z, Sato F, Berki AT, Meltzer SJ, et al. (2006). A genome-wide search identifies epigenetic silencing of somatostatin, tachykinin-1, and 5 other genes in colon cancer. *Gastroenterology*

131:797-808.

Mund C, Beier V, Bewerunge P, Dahms M, Lyko F, and Hoheisel JD. (2005) Array-based analysis of genomic DNA methylation patterns of the tumour suppressor gene p16INK4A promoter in colon carcinoma cell lines. *Nucleic Acids Res*. **33**:e73.

Munson K, Clark J, Lamparska-Kupsik K, and Smith SS. (2007). Recovery of bisulfite-converted genomic sequences in the methylation-sensitive QPCR. *Nucleic Acids Res.* **35**:2893-2903.

Nygren AO, Ameziane N, Duarte HM, Vijzelaar RN, Waisfisz Q, Hess CJ, Schouten JP, and Errami A. (2005). Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. *Nucleic Acids Res.* **33**:e128.

Oakeley EJ, Podestà A, and Jost JP. (1997). Developmental changes in DNA methylation of the two tobacco pollen nuclei during maturation. *Proc. Natl. Acad. Sci.* **94**:11721-11725.

Oakeley EJ, Schmitt F, and Jost JP. (1999). Quantification of 5-methylcytosine in DNA by the chloroacetaldehyde reaction. *BioTechniques* **27**:744-752.

Oakeley EJ. (1999). DNA methylation analysis: a review of current methodologies. *Pharmacol. Ther.* **84**:389-400.

Odom DT, Zizlsperger N, Gordon DB, Bell GW, Rinaldi NJ, Murray HL, Volkert TL, Schreiber J, Rolfe PA, Gifford DK et al. (2004). Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**:1378-1381.

Okamoto A, Tainaka K, and Kamei T. (2006). Sequence-selective osmium oxidation of DNA: efficient distinction between 5-methylcytosine and cytosine. *Org. Biomol.*

Chem. **4**:1638-1640.

Olek A, Oswald J, and Walter J. (1996). A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res.* **24**:5064-5066.

Ordway JM, Bedell JA, Citek RW, Nunberg A, Garrido A, Kendall R, Stevens JR, Cao D, Doerge RW, Korshunova Y, et al. (2006). Comprehensive DNA methylation profiling in a human cancer genome identifies novel epigenetic targets. *Carcinogenesis* **27**:2409-2423.

Pack SP, Kamisetty NK, Nonogawa M, Devarayapalli KC, Ohtani K, Yamada K, Yoshida Y, Kodaki T, and Makino K. (2007). Direct immobilization of DNA oligomers onto the amine-functionalized glass surface for DNA microarray fabrication through the activation-free reaction of oxanine. *Nucleic Acids Res.* **35**:e110.

Panning B and Jaenisch R. (1998). RNA and the epigenetic regulation of X chromosome inactivation. *Cell* **93**:305-308.

Pastinen T, Kurg A, Metspalu A, and Peltonen L. (1997). Minisequencing: A Specific Tool for DNA Analysis and Diagnostics on Oligonucleotide Arrays. *Genome Res*. **7**: 606-614.

Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, and Syvänen AC. (2000). A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res.* **10**:1031-1042.

Pastinen T, Perola M, Ignatius J, Sabatti C, Tainola P, Levander M, Syvänen AC, and Peltonen L. (2001). Dissecting a population genome for targeted screening of disease mutations. *Hum. Mol. Genet.* **10**:2961-2972.

Paulsen M and Ferguson-Smith AC. (2001). DNA methylation in genomic imprinting,

development, and disease. *J. Pathol.* **195**:97–110.

Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, and Fodor SP. (1994). Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci.* **91**:5022–5026.

Peng DF, Kanai Y, Sawada M, Ushijima S, Hiraoka N, Kitazawa S, and Hirohashi S. (2006). DNA Methylation of Multiple Tumor-related Genes in Association with Overexpression of DNA Methyltransferase 1 (DNMT1) during Multistage Carcinogenesis of the Pancreas, *Carcinogenesis* **27**:1160-1168.

Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, and Fischer RL. (2007). DNA demethylation in the Arabidopsis genome. *Proc. Natl. Acad. Sci.* **104**:6752-6757.

Peters I, Rehmet K, Wilke N, Kuczyk MA, Hennenlotter J, Eilers T, Machtens S, Jonas U, and Serth J. (2007). RASSF1A promoter methylation and expression analysis in normal and neoplastic kidney indicates a role in early tumorigenesis. *Mol. Cancer.* **6**:49.

Piotrowski A, Benetkiewicz M, Menzel U, Ståhl TD, Mantriprogada K, Grigelionis G, Buckley PG, Jankowski M, Hoffman J, Bata D, et al. (2006). Microarray-based survey of CpG islands identifies concurrent hyper- and hypomethylation patterns in tissues derived from patients with breast cancer. *Genes Chromosomes Cancer.* **45:** 656-667.

Podestà A, Castiglione MR, Avanzi S, and Montagnoli G. (1993). Molecular geometry of antigen binding by a monoclonal antibody against 5-methylcytidine. *Int. J. Biochem.* **25**:929-933.

Pourmand N, Caramuta S, Villablanca A, Mori S, Karhanek M, Wang SX, and Davis RW. (2007). Branch migration displacement assay with automated heuristic analysis for discrete DNA length measurement using DNA microarrays. *Proc. Natl. Acad. Sci.* **104**:6146-6151.

Pozhitkov A, Noble PA, Domazet-Loso T, Nolte AW, Sonnenberg R, Staehler P, Beier M, and Tautz D. (2006). Tests of rRNA hybridization to microarrays suggest that hybridization characteristics of oligonucleotide probes for species discrimination cannot be predicted.   *Nucleic Acids Res*. **34**:e66.

Raizis AM, Schmitt F, and Jost JP. (1995) A bisulfite method of 5-methylcytosine mapping that minimizes template degradation. *Anal. Biochem.* **226**:161-166.

Rakyan VK, Hildmann T, Novik KL, et al. (2004). DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol.* 2:e405.

Rand K, Qu W, Ho T, Clark SJ, and Molloy P. (2002). Conversion-specific detection of DNA methylation using real-time polymerase chain reaction (ConLight-MSP) to avoid false positives. *Methods* **27**:114–120.

Rashid A, Shen L, Morris SJ, Issa JP, and Hamilton SR. (2001). CpG island methylation in colorectal adenomas. *Am. J .Pathol.* **159:**1129-135.

Rauch T, Wang Z, Zhang X, Zhong X, Wu X, Lau SK, Kernstine KH, Riggs AD, and Pfeifer GP. (2007). Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay. *Proc. Natl. Acad. Sci.* **104**:5527-5532.

Rauch TA, Zhong X, Wu X, Wang M, Kernstine KH, Wang Z, Riggs AD, and Pfeifer GP.

(2008). High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc. Natl. Acad. Sci.* **105**:252-257.

Reinders J, Delucinge Vivier C, Theiler G, Chollet D, Descombes P, and Paszkowski J. (2008). Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. *Genome Res.* **18**:469-476.

Rinn JL, Euskirchen G, Bertone P, Martone R, Luscombe NM, Hartman S, Harrison PM, Nelson FK, Miller P, Gerstein M, et al. (2003). The transcriptional activity of human Chromosome 22. *Genes Dev.* **17**:529-540.

Robertson KD. (2001). DNA methylation, methyltransferases, and cancer. *Oncogene* **20**:3139-3155.

Robertson KD. (2005). DNA methylation and human disease. *Nat. Rev. Genet.* **6**:597-610.

Ronaghi M, Karamohamed S, Pettersson B, Uhlen M, and Nyren P. (1996). Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**:84-89.

Ronaghi M, Uhlen M, and Nyren P. (1998). A sequencing method based on real-time pyrophosphate. *Science* **281**:363-365.

Rogers YH, Jiang-Baucom P, Huang ZJ, Bogdanov V, Anderson S, and Boyce-Jacino MT. (1999). Immobilization of oligonucleotides onto a glass support via disulfide bonds: A method for preparation of DNA microarrays. *Anal. Biochem.* **266**:23–30.

Rougier N, Bourc'his D, Gomes DM, Niveleau A, Plachot M, Pàldi A, and Viegas-Péquignot E. (1998). Chromosome methylation patterns during

mammalian preimplantation development. *Genes Dev.* **12**:2108-2113.

Rouillard JM, Erson AE, Kuick R, Asakawa J, Wimmer K, Muleris M, Petty EM, and Hanash S. (2001). Virtual genome scan: a tool for restriction landmark-based scanning of the human genome. *Genome Res.* **11**:1453–1459.

Royce TE, Rozowsky JS, Bertone P, Samanta M, Stolc V, Weissman S, Snyder M, and Gerstein M. (2005). Issues in the analysis of oligonucleotide tiling microarrays for transcript mapping. *Trends Genet.* **21**:466-475.

Rusha LJ and Plass C. (2002). Restriction landmark genomic scanning for DNA methylation in cancer: past, present, and future applications. *Anal. Biochem.* **307**:191–201.

Sadri R and Hornsby PJ. (1996). Rapid analysis of DNA methylation using new restriction enzyme sites created by bisulfite modification. *Nucleic Acids Res.* **24**:5058–5059.

Sandoval Guerrero K, Revilla Vazquez A, Segura-Pacheco B, and Duenas-Gonzalez A. (2005). Determination of 5-methyl-cytosine and cytosine in tumor DNA of cancer patients *Electrophoresis* **26**:1057-1062.

Sanger F, Nicklen S, and Coulson AR. (1977). DNA **sequencing** with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74:**5463-5467.

Sato N, Fukushima N, Maitra A, Matsubayashi H, Yeo CJ, Cameron JL, Hruban RH, and Goggins M. (2003). Discovery of novel targets for aberrant methylation in pancreatic carcinoma using high-throughput microarrays. *Cancer Res.* **63**:3735-3742.

Sato F and Meltzer SJ. (2005). CpG island hypermethylation in progression of

esophageal and gastric cancer. *Cancer* **106**:483-493.

Schatz P, Schuster M, Berlin K, Tost J, and Gut IG. (2003). Analysis and accurate quantification of CpG methylation by MALDI mass spectrometry. *Nucleic Acids Res.* **31**:e50.

Schena M, Shalon D, Davis RW, and Brown PO. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**:467-470.

Schmitt F, Oakeley EJ, and Jost JP. (1997). Antibiotics induce genome-wide hypermethylation in cultured Nicotiana tabacum plants. *J. Biol. Chem.* **272**:1534-1540.

Schöb H and Grossniklaus U. (2006). The first high-resolution DNA "methylome". *Cell* **126**:1025-1028.

Schuebel KE, Chen W, Cope L, Glöckner SC, Suzuki H, Yi JM, Chan TA, Van Neste L, Herman JG, Baylin SB, et al. (2007). Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. *PLoS. Genet.* **3**:1709-1723.

Schumacher A, Kapranov P, Kaminsky Z, Flanagan J, Assadzadeh A, Yau P, Virtanen C, Winegarden N, Cheng J, Gingeras T, et al. (2006). Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res.* **34**:528 -542.

Shann YJ, Cheng C, Chiao CH, Chen DT, Li PH, and Hsu MT. (2008). Genome-wide mapping and characterization of hypomethylated sites in human tissues and breast cancer cell lines. *Genome Res.* **18**:791-801.

Shapiro R, Servis RE, and Welcher M. (1970). Reactions of uracil and cytosine

derivatives with sodium bisulphite. A specific deamination method. *J. Am. Chem. Soc.* **92**:422–424.

Shaw RJ, Akufo-Tetteh EK, Risk JM, Field JK, and Liloglou T. (2006). Methylation enrichment pyrosequencing: combining the specificity of MSP with validation by pyrosequencing. *Nucleic Acids Res.* **34**:e78.

Shibata H, Yoshino K, Muramatsu M, Plass C, Chapman VM, and Hayashizaki Y. (1995). The use of restriction landmark genomic scanning to scan the mouse genome for endogenous loci with imprinted patterns of methylation. *Electrophoresis* **16**:210-217.

Shoemaker DD, Lashkari DA, Morris D, Mittmann M, and Davis RW. (1996). Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nat. Genet*. **14**:450-456.

Shumaker, JM, Metspalu A, and Caskey CT. (1996). Mutation detection by solid phase primer extension. *Hum. Mutat.* **7**:346-354.

Singer-Sam J, Roberts-Ems J, and Riggs AD. (1979). Methylation of mouse liver DNA studied by means of the restriction enzymes msp I and hpa II. *Science* **203**:1019–1021.

Singer-Sam J, Grant M, LeBon JM, Okuyama K, Chapman V, Monk M, and Riggs AD. (1990). Use of a *Hpa*II-polymerase chain reaction assay to study DNA methylation in the Pgk-1 CpG island of mouse embryos at the time of X-chromosome inactivation. *Mol. Cell. Biol.* **10**:4987-4989.

Singh-Gasson S, Green RD, Yue Y, Nelson C, Blattner F, Sussman MR, and Cerrina F. (1999). Maskless fabrication of light-directed oligonucleotide microarrays using

a digital micromirror array. *Nat. Biotechnol.* **17**:974–978.

Smiraglia DJ and Plass C. (2002). The study of aberrant methylation in cancer via restriction landmark genomic scanning. *Oncogene* **21**:5414-5426.

Smiraglia DJ, Kazhiyur-Mannar R, Oakes CC, Wu YZ, Liang P, Ansari T, Su J, Rush LJ, Smith LT, Yu L, et al. (2007). Restriction landmark genomic scanning (RLGS) spot identification by second generation virtual RLGS in multiple genomes with multiple enzyme combinations. *BMC Genomics* **8**:446.

Snijders AM, Nowak N, Segraves R, Blackwood S, Brown N, Conroy J, Hamilton G, Hindle AK, Huey B, Kimura K, et al. (2001). Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.* **29**:263-264.

Song JY, Park HG, Jung SO, and Park J. (2005). Diagnosis of HNF-1alpha mutations on a PNA zip-code microarray by single base extension. *Nucleic Acids Res.* **33**:e19.

Song L, James SR, Kazim L, and Karpf AR. (2005). Specific method for the determination of genomic DNA methylation by liquid chromatography-electrospray ionization tandem mass spectrometry. *Anal. Chem.* **77**:504-510.

Sonoki S, Lin J, and Hisamatsu S. (1998). Determination of mevalonic acid in human urine by liquid chromatography with precolumn derivatization and fluorescence detection. *Anal. Chim. Acta.***365**:213-217.

Sotgia S, Carru C, Franconi F, Fiori PB, Manca S, Pettinato S, Magliona S, Ginanneschi R, Deiana L, and Zinellu A. (2008). Rapid quantification of total genomic DNA methylation degree by short-end injection capillary zone electrophoresis. *J. Chromatogr. A.* **1185**:145-150.

Southern EM, Maskos U, and Elder JK. (1992). Analyzing and comparing nucleic acid sequences by hybridization to arrays of Oligonucleotidenucleotides: evaluation using experimental models. *Genomics* **13**:1008-1017.

Stirzaker C, Millar DS, Paul CL, Warnecke PM, Harrison J, Vincent PC, Frommer M, and Clark SJ. (1997). Extensive DNA methylation spanning the Rb promoter in retinoblastoma tumors. *Cancer Res.* **57**:2229-2237.

Sun X, Ding H, Hung K, and Guo B. (2000). A new MALDI-TOF based mini-sequencing assay for genotyping of SNPS. *Nucleic Acids Res*. **28**:e68.

Suzuki H, Gabrielson E, Chen W, Anbazhagan R, van Engeland M, Weijenberg MP, Herman JG, and Baylin SB. (2002). A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. *Nat. Genet.* **31**:141-149.

Takahashi T, Suzuki M, Shigematsu H, Shivapurkar N, Echebiri C, Nomura M, Stastny V, Augustus M, Wu CW, Wistuba I, et al. (2005). Aberrant methylation of Reprimo in human malignancies. *Int. J. Cancer*, **115**:503–510.

Takahashi T, Shigematsu H, Shivapurkar N, Reddy J, Zheng Y, Feng Z, Suzuki M, Nomura M, Augustus M,   Yin J, et al. (2006). Aberrant promoter methylation of multiple genes during multistep pathogenesis of colorectal cancers. *Int. J. Cancer*, **118**:924–931.

Tanaka K, Tainaka K, Umemoto T, Nomura A, and Okamoto A. (2007). An osmium-DNA interstrand complex: application to facile DNA methylation analysis. *J. Am. Chem. Soc.* **129**:14511-14517.

Tanaka K, Tainaka K, Kamei T, and Okamoto A. (2007). Direct labeling of

5-methylcytosine and its applications. *J. Am. Chem. Soc.* **129**:5612-5620.

Tanaka K and Okamoto A. (2007). Degradation of DNA by bisulfite treatment. *Bioorg. Med. Chem. Lett.* **17**:1912-1915.

Taylor KH, Kramer RS, Davis JW, Guo J, Duff DJ, Xu D, Caldwell CW, and Shi H. (2007). Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. *Cancer Res.* **67**:8511–8518.

Tompa R, McCallum CM, Delrow J, Henikoff JG, van Steensel B, and Henikoff S. (2002). Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr. Biol.* **12**:65-68.

Tost J, Dunker J, and Gut IG. (2003). Analysis and quantification of multiple methylation variable positions in CpG islands by Pyrosequencing. *Biotechniques* **35**:152-156.

Toyooka S, Toyooka KO, Maruyama R, Virmani AK, Girard L, Miyajima K, Harada K, Ariyoshi Y, Takahashi T, Sugio K, et al. (2001). DNA Methylation Profiles of Lung Tumor. *Mol. Cancer Ther.* **1**:61–67.

Toyota M, Ho C, Ahuja N, Jair KW, Li Q, Ohe-Toyota M, Baylin SB, and Issa JP. (1999). Identification of differentially methylated sequences in colorectal cancer by methylated CpG island amplification. *Cancer Res.* **59**:2307-2312.

Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, and Issa JP. (1999). CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci,* **96**:8681–8686.

Tran RK, Henikoff JG, Zilberman D, Ditt RF, Jacobsen SE, and Henikoff S. (2005). DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes. *Curr. Biol.* **15**:154-159.

Tusnady GE, Simon I, Varadi A, and Aranyi T. (2005). BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Res.* **33**:e9.

Uhlmann K, Brinckmann A, Toliat MR, Ritter H, and Nürnberg P. (2002). Evaluation of a potential epigenetic biomarker by quantitative methyl-single nucleotide polymorphism analysis. *Electrophoresis* **23**:4072-4079.

Venter JC, Adams MD, Myers EW, et al. (2001). The sequence of the human genome. *Science* **291**:1304–51.

Wagner I and Capesius I. (1981). Determination of 5-methylcytosine from plant DNA by high-performance liquid chromatography. *Biochim. Biophys. Acta* **654**:52-56.

Walsh CP, Chaillet JR, and Bestor TH. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* **20**:116-117.

Wang DR and Tang D. (2008). Hypermethylated SFRP2 gene in fecal DNA is a high potential biomarker for colorectal cancer noninvasive screening. *World J. Gastroenterol.* **14**:524-531.

Warnecke PM and Bestor TH. (2000). Cytosine methylation and human cancer. *Curr. Opin. Oncol.* **12**:68–73.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, and Schübeler D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* **37**:853-862.

Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, and Schübeler D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**:457-466.

Winzeler EA, Shoemaker DD, Boeke JD, Bussey H, Friend S, Lockhardt D, Snyder M, Philippsen P, Johnston M, and Davis R. (1999). Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**:901-906.

Wirtz M, Stach D, Kliem HC, Wiessler M, and Schmitz OJ. (2004). Determination of the DNA methylation level in tumor cells by capillary electrophoresis and laser-induced fluorescence detection. *Electrophoresis* **25**:839-845.

Wolber PK, Collins PJ, Lucas AB, De Witte A, and Shannon KW. (2006). The Agilent in situ-synthesized microarray platform. *Methods Enzymol.* **410**:28-57.

Wrobel G, Schlingemann J, Hummerich L, Kramer H, Lichter P, and Hahn M. (2003). Optimization of high-density cDNA-microarray protocols by 'design of experiments'. *Nucleic Acids Res.* **31**:e67.

Wu J, Issa JP, Herman J, Bassett DE, Nelkin BD, and Baylin SB. (1993). Expression of an exogenous eukaryotic DNA methyltransferase gene induces transformation of NIH 3T3 cells. *Proc. Natl. Acad. Sci.* **90**:8891-8895.

Xiong Z and Laird PW. (1997). COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res*. **25**:2532–2534.

Yamada Y, Jackson-Grusby L, Linhart H, Meissner A, Eden A, Lin H, and Jaenisch R. (2005). Opposing effects of DNA hypomethylation on intestinal and liver carcinogenesis. *Proc. Natl. Acad. Sci.* **102:**13580–13585.

Yan PS, Chen CM, Shi H, Rahmatpanah F, Wei SH, Caldwell CW, and Huang TH. (2001). Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays.*Cancer Res.* **61**:8375-8380.

Yoder JA, Walsh CP, and Bestor TH. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**:335–340.

Yu J, Zhu T, Wang Z, Zhang H, Qian Z, Xu H, Gao B, Wang W, Gu L, Meng J, et al. (2007). A novel set of DNA methylation markers in urine sediments for sensitive/specific detection of bladder cancer. *Clin. Cancer Res.* **13**:7296-7304.

Yuan E, Haghighi F, White S, Costa R, McMinn J, Chun K, Minden M, and Tycko B. (2006). A single nucleotide polymorphism chip-based method for combined genetic and epigenetic profiling: validation in decitabine therapy and tumor/normal comparisons. *Cancer Res.* **66**:3443 -3451.

Zeschnigk M, Bohringer S, Price EA, Onadim Z, Masshofer L, and Lohmann DR. (2004). A novel real-time PCR assay for quantitative analysis of methylated alleles (QAMA): analysis of the retinoblastoma locus. *Nucleic Acids Res*. **32**:e125.

Zhang W, Bauer M, Croner RS, Pelz JO, Lodygin D, Hermeking H, Stürzl M, Hohenberger W, and Matzel KE. (2007). DNA stool test for colorectal cancer: hypermethylation of the secreted frizzled-related protein-1 gene. *Dis. Colon. Rectum.* **50**:1618-1626.

Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE, et al. (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* **126**:1189-1201.

Zilberman D. (2007). The human promoter methylome. *Nat. Genet*. **39**:442-443.

Zilberman D, Gehring M, Tran RK, Ballinger T, and Henikoff S. (2007). Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.* **39**:61-69.

Zilberman D and Henikoff S. (2007). Genome-wide analysis of DNA methylation patterns. *Development* **134**:3959-3965.

Zluvova J, Janousek B, and Vyskot B. (2001). Immunohistochemical study of DNA methylation dynamics during plant development. *J. Exp. Bot.* **52**:2265-2273.

APPENDIX

The sequences of 32 Zip-Code

1	TGGGTCCACAGTACCGCTGCAGA
2	CCGTGGGAGATTAGGTGGCTCAGA
3	GGGAATGGAGGTGGGAACGAGACA
4	CGCAAAGCAGACACAGGGTCGATT
5	GGGAGGCTGCTGTCCTTTCGATCA
6	AGACGCACCGCAACAGGCTGTCAA
7	TTGCAACGGGCTGGTCAACGTCAA
8	GCACCGATATGGAGACCGCAGACA
9	GGCTCGCAGGCTGGCTCATCCTAA
10	GGCTGCACCCGTTGAGGCACATCA
11	GTCCGTCCATGGCAAGCGTGATCA
12	ACAGCGTGTTCGTTGCTTGCATCA
13	CATCGCTGCAAGTACCGCACTCAA
14	GCACGGGAGCTGACGACGTGTCAA
15	CAAGGCACGTCCCAGACGCATCAA
16	CGCAGGTCGCTGCGTGTCCTGATT
17	GTCCCAAGTTGCGGCTCACTTTCG
18	AACGGGGAAGGTTGAGCGTGACAG
19	ATGGCGATGGTCCACTCGCAATCA
20	TTCGCCGTCGTGTAGGCTTTTCAA
21	CATCCAAGGTCCGACACGCAACGA
22	CGCATTCGATGGACAGGACATTCG
23	AAGCAAGCCAAGGTATGGCTTTGC
24	ACGAGGGATACCCGCAAACGATCA
25	CGTGGCTGACTCGCTGCGATGACA
26	CGCAATGGTAGGTGAGCAAGCAGA
27	ATGGCCGTGCTGGGGACAAGTCAA
28	GCTGGCTGGCACGCACCAGAATCA
29	CACTGCACACGAAACGGCACACAG
30	CGTGCGCACACTCACTGTCCTTCG
31	TTCGGCTGTTCGTAGGCAAGAGGT
32	GGCTGGGACGTGCAGACCGTTCAA

136