Multiplexed profiling of candidate genes for CpG island methylation status using a flexible PCR/LDR/Universal Array assay

Yu-Wei Cheng, Carrie Shawber, Dan Notterman, et al.

Genome Res. 2006 16: 282-289
Access the most recent version at doi:10.1101/gr.4181406

Multiplexed profiling of candidate genes for CpG island methylation status using a flexible PCR/LDR/Universal Array assay

Yu-Wei Cheng,[1] Carrie Shawber,[2] Dan Notterman,[3] Philip Paty,[4] and Francis Barany[1,5]

[1]*Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, New York 10021, USA;* [2]*Department of OB/GYN, Columbia University Medical Center, New York, New York 10032, USA;* [3]*Departments of Pediatrics and Molecular Genetics, University of Medicine and Dentistry of New Jersey (UMDNJ)-Robert Wood Johnson Medical School, New Brunswick, New Jersey 08901, USA;* [4]*Department of Surgery, Colorectal Surgery Service, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA*

DNA methylation in CpG islands is associated with transcriptional silencing. Accurate determination of cytosine methylation status in promoter CpG dinucleotides may provide diagnostic and prognostic value for human cancers. We have developed a quantitative PCR/LDR/Universal Array assay that allows parallel evaluation of methylation status of 75 CpG dinucleotides in the promoter regions of 15 tumor suppressor genes (*CDKN2B*, *CDKN2A*, *CDKN2D*, *CDKN1A*, *CDKN1B*, *TP53*, *BRCA1*, *TIMP3*, *APC*, *RASSF1*, *CDH1*, *MGMT*, *DAPK1*, *GSTP1*, and *RARB*). When compared with an independent pyrosequencing method at a single promoter, the two approaches gave good correlation. In a study using 15 promoter regions and seven blinded tumor cell lines, our technology was capable of distinguishing methylation profiles that identified cancer cell lines derived from the same origins. Preliminary studies using 96 colorectal tumor samples and 73 matched normal tissues indicated CpG methylation is a gene-specific and nonrandom event in colon cancer. This new approach is suitable for clinical applications where sample quantity and purity can be limiting factors.

[Supplemental material is available online at www.genome.org.]

Aberrant methylation of CpG dinucleotides in the 5′ regulatory region of genes often results in transcriptional inactivation and has been implicated in aging, heart and neurodegenerative diseases, as well as in the pathogenesis of various types of cancers (Feinberg and Vogelstein 1983; Gardiner-Garden and Frommer 1987; Post et al. 1999; Baylin and Herman 2000; Robertson and Wolffe 2000; Warnecke and Bestor 2000; Feinberg 2001; Jones and Baylin 2002; Cui et al. 2003). There is a growing interest in understanding the correlation between aberrant DNA methylation and tumorigenesis (Huang et al. 1999; Toyota et al. 1999; Costello et al. 2000; Yamashita et al. 2003) in order to facilitate disease marker discovery, diagnostic tool development, and the study of chemotherapeutic response (Laird 2003).

Current methods for detecting 5-methylcytosine can be divided into three major approaches (Laird 2003): (1) profiling methylation globally, (2) identifying methylation patterns at a cluster of CpG sites, and (3) determining methylation levels at individual CpG dinucleotides. Each category offers a different perspective for studying DNA methylation. In general, global screening methods rely on methylation-sensitive restriction enzyme digestion and provide opportunities for new epigenetic marker discovery (Huang et al. 1999; Toyota et al. 1999; Costello et al. 2000; Yamashita et al. 2003). Methylation-specific PCR (MSP) (Herman et al. 1996) and variations of this procedure (Cottrell and Laird 2003; Zeschnigk et al. 2004) were introduced to study the methylation pattern of a few closely neighboring CpG sites. Since DNA methylation is believed to be an early event during carcinogenesis (Laird 1997), the high sensitivity of MSP is suitable for use as an early detection tool on known epigenetic markers (Hoque et al. 2004). For quantitative assessment of individual CpG dinucleotide methylation status, the commonly used methods, including bisulfite treatment, were followed by sequencing (e.g., bisulfite sequencing and pyrosequencing) (Frommer et al. 1992; Uhlmann et al. 2002; Dupont et al. 2004; Yang et al. 2004), primer extension (e.g., SNuPE) (Gonzalgo and Jones 1997), restriction enzyme digestion (e.g., COBRA) (Xiong and Laird 1997), or real-time PCR (Zeschnigk et al. 2004). These assays provide quantitative profiling or detailed analysis of 5-methylcytosine distribution. The quantitative information generated is currently being used for correlating disease-specific methylation markers to clinical outcomes and facilitating the discovery of anti-tumorigenic drugs (Cheng et al. 2004; Issa 2004). However, the current methods analyze CpG methylation status one gene at a time and have limited multiplexing capability. Bisulfite sequencing provides the most comprehensive data on methylation status at every CpG but requires subcloning and sequence analysis of 10–20 individual clones. Higher throughput has been achieved by combining bisulfite-PCR with microarray technology utilizing oligonucleotide probes designed to form a perfect match with either methylated or unmethylated alleles within the target sequences (Adorjâan et al. 2002; Balog et al. 2002; Gitan et al. 2002). This allows parallel evaluation of CpG methylation status at numerous CpG sites across multiple genomic regions of interest. However, bisulfite treatment renders genomic DNA into AT-rich sequences, which exacerbates nonspecific and mismatch hybridizations due to differences in annealing temperatures between different probe sequences. In ad-

[5]Corresponding author.
E-mail barany@med.cornell.edu; fax (212) 746-8104.
Article published online ahead of print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.4181406.

16:282–289 ©2006 by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/06; www.genome.org

dition, probes containing two or more CpG dinucleotides may lack the sensitivity to distinguish partially methylated sequences from those that are fully methylated in heterogeneous clinical samples.

We seek to develop a robust assay for clinical application that provides quantitative methylation levels for multiple CpG dinucleotides in a given genomic region, as well as allowing specific evaluation of many genes in parallel. Such an assay can provide a representational CpG methylation profile of candidate genomic regions, and this profile information may be useful for disease stratification or as predictors of therapeutic response. This work presents a new method that aims to substantially improve quantitative microarray-based methylation detection to meet these needs. As illustrated in Figure 1, combining PCR, ligase detection reaction (LDR), and universal Array (where zip-code sequences appended to LDR primers, guide products to zip-code complements on an array) (Gerry et al. 1999; Favis et al. 2000) allows multiplexing and provides high specificity and accuracy. A detailed, quantitative methylation profile of essentially any set of CpG dinucleotides can be determined by using this assay. Fifteen tumor suppressor genes commonly linked to transcriptional silencing in various human cancers were chosen and the methylation status of their promoter regions evaluated (http://www.mdanderson.org/departments/methylation). Up to six CpG dinucleotides per promoter regions were investigated and a total of 75 CpG dinucleotides queried per sample.

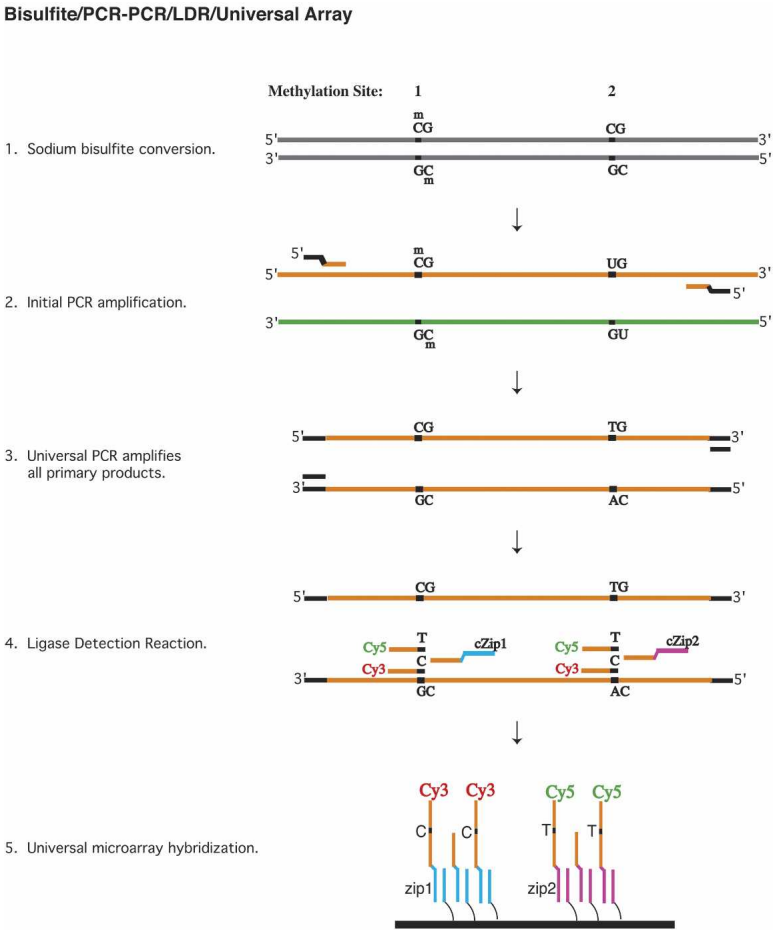

Figure 1. Schematic diagram of the assay. Two hypothetical CpG dinucleotide sites 1 and 2 are designated as methylated and unmethylated, respectively. Sodium bisulfite converts unmethylated, but not methylated, cytosines into uracils. This conversion renders the genomic DNAs into two asymmetrical, noncomplementary strands, and only one designated bisulfite-modified strand (highlighted in orange) is amplified and analyzed. In the initial amplification, PCR primers are designed with a gene-specific 3′ portion and an upstream universal sequence (highlighted in black). This universal sequence is used as a PCR primer in the subsequent PCR to simultaneously amplify all the primary amplicons (for ease of illustration, only one amplicon is shown). LDR is performed in a multiplex fashion with three primers (two discriminating and one common primers) interrogating each of the selected CpG sites. The discriminating primers contain a 5′ fluorescent label and a 3′ discriminating nucleotide to determine either methylated (with 5′ Cy3 and 3′ C) or unmethylated (with 5′ Cy5 and 3′ T) cytosines. The common primers bear a 5′ phosphate and a 3′ unique zip-code complement sequence (e.g., cZip1 and cZip2). Ligation occurs only if the nucleotides at the ligation junction are perfectly base-paired with a complementary template and the ligation products are captured onto a Universal microarray with prespotted zip-codes (addresses). For example, address Zip1 identifies methylated cytosine in methylation site 1, and address Zip2 identifies unmethylated cytosine in methylation site 2. Three Universal microarray addresses are assigned for each promoter region, and each address is double-spotted to ensure the quality of array fabrication and oligonucleotide hybridization efficiency.

Results

Determining assay specificity and quantitative accuracy of bisulfite-PCR/LDR/Universal Array

The general design of the assay is illustrated in Figure 1. Genomic DNAs were treated with sodium bisulfite to convert unmethylated, but not methylated, cytosines into uracils. We have modified the standard bisulfite protocol to ensure a thorough deamination of unmethylated cytosines and increase DNA recovery (Boyd and Zon 2004). Gene-specific PCR primers bearing 5′ universal tails were designed to flank each promoter region. A second, universal PCR step allows approximately equal fragment amplification of all sequences amplified in the primary PCR. Since PCR is not the final readout in this assay, primer design is flexible and less constrained by sequence context and is independent of CpG dinucleotide methylation status. Three LDR primers were used to determine the methylation status of each CpG dinucleotide. LDR primers were designed to tolerate mismatched base pairs at internal CpG sites and allow hybridization to fully and partially methylated sequences, as well as unmethylated sequences. A high fidelity Tth ligase (Luo et al. 1996), only ligates the upstream (discriminating) and downstream (common) primers when the 3′ discriminating nucleotide at the junction is complementary to the DNA template. This feature allows accurate, quantitative detection of targeted CpG dinucleotides regardless of the presence of internal CpG dinucleotides within the primer sequences. For example (Fig. 1), at the methylated CpG site 1, only Cy3-C–labeled ligation products are formed, whereas only Cy5-T–labeled ligation

products are formed at the unmethylated CpG site 2. Unique complementary zip-code sequences on the 3′ ends of common primers guide LDR products to their corresponding zip-codes on a Universal Array (Gerry et al. 1999; Favis et al. 2000). The zip-codes are unique sequences designed with a constant T_m and have no homology to either the target sequence or to other sequences in the genome. This design eliminates false signals due to nonspecific binding and mismatch hybridizations.

The assay was validated on genomic DNAs extracted from two commonly used colorectal (HCT15) and prostate cancer (LNCaP) cell lines. Promoter regions chosen in this study have included 15 tumor suppressors (*CDKN2B* [formerly known as $p15^{INK4b}$], *CDKN2A* [formerly known as $p16^{INK4a}$], *CDKN2D* [formerly known as $p14^{ARF}$], *CDKN1A* [formerly known as $p21^{CIP}$], *CDKN1B* [formerly known as $p27^{KIP}$], *TP53* [formerly known as *p53*], *BRCA1*, *TIMP3*, *APC*, *RASSF1*, *CDH1* [formerly known as *ECAD*], *MGMT*, *DAPK1* [formerly known as *DAPK*], *GSTP1*, and *RARB* [formerly known as *RARβ*]) and one hemi-methylated imprinted gene (*SNRPN*, as an internal control). The methylation profiles of the candidate promoter regions were determined by bisulfite sequencing, which revealed *CDKN2B*, *CDKN2A*, *CDKN2D*, *CDKN1A*, *CDKN1B*, *TP53*, *BRCA1*, *DAPK1*, *CDH1*, *MGMT*, and *TIMP3* were unmethylated in LNCaP, while *CDKN2A*, *CDKN2D*, *MGMT*, *RARB*, and *RASSF1* were methylated in HCT15 among 15 tumor suppressor genes. The initial LDR/Universal Array assay was designed to evaluate methylation status of three CpG sites per promoter region. LDR primers detecting methylated and unmethylated cytosines were validated by using in vitro methylated (SssI methylase) and untreated normal human lymphocyte genomic DNAs, respectively (data not shown). Following bisulfite treatment, genomic DNA of each cell line sample was multiplex PCR amplified, and the pooled PCR products were subjected to LDR/Universal Array analysis (Fig. 2A). We tested the assay specificity by using LNCaP DNA and subsets of LDR primers that detect only unmethylated cytosines (Fig. 2B; data not shown). The capture of Cy5 fluorescence signals only at the designated zip-code addresses for each LDR primer set indicated that LDR/Universal Array did not generate nonspecific ligation products and that mismatch hybridization was absent. To further demonstrate the assay's accuracy, LDR primers that detected only methylated cytosines were used to investigate a total of 48 CpG sites simultaneously for each cell line (Fig. 2C). Our data are consistent with the bisulfite sequencing results, indicating an accurate methylation profile was obtained. Different levels of fluorescence intensity were observed at several zip-code addresses. These variations suggested that the targeted CpG dinucleotides may have different methylation levels within the same promoter regions (e.g., *RASSF1* and *TIMP3*).

To determine if the assay could be quantitative, genomic DNA from HCT15 (carrying methylated CpG dinucleotides) was mixed with normal human lymphocytes (carrying unmethylated alleles), such that the test samples contained 0%, 20%, 40%, 60%, 80%, and 100% of HCT15 DNA. These mixtures were subjected to bisulfite-PCR/LDR/Universal Array analysis (Fig. 3A; data not shown). The average fluorescence intensity representing either methylated (Cy3) or unmethylated (Cy5) alleles from each double-spotted zip-code address was used to calculate the methylation ratio of Cy3/(Cy3 + Cy5). Each experiment was repeated at least twice and produced consistent results. Most of the CpG dinucleotides we evaluated have R^2 values between 0.98 and 0.89. Those CpG sites that gave lower R^2 values are likely due to inefficient competition between LDR primers targeted toward

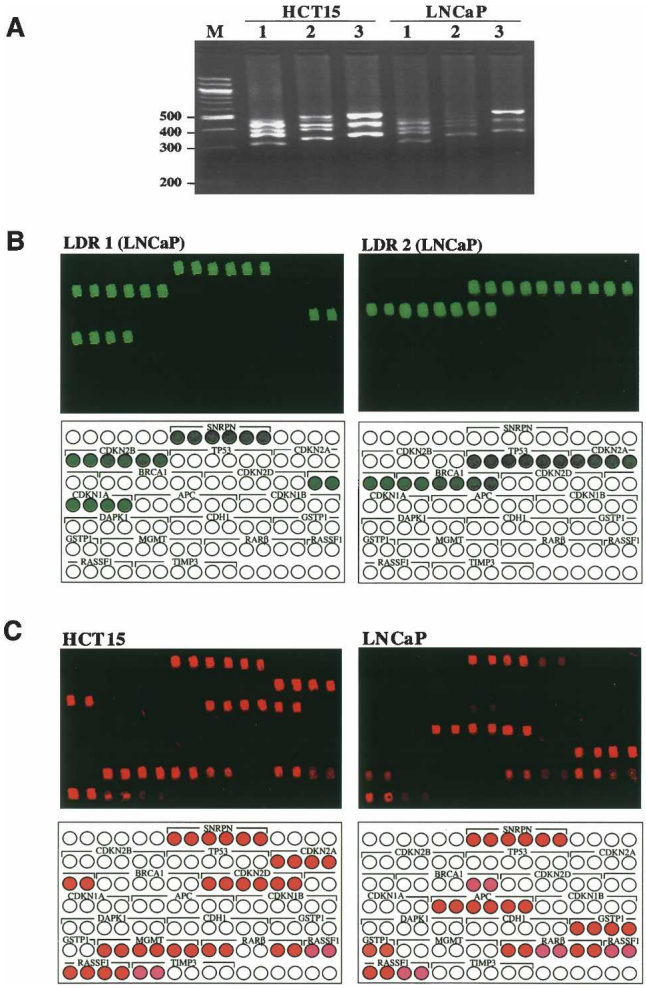

Figure 2. Representative bisulfite-PCR/LDR/Universal Array analysis of 16 promoter regions of cell lines HCT15 and LNCaP. (*A*) For the ease of demonstration, either five or six promoter regions were amplified in one PCR, and a total of 16 genes were simultaneously analyzed. The gene names and the corresponded PCR fragments are as follows: (lane *1*) *CDKN2B* (317 bp), *CDKN2A* (363 bp), *CDKN1A* (391 bp), *CDKN1B* (426 bp), *SNRPN* (442 bp), and *BRCA1* (459 bp); (lane *2*) *CDKN2D* (346 bp), *TIMP3* (404 bp), *APC* (433 bp), *RASSF1* (474 bp), and *CDH1* (513 bp); and (lane *3*) *MGMT* (362 bp), *TP53* (418 bp), *DAPK1* (434 bp), *GSTP1* (507 bp), and *RARB* (522 bp). (*B*) LDR/Universal Array analysis of the unmethylated cytosines in LNCaP amplicons. All PCR products were pooled as LDR templates, but only selected LDR primers were used in each reaction (LDR set1: *SNRPN*, *CDKN2B*, *CDKN1A*; LDR set2: *CDKN2A*, *TP53*, *BRCA1*). The subset of promoter regions that were interrogated in each LDR are depicted in the diagram (green circles) under each array image. The Cy5-labeled LDR products (false color green, designed for unmethylated cytosines) were captured on Universal Arrays. (*C*) All PCR products of each sample were pooled and subjected to LDR/Universal Array assay. Only Cy3-labeled LDR primers (false color red) were used in this assay to detect methylated cytosines. The diagram *under* each array image depicts the correlated zip-codes (circles) that were assigned to represent the CpG methylation status in each of the 16 promoter regions. Each zip-code was double-spotted on the array to ensure fabrication quality. Red and empty circles represent methylated and unmethylated CpG sites, respectively. Pink circles represent those CpG dinucleotides that have lower level of methylation. The PCR and LDR primer sequences and their concentrations used in these experiments were listed in the Supplemental Tables 1, 2, and 3.

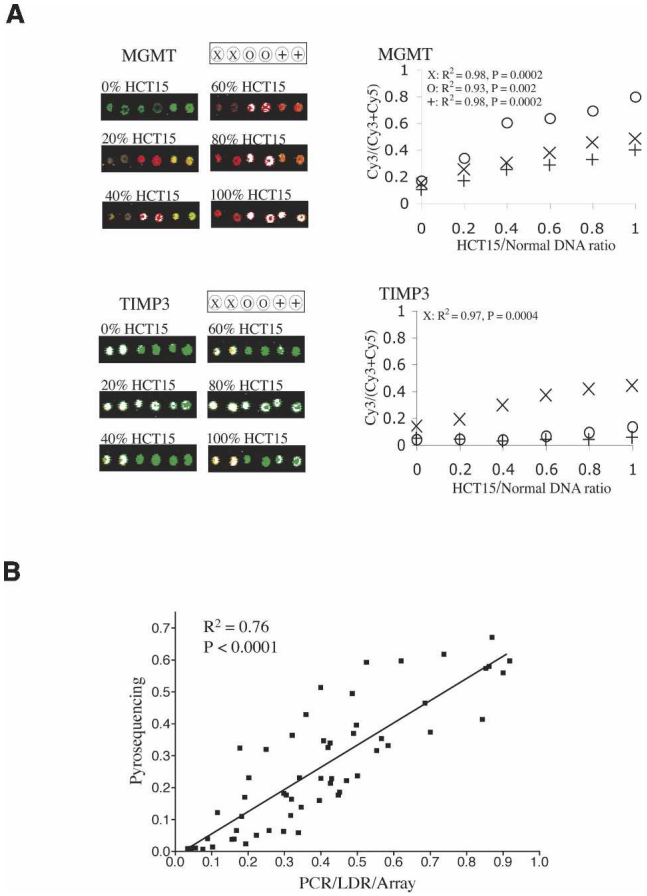

Figure 3. The quantification curves of the assay. Genomic DNAs of HCT15 and normal human lymphocytes were mixed in 0%, 20%, 40%, 60%, 80%, and 100% ratios and subjected to bisulfite-PCR/LDR/Universal Array analysis. (*A*) Representative array images are shown scanned in both Cy3 and Cy5 channels. False color red (Cy3) and green (Cy5) represent the methylated and unmethylated alleles of CpG dinucleotides, respectively. Color composites of the two channels reflect the methylation levels. Each zip-code was double-spotted on the array to ensure fabrication quality. MGMT and TIMP3 were used as examples to show the assay linearity measured at individual CpG dinucleotides. The plotted value at *y*-axis represents the fluorescence intensity Cy3/(Cy3 + Cy5) ratio. The value at *x*-axis represents the percentage of HCT15 mixed with normal human lymphocyte genomic DNAs. The R^2 and *P*-values of each linear regression line were calculated, although the lines were omitted in the plots for visual clarity. Nearly no methylation was observed at two of the CpG sites of TIMP3 resulting in poor statistical correlation (circles: $R^2 = 0.81$, $P = 0.02$; crosses: $R^2 = 0.08$, $P = 0.01$). The experiments were repeated three times with different sample preparations and array hybridizations. The PCR and LDR primer sequences and their concentrations used in these experiments are listed in Supplemental Tables 3 and 4. (*B*) Comparison of the MGMT methylation level from 15 colorectal carcinomas using pyrosequencing technology and bisulfite/PCR/LDR/Universal Array. Three CpG sites were evaluated per DNA sample. The plotted value at *y*-axis represents the percentage of methylated cytosines in the tumor samples as obtained from pyrosequencing. The value at *x*-axis represents the ratio of fluorescence intensities Cy3/(Cy3 + Cy5). The mixed genomic DNAs of HCT15 and normal human lymphocytes shown in *A* were included as controls.

unmethylated and methylated alleles and can be resolved by redesigning the LDR primers to have a higher melting temperature. Our analysis confirmed the different percentage of methylation at each CpG dinucleotide and suggested that methylation level is not 100% at each CpG site in tumor cell line DNA. For example, for HCT15, a medium level of methylation was observed at the first CpG dinucleotide and <10% methylation at the other two CpG dinucleotides in TIMP3. By comparing the ratio of (methylated): (methylated + unmethylated) DNA in different cell lines, we could extrapolate the CpG methylation level at a given position. Alternatively, SssI could be used to methylate DNA in vitro to completion to generate standard curves for calibration (data not shown). We tested assay sensitivity by mixing tumor cell line DNA with normal human lymphocyte DNAs. A preliminary study suggested that the unbiased PCR primer design was sufficient to detect the presence of methylated alleles, even when diluted down to 1% in unmethylated alleles (Supplemental Fig. 1). Nevertheless, in the colorectal cancer study, we currently use 10%–15% as a cut-off for our scoring criteria. Overall, our data demonstrate that bisulfite-PCR/LDR/Universal Array approach is a quantitative and sensitive method for the measurement of DNA methylation.

Comparison of the method with a pyrosequencing assay using clinical tumor samples

To further evaluate the quantitative accuracy and clinical utility, we analyzed *MGMT* methylation level at three CpG dinucleotides per sample on a total of 15 colorectal tumors, using both our assay as well as by pyrosequencing. Genomic DNAs of these tumors were bisulfite treated, amplified by using multiplex PCR, and analyzed by LDR/Universal Array. Alternatively, *MGMT* was uniplex amplified from the same tumor DNAs for pyrosequencing. Three sequencing primers were designed to investigate the methylation level of the same CpG dinucleotides that were studied by LDR/Universal Array. The comparison revealed a high correlation between these two methods (Fig. 3B). The few samples where results varied may be due to variations in array fabrication, differences in efficiency of multiplex versus uniplex PCR amplification, and, most likely, the pyrosequencing primer design. Two of the sequencing primers contained two and three CpG sites, respectively, and generated a biphasic plot while establishing standard curves. This may have led to bias in the quantification seen by pyrosequencing. Nevertheless, this highly significant correlation suggests bisulfite/PCR/LDR/Universal Array is an accurate method for quantitative analysis of heterogeneous clinical samples.

Application of the method to cancer cell lines and heterogeneous clinical samples

We performed a blinded study to determine methylation profiles of several cancer cell lines (five colorectal, one breast, and one prostate) with this approach (Fig. 4). Three CpG sites per promoter region were evaluated. Methylation was not observed in the promoter regions of *CDKN2B*, *CDKN1A*, *CDKN1B*, *TP53*, and *BRCA1* among all the tested cell line DNAs. These results suggested that promoter methylation is not a random event. All cell lines under the blinded study have very distinct methylation profiles among the 15 tumor suppression genes except for two pairs of cell lines: HCT15/DLD-1 and HT29/WiDr, which share almost identical promoter methylation patterns. Each pair of cell lines HCT15/DLD-1 and HT29/WiDr was established from the same colon carcinomas (additional cell line information is described in American Type Culture Collection [ATCC] Web site http://www.atcc.org) (Chen et al. 1995). This observation suggested that colorectal cancer cell lines derived from the same tumor have essentially identical methylation profiles that are

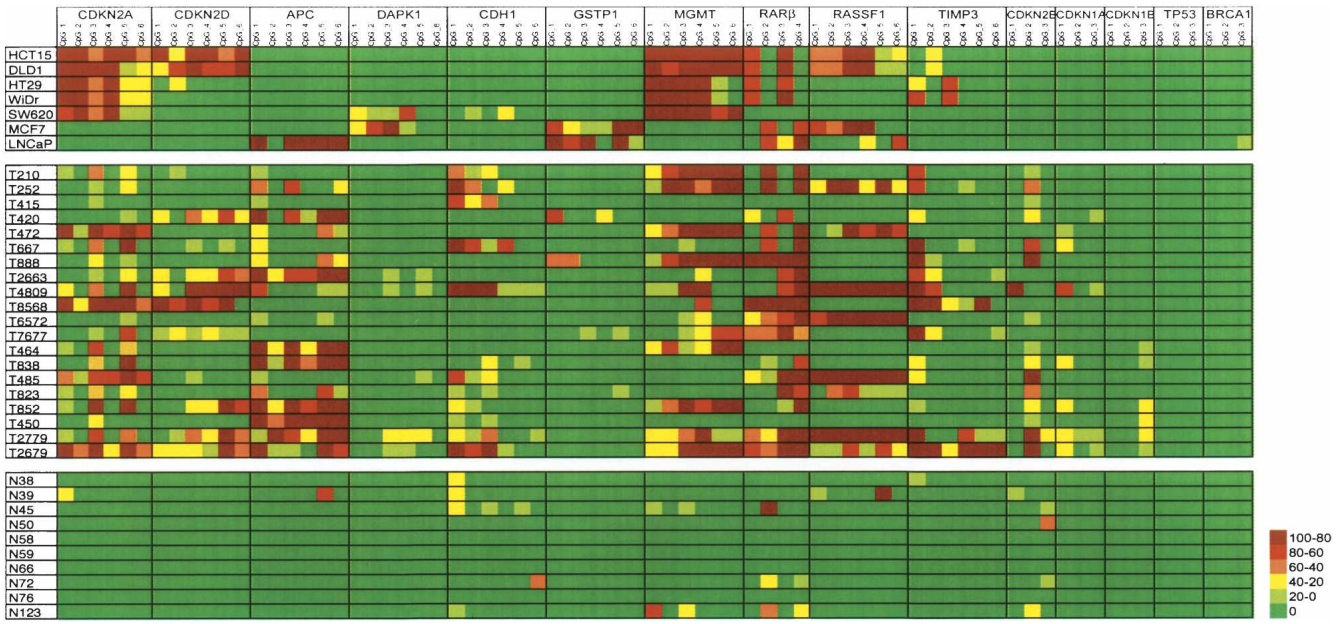

Figure 4. Selected examples of DNA methylation profiles of cancer cell lines and clinical samples. Five colorectal (HCT15, DLD-1, HT29, WiDr, and SW620), one breast (MCF7), one prostate (LNCaP) cancer cell lines, 20 primary colorectal cancer (T series), and 10 adjacent normal tissues (N series) were analyzed. Six CpG sites per promoter region were analyzed in each sample except *CDKN2B*, *CDKN1A*, *CDKN1B*, *TP53*, and *BRCA1*. Standard curves were not established for the genes mentioned above since hypermethylation is not reported in the literature or observed in our clinical sample study. The standard curves should be established when applying this assay to other tumor types such as breast cancer. Around 10%–15% established CpG methylation standard curves gave lower R^2 values (between low 80s and 70s), including *CDKN2A* (CpG_3 and CpG_6), *CDKN2D* (CpG_1), *GSTP1* (CpG_3), *DAPK1* (CpG_3), *RASSF1* (CpG_4), and *TIMP3* (CpG_2). Only four CpG sites of RARB were shown due to defects on zip-code addresses during array fabrication. The color scale represents the percentage of methylation levels determined from the standard curves at each CpG dinucleotide. Notice that for the cell lines HCT15/DLD-1 and HT29/WiDr, each pair is derived from a same tissue origin reflected in their identical methylation patterns. The PCR and LDR primer sequences and their concentrations used in these experiments are listed in the Supplemental Tables 3, 4, and 5.

distinct from cell lines derived from a different patient. To ensure the accuracy of scoring a candidate gene promoter as hypermethylated, the methylation status of three additional CpG sites per promoter was examined (examples are shown in Supplemental Fig. 2) for the 10 genes that were found to be methylated in the cell line study (*CDKN2A*, *CDKN2D*, *TIMP3*, *APC*, *RASSF1*, *CDH1*, *MGMT*, *DAPK1*, *GSTP1*, and *RARB*). Within each promoter, the six CpG sites examined were dispersed evenly throughout the amplified genomic sequence. Some of the CpG sites investigated were within 50 bases and resulted in the design of overlapping LDR primers. Nevertheless, we have found that LDR efficiency was not altered by the overlapping primer design (data not shown). Thus, the cytosines of all 75 CpG sites for a given sample were interrogated independently at the same time.

In a pilot study, we have profiled methylation status of 96 colorectal tumor samples and 73 matched normal tissues (selected data are shown in Fig 4; Supplemental Fig. 2). Our preliminary analysis shows a rich variety of methylation profiles. Seven promoter regions (*CDKN2A*, *CDKN2D*, *APC*, *MGMT*, *RARB*, *RASSF1*, and *TIMP3*) showed statistically significant increased methylation in tumor compared to normal tissues. Promoter regions of *CDKN1A*, *CDKN1B*, *TP53*, and *BRCA1* revealed little or no methylation. Thus, methylation profiles of clinical samples were similar to those observed in the colorectal cancer cell lines. One of the CpG dinucleotides of *CDKN2B* was frequently methylated in the tumor samples. The biological significance of this methylation remains to be investigated by examining additional CpG dinucleotides to determine *CDKN2B* promoter hypermethylation status. Nevertheless, these results reaffirm that methylation in colorectal cancer is gene specific and nonrandom. Additional tumor samples will be examined to provide sufficient power to determine the correlation between promoter methylation and the tissue pathological-clinical information.

Discussion

We have presented an accurate and quantitative assay that provides a representational CpG methylation profile from colorectal cancer samples, where stromal cell infiltration is often seen. This assay simultaneously determines the DNA methylation status of multiple gene promoters, querying a total of 75 CpG dinucleotide sites per sample. Genomic DNAs isolated from seven cancer cell lines and 169 colon tumor samples were tested. Cell lines derived from the same tumor have essentially identical methylation profiles. The percentage of CpG dinucleotide methylation of MGMT promoter in clinical samples was compared by our assay with those derived from pyrosequencing and resulted in a high correlation.

The candidate promoter regions, involving genes in cell cycle regulation, DNA repair, and tumor metastasis and invasion, were previously reported in the literature as associated with abnormal gene silencing in tumors or cancer cell lines. Although the percentage of methylated promoters in a cohort may vary due to the sample source and the assays used in determining DNA methylation status, our preliminary analysis of the colorectal tumor methylation profile gave consistent results as those published previously (Esteller et al. 2001). For example, in agreement with our own findings, it has been shown that there is

essentially no hypermethylation in *CDKN2B*, *TP53*, *BRCA1*, and *GSTP1* promoter regions, while *CDKN2D*, *CDKN2A*, *MGMT*, and *APC* were methylated to a higher level among all the samples tested (Esteller et al., 2001).

Our assay allows virtually any CpG site in the promoter and first intron regions for more than a dozen genes to be analyzed simultaneously. There is increasing evidence that genes such as *MLH1* and *RASSF1A* exhibit an increasing gradient of methylation from the promoter proximal region to the first exon (Deng et al. 1999; Yan et al. 2003). To avoid the bias of scoring a hyper (or hypo)-methylated promoter and linking it to its disease state, multiple CpG sites across a larger window of the genomic region should be investigated in each assay. Studies done with MSP- and restriction enzyme–based methods only reveal the methylation pattern of small sequence regions; additional sequence contexts may be needed to sufficiently determine the promoter methylation status. Our approach interrogated multiple CpG dinucleotides that were evenly distributed over 300–500 bases. For paraffin-embedded tissue, two or more adjacent shorter PCR amplicons should be designed to overcome the poor amplification typically observed in these types of samples (Supplemental Fig. 3). Moreover, the LDR efficiency was not affected, even when some of the LDR primers were designed with overlapping sequences. This technique provides a detailed mapping of the methylation profile in each promoter CpG island locus that may correlate with transcriptional silencing during disease progression.

The bisulfite-PCR/LDR/Universal Array approach provides several advantages over existing methods for the analysis of DNA methylation patterns. First, there are two levels of specificity facilitated by gene-specific primers, initially during PCR and subsequently during LDR. Given the numerous duplications annotated and suspected in the human genome, this approach enhances the ability to reliably target the CpG islands in the locus of interest. Second, LDR allows accurate identification of low abundance nucleotide alterations with a remarkable accuracy due to the high fidelity of Tth ligase (Luo et al. 1996). This unique feature eliminates the concern of mismatch hybridization due to partial methylation of internal CpGs of the LDR primer sequences and allows LDR primers to be placed at essentially any CpG dinucleotide of interest. Third, the unique zip-code sequences were designed with similar T_m across the platform and have no sequence homology in the human genome. This feature allows only ligated LDR products to be captured, thus avoiding background signals. As the Universal Array can be easily expanded (Gerry et al. 1999), additional CpG sites or genes can be evaluated in a single assay. Fourth, this assay has the potential to detect low abundance methylated alleles. By redesigning the multiplex gene-specific PCR primers to be methyl specific, a detection of at least 0.1% was achieved, albeit with reduced quantitative dynamic range (Supplemental Fig. 1). An inherent problem with many DNA amplification techniques is that greater detection sensitivity comes at the cost of increased false positives. Although methylation assays relying solely on PCR as a readout tool offer sensitive detection, they are prone to false positives resulting from the AT richness or incomplete deamination of bisulfite-treated DNAs. Consequently, such assays would be limited in their multiplex capability. Our assay confirmed promoter methylation status via six CpG sites within a PCR fragment; this approach offers high specificity and accuracy while avoiding false positives. Moreover, each module of the PCR/LDR/Universal microarray approach is ideal for multiplexing and can be automated by using a liquid handing system to increase throughput. A recent publication has reported the detection of dozens to hundreds of possible mutations by using multiplex PCR/LDR in a single-tube format (Favis et al. 2004). The capture of multiplex LDR products onto an array format provides an efficient "modular" readout and substantially increases assay throughput. Universal microarray experiments in our laboratory are now performed in an array-of-arrays format, where 64 array hybridizations are carried out simultaneously (data not shown). This array-of-arrays approach drastically reduces the cost of array fabrication, minimizes the variation during hybridization, and increases throughput.

In summary, we present a robust and accurate method that determines cytosine methylation at any selected set of CpG dinucleotides in the genome. Importantly, this new method allows the evaluation of methylation level at individual CpG sites. Quantitative values for this parameter may facilitate stratification of tumors, since based on the degree of methylation, it may be possible to estimate disease progression. In addition, the ability to quantify methylation at specific sites in advanced tumors may provide information on tumor heterogeneity. These data will enable clinical decisions related to individualized treatment strategies. Our goal is to expand this prototype assay into a focused array platform that investigates 30–50 frequently methylated tumor suppressor promoters observed in several different human carcinomas. Around 10–15 individual CpG dinucleotides evenly distributed over a larger window of each promoter region will be interrogated. We anticipate that such a focused platform will facilitate the development of DNA-based molecular markers for disease diagnosis and prognosis and will be suitable for routine clinical use.

Methods

Cell line culture, tumor samples, and DNA extraction

Normal human lymphocyte genomic DNA was purchased from Roche. Colorectal, breast, and prostate cancer cell lines were obtained from American Type Culture Collection and cultured under the ATCC-recommended media conditions. Fresh frozen primary colorectal adenocarcinomas were obtained from Memorial Sloan Kettering Cancer Center under Institutional Review Board (IRB)-approved protocols. Genomic DNAs were extracted by using the DNeasy Tissue Kit (Qiagen) according to the manufacturer's guidelines.

Sodium bisulfite treatment of genomic DNAs

Typically, 1 µg genomic DNA was denatured in 40 µL of 0.2 N NaOH by incubating for 10 min at 37°C before addition of 30 µL of freshly prepared 10 mM hydroquinone and 520 µL of 3 M sodium bisulfite. The reaction was incubated for 20 min at 50°C, 15 sec at 85°C for 48 cycles (16 h). The DNA clean-up procedure was as follows: (1) the total reaction volume (~600 µL) was transferred to a Microcon NCO30 filter (Millipore) and centrifuged at 13000*g* for 16 min; (2) 500 µL deionized H_2O were added to the upper chamber, centrifuged at 13,000*g* for 7 min, the filtrate discarded, and the wash repeated twice; (3) 500 µL of 0.3M NaOH were added to the upper chamber, incubated for 5 min at room temperature and then centrifuged at 13,000*g* for 8 min; (4) 500 µL deionized H_2O were added to the upper chamber, centrifuged at 13,000*g* for 8 min, the filtrate discarded, and the wash repeated; and (5) the filter was inverted to collect the bisulfite-converted DNA. An appropriate volume of water (if needed) was

used to rinse the upper chamber to recover DNA in a final volume of 20 µL.

Multiplex PCR amplification

The multiplex PCR consists of two stages. PCR stage I (12.5 µL) contained 1.5 µL bisulfite-modified DNA, 400 µM of each dNTP, 1× AmpliTaq Gold PCR buffer, 4 mM $MgCl_2$, and 1.25 U AmpliTaq Gold polymerase (Applied Biosystems). Mineral oil was added prior to thermal cycling. The PCR stage I conditions were as follows: 10 min at 95°C; 15 cycles of 30 sec at 94°C, 1 min at 60°C, and 1 min at 72 °C; followed by a final extension step of 5 min at 72°C. PCR stage II (12.5 µL) contained 400 µM of each dNTP, 1× AmpliTaq Gold PCR buffer, 4 mM $MgCl_2$, 12.5 pmol universal primer (UniB2, see Supplemental Table 1), and 1.25 U AmpliTaq Gold polymerase. The 12.5 µL reaction mixture was added through the mineral oil to the completed stage I PCR. The PCR stage II conditions were as follows: 10 min at 95°C; 30 cycles of 30 sec at 94°C, 1 min at 55°C, 1 min at 72°C; followed by a final extension step of 5 min at 72°C. *Taq* DNA polymerase was inactivated by adding 1.25 µL Proteinase K (20 mg/mL, Qiagen) to the completed stage II PCR, incubating for 10 min at 70°C and 15 min at 90°C. Before pooling the PCR products for LDR assay, the presence of amplicons was confirmed by electrophoresis on a 3% agarose gel.

LDR, Universal Array hybridization, and data analyses

A typical LDR (20 µL) contained 20 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 100 mM KCl, 10 mM DTT, 1 mM NAD, 25 fmol wild-type Tth ligase (Zirvi et al. 1999), 500 fmol of each LDR primer, and 5–10 ng of each PCR amplicon. The LDR conditions were as follows: 3 min at 95°C; 25 cycles of 30 sec at 95°C and 4 min at 60°C. The LDR reaction was diluted with an equal volume of 2× hybridization buffer (600 mM MES at pH 6.0, 20 mM $MgCl_2$, 0.2% SDS), denatured for 3 min at 95°C and plunged on ice. The Universal Arrays were pre-equilibrated with 1× hybridization buffer at room temperature for at least 15 min. Coverwells (Grace Bio-Labs) were attached to arrays and filled with 40 µL denatured LDR reactions. The assembled arrays were incubated in a rotating hybridization oven for 60 min at 65°C. After hybridization, the arrays were washed in 300 mM Bicine (pH 8.0), 0.1% SDS for 10 min at 60°C. An updated version with 384 addresses will accommodate all the LDR products. Each array was scanned by using a Perkin Elmer ProScanArray under the same laser power and PMT within the linear dynamic range. The Cy3 and Cy5 dye bias was determined by measuring the fluorescence intensity of an equal mole of Cy3- and Cy5-labeled LDR primers manually deposited on a slide surface. This fluorescence intensity ratio ($W = I_{Cy3}/I_{Cy5}$) was used to normalize the label bias when calculating the methylation ratio Cy3/(Cy3 + Cy5). MetaMorph Imaging System (Universal Imaging) was used to create images depicting the Cy3 (red) and the Cy5 (green).

Oligonucleotide design and synthesis

Oligonucleotides were obtained from IDT or synthesized in-house on an ABI 394 DNA Synthesizer (PE Biosystems) using standard phosphoramidite chemistry (Khanna et al. 1999). Spacer phosphoramidite C18, 3′-amino-modifier C3 CPG, C3 spacer, and Cy3, Cy5, and standard phosphoramidites were purchased from Glen Research. All other reagents were purchased from PE Biosystems. The zip-code oligonucleotides were synthesized on a 3′ amino modifier C3 column with a spacer C18 inserted before the first base. The common LDR primers were synthesized with 5′ phosphates and 3′ C3 spacers as blocking groups. Oligonucleotides with cyanine labels were cleaved from the CPG supports and deprotected according to manufacturer's recommendations. Both labeled and unlabeled LDR oligonucleotides were purified and desalted on SuperPure columns (Biosearch Technologies) according to the manufacturer's instructions, then spin-dried (Speed-Vac) and stored at −20°C. For those primers that inevitably covered CpG dinucleotides in the body of their sequences, the nucleotides that base paired with cytosines in CpG dinucleotide were synthesized in two ways. One was to use nucleotide analogs dK or dP in the primers' syntheses. The pyrimidine derivative dP base pairs with either A or G, while the purine derivative dK base pairs with either C or T at similar efficiency. Alternatively, to reduce the cost of primer synthesis, those nucleotide positions with analogs dK or dP incorporated were substituted by nucleotides dG or dC, respectively. For example, the substituted nucleotide dG in a PCR primer formed either Watson-Crick base pair with C (methylated) or wobble base pair with U (unmethylated) on the bisulfite-modified DNA template.

Universal Array fabrication

Polymer-coated slides were fabricated as previously described (Gerry et al. 1999; Favis et al. 2000) or were purchased (CodeLink slides) from Amersham Biosciences. Universal Arrays were spotted by using a Pixsys5500 robot with a quill-type spotter in a controlled humidity chamber (Cartesian Technologies). Zip-code oligonucleotides each with a unique 24-mer sequence were prepared by mixing 5 µL of 1000 µM stock oligonucleotides with 5 µL of 0.4 M K_2HPO_4/KH_2PO_4 (pH 8.5) in 384 conical well spotting plates. Arrays were printed under relative humidity 60%–70%. To ensure that all the zip-codes were spotted without cross-contamination during array fabrication, one out of 10 slides on average was subjected to quality control by hybridizing fluorescein-labeled zip-code complements targeting a combination of rows or columns of zip-code addresses. A batch of fabricated arrays passed the quality control only when specific fluorescein signals were present on all the targeted rows and columns without extraneous signals on the adjacent, unexpected neighboring addresses.

Pyrosequencing

A promoter sequence of MGMT was PCR amplified by using 1 µL bisulfite-modified DNA, 400 µM of each dNTP, 1× AmpliTaq Gold PCR buffer, 4 mM $MgCl_2$, 0.2 µM PCR primers (5′-GGTTTTAGGAGGGGAGAGATT-3′ and 5′-CCTAACCCRAATAACCCTTC-3′), and 1.25 U AmpliTaq Gold polymerase (Applied Biosystems). The PCR condition was as follows: 15 min at 94°C; 45 cycles of 15 sec at 95°C, 30 sec at 58°C, and 15 sec at 72°C; followed by a final extension step for 5 min at 72°C. Three sequencing primers (5′-GTAGTAGTTTAGAGTAGGAT-3′, 5′-TTTTAGAGAGTTTTTAGGAT-3′ and 5′-AAATTAAGGTATAGAGTTTT-3′) were designed to determine the CpG dinucleotide methylation levels. The primers were designed and experiments were performed by Biotage.

Acknowledgments

We thank the Neil Bander laboratory for providing prostate cancer cell lines for DNA extraction; Reyna Favis, Maneesh Pingle, and Sarah Giardina for critical reading of the manuscript; and Jurg Ott, Sandra Barral, and Robert Westphalen for statistic analysis of the clinical sample data. We also thank Reyna Favis, Norman Gerry, Jianmin Huang, Brian Kirk, Maneesh Pingle, Richard Shattock, Hanna Pincas, Manny Bacolod, Kathy Granger, and Ali Gure for insightful discussion and technical assistance. Work in

the Barany laboratory is sponsored by the National Cancer Institute (P01-CA65930) and in part by a research grant from Applied Biosystems, for which F.B. also serves as a consultant.

References

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

Balog, R.P., de Souza, Y.E., Tang, H.M., DeMasellis, G.M., Gao, B., Avila, A., Gaban, D.J., Mittelman, D., Minna, J.D., Luebke, K.J., et al. 2002. Parallel assessment of CpG methylation by two-color hybridization with oligonucleotide arrays. *Anal. Biochem.* **309:** 301–310.

Baylin, S.B. and Herman, J.G. 2000. DNA hypermethylation in tumorigenesis: Epigenetics joins genetics. *Trends Genet.* **16:** 168–174.

Boyd, V.L. 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.

Chen, T.R., Dorotinsky, C.S., McGuire, L.J., Macy, M.L., and Hay, R.J. 1995. DLD-1 and HCT-15 cell lines derived separately from colorectal carcinomas have totally different chromosome changes but the same genetic origin. *Cancer Genet. Cytogenet.* **81:** 103–108.

Cheng, J.C., Yoo, C.B., Weisenberger, D.J., Chuang, J., Wozniak, C., Liang, G., Marquez, V.E., Greer, S., Orntoft, T.F., Thykjaer, T., et al. 2004. Preferential response of cancer cells to zebularine. *Cancer Cell* **6:** 151–158.

Costello, J.F., Frèuhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomèaki, P., Lang, J.C., et al. 2000. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat. Genet.* **24:** 132–138.

Cottrell, S.E. and Laird, P.W. 2003. Sensitive detection of DNA methylation. *Ann. N.Y. Acad. Sci.* **983:** 120–130.

Cui, H., Cruz-Correa, M., Giardiello, F.M., Hutcheon, D.F., Kafonek, D.R., Brandenburg, S., Wu, Y., He, X., Powe, N.R., and Feinberg, A.P. 2003. Loss of IGF2 imprinting: A potential marker of colorectal cancer risk. *Science* **299:** 1753–1755.

Deng, G., Chen, A., Hong, J., Chae, H.S., and Kim, Y.S. 1999. Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression. *Cancer Res.* **59:** 2029–2033.

Dupont, J.M., Tost, J., Jammes, H., and Gut, I.G. 2004. De novo quantitative bisulfite sequencing using the pyrosequencing technology. *Anal. Biochem.* **333:** 119–127.

Esteller, M., Corn, P.G., Baylin, S.B., and Herman, J.G. 2001. A gene hypermethylation profile of human cancer. *Cancer Res.* **61:** 3225–3229.

Favis, R., Day, J.P., Gerry, N.P., 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.

Favis, R., Huang, J., Gerry, N.P., 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, A.P. 2001. Methylation meets genomics. *Nat. Genet.* **27:** 9–10.

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

Frommer, M., McDonald, L.E., Millar, D.S., Collis, C.M., Watt, F., Grigg, G.W., Molloy, P.L., and Paul, C.L. 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.

Gardiner-Garden, M. and Frommer, M. 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196:** 261–282.

Gerry, N.P., Witowski, N.E., Day, J., Hammer, R.P., 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, R.S., Shi, H., Chen, C.M., Yan, P.S., and Huang, T.H. 2002. Methylation-specific oligonucleotide microarray: A new potential for high-throughput methylation analysis. *Genome Res.* **12:** 158–164.

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

Herman, J.G., Graff, J.R., Myèohèanen, S., Nelkin, B.D., and Baylin, S.B. 1996. Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands. *Proc. Natl. Acad. Sci.* **93:** 9821–9826.

Hoque, M.O., Begum, S., Topaloglu, O., Jeronimo, C., Mambo, E., Westra, W.H., Califano, J.A., and Sidransky, D. 2004. Quantitative detection of promoter hypermethylation of multiple genes in the tumor, urine, and serum DNA of patients with renal cancer. *Cancer Res.* **64:** 5511–5517.

Huang, T.H., Perry, M.R., and Laux, D.E. 1999. Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet.* **8:** 459–470.

Issa, J.P. 2004. CpG island methylator phenotype in cancer. *Nat. Rev. Cancer* **4:** 988–993.

Jones, P.A. and Baylin, S.B. 2002. The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* **3:** 415–428.

Khanna, M., Park, P., Zirvi, M., Cao, W., Picon, A., Day, J., Paty, P., and Barany, F. 1999. Multiplex PCR/LDR for detection of K-ras mutations in primary colon tumors. *Oncogene* **18:** 27–38.

Laird, P.W. 1997. Oncogenic mechanisms mediated by DNA methylation. *Mol. Med. Today* **3:** 223–229.

———. 2003. The power and the promise of DNA methylation markers. *Nat. Rev. Cancer* **3:** 253–266.

Luo, J., Bergstrom, D.E., and Barany, F. 1996. Improving the fidelity of *Thermus thermophilus* DNA ligase. *Nucleic Acids Res.* **24:** 3071–3078.

Post, W.S., Goldschmidt-Clermont, P.J., Wilhide, C.C., Heldman, A.W., Sussman, M.S., Ouyang, P., Milliken, E.E., and Issa, J.P. 1999. Methylation of the estrogen receptor gene is associated with aging and atherosclerosis in the cardiovascular system. *Cardiovasc. Res.* **43:** 985–991.

Robertson, K.D. and Wolffe, A.P. 2000. DNA methylation in health and disease. *Nat. Rev. Genet.* **1:** 11–19.

Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.P. 1999. CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci.* **96:** 8681–8686.

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

Warnecke, P.M. and Bestor, T.H. 2000. Cytosine methylation and human cancer. *Curr. Opin. Oncol.* **12:** 68–73.

Xiong, Z. and Laird, P.W. 1997. COBRA: A sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* **25:** 2532–2534.

Yamashita, K., Dai, T., Dai, Y., Yamamoto, F., and Perucho, M. 2003. Genetics supersedes epigenetics in colon cancer phenotype. *Cancer Cell* **4:** 121–131.

Yan, P.S., Shi, H., Rahmatpanah, F., Hsiau, T.H., Hsiau, A.H., Leu, Y.W., Liu, J.C., and Huang, T.H. 2003. Differential distribution of DNA methylation within the RASSF1A CpG island in breast cancer. *Cancer Res.* **63:** 6178–6186.

Yang, A.S., Estâecio, M.R., Doshi, K., Kondo, Y., Tajara, E.H., and Issa, J.P. 2004. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. *Nucleic Acids Res.* **32:** e38.

Zeschnigk, M., Bèohringer, S., Price, E.A., Onadim, Z., Masshèofer, L., and Lohmann, D.R. 2004. A novel real-time PCR assay for quantitative analysis of methylated alleles (QAMA): Analysis of the retinoblastoma locus. *Nucleic Acids Res.* **32:** e125.

Zirvi, M., Bergstrom, D.E., Saurage, A.S., Hammer, R.P., and Barany, F. 1999. Improved fidelity of thermostable ligases for detection of microsatellite repeat sequences using nucleoside analogs. *Nucleic Acids Res.* **27:** e41.

Received May 25, 2005; accepted in revised form October 6, 2005.