

NIH Public Access

Author Manuscript

Anal Biochem. Author manuscript; available in PMC 2012 October 31.

Published in final edited form as: *Anal Biochem.* 2002 October 15; 309(2): 301–310.

Parallel assessment of CpG methylation by two-color hybridization with oligonucleotide arrays

Robert P. Balog^{a,b}, Y. Emi Ponce de Souza^b, Hue M. Tang^b, Gina M. DeMasellis^b, Boning Gao^c, Adrian Avila^c, Desmond J. Gaban^b, David Mittelman^b, John D. Minna^c, Kevin J. Luebke^{a,*}, and Harold R. Garner^{a,b}

^aCenter for Biomedical Inventions, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390-8573, USA

^bMcDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390-8591, USA

^cThe Hamon Center for Therapeutic Oncology Research, University of Texas Southwestern Medical Center at Dallas, Dallas, TX 75390-8593, USA

Abstract

We have developed a method for the parallel analysis of multiple CpG sites in genomic DNA for their state of methylation. Hypermethylation of CpG islands within the promoters and 5' exons of genes has been found to be a mechanism of transcriptional inactivation associated with a variety of tumors. The method that we developed relies on the differential reactivity of methylated and unmethylated cytosines with sodium bisulfite, which exclusively converts unmethylated cytosines to deoxyuracils. The resulting sequence changes are determined with single-nucleotide resolution by hybridization to an oligonucleotide array. Cohybridization with a reference sample containing a different label provides an internal standard for assessment of methylation state. This method provides advantages in parallelism over existing methods of methylation analysis. We have demonstrated this technique with a region from the promoter of the tumor suppressor gene p16, which is hypermethylated in many cancers.

Keywords

Hypermethylation; CpG island; Oligonucleotide array; Sodium bisulfite; Tumor suppressor

Methylation of cytosines in CpG dinucleotides is an important mechanism of transcriptional regulation. It is involved in a diversity of normal biological processes such as X chromosome inactivation [1] and transcriptional regulation of imprinted genes [2]. Aberrant methylation of cytosines can also effect transcriptional inactivation of certain tumor suppressor genes, associated with a number of human cancers [3,4]. Cytosine methylation in CpG-rich areas (CpG islands) located in the promoter regions of some genes is of special regulatory importance. Therefore, wide-scope mapping of methylation sites in CpG islands is important for understanding both normal and pathological cellular processes. Furthermore, methylation of certain sites might serve as an important marker for early diagnosis and treatment decisions of some cancers [5,6].

^{© 2002} Elsevier Science (USA). All rights reserved.

^{*}Corresponding author. Fax: +1-214-648-4156. kevin.luebke@utsouthwestern.edu (K.J. Luebke).

A variety of methods have been used to identify sites of DNA methylation. One common method has relied on the inability of restriction endonucleases to cleave sequences that contain one or more methylated cytosines [7]. Genomic DNA is fragmented with methylation-sensitive restriction enzymes, and cleavage at the site of interest is probed by Southern blotting, PCR, or hybridization to a CpG island array [8]. This method is limited to sites that fall within the recognition sequences of methylation-sensitive restriction enzymes.

Alternative methods rely on the differential chemical reactivities of cytosine and 5-methyl cytosine with reagents such as sodium bisulfite, hydrazine, or permanganate. Treatment with sodium bisulfite can be used to convert methylated and unmethylated DNA to different sequences [9,10] (Fig. 1a). Under appropriate conditions, unmethylated cytosines in DNA react with sodium bisulfite to yield deoxyuracil, which behaves as thymidine in enzymatic template-directed polymerization. Methylated cytosines, however, are unreactive and behave as cytosine in enzymatic template-directed polymerization.

The sequence differences resulting from bisulfite treatment can be assessed in several ways. Methylation-specific PCR uses a set of primers specific to the sequences resulting from bisulfite treatment of either methylation state at a given site [11]. One potential site of methylation is probed at a time in this assay. Standard sequencing by primer extension following bisulfite treatment is commonly used as a way of assessing methylation status of DNA [12–14]. This approach allows parallel analysis of multiple proximal cytosines in one assay, but it frequently requires laborious cloning and sequencing of individual inserts.

Here we describe a new method for mapping individual sites of CpG methylation in genomic DNA. The method makes use of oligonucleotide arrays [15] to determine the sequence of a DNA sample [16–18] after treatment with sodium bisulfite. Oligonucleotide arrays provide high-throughput, sensitive detection of variations in DNA sequence at the level of single nucleotides [18]. We report here that their application to analysis of CpG methylation provides a method that is sensitive to heterogeneous methylation within a region and detects methylated DNA in a background of unmethylated DNA. It will enable the parallel and simultaneous analysis of many individual potential sites of methylation in widely separated regions of the genome.

Materials and methods

Array fabrication

Corning 1 × 3-inch glass microscope slides were cleaned and coated with 3glycidoxypropyltrimethoxysilane (Aldrich) and polyethlyeneglycol (M_a 300; Aldrich) as described by Maskos and Southern [19]. Slides were stored in a dessicator at room temperature until use. In preparation for microarray fabrication, the synthesis area of a slide was reacted with a 1:1 (vol:vol) mixture of 0.1M protected linker phosphoramidite (MeNPOC–hexaethylene glycol β -cyanoethyl phosphoramidite) [20,21] and tetrazole in acetonitrile (Annovis, Aston, PA). The mixture was allowed to react for 2 min with the glass surface and then washed with acetonitrile.

An array of oligonucleotide probes was synthesized in situ on the resulting surface using light-directed phosphoramidite synthesis. MeNPOC-protected phosphoramidites were used in the synthesis [20]. Light for each photochemical deprotection step was spatially addressed with a Texas Instruments Digital Light Processor (DLP). The DLP was illuminated with the 365 nm peak from a 200-W Hg/Xe arc lamp. Illumination of the DLP and projection of the reflected image were accomplished with a custom optical system designed by Brilliant Technologies (Denton, TX). The image of the DLP was projected onto the reactive surface without magnification. The DLP was coordinated with a home-built fluidics system for

automated DNA synthesis [22]. Custom software generated the patterns of illumination required to fabricate the desired array of oligonucleotides. Final deprotection of the synthesized array was with a 1:1 (vol:vol) solution of ethylenediamine and ethanol for 2 h at room temperature.

Preparation of DNA and amplification of promoter regions

Cell lines H1299 and H69 were established as described by Phelps and co-workers [23] and have been deposited in the American Type Culture Collection. The cells were cultured in RPMI 1640 (Invitrogen) supplemented with 5% fetal bovine serum. Genomic DNA was purified from these cell lines as described by Maruyama et al. [24]. The extracted, purified DNA was treated with sodium bisulfite as described previously [11]. The *p16* promoter region was amplified in a PCR using 50 ng sodium bisulfite-treated genomic DNA as template and the primers 5' [Cy3 or biotin]TTAGAGGATTTGAG GGAT3' and 5'AAAACTCCATACTACTCC3'. Primers were purchased from Operon Technologies (Alameda, CA). A touchdown method was used for the first 14 cycles of amplification, starting at an annealing temperature of 68 °C and decreasing the annealing temperature 1 °C per cycle. Amplification was continued for an additional 30 cycles with an annealing temperature of 55 °C. Denaturation and extension were carried out at 94 and 72 °C, respectively. The product of this amplification was used as the template for a second set of PCRs using the same protocol. The products were desalted (NAP column; Amersham Pharmacia Biotech) and precipitated with ethanol and sodium acetate prior to being dissolved in hybridization buffer.

Array hybridization

The hybridization mixture contained, $0.1-1\mu$ M labeled analyte sample, $0.1-1\mu$ M labeled reference sample (in this case, DNA from a sample known to be unmethylated in the analyzed region), 1μ M Control Oligo 1 (5'

[Cy3]CTTGGCTGTCCCAGAATGCAAGAAGCC

CAGACGGAAACCGTAGCTGCCCTGGTAGGTT TT), and 1 μ M Control Oligo 2 (5' [Cy3]TATATCAA AGCAGTAAGTAG) in 3M tetramethyl ammonium chloride, 0.05% Trition X-100, 1mM EDTA, 10mM Tris–HCl, pH 7.5. Control oligos 1 and 2 are derived from the gene *p53* and the HIV genome, respectively, and bind to control features on the array as indicators of array quality and performance. The sample was applied to the array surface under a 22 × 22-mm cover slip. Hybridization was carried out in a closed chamber containing a pool of hybridization buffer. The array with sample was heated to 95 °C for 20 min followed by warming at 60 °C for 1 h. After hybridization, the array was washed three times with 6× SSPE (Sigma), 0.09% Tween, followed by three washes with 0.8× SSPE, 0.01% Tween at room temperature. After this wash, the array was dried centrifugally, stained with 2µ.g/mL of Cy5–Streptavidin (Amersham Pharmacia) for 5 min at room temperature, and washed with 6× S S PE, 0.09% Tween. Finally, the array was scanned using an Axon Genepix 3000 scanner to detect Cy3 and Cy5 fluorescence intensity. The signal intensity for each feature was determined using custom analysis software.

TA cloning and sequencing

The 190-bp amplicon of sodium bisulfite-treated DNA was cloned into plasmid pCR 2.1 using a TA cloning kit (Invitrogen, Carlsbad, CA) and vendor protocols. Plasmid was isolated from 18 individual colonies and the insert sequenced using an ABI3100 sequencer with T7 and M13 primers and dye-terminated DNA sequencing protocols.

Construction of 190-bp duplex for heterogeneous methylation study

A 190-bp duplex with simulated methylation at position 25 was created. The following oligonucleotides were obtained from Operon Technologies: Oligo A (5'CCACCCTCTAATAACCAACCAACCCTCCTC TTTCTTCCTCCAATACTAACAAAAAAACCCCCCT CCAACCCTATCCCTCAAATCCTCTAA), Oligo B (5'GTGTGTTTGGTGGTTGCGGAGAGGGGGAG AGGATTTGAGGGATAGGGTTG GAGGGGGTTTTTTTGTTAGTATTGGAGGAAGA AAGAGGAGGGGTTGGTTGGTTATTAGAGGGT GGGGTGGATTGT), and Oligo D (5'AAAACTCCAT ACTACTCCCCACCACCAACTCCATACTACTCCC CACCACCCACTACCTACTCTCCCCCTCTCCGCA ACCACCAAACACACACAATCCACC). Oligos A and B (70 pmol each) were phosphorylated with polynucleotide kinase (New England BioLabs). The phosphorylated DNA was phenol extracted, chloroform extracted, and then ethanol precipitated. Phosphorylated Oligo A was annealed with Oligo C, and phosphorylated Oligo B was annealed with Oligo D. The resulting duplexes were mixed in equimolar amounts and ligated with T4 ligase at 14 °C overnight. The resulting 190-bp duplex (sequence in Fig. 2, with all cytosines except cytosine 25 converted to thymidine) was amplified as described above for the *p16* promoter region.

Results

Assay for methylation by hybridization to an array of oligonucleotide probes

The essential features of the assay we have developed are shown schematically in Fig. 1. A sample of genomic DNA is treated with sodium bisulfite under conditions that convert unmethylated cytosines to deoxyuridines. Methylated cytosines remain unconverted (Fig. 1a). At least one region of interest is amplified by PCR, which recapitulates the deoxyuracils in the template as thymidines. The product is labeled during amplification with an easily detectable tag such as a fluorophore. The presence of a cytosine or a thymidine at each position corresponding to a site of potential methylation is assayed by hybridization to a set of complementary oligonucleotide probes covalently bound to a substrate (Fig. 1b). Each probe for a given position is identical, except for a center base substitution (A, C, G, or T) used to determine the analyte sequence by hybridization. Many different CpG sites are simultaneously queried with an array of many oligonucleotide probes.

As a test of this method, we have examined the methylation state of a region of the promoter for the tumor suppressor gene p16. Hypermethylation of this promoter is known to repress transcription of p16 and is associated with a number of cancers [5,25,26]. We treated samples of genomic DNA from lung tumor cell lines with sodium bisulfite and amplified and labeled a 190-bp region of the p16 promoter. The sequence of the 190-bp of interest (prior to treatment with sodium bisulfite) is shown in Fig. 2 (GenBank Accession No. AL449423, bp 65,535–65,724). The amplified DNA was analyzed by hybridization to an array of oligonucleotide probes, each 21 bases in length, synthesized directly on a glass surface by light-directed methods [15,20,21]. Spatially patterned illumination for the photodeprotection step of the synthesis was accomplished using a digital micromirror device [22,27,28].

The result of hybridization and scanning of four probes designed to query a single cytosine (cytosine number 1) is shown in Fig. 3. The DNA analyzed with the Cy5 label was from a lung tumor cell line (H1299) in which all of the CpG dinucleotides in the analyzed 190- base

region were previously found to be methylated (S. Zochbauer-Muller, unpublished). We independently confirmed this methylation state using dye-terminated sequencing of bisulfite-treated DNA. The feature with the highest signal of the four features shown is the one probing for a cytosine (the variable base in the probe is a guanine). The ratio of the signal for this feature to the next highest signal (in the feature probing for a guanine) is 2.8, identifying the base in the analyte as a cytosine. A cytosine at this position was anticipated as the outcome of bisulfite treatment of the methylated base.

The comparison most relevant to detection of methylation is between the signal in the feature that probes for a cytosine at each position and the signal in the feature that probes for a thymidine at the same position in the bisulfite-treated DNA. The ratio of these signals (C:T) is listed for each of the cytosines in the analyzed sequence in Table 1. Cytosines outside of CpG dinucleotides, which are not methylated, serve as an internal indicator for the effectiveness of the bisulfite treatment in converting unmethylated cytosines to deoxyuracils and for the discrimination between cytosines and thymidines by the probes on the array. The ratio of signals in those features ranges from 0.24 to 1.09. Because our independent sequence analysis of the bisulfite- treated DNA confirmed complete conversion of all unmethylated cytosines to deoxyuracils, values near 1.0 for this ratio are likely due to imperfect discrimination between cytosine and thymidine in hybridization. At cytosine 1, the position queried by the probes shown in Fig. 3, the ratio of signals (C:T) is 3.57. The values range from 1.91 to 13.8 for cytosines in CpG dinucleotides (Table 1), in all cases considerably higher than the highest ratio of signals for the unmethylated cytosines.

To provide an objective standard for discrimination between methylated and unmethylated cytosines and to facilitate visualization of *changes* in methylation state, a reference sequence containing a different label was cohybridized with the array. As a model reference sequence, we have used DNA from a different lung tumor cell line (small cell lung cancer H69) in which the p16 promoter has been found to be unmethylated at each CpG in the 190-base region of interest (S. Zochbauer- Muller, unpublished). We have also confirmed these results using dye-terminated sequencing of bisulfite- treated DNA. The same 190-base region (Fig. 2) of H69 was amplified with a primer labeled with Cy3.

The result for cytosine number 1 is shown in Figs. 3b and c. The probe for thymidine has the highest signal intensity, and the C:T ratio for the reference strand is 0.52 at this position. A useful method for judging changes in methylation state is to compare the C:T ratio for a set of probes with the analyte fluorophore to the C:T ratio for the same probes with the reference fluorophore. In Fig. 3 the ratio of sample fluorophore (Cy5) C:T ratio to reference fluorophore (Cy3) C:T ratio is 6.8. Using a ratio of ratios in this manner reduces the effects of imperfect hybridization specificity on the results.

The ratio of ratios was computed for each cytosine in the original sequence and is listed in Table 1. Cytosines not part of a CpG were used as an internal standard for unmethylated positions. The ratio of signal ratios for these cytosines had a mean of 1.59 and a standard deviation of 0.49 (n = 20) and were distributed normally, allowing calculation of a Z score (see legend to Table 1). In the H1299 sample, the values for all 16 cytosines in CpGs were at least four standard deviations from the mean of values for cytosines not in CpGs (Fig. 4a, Z scores listed in Table 1). This experiment was performed in six replicates (three separate hybridizations on arrays that contained duplicate sets of probes) with equivalent results (Z scores greater than 4.0 for all cytosines in CpG dinucleotides) using different preparations of bisulfite-treated DNA in each hybridization. Two different preparations of genomic DNA were used for each of the analytes and reference samples in different hybridizations. An experiment in which the dye labels were reversed between the analyte and the reference samples also yielded equivalent results (data not shown).

Specificity for detection of heterogeneous methylation

The region of the *p16* promoter that we have studied is uniformly methylated at all CpG sites in the H1299 cell line. However, promoter regions are frequently not uniformly methylated. This nonuniformity of methylation can have important biological consequences, because methylation of all CpG sites within a promoter region does not have equal effect on transcription [12]. Thus, the ability for the assay to independently discriminate methylation states at different CpG sites is essential.

To test the ability of the assay to detect methylation at an individual site and to further define the threshold for assignment of methylation state, we created a 190-bp test duplex by chemical synthesis and ligation. One strand of the duplex is identical in sequence to bisulfite-treated H69 genomic DNA, except the position of the 25th cytosine simulates methylation by being a cytosine rather than a thymidine. The test duplex was labeled by amplification with a labeled primer, and bisulfite-treated DNA from H69 lung tumor cells was amplified and labeled for use as a reference sequence. Cohybridization of the analyte and reference samples to the array resulted in the ratios of analyte (C:T) to reference (C:T) listed in Table 1 for all 36 cytosines.

The site of simulated methylation had an analyte (C:T):reference (C:T) ratio of 2.38, nearly eight standard deviations (Z score = 7.97) from the mean of that ratio for the cytosines not in CpG dinucleotides $(1.13 \pm 0.16; n = 20)$. This ratio for the other cytosines in CpGs ranged from 0.91 to 1.64. These differed from the mean for the internal standard cytosines by -1.8 to 3.6 standard deviations (Fig. 4b and Table 1). Thus, the authentic cytosine could be clearly distinguished from the other potential positions of methylation by its considerably larger variation from the internal standards. The range of ratios for the positions simulating unmethylated CpGs suggests a threshold Z score of greater than 3.6 (i.e., greater than 3.6 standard deviations from the mean of the internal standards) to indicate a genuine difference from an unmethylated cytosine. Six replicates of this experiment gave equivalent results, though in one of the six replicates the Z score for cytosine 30 was 4.3, above the threshold of 3.6.

Detection of methylated DNA in the presence of unmethylated DNA

Biological samples of genomic DNA often include individual CpG sites that are partially but not exhaustively methylated [14]. Thus, the ability to detect methylated cytosines within analytes that contain a significant amount of DNA that is not methylated at the queried positions is desirable. To test this ability, we performed the array hybridization assay with a mixture of samples prepared from methylated and unmethylated DNA.

The 190-base region shown in Fig. 2 was amplified separately from bisulfite-treated samples of genomic DNA from H1299 and H69. The amount of amplified DNA from each sample was estimated by visualization on an agarose gel, and the amplified samples were mixed in a ratio of approximately 20:80 (H1299:H69). This mixture approximates a sample in which 20% of each CpG is methylated. The mixture was labeled by an additional amplification with a labeled primer. A reference sample (derived purely from H69) was also amplified and labeled, and the analyte mixture and reference were cohybridized to the methylation probe array.

The results of this hybridization are summarized in Table 1. Of the 16 cytosines in CpG dinucleotides, 8 had Z scores greater than 3.6, identifying them as partially methylated (Fig. 4c). The remaining 8 could not be distinguished from bases converted entirely to deoxyuracils by treatment with bisulfite.

Discussion

This study demonstrates the utility of array-based sequence analysis to the parallel detection of methylation in CpG islands. The differential reactivity of bisulfite with cytosine and 5methylcytosine forms the basis of several techniques for the assessment of DNA methylation; however, new approaches to the read-out of the sequence that results from treatment with bisulfite are desirable. The need for high-throughput methods is highlighted by the prevalence of CpG islands in the genome. Computer analysis of the March 2001 Unigene build revealed that 32,597 of the 92,152 clusters contained CpG islands. Of the 14,968 clusters with annotation, 10,438 have CpG islands. These islands in the annotated clusters comprise 4,398,560 bp in 5' noncoding regions, 7,074,411 bp in coding regions, and 492,323 bp in 3' noncoding regions. A high-throughput method such as described here will be necessary to interrogate even a small fraction of these sites in a given experiment. Sequence analysis by hybridization to oligonucleotide arrays is an approach that affords a high degree of parallelism and flexibility.

Another group has recently reported use of oligonucleotide arrays to probe the methylation states of a small number of CpG dinucleotides from each of a large number of promoters, introns, and exons (from 56 genes) in parallel [29]. This genome-wide survey of methylation sites using spotted (i.e., not synthesized in situ, directly on the glass substrate) oligonucleotide probes and a single-color analysis allowed accurate determination of tumor classes in samples for which there was not prior knowledge of tumor class. That study demonstrated the level of parallelism that can be achieved across widely disparate regions of the genome in a single assay, a clear advantage over the methodologies that are currently in widespread use. The present report comprehensively describes the performance of such an assay at all individual CpG sites within a single amplified region and describes the use of a distinctly labeled reference sample to create an objective standard for judging methylation state.

The success of this assay relies on discrimination between a cytosine and a thymidine in the array hybridization. However, in experiments reported here, the specificity of this discrimination varied considerably, both in a context-dependent way within one experiment and from one experiment to another. For example, the C:T ratio at positions that were confirmed independently by sequencing to be thymidines (i.e., unmethylated and completely converted by bisulfite) was occasionally larger than 1.0. This variable specificity may be due in part to the relatively high stability of the G \cdot T mismatch in some sequence contexts [30] and cross-hybridization with homologous sequences in the analyte. Variation in specificity at a given base from one experiment to another may be due to small variations in sample composition and in hybridization and wash conditions. This variability is most directly accounted for experimentally by comparison to a cohybridized sample of reference methylation state.

The comparison to a sample of reference methylation state is especially useful, because information about differences in methylation state is often sought. In the demonstration described here, the difference between the analyte sample and a sample known to be unmethylated was assayed. However, many comparisons are possible, such as DNA from diseased tissue compared to a matched sample from healthy tissue or DNA from tissue at different points along a disease progression. As is apparent in Fig. 3c, cohybridization with a reference sample containing a different label facilitates visualization of changes in methylation state; the presence of two colors in one set of four probes is visually obvious.

After the context dependence of variability is accounted for, other aspects of experimental variability can be assessed using the known unmethylated positions as internal standards.

The Z scores calculated here offer a measure of the statistical significance of the difference between the analyte to reference ratio of a given interrogated cytosine and those known to be unmethylated. The use of an empirically determined threshold Z score to judge methylation state is analogous to the use of an empirically determined threshold signal ratio to identify nucleotides in standard array-based sequence analysis [31]. Under the experimental conditions that we have used, the Z score that we calculate is clearly correlated with methylation state, and a single cytosine corresponding to a uniquely methylated position is distinguished from the unmethylated cytosines.

Other workers have used arrays of small numbers of spotted oligonucleotide probes to analyze the collective methylation state of clusters of CpG dinucleotides [32]. These workers used a single-color hybridization and a calibration curve for each probe to determine methylation state for each cluster. They concluded that analyte cross-reactivity obstructed independent analysis of individual cytosines. The CpG-rich regions of interest for methylation analysis are highly repetitive. They are also depleted in adenines, and after treatment with bisulfite, they can be depleted in cytosines. This low sequence complexity limits the uniqueness of probes that interrogate different base positions, making crosshybridization likely. However, our results show that it is possible to detect methylation at an individual cytosine by hybridization to probes synthesized in situ using internal controls such as cytosines outside of CpG dinucleotides and a cohybridized reference sample. Future experiments will test the generality with which this assay can interrogate independent sites for methylation, but the results of this work indicate that many or most sites can be interrogated independently.

At many sites, the array hybridization assay was able to unambiguously detect as little as 20% methylation in a background of unmethylated DNA. Furthermore, the predominance of Z scores close to the threshold for assignment of methylation could be interpreted as a likely indicator of low levels of methylation at the several sites that fall just below that threshold. The criterion for detection of methylation developed in this assay is stringent, the threshold Z score being defined to make the probability of false positives low. However, the Z scores for different cytosines display different sensitivities to the extent of methylated, small amounts of methylation will be more difficult to detect with confidence. Sensitive detection of small amounts of methylation at certain cytosines requires specific calibration of the assay at those cytosines. Although this assay is not best used to rule out small amounts of methylation or to quantitate extent of methylation, its sensitivity as an indicator of methylation is comparable to (or better than) that of standard dye-terminated sequencing of bulk mixtures of bisulfite-treated DNA.

The probes for any given cytosine often include bases complementary to cytosines in other CpG dinucleotides. In general, the methylation states of these proximal CpG sites will be unknown. Thus, there is uncertainty about the sequence that perfectly complements the analyte beyond the queried position. The number of possible complementary probe sequences is 2^n , where *n* is the additional number of CpG dinucleotides in the probe. For example a probe containing the queried cytosine in a CpG dinucleotide and three other cytosines in CpG dinucleotides would require eight (i.e., 2^3) alternate sets of probes to query all possible combinations of methylation states.

The results described here were obtained from experiments in which the probe sequences were designed with the assumption that all cytosines outside of the one being queried were unmethylated. However, we included the alternate probe sets in the arrays that we synthesized to evaluate their usefulness in the assay. Signal ratios, analyte to reference ratios, and Z scores were calculated using the brightest set of features in each channel, using

the brightest set of features in the analyte channel, and using the brightest set of features in the reference channel (data not shown).

Although the brightest feature in any channel was frequently consistent with the known methylation state of the proximal CpG dinucleotides in that sample, that correspondence was not always found. Exceptions are likely to have resulted from insensitivity to mismatches near the ends of the probes coupled with small random variations in hybridization efficiency at different positions on the array. In all cases when the feature set used was selected based on signal intensity, the signal ratios, analyte to reference ratios, and Z scores were more poorly correlated with known methylation state than when the set of probes designed assuming absence of methylation was used. Nevertheless, these methods of analysis bear further investigation for their potential usefulness in refinement of the assay.

Additional probes might also be included to interrogate the other possible strands of DNA that reflect methylation status of a region. After bisulfite treatment, the two strands of genomic DNA are no longer mutually complementary. Amplification of each produces two complementary strands of different sequence. Thus, information about the methylation state of the initial sequence is contained in four different sequences of DNA, each of which can be analyzed independently on the same array.

Though the additional probes described above may enhance the assay, this work demonstrates that as few as two array features can be used to effectively probe each cytosine in a region of interest. Thus, using light-directed methods of high feature density array synthesis, hundreds of thousands of features can be created on a single array to probe, in parallel, hundreds of thousands of potential methylation sites in widely dispersed regions of the genome. Methods of array synthesis that allow high feature densities and facile changes in probe content, such as the method used in this report using a micromirror array, will make this technique particularly valuable for the de novo discovery of sites of aberrant methylation states.

Acknowledgments

This work was supported by Grants, P50 CA70907 and R33 CA81656 from the National Cancer Institute (NCI) and by a grant from the Donald W. Reynolds Foundation. R.P.B. was supported by a McDermott Center Human Genomics training grant. We are grateful to Yuan Qi for computational analysis of CpG island occurrence and Glenn McGall of Affymetrix for generously supplying phosphoramidites.

References

- Riggs A, Pfeifer G. X-chromosome inactivation and cell memory. Trends Genet. 1992; 8:169–174. [PubMed: 1369742]
- 2. Tremblay K, Saam J, Ingram R, Tilghman S, Bartolomei M. A paternal-specific methylation imprint marks the alleles of the mouse H19 gene. Nat Genet. 1995; 9:407–413. [PubMed: 7795647]
- 3. Sekido Y, Fong KM, Minna JD. Progress in understanding the molecular pathogenesis of human lung cancer. Biochim Biophys Acta. 1998; 1378(1):F21–F59. [PubMed: 9739759]
- 4. Rountree M, Bachman K, Herman J, Baylin S. DNA methylation, chromatin inheritance, and cancer. Oncogene. 2001; 20(24):3156–3165. [PubMed: 11420732]
- Zochbauer-Muller S, Fong KM, Virmani AK, Geradts J, Gazdar AF, Minna JD. Aberrant promoter methylation of multiple genes in non-small cell lung cancers. Cancer Res. 2001; 61(1):249–255. [PubMed: 11196170]
- 6. Burbee DG, Forgacs E, Zochbauer-Muller S, Shivakumar L, Fong K, Gao B, Randle D, Kondo M, Virmani A, Bader S, Sekido Y, Latif F, Milchgrub S, Toyooka S, Gazdar AF, Lerman MI, Zabarovsky E, White M, Minna JD. Epigenetic inactivation of RASSF1A in lung and breast cancers and malignant phenotype suppression. J Natl Cancer Inst. 2001; 93(9):691–699. [PubMed: 11333291]

- Issa J, Ottaviano Y, Celano P, Hamilton S, Davidson N, Baylin S. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. Nat Genet. 1994; 7:536–540. [PubMed: 7951326]
- Yan PS, Chen C-M, Shi H, Rahmatpanah F, Wei SH, Caldwell CH, Huang TH-M. Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. Cancer Res. 2001; 61:8375–8380. [PubMed: 11731411]
- 9. Wang R, Gehrke C, Ehlich M. Comparison of bisulfite modification of 5-methyldeoxycytidine and deoxycytidine residues. Nucleic Acids Res. 1980; 8:4777–4790. [PubMed: 7443525]
- Frommer M, McDonald L, Millar D, Collis C, Watt F, Grigg G, Molloy P, Paul C. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci USA. 1992; 89:1827–1831. [PubMed: 1542678]
- Herman J, Graff J, Myohanen S, Nelkin B, Baylin S. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci USA. 1996; 93:9821–9826. [PubMed: 8790415]
- Deng G, Chen A, Hong J, Chae HS, Kim YS. Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression. Cancer Res. 1999; 59(9): 2029–2033. [PubMed: 10232580]
- Melki J, Vincent P, Clark S. Concurrent DNA hypermethylation of multiple genes in acute myeloid leukemia. Cancer Res. 1999; 59:3730–3740. [PubMed: 10446989]
- 14. Xu X, Wu L, Du F, Davis A, Peyton M, Tomizawa Y, Maitra A, Tomlinson G, Gazdar A, Weissman B, Bowcock A, Baer R, Minna J. Inactivation of human SRBC, located within the 11p15.5-p15.4 tumor suppressor region, in breast and lung cancers. Cancer Res. 2001; 61:7943– 7949. [PubMed: 11691816]
- Pease AC, Solas D, Sulivan EJ, Cronin MT, Holmes CP, Fodor SPA. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci USA. 1994; 91:5022–5026. [PubMed: 8197176]
- Hacia JG, Collins FS. Mutational analysis using oligonucleotide microarrays. J Med Gen. 1999; 36(10):730–736.
- Hacia JG, Sun B, Hunt N, Edgemon K, Mosbrook D, Robbins C, Fodor SP, Tagle DA, Collins FS. Strategies for mutational analysis of the large multiexon ATM gene using high-density oligonucleotide arrays. Genome Res. 1998; 8(12):1245–1258. [PubMed: 9872980]
- Hacia J, Brody L, Chee M, Fodor S, Collins F. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-color fluorescence analysis. Nature Genet. 1996; 14:441–447. [PubMed: 8944024]
- Maskos U, Southern E. Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesized in situ. Nucleic Acids Res. 1992; 20:1679–1684. [PubMed: 1579459]
- McGall G, Barone A, Diggeleman M, Fodor S, Gentalen E, Ngo N. The efficiency of light-directed synthesis of DNA arrays on glass substrates. J Am Chem Soc. 1997; 119:5081–5090.
- McGall GH, Fidanza JA. Photolithographic synthesis of high-density oligonucleotide arrays. Methods Mol Biol. 2001; 170:71–101. [PubMed: 11357690]
- 22. Luebke, KJ.; Balog, RP.; Mittelman, D.; Garner, HR. Digital optical chemistry: a novel system for the fabrication of custom oligonucleotide arrays. In: Kordal, R.; Usmani, AM.; Law, WT., editors. Microfabricated Sensors: Application of Optical Technology for DNA Analysis. American Chemical Society; Washington, DC: 2002. p. 87-106.
- Phelps R, Johnson B, Ihde D, Gazdar A, Carbone D, McClintock P, Linnoila I, Matthews M, Bunn PJ, Carney D, Minna J, Mulshine J. NCI-Navy medical oncology branch cell line data base. J Cell Biochem Suppl. 1996; 24:32–91. [PubMed: 8806092]
- Maruyama R, Toyooka S, Toyooka KO, Harada K, Virmani AK, Zochbauer-Muller S, Farinas AJ, Vakar-Lopez F, Minna JD, Sagalowsky A, Czerniak B, Gazdar AF. Aberrant promoter methylation profile of bladder cancer and its relationship to clinicopathological features. Cancer Res. 2001; 61(24):8659–8663. [PubMed: 11751381]

- Herman J, Merlo A, Mao L, Lapidus R, Issa J, Davidson N, Sidransky D, Baylin S. Inactivation of the CDKN2/p167/MTS1 gene is frequently associated with aberrant methylation in all common human cancers. Cancer Res. 1995; 55:4525–4530. [PubMed: 7553621]
- 26. Merlo A, Herman J, Mao L, Lee D, Gabrielson E, Burger P, Baylin S, Sidransky D. 5' CpG island methylation is associated with transcriptional silencing of the tumor suppressor p16/CDKN2/ MTS1 in human cancers. Nat Med. 1995; 1:686–692. [PubMed: 7585152]
- Jaklevic J, Garner H, Miller G. Instrumentation for the human genome project. Annu Rev Biomed Eng. 1999; 1:649–678. [PubMed: 11701503]
- Singh-Gasson S, Green RD, Yue Y, Nelson C, Blattner F, Sussman MR, Cerrina F. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. Nat Biotechnol. 1999; 17(10):974–978. [PubMed: 10504697]
- 29. Adorjan P, Distler J, Lipscher E, Model E, Muller J, Pelet C, Braun A, Florl A, Maier S, Muler V, Otto T, Scholz C, Schulz W, Seifert H, Schwope I, Ziebarth H, Berlin K, Piepenbrock C, Olek A. Tumour class prediction and discovery by microarray-based DNA methylation analysis. Nucleic Acids Res. 2001; 30(5):e21. [PubMed: 11861926]
- Allawi H, SantaLucia JJ. Thermodynamics and NMR of internal G•T mismatches in DNA. Biochemistry. 1997; 36:10581–10594. [PubMed: 9265640]
- 31. Kozal M, Shah N, Shen N, Yang R, Fucini R, Merigan T, Richman D, Morris D, Hubbell E, Chee M, Gingeras T. Extensive polymorphisms observed in HIV-clade B protease gene using high-density oligonucleotide arrays. Nat Med. 1996; 2(7):753–759. [PubMed: 8673920]
- Gitan R, Shi H, Chen CM, Yan P, Huang T. Methylation specific oligonucleotide array: a new potential for high-throughput methylation analysis. Genome Res. 2001; 12:158–164. [PubMed: 11779841]



Fig. 1.

Assay for CpG methylation by treatment with sodium bisulfite and sequence analysis with an oligonucleotide array. (a) Treatment with sodium bisulfite converts all unmethylated cytosines to deoxyuracils while methylated cytosines remain unconverted. In this sequence, four cytosines are unmethylated and converted to deoxyuracils while one cytosine, denoted as methylated with a superscript Me, remains a cytosine. (b) Sequence analysis of a labeled representative of the bisulfite-treated DNA by hybridization to an array of oligonucleotides. The oligonucleotide probes are covalently bound to a substrate. The central base of each probe for a given position is varied to test for the identity of the base by hybridization. The probe with which the most label is associated identifies the base at the central position. A cytosine at the probed position indicates methylation that prevented conversion by sodium bisulfite.



Fig. 2.

The sequence of the 190-bp region of the p16 promoter studied in this work. After treatment with bisulfite, the strand shown was amplified using the indicated primers and labeled. This region contains 36 cytosines (indicated with capital letters). The numbers correspond to Table 1. Of these, 16 cytosines are within CpG dinucleotides (shown in red, capital, and underlined)) and 20 cytosines are not within CpG dinucleotides.



Fig. 3.

Four probes from an array to analyze the methylation state of a region of the promoter for *p16*. The array was hybridized as described in the text, washed, and scanned for fluorescence. Each 21-nucleotide probe is complementary to the sequence surrounding cytosine number 1, with a different base for each probe in apposition to cytosine number 1. For example, the probe for A has a thymidine in that central position. The four probe features shown were on an array comprising 4620 probe features, which included probes for each position in the 190-nucleotide analyte sequence and control features and features for testing probe design principles (see Discussion). (a) Fluorescence scan of the Cy5 (analyte) channel of the array. (b) Fluorescence scan of the Cy3 (reference) channel of the array. (c) Overlay of the analyte and reference channels demonstrating the appearance of a methylated site compared with an unmethylated reference. The rectilinear pattern within each feature is due to resolution of the effect of individual micromirrors in the synthesis of DNA probes.





Fig. 4.

Histogram plots showing Z scores for each cytosine in a CpG dinucleotide analyzed, numbered as in Fig. 2. The Z score is defined as described in the legend to Table 1. The threshold for calling methylation is set to 3.6, indicated by the horizontal line at that value. In each case the reference sample was derived from unmethylated DNA. (a) Results of analysis in which the analyte was derived from uniformly methylated DNA. (b) Results of analysis in which the analyte was derived from a synthetic duplex simulating unique methylation at cytosine number 25. (c) Results of analysis in which the analyte was derived from a 80% unmethylated DNA.

Table 1

Signal intensity ratios for each analyzed cytosine

H1299 and H69d					25th C duplex ^e				20:80 Mixturef			
Cytosine number <i>§</i>	C:T ratio analyte ^A	C:T ratio reference ^a	Analyte (C:T)/Reference (C:T) $m{b}$	Z Score ^C	C:T ratio analyte <i>d</i>	C:T ratio reference ^d	Analyte (C:T)/Reference (C:T) $m{b}$	Z score ^C	C:T ratio analyte ^d	C:T ratio reference <i>d</i>	Analyte (C:T)/Reference (C:T) $m{b}$	Z score ^C
1	3.57	0.52	6.80	10.7	0.86	0.88	66.0	-0.90	0.99	0.52	1.92	4.61
2	0.46	0.54	0.85	-1.50	0.74	0.69	1.08	-0.29	0.70	0.70	1.00	1.01
3	0.44	0.36	1.23	-0.72	0.75	0.75	1.00	-0.82	0.39	0.32	1.20	1.80
4	0.39	0.29	1.34	-0.50	0.87	0.86	1.01	-0.76	0.44	0.36	1.22	1.88
5	13.8	0.39	35.7	69.7	0.90	0.89	1.01	-0.75	1.16	0.49	2.35	6.29
9	0.24	0.22	1.13	-0.94	1.07	0.96	1.12	-0.08	0.32	0.64	0.5	-0.97
7	0.34	0.36	0.94	-1.33	1.01	0.99	1.01	-0.72	0.50	0.76	0.65	-0.37
8	0.36	0.41	0.88	-1.45	0.70	0.58	1.22	0.58	0.36	0.62	0.58	-0.64
6	0.33	0.27	1.23	-0.73	0.68	0.65	1.05	-0.50	0.34	0.64	0.53	-0.85
10	9.28	0.41	22.5	42.8	0.82	0.68	1.20	0.46	1.43	0.67	2.15	5.51
=	0.93	0.53	1.76	0.36	0.85	0.88	0.97	-1.00	0.62	06.0	0.69	-0.20
12	1.09	0.48	2.29	1.44	1.01	0.72	1.41	1.79	0.70	0.55	1.28	2.08
13	0.65	0.52	1.23	-0.69	0.85	0.76	1.11	-0.10	0.61	0.93	0.66	-0.35
14	0.65	0.51	1.23	-0.60	0.83	0.80	1.05	-0.52	0.51	0.68	0.74	-0.02
15	1.08	0.60	1.81	0.44	0.92	0.93	660	-0.87	0.61	0.98	0.62	-0.48
16	3.55	0.54	6.64	10.3	0.94	0.72	1.30	1.12	1.90	0.86	2.21	5.71
17	0.27	0.11	2.44	1.75	0.62	0.56	1.11	-0.11	0.20	0.51	0.39	-1.41
18	1.99	0.46	4.34	5.62	0.9	1.06	0.85	-1.76	0.50	0.42	1.19	1.73
19	2.36	0.60	3.91	4.75	1.10	0.76	1.45	2.08	1.04	0.57	1.83	4.25
20	1.91	0.53	3.63	4.18	1.01	0.82	1.23	0.68	1.99	1.04	1.92	4.58
21	0.40	0.18	2.27	1.39	0.51	0.45	1.14	0.08	0.35	0.62	0.57	-0.69
22	3.11	0.69	4.54	6.05	0.82	0.71	1.16	0.24	2.17	1.39	1.56	3.19
23	3.38	0.59	5.73	8.46	1.07	0.68	1.56	2.77	2.20	1.41	1.59	3.32
24	0.45	0.27	1.68	0.20	0.60	0.49	1.22	0.62	0.34	0.49	0.70	-0.17
25	3.55	0.52	6.81	10.7	1.48	0.62	2.38	7.97	1.12	0.74	1.51	2.99
26	0.62	0.29	2.11	1.07	0.81	0.75	1.08	-0.29	0.69	0.78	0.89	0.59
27	0.46	0.29	1.58	-0.01	0.7	0.74	0.94	-1.17	0.49	0.87	0.56	-0.73
28	2.88	0.52	5.52	8.02	1.00	0.89	1.12	-0.04	1.24	0.63	1.98	4.82
29	2.11	0.43	4.85	6.66	0.93	0.58	1.59	2.95	0.93	0.96	0.96	0.85
30	3.40	0.42	8.09	13.3	1.03	0.62	1.67	3.47	0.91	1.11	0.82	0.29
31	0.70	0.38	1.87	0.57	0.77	0.58	1.32	1.23	0.59	0.73	0.81	0.25
32	0.60	0.34	1.75	0.33	0.79	0.50	1.57	2.82	0.53	0.67	0.80	0.21

H1299 and H69 d					25th C duplex ^e				20:80 Mixtur&			
Cytosine number <i>8</i>	C:T ratio analyte ^d	C:T ratio reference a	Analyte (C:T)/Reference (C:T) $m{b}$	Z Score ^C	C:T ratio analyte ^d	C:T ratio reference ^d	Analyte (C:T)/Reference (C:T) $m{b}$	Z score ^C	C:T ratio analyte ^đ	C:T ratio reference ^d	Analyte (C:T)/Reference (C:T) $m{b}$	Z score ^C
33	0.37	0.18	2.04	0.93	0.57	0.50	1.14	60.0	0.30	0.59	0.51	-0.93
34	2.14	0.52	4.10	5.13	0.82	0.63	1.30	1.09	1.16	0.63	1.85	4.33
35	2.11	0.44	4.77	6.51	1.21	0.72	1.69	3.55	1.31	1.33	0.98	0.93
36	4.48	0.49	9.15	15.5	1.18	0.80	1.47	2.20	2.28	1.66	1.38	2.48
c.												
^a Mean fluorescenc	se signal in the regi	ion defined by the pro-	obe for cytosine at a position	divided by	the mean fluoresc	ence signal in the reg	gion defined by the probe for	thymidine .	at the same positiv	on.		
$p_{\tilde{c}} = p_{\tilde{c}}$	•					•	, , , ,					

C:T ratio at a probed position for the fluorescence channel corresponding to the analyte sample divided by the C:T ratio at the same position for the fluorescence channel corresponding to the reference sample.

 c^2 score = $(R - R_0)/S$, where R = analyte (C:T)/reference (C:T) at a given position, R_0 =mean analyte (C:T)/reference (C:T) for all cytosines not in CpGs, and S = standard deviation in the mean analyte (C:T)/reference (C:T) for all cytosines not in CpGs.

d Analysis of sample derived from fully methylated DNA from lung tumor cell line H1299 with reference derived from unmethylated DNA from lung tumor cell line H69.

 e^{a} Analysis in which the analyte was derived from a synthetic duplex simulating unique methylation at cytosine number 25.

f Analysis in which the analyte was derived from a mixture of approximately 20% methylated DNA and 80% unmethylated DNA.

 $\ensuremath{\mathcal{E}}$ Cytosines in CpG islands are in shaded rows.

Balog et al.

NIH-PA Author Manuscript