Applications of CpG Island Microarrays for High-Throughput Analysis of DNA Methylation

Pearly S. Yan, Chuan-Mu Chen,* Huidong Shi, Farahnaz Rahmatpanah, Susan H. Wei and Tim Hui-Ming Huang

Department of Pathology and Anatomical Sciences, Ellis Fischel Cancer Center, University of Missouri, Columbia, MO 65203 and *Department of Zoology, Life Science College, National Chung Hsing University, Taichung 402, Taiwan, Republic of China

ABSTRACT Differential methylation hybridization (DMH) is a high-throughput microarray technique designed to identify changes in DNA methylation patterns commonly observed in cancer and other disease states. The DMH methodology comprises three fundamental components: the arraying of CpG island clones on glass slides, the preparation of the sample amplicons under investigation, and the hybridization of amplicon targets onto the CpG island microarray. Herein, we outline the DMH protocol and illustrate its utility and the validation approaches used in analyzing the hypermethylation profile of breast tumor and normal samples.

KEY WORDS: • DNA methylation • CpG islands • microarray

There is accumulated evidence that dietary methyl donors such as folic acid can affect the level of DNA methylation in humans (1). DNA methylation is vital in establishing, defining, or stabilizing the various cell types in the developing embryo (2). Methylation in specific guanine cytosine (GC) rich regions, called cytosine guanine dinucleotide (CpG) islands (3), is frequently associated with transcriptional silencing of genes on the inactive X chromosome (4) and imprinted genes (5). In addition, CpG methylation may serve to repress undesired expression of repetitive elements, transposons, and endogenous retroviral sequences in the cell (6). It is, therefore, easy to see the potential consequences of perturbation of DNA methylation by nutrients, such as folic acid, vitamin B-12, methionine, choline, and betaine, which all play critical roles in maintaining the methylation status via S-adenosylmethionine (SAM) and other one-carbon metabolites (7). Findings from epidemiological studies, clinical studies, and experimental studies indicate that deficiency in dietary sources of methyl donors are associated with an increased risk of cancer in the colon (8), the breast (9), the cervix (10), the lung (11), the stomach (12), the pancreas, and multiple myeloma (13). It is reasonable to hypothesize that a reduced level of methyl donors would perturb the one-carbon metabolism, leading to abnormal DNA methylation in the genome, as is often observed in tumor cells. It is interesting to note that these changes are not uniform across the tumor genome. It is not unusual to detect an overall loss in methylation while there are pockets of increase in DNA methylation. Loss of methylation or hypomethylation usually occurs in the normally methylated CpG sites. This global hypomethylation may promote genomic instability or chromosomal rearrangements seen in some cancers (14). If hypomethylation occurs to occur in growth-inducing genes such as oncogenes, their overexpression may contribute to tumorigenesis. Conversely, the normally unmethylated CpG islands (CGI) can become hypermethylated in neoplastic cells. DNA methylation may silence genes via alteration of chromatin structure in promoter CGI, preventing normal interaction between these regions and transcriptional machinery (15). If hypermethylation occurs in tumor suppressor genes, their inactivation may lead to formation of cells with increased carcinogenic potential. To complicate matters further, genetic polymorphisms, such as those associated with methylentetrahydrofolate reductase (EC 1.5.1.20), are found to be associated with an altered risk of colorectal...
cancer in humans, probably through modification of SAM bioavailability (16). To reduce complexity, one can perform well-defined studies using animal models or cultured cell lines to examine the link between dietary methyl donors and DNA methylation. Tissue or cell line DNA from these studies can then be characterized by methylation analysis. Herein, we describe a microarray technique, differential methylation hybridization (DMH), developed in our laboratory that allows one to investigate the methylation status of multiple CGI in the human genome. This paper highlights the pertinent features of our DMH assay in its present microarray format, describes the steps to validate the assay, presents DNA methylation data generated from our CGI microarray, and outlines potential ways to use the CGI microarray.

MATERIALS AND METHODS

The DMH assay encompasses three components: the printing of the CGI microarray, the preparation of the fluorescently-labeled targets, and the analysis of the scanned image.

Preparation of the CGI microarray

The schematic for the preparation of the CGI microarray is shown in Fig. 1. Briefly, the resource material for printing the microarray was obtained from the UK Human Genome Mapping Project Resource Centre, Hinxton, Cambridge. This CGI library originates from Dr. Sally Cross and coworkers (17). Male human genomic DNA was restricted with a four-base frequent cutter MseI. The recognition site of this restriction enzyme [thymine/thymine/adenine/adenine four-base sequence (TTAA)] rarely occurs in GC-rich regions; thus, most GC-rich CGI remained intact whereas the bulk of the genomic DNA was removed by restriction enzyme cleavage. The restricted product, including the CpG fragments of promoter regions and the first exon of many transcripted genes, were subsequently subjected to in vitro methylation. The reaction mixture was passed through a methyl-CpG-binding protein 2 column to enrich the low copy and single-copy CGI fragments. One should note that not all promoter CGI are flanked by MseI restriction sites; examples of such promoter CGI include p16\(^{INK4a}\), p15\(^{INK4b}\), and tissue inhibitor of metalloproteinase 3 (TIMP3) (18). As a result, these fragments will not be present in the current version of our CGI microarray. We are in the process of generating other versions of CGI microarrays that will allow us to probe the methylation status of these important fragments. Nonetheless, the present version of the CGI microarray does generate an impressive amount of information and data.

Individual clones from the CGI library were organized in 96-well culture plates. Two more copies of each of these plates were generated in glycerol stocks and stored in separate freezers to safeguard against unforeseen clone mix-ups and contaminations. The arrayed probes were polymerase chain reaction (PCR) products generated by transferring a portion of the clones from 96-well culture plates to 96-well PCR plates using a 96-pin replicator, as described previously (19). The primers for the colony PCR reactions were those provided with the CGI library. The amplification efficiencies can be rapidly determined with the aid of a robotic liquid-handling device and a 96-well format gel electrophoresis system. A representative image of 96-well PCR products is shown in Fig. 1. Our preliminary hybridization trials showed that purification of PCR products is not a prerequisite for strong probe/target hybridization. Our choice of arrayer, namely the Affymetrix/GMS 417 Arrayer, permits the dotting of unpuriﬁed target material, and the analysis of the scanned image.

Preparation of fluorescently labeled targets

To illustrate the usefulness of the DMH assay, methylation proﬁles of paired breast tumor and normal samples from patients undergoing mastectomy at Ellis Fischel Cancer Center (Columbia, MO) are described and presented. The collection of breast tumor and normal samples was in compliance with our institutional review board. Routinely, all DNA used in amplicon generation is extracted from samples using the QiaAmp Tissue Kit (Qiagen, Valencia, CA). To ensure successful preparation of amplicons or the amplified target material, the quality and the quantity of the DNA has to be ascertained with care. As such, the spectrophotometric readings of extracted DNA are verified on a low percentage agarose gel. The amplicon preparation steps are outlined in Fig. 2. First, DNA (1–2 μg) is restricted with MseI, the same four-base restriction enzyme that generates the CGI...
fragments arrayed on the glass slides. Once again, this enzyme restricts bulk DNA into small fragments (<200 bp). Because its recognition site (TATAA) rarely occurs in GC-rich regions, most GC-rich CGI remain intact after this restriction. The cleaved ends of DNA are ligated to unphosphorylated linkers (19). The use of universal linkers, rather than specific primers, allows us to amplify all methylated fragments in patient tumor and normal amplicons that are flanked by an MstI recognition site. The ligated DNA is then restricted either singly or consecutively with up to two four-base methylation-sensitive endonucleases: Bsu3I, HpaII, and HhaI. Genomic fragments containing methylated sites are protected from the digestion and can be amplified by linker-PCR. Many more of these fragments will be present in the tumor compared with normal tissue as illustrated in our previous findings (19). On the other hand, unmethylated sites are cut and cannot be amplified and most linker-ligated MstI fragments in the normal controls will succumb to this fate.

The basis of the DMH assay builds on the use of methylation-sensitive endonucleases to discriminate between the methylation profile of a tumor sample from its paired normal control. For this scheme to work, we need to ascertain the presence of restriction enzyme recognition sites in the CGI clones. We randomly selected 50 CGI clones from those that are arrayed on the CGI microarray. Over 90% of these clones contain at least one of the three restriction sites. This information prompted us to reevaluate our original DMH strategy (19) whereby a subtractive hybridization step, which is both costly and laborious to perform, is implemented to reduce the presence of repetitive sequences in the prepared amplicons. We reasoned that the majority of the CpG sites present in these repetitive fragments would be unmethylated in the normal genome and possibly hypomethylated in the tumor genome. Therefore, being able to remove by enzymatic restriction close to 90% of the repetitive sequences would eliminate the necessity of having a subtractive hybridization step. We compared the hybridization intensities of both high- and low-copy number CGI loci prepared with and without this step and found them to be quite similar. As such, our revised DMH assay has two methylation-sensitive restriction steps (to increase coverage and to safeguard against possible incomplete digestion) in place of the subtractive hybridization step, followed by one methylation-sensitive restriction step. An added advantage of this modification is that it removes a tedious step from the DMH assay and makes the whole protocol amenable to a high-throughput sample preparation scheme.

The last step of the amplicon preparation is linker-PCR amplification. We chose to amplify the linker-ligated fragments for only 20 cycles. At this level of amplification, especially if most unmethylated repetitive sequences are removed by two methylation-sensitive endonucleases, the PCR products of most templates should be in the linear amplification range. At this stage of sample preparation, a portion of the amplicon is reserved for validating the hypermethylation cutoff using the Southern hybridization approach as described in a subsequent section. The remaining amplicon is purified and ready for indirect fluorescent labeling.

The aminoallyl labeling step, the fluorescent dye coupling step, and the labeled target purification step are essentially as described by DeRisi et al. The aminoallyl labeling step, the fluorescent dye coupling step, and the labeled target purification step are essentially as described by DeRisi et al. Due to the complexity of our amplicons and the size of the CGI, we found the combined tumor/normal control pair with 8 μg DNA, 180 pmol Cy5, and 150 pmol Cy3 would give strong hybridization signals. The hybridization of 7776 spots is carried out under a 24 × 50 mm cover glass sealed tightly within a moistened hybridization chamber in a 60°C water bath for up to 15 h. The posthybridization washing steps are essentially those described by DeRisi et al. A representative CGI microarray image from the breast tumor study is shown in Fig. 3. The scanned image is analyzed using the GenePix 3.0 software. The Cy5/Cy3 ratios (i.e., the hybridization intensity from the test amplicons to the hybridization intensity from the normal amplicons) from each image are normalized guided by both the average global Cy5/Cy3 ratio from each image and the Cy5/Cy3 ratios from 10 internal controls (clones without restriction cut sites and their copy numbers remain the same in tumor and normal controls).
Validation of DMH microarray data

The DMH microarray data are verified by standard molecular biology assays. The two assays routinely used in our laboratory for this purpose are the Southern analysis using either genomic DNA or DMH amplicons as the hybridization templates, and methylation-specific PCR (20). As indicated earlier, we usually reserve a portion of the DMH amplicons for the preparation of Southern filters. Fragment probes of CGI shown to be important by clustering algorithms are synthesized and used for Southern hybridization. This type of validation is important in that it affirm the cutoff ratio can accurately assign the methylation status of each studied locus. Using genomic DNA as the hybridization template in a Southern hybridization is a useful confirmation tool. Fragment probes similar to those mentioned previously are used for hybridization. The presence or absence of appropriate hybridization bands confirms the DMH microarray data. The disadvantage of this classical method is that it takes 7–10 µg DNA from each sample pair to prepare the Southern filters. This amount of DNA is not practical for small surgical samples. A more reasonable assay to verify the DMH data is by methylation-specific PCR. Here, only 1–2 µg DNA is needed for the initial bisulfide conversion. Primers representing the methylated and the unmethylated alleles are synthesized. PCR products from these primers are separated on an agarose gel. The presence or absence of bands in the methylated lane signify the methylation status of the interrogated CGI locus.

Fig. 4 depicts representative data from the breast tumor study. Methylated MseI fragments, mostly present in tumor samples, were protected from HhaI restriction [recognition sites: guanine/cytosine/guanine/cytosine four-base sequence (GGCG)] and were labeled with Cy5. Panel A shows the DMH image of CpG clone SC76F1 from eight patients, together with the corresponding adjusted ratios of normalized and adjusted Cy5/Cy3 ratios of eight breast tumor samples.

![Figure 3](image-url)  
**Figure 3** Representative results of differential methylation hybridization. To illustrate the usefulness of the DMH assay, breast tumor and normal amplicons were prepared as described in the text, fluorescently labeled with Cy5 and Cy3, respectively, and cohybridized to a CpG island microarray containing 7776 CpG island tags. The hybridization output is measured in intensities of the two fluorescence reporters; red signifying tumor specific, green signifying normal sample specific, and yellow signifying overlay and normal samples. The inset is an expanded view of the box area. Yellow spots indicate equal amounts of bound DNA from each amplicon, signifying no methylation differences between tumor and normal genomes. Spots hybridized predominantly with tumor amplicon, but not with normal amplicon, would appear red and are indicative of hypermethylated CpG island loci, present in the tumor genome. Green spots indicate a greater presence of the normal amplicon on the CpG island loci, depicting loss of methylation in the tumor genome. DMH, differential methylation hybridization; CpG, cytosine guanine dinucleotide.

normal samples). CGI tags whose signal intensities were slightly above the background or were devoid of hybridization signals represent the unmethylated loci in both tumor and normal samples; their genomic fragments were removed by enzymatic restriction using the methylation-sensitive endonucleases before linker-PCR. Yellow spots (normalized Cy5/Cy3 = 1) represent equal amounts of bound DNA from each amplicon, indicating no methylation differences between tumor and normal genomes. In some instances, these yellow spots represent CGI tags that do not contain the internal Hhail, HpaII, or BsrUI recognition sites and have equal copy numbers in both tumor and control DNA. CGI tags hybridized predominately with the tumor amplicon, but not with the normal amplicon, appear as red spots. A factor of 1.5 is then applied to the normalization factor to reduce the false identification rate. It should be noted that the magnitude of the Cy5/Cy3 ratio does not necessarily reflect the extent of hypermethylation, as the target preparation is PCR based. Less frequently, we also encountered green spots (normalized Cy5/Cy3 ≤0.5) by DMH, denoting the presence of hypomethylated sequences in the tumor genome. Sequence analysis indicated that most of the green spots in the microarray panel are repetitive elements, which are often methylated in nontumor cells. Single- or low-copy number CGI loci identified as hypermethylated or hypomethylated in each tumor/normal pair are flagged. Compilation of the hypermethylated loci across the studied samples are then analyzed by various clustering algorithms to reveal the outcome of CGI methylation on patient prognosis or tumor progression. We have successfully used a hierarchical clustering algorithm (http://rana.stanford.edu/clustering) and a nonhierarchical clustering algorithm (Partek Fuzzy C-means protocol) to analyze the methylation data set. Clustering results generated by these algorithms are further refined by a new approach called batching sequential forward selection (developed by Dr. Chi-Ren Shyu, a collaborator in the Computer Science Department, University of Missouri, Columbia, MO), which can better interpret high-dimensional microarray data and allow ease of data visualization.

**FIGURE 4** Methylation analysis of CpG island locus SC76F1. (A) Normalized and adjusted Cy5/Cy3 ratios of eight breast tumor samples. (B) The positions of the methylation-sensitive HhaI sites and the probe in the SC76F1 region are indicated. (C) Breast tumor DNA (10 µg) was treated with MseI and HhaI and subject to Southern analysis. C, control DNA digested with MseI only; N, normal DNA digested with MseI and HhaI. The presence of a 770-bp band indicates hypermethylated DNA fragments, the presence of a 660-bp band indicates partially methylated DNA fragments, whereas the presence of a 550-bp band indicates unmethylated DNA fragments. Cpg, cytosine guanine dinucleotide.
fluorescent intensities between the two dyes (i.e., Cy5/Cy3 ratios). The SC76F1 sequence matched a CGI region located within the promoter of the glypican 3 (GPC3) gene at chromosome Xq26 (21). This gene encodes a glypican integral membrane protein. We determined the methylation of the GPC3 promoter in eight patient tumor samples and one normal sample. The expected sizes of the unmethylated, partially methylated, and hypermethylated fragments are shown in panel B. The Southern analysis data using genomic DNA as templates, as shown in panel C, confirmed the DMH microarray data in this Cpg locus.

The CGI microarray and the DMH methylation assay are validated and in place for querying multiple CGI loci at a time. At the present configuration, due to the need of 1–2 μg DNA as starting material, cell culture studies, animal feeding trials, or profiling of tumors with a large sample size are good candidates for DMH microarray experiments. Once the protocol is set up in a laboratory, many samples can be processed and hybridized in a short time. The high-throughput nature of microarray studies permits researchers to adequately replicate their experiments, thereby increasing the confidence in the resultant methylation profiles. We are hopeful that this methylation analysis method will be implemented by researchers who are presently performing methyl acceptance assays whereby the ability of DNA to incorporate [3H]methyl groups from labeled SAM in an in vitro assay is measured (22,23). We feel the ability to examine the methylation status of multiple CGI loci would provide tremendous insight into how DNA methylation is influenced by dietary methyl donors.

LITERATURE CITED