illumina®

GoldenGate® Assay for Methylation and BeadArrayTM Technology

1. Introduction

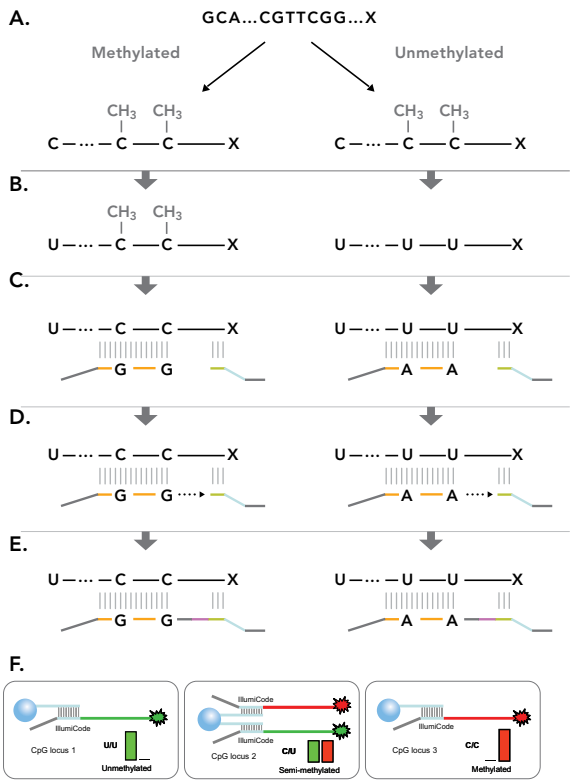
The GoldenGate Assay for Methylation allows measurement of the DNA methylation status of up to 1,536 targeted CpG sites in 96 samples simultaneously. A description of the GoldenGate Assay for Methylation will be provided in the following text, including: (1) assay design, (2) assay controls, (3) assay performance such as reproducibility and sensitivity, (4) examples of applications, (5) user interface, and (6) data analysis. Also described are how results obtained from the GoldenGate Assay for Methylation correlate to those obtained with other common methods of methylation detection such as methylation-specific PCR (MSP), bisulfite sequencing, pyrosequencing, and MethyLight.

2. Background

DNA methylation plays a critical role in the regulation of gene expression in development and differentiation, and diseases such as multiple sclerosis, diabetes, schizophrenia, aging, and cancer. The ability to access the epigenetic information for a large number of genes or the entire genome should greatly facilitate the understanding of the nature of gene regulation in cells, as well as the epigenomic mechanism of interactions between cells and the environment. This ability should also have significance for studies of human epigenetic disorders and assisted reproduction. Microarray-based DNA methylation profiling technologies have been developed to meet this goal. These methods can be categorized into three main classes of methylation status interrogation: (1) discrimination of bisulfite-induced C to T transition, (2) cleavage of genomic DNA by methylation-sensitive restriction enzymes, and (3) immunoprecipitation with methyl-binding protein or antibodies against methylated cytosines. Each of these methods has its own limitations. For example, methylation-sensitive restriction enzymes can not interrogate every CpG site, while the immunoprecipitation method can not provide methylation information at single-base resolution for any targeted sequence. For bisulfite-based approaches, the challenges lie in the reduced genome complexity after bisulfite conversion of the genomic DNA. Target-specific probe selection and hybridization specificity remain as the main technical hurdles.

The GoldenGate Assay, a high-throughput single nucleotide polymorphism (SNP) genotyping system1, was adapted for DNA methylation detection to identify bisulfite-converted genomic DNA2. The GoldenGate Assay for Methylation combines a microbead-based array platform, a high level of assay multiplexing, and scalable automation for sample handling and data processing. Unlike restriction enzyme-based methods, assay probes can be specifically designed for many of the CpG sites in the genome, and assay oligos can be designed to interrogate either the Watson or Crick strand at each CpG site. Different from other direct hybridization approaches, our method incorporates an allele-specific extension and oligonucleotide ligation step, which enables biochemical discrimination and assay specificity while allowing multiplexed profiling of CpG methylation status in several hundred genes. This technology has been successfully used to analyze methylation profiles of 1,536 CpG sites from 371 genes in cancer cell lines, lung cancers, and normal tissues, and to identify a panel of adenocarcinoma-specific methylation markers2. It has also been used to identify a unique epigenetic signature for human embryonic stem cells3. These proof-of-principle results effectively demonstrate the utility of the Illumina GoldenGate Methylation Platform for biomarker discovery and validation.

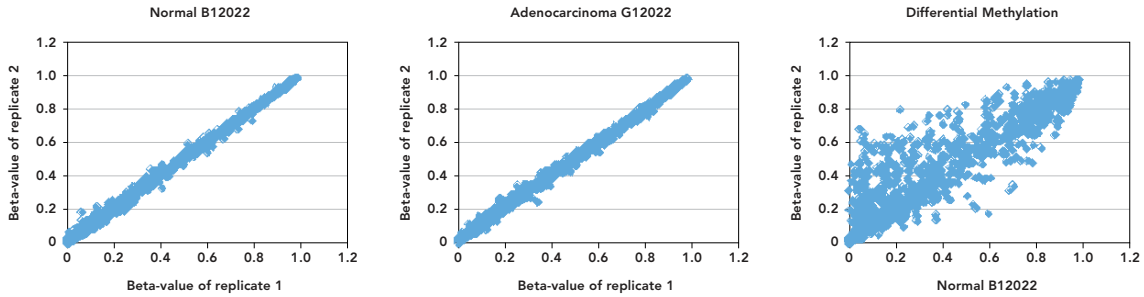
Figure 1: DNA Methylation Assay Scheme



(a) The top strand of the gDNA sequence is shown.*
(b) Only unmethylated cytosines are converted to uracils after bisulfite conversion.
(c) Two probe pairs are designed for each CpG site: an allele-specific oligo (ASO in orange) and locus-specific oligo (LSO in green) probe pair for the methylated state of the CpG site and a corresponding ASO-LSO pair for the unmethylated state. Pooled oligos are annealed to the target. (d) Extension occurs from the matched ASO toward the LSO. (e) Ligation (purple) of the extended ASOs creates PCR templates. The ligated products are then PCR amplified with fluorescently-labeled primers, and hybridized to an array. The IlluminaCode (blue) incorporated within the PCR product is hybridized to a complementary sequence on the array (f). The two fluors enable identification of methylated and unmethylated loci. CpG locus one, two, and three above show unmethylated, semi-methylated, and fully methylated sites, respectively.

*If other CpG sites are present in close proximity to the target CpG site, it is assumed that they have the same methylation status as the site of interest.

Figure 2: Methylation Assay Reproducibility And Differential Methylation Detection



Comparison of methylation profiles between lung cancer and matching normal tissue. The b value (i.e., the methylation ratio measured for all 1,536 CpG sites) obtained from one replicate experiment is plotted against that obtained from another technical replicate experiment. DNA samples derived from lung adenocarcinoma G12022 (center) and its matching normal tissue (left) demonstrate the reproducibility of replicated assays. The right panel shows the comparison between the normal and corresponding adenocarcinoma samples.

3.Technology

Implementation of the methylation profiling assay on the SNP genotyping platform

The GoldenGate Assay for SNP genotyping was adapted for DNA methylation detection. Non-methylated cytosines (C) are converted to uracils (U) when treated with bisulfite, while methylated cytosines remain unchanged. Because the hybridization properties of uracil are similar to that of thymine (T), the detection of the methylation status of a particular cytosine can be carried out following bisulfite treatment by using a genotyping assay for a C/T polymorphism.

Four probes are designed for each CpG site: two allele-specific oligos (ASO) and two locus-specific oligos (LSO). Each ASO-LSO oligo pair corresponds to either the methylated or unmethylated state of the CpG site (Figure 1). The gap between the ASO and LSO allows difficult sequences or ambiguous bases in CpG islands of interest to be avoided. This flexibility is particularly important for methylation studies because of a decrease in sequence complexity as a result of bisulfite treatment. If other CpG sites are present in close proximity to the target CpG site, we make the assumption that they have the same methylation status as the site of interest. This design hypothesis is based on previous reports and our own bisulfite sequencing results, in which a majority (> 90%) of the adjacent CpG sites were shown to be co-methylated or co-unmethylated[4,5]. Therefore, the GoldenGate Assay for Methylation provides not only a methylation measurement at the targeted CpG site, but also “inferred” methylation information at the adjacent underlying CpG sites.

This design strategy is used in methylation-specific PCR primer design[6] and other microarray-based DNA methylation analysis[7]. While many CpG sites exist within each CpG island, we only select those for which robust assays can be designed, i.e., assays with a design score above a predetermined threshold.

The EZ DNA methylation kit (Zymo Research, Orange, CA) is used for bisulfite conversion of all DNA samples, according to the manufacturer’s recommendations. For each sample, 500 ng to 2 µg of genomic DNA can be used for each bisulfite conversion. An aliquot of the converted DNA (corresponding to 250 ng starting gDNA) is then used to assay up to 1,536 CpG sites simultaneously on an array. The subsequent assay procedures are similar to those described previously for standard SNP genotyping1. For more details, please see the GoldenGate Assay for Methylation System User Guide (Illumina Part No. #11228975).

The methylation status of an interrogated CpG site is determined by calculating the intensity of the methylated (M) and unmethylated (U) alleles. Background intensity computed from a set of negative controls is subtracted from each analytical data point. The ratio of fluorescent signals is then computed from the two alleles, where methylation value (b) = Max(M, 0)/[Max(M, 0) + U(M, 0) + 100].

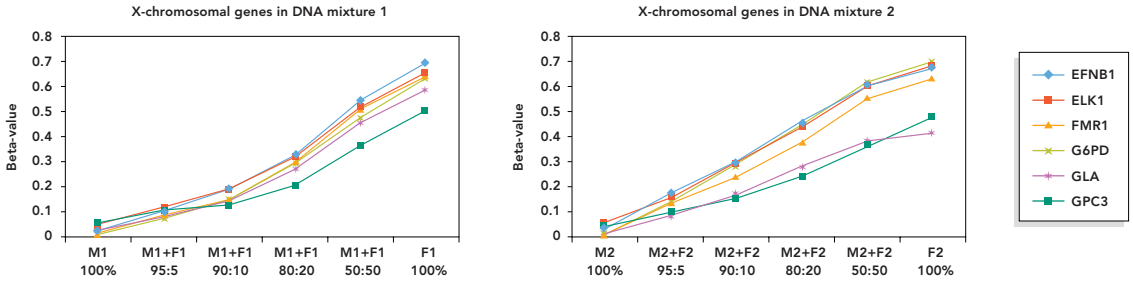
b provides a continuous measure of levels of DNA methylation in samples, ranging from 0 in the case of completely unmethylated sites to 1 in completely methylated sites. An absolute value is used in the denominator of the formula, as a compensation for any negative values which may arise from global background subtraction (i.e., over-subtraction). A constant bias of 100 is added to the denominator to regularize b when both U and M values are small.

It is important to note that b only provides a relative but not an absolute methylation level measurement in a sample. For example, for any given CpG site, the two ASO-LSO oligo pairs may not hybridize to their corresponding target sequences at exactly the same efficiency; therefore, when b = 0.5, it may not mean that methylation level is 50%. Furthermore, b may not be absolutely zero at some assayed CpG sites even when they are not methylated, presumably because of some degree of cross-hybridization of the assay oligos to other non-specific genomic regions. The lack of absolute quantification of this assay, however, does not undermine its capability for differential methylation detection.

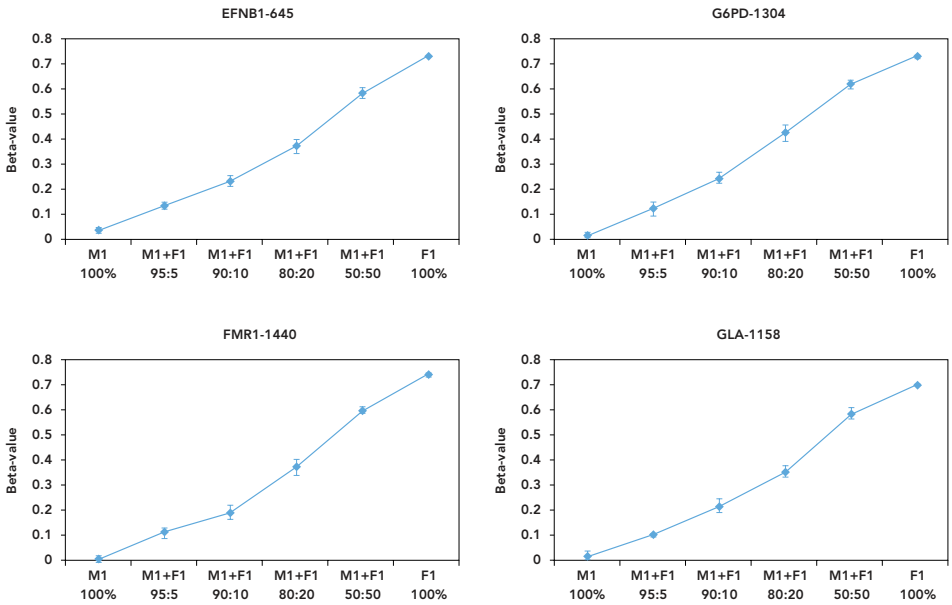
Methylation Assay Reproducibility and Differential Methylation Detection

We obtained highly reproducible DNA methylation profiles between technical replicates (Figure 2), with an average r2 of 0.98 ± 0.02 when the b values were compared. We observed that the standard devia-

Figure 3: Methylation Detection In GDNA Mixtures



A. Female genomic DNA was diluted with male genomic DNA at ratios of 5:95, 10:90, 20:80 and 50:50. Two sets of mixtures were made and measured: M1 (male NA10923)/F1 (female NA10924) and M2 (male NA07033)/F2 (female NA06999). Methylation levels of six X-chromosome linked genes were calculated as the average of several (1-5) CpG sites for each gene.



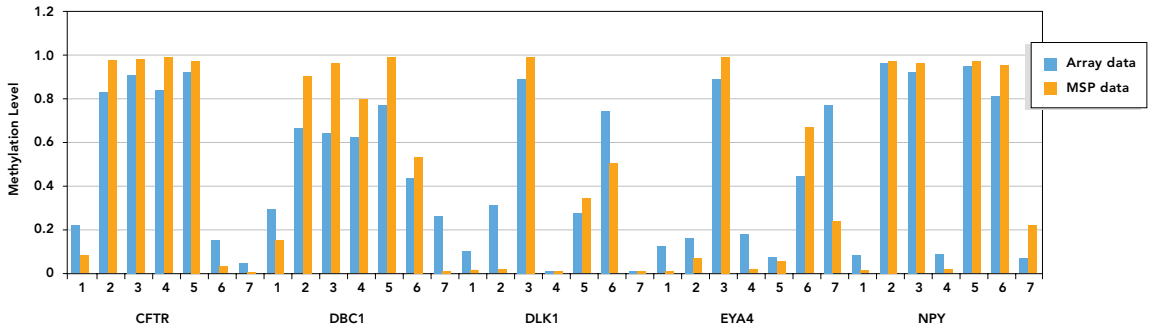
B. Methylation profiles of individual CpG sites (only four are shown). Error bars represent the standard deviation of b values calculated from four replicate experiments performed with the first set of mixtures (NA10923/NA10924).

tion of the b value obtained for all the 1,536 CpG sites across four replicates was less than 0.06 in 99% of cases. Therefore, we estimate that our method can discriminate levels of methylation (b) that differ by as little as 0.17 (1.96 x $\sqrt{2}$ x 0.06). The Illumina BeadStudio software can compare the b values between samples or sample groups and perform a differential methylation analysis. The difference in methylation levels between samples in a pair is quantified by a "Diff Score;" a Diff Score equal to 20, 30, or 40 corresponds to a significance p value equal to 0.01, 0.001, or 0.0001, respectively. To estimate the false positive rate for differential methylation detection, we measure methylation levels at 1,536 CpG sites for the same sample in replicate experiments, and usually less than 1% of CpG sites show up with a Diff Score ≥ |20|.

Sensitivity of the Methylation Assay

To assess the assay sensitivity, we measured the methylation status of six X-linked housekeeping genes–EFNB1, ELK1, FMR1, G6PD, GPC3, and GLA–together with several hundred genes in male and female genomic DNA samples. Female genomic DNA was diluted into male genomic DNA at ratios of 5:95, 10:90, 20:80 and 50:50, prior to bisulfite conversion. Methylation levels from 5:95 and 0:100 mixtures could be unambiguously distinguished from each other (i.e., the maximum b value in the 0:100 mixtures was less than the minimum b value in the 5:95 mixtures; Figure 3A, 3B). This indicates that the GoldenGate Assay for Methylation can detect as little as 2.5% methylation for well-performing CpG sites. On average, the assay can detect change in b of ≥ 0.2 with 95% confidence, which is consistent with our previous estimation.

Figure 4: Comparison Of Methylation-Specific PCR And Array-Based Methylation Data



MSP was used to confirm the methylation status of CpG sites within the promoter regions of five genes that showed distinct methylation profiles in one normal lung tissue and six lung cancer cell lines from array-based methylation analysis (1: normal lung tissue, 2: NCI-H69, 3: NCI-H526, 4: NCI-H358, 5: NCI-H1299, 6: NCI-H1395, and 7: NCI-H2126). Methylation level is represented as bars, blue for microarray data and orange for MSP data.

GoldenGate Methylation Data Validation

Methylation-specific PCR (MSP) has been widely used to monitor the methylation status of individual genes6,8. MSP was used to confirm the methylation measurements generated by our microarray analysis. MSP primers specific to either methylated or unmethylated DNA were designed to target corresponding CpG sites within the promoter regions of CFTR, DBC1, DLK1, EYA4, and NPY genes. Bisulfite-treated genomic DNA samples derived from normal lung tissue and cancer cell lines were analyzed using real-time MSP. Of the 35 MSP data points, 34 were highly concordant with the methylation status determined by the microarray analysis (Figure 4), with a Spearman correlation coefficient, r = 0.891.

In addition, eight CpG sites were validated, which showed elevated methylation in our adenocarcinoma samples using bisulfite sequencing[9]. DNA samples from two normal and four adenocarcinoma samples were treated with bisulfite and regions of interest were amplified by PCR. PCR fragments were cloned and individual colonies were picked for sequencing. Twelve cloned fragments were sequenced for each CpG site in each selected sample. In all cases, an increase in methylation in cancer samples compared to normal samples was observed. Though the absolute levels of methylation detected by the two different methods were somewhat different, a strong correlation was obtained between these two data sets, with a Spearman correlation coefficient (*r*) equal to 0.70[1].

These results confirm the overall validity of our method, they demonstrate that the GoldenGate Assay for Methylation can reliably detect methylation differences in clinical samples for more than 1,500 CpG sites and that the assay may be used for both marker discovery and validation.

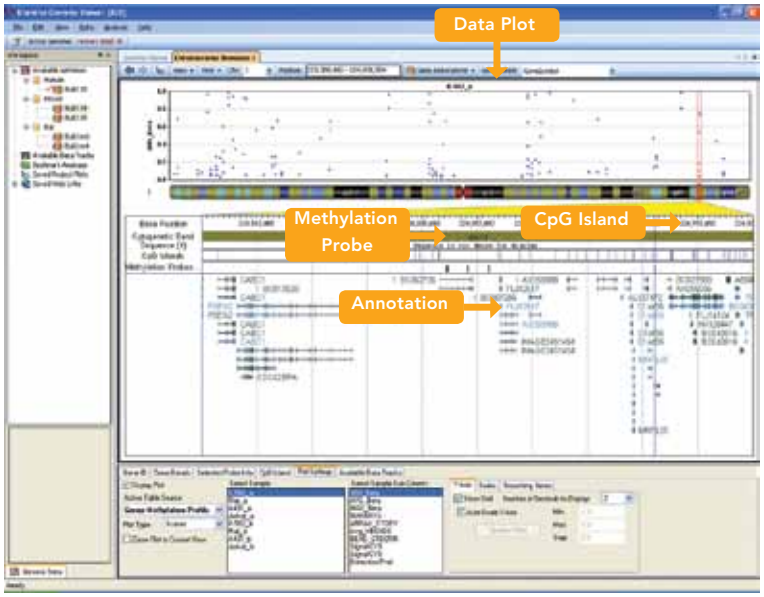
Internal Controls for the GoldenGate Assay for Methylation

Plasmids pUC19, pACYC184, and phage fX174 were used as internal control DNA samples. These DNA samples (unmethylated, in vitro methylated or mixed at a 1:1 ratio) were spiked into human genomic DNA at a 1:1 molar ratio (at approximately 2–4 pg plasmid DNA / 1 mg gDNA), and were used in every experiment to monitor both bisulfite conversion efficiency and accuracy of methylation detection.

Currently, the GoldenGate Assay for Methylation has nine controls which help to monitor each step of the assay process, from bisulfite conversion to the analytical hybridization of the assay onto the universal array.

1. **Bisulfite conversion controls test for the presence of unconverted genomic DNA in assay samples. Primers are designed to the same DNA locus, with one pair targeting converted, and the other pair targeting unconverted DNA sequence.**
2. **The first hybridization controls test the annealing specificity of ASOs. These control ASOs have different melting temperatures for the same DNA locus.**
3. **The allele-specific extension controls test the extension efficiency of properly matched versus mismatched allele-specific oligos (ASOs).**
4. **The extension gap control tests the efficiency of extending 15 bases from the 3' end of the allele-specific oligo to the 5' end of the locus-specific oligo.**
5. **The second hybridization controls test the hybridization of single-stranded assay products to IllumiCode Sequences on the array beads.**
6. **Negative controls define the methylation assay background. LSO probes can hybridize to bisulfite-converted DNA. ASO sequences are randomly permutated and should not hybridize to the DNA template. As a result, an amplifiable target should not be formed, and the signal from these IllumiCodes should be low.**
7. **Assay intensity shows average signal intensity for all assay probes on the array.**
8. **Gender-specific methylation controls are designed against X-linked genes. Cy3 and Cy5 signal should be detected in females, and only Cy3 signal should be detected in males.**
9. **The PCR contamination detection controls are divided into**

Figure 5: Illumina Chromosome Browser (ICB)



Use the ICB to view genomic coordinates of probes (SNP, GX, DASL, or methylation) along with data plot, gene, sequence, and CpG island information.

four types. Only one type is added to each oligo pool for assay for methylation (OMA) tube. When a single OMA is run, only one contamination control type should have high signal. If two or more contamination control types have high signal, the significant contamination may have occurred.

For detailed information about the assay controls, please refer to the GoldenGate Assay for Methylation System Manual (Illumina Part No. #11228975).

BeadStudio: Visual Data Display and Data Analysis

Heat Map, Differential Methylation Analysis, and Correlation between Methylation and Gene Expression

The BeadStudio Methylation Module has numerous features that enable researchers to view vast amounts of data in a single graph. These tools include the Illumina Genome Viewer (IGV), the Illumina Chromosome Browser (ICB), heat map, scatter plot, and line plot. The ICB can display information such as chromosomal coordinates, percent of GC content, CpG islands, β values, and lists which methylation probes are implicated in a region of interest (Figure 5).

Figure 6 shows the utility of the BeadStudio's heat map capability. The heat map displays methylation profiles of bisulfite-converted gDNA from 288 samples, including normal tissue samples and both cancerous and normal cell lines arrayed in duplicate to assess the variation within the 1,505 CpG sites. The methylation patterns of the normal samples versus cancer samples are readily distinguishable.

The ability to directly correlate gene expression with methylation profiles using the BeadStudio Methylation Module adds another dimension to understanding how methylation impacts gene expression. Data from Illumina Expression BeadChips or the DASL® Assay can be imported into the BeadStudio Methylation Module so that expression levels of genes present in both data sets can be compared and displayed.

Standard Cancer Methylation Panel

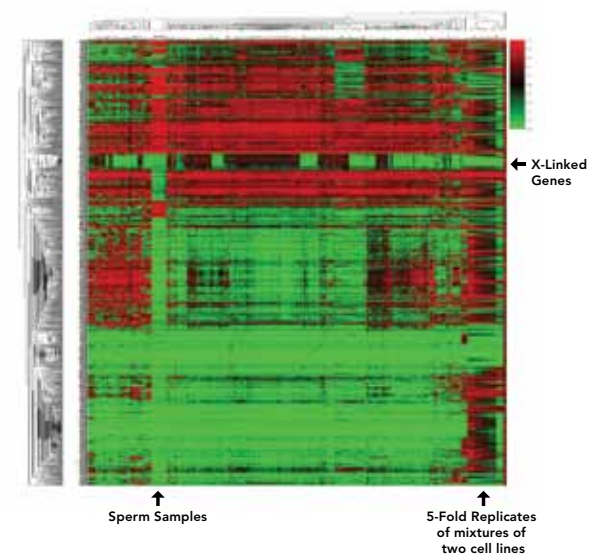
The standard Methylation Cancer Panel I interrogates 1,505 CpG sites, selected from 807 genes, where 231 genes contain one CpG site per gene, 463 genes contain two CpG sites and 114 genes have three or more CpG sites. These genes (download from www.illumina.com) were selected based on their biological relevance. They include tumor suppressor genes and oncogenes; genes that are involved in cancer development, for example: DNA repair genes; metastasis-inhibitor genes; genes regulated by various signaling pathways and/or responsible for altered cell growth, differentiation, and apoptosis; genes considered to be targets for oncogenic transformation; x-linked genes and imprinted genes; and previously reported differentially-methylated genes. All of these genes are represented in our Human Whole-Genome Expression BeadChips.

A 2 kb sequence—1.5 kb upstream and 500 bp downstream from the transcription starting site (TSS)—was extracted for each target gene based on human genome RefSeq build 36. Initially, primers for over 1,000 genes at 3 CpG loci per gene, where possible, were designed and tested. Assay probes were selected with the best design characteristics, the best performance in experimental testing using methylated and unmethylated reference samples, and a comprehensive panel of 46 normal and cancer tissues and cell lines covering a broad range of tissues of origin such as breast, colon, prostate, ovary, liver, lung, blood, and others. Probes demonstrating differential methylation between samples with Db of at least 0.2 between methylated and un-

methylated DNA reference samples were selected for the final assay.

The unmethylated reference samples were generated by genome-wide amplification of human genomic DNA. After amplification, DNA methylation was diluted > 1000-fold, effectively rendering the amplified genomic DNA unmethylated. The methylated templates were generated by in vitro methylation using a CpG-methylase, SssI.

Figure 6: Cluster Analysis Of DNA Methylation Data For 1,505 CPG LOCI On 288 DNA Samples



Two-dimensional cluster analysis of DNA methylation was performed using data obtained for 1,505 different reactions on 288 human DNA samples using the Illumina platform, representing 433,440 quantitative DNA methylation measurements. High levels of DNA methylation are indicated in red, while low levels are indicated in green. Samples included primary tumor DNA from breast, colorectal, lung, pancreas, and ovary cancers; histologically normal breast, lung, and colorectal tissues; colorectal and ovary cancer cell lines, sperm DNA samples, and various controls and replicate samples, including two cell lines mixed in five fixed ratios (0%, 25%, 50%, 75%, and 100%) and repeated a total of five times. These samples are distinctly visible on the right side of the heat map. Sperm samples also display a characteristic methylation profile, and are recognizable as a distinct cluster. X-linked genes clearly identify female samples.

4. Conclusion

The GoldenGate Assay for Methylation is a highly reproducible and multiplexed method for the high-throughput quantitative measurement of DNA methylation. This method provides not only a discrete measure of positive versus negative DNA methylation, but also a continuous measure of levels of DNA methylation. For a 17% difference in absolute methylation level (e.g., 10% vs. 27%), signals are expected to have largely non-overlapping distributions. Importantly, the assay can detect as little as 2.5% methylation status for some CpG sites. Unlike restriction enzyme-based methods, assay probes can be designed specifically for many of the CpG sites in the genome, and assay oligos can be designed to interrogate the Watson strand, the Crick strand, or both strands at each CpG site. Assay end products are processed using a Sentrix® Universal Array Matrix (SAM). As a result, gene or CpG sets can be refined iteratively, because no custom pre-defined arrays need to be developed. This method can detect changes in methylation status at up to 1,536 different CpG sites simultaneously, using only 500 ng of genomic DNA for duplicate assays. Coupled with integrated genomic analysis tools, the Illumina GoldenGate Methylation platform will provide powerful insight into epigenetic mechanisms of gene regulation.

References

1. Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, et al. (2003) Highly parallel SNP genotyping. Cold Spring Harbor Symposia on Quantitative Biology 68: 69-78.
2. Bibikova M, Lin Z, Zhou L, Chudin E, Wickham Garcia E, et al. (2006) High-throughput DNA methylation profiling using universal bead arrays. Genome Res 16(3): 383-93.
3. Bibikova M, Chudin, E, Wu B, Zhou L, Wickham Garcia E, et al. (2006) Human embryonic stem cells have a unique epigenetic signature. Genome Res 16(9): 1075-83.
4. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, et al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. Nat Genet 38(12): 1378-85.
5. Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, et al. (2004) DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. PLoS Biol 2(12): e405.
6. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB (1996) Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proc Natl Acad Sci U S A 93(18): 9821-6.
7. Adorjan P, Distler J, Lipscher E, Model F, Muller J, et al. (2002) Tumour class prediction and discovery by microarray-based DNA methylation analysis. Nucleic Acids Res 30(5): e21.
8. Eads CA, Danenberg KD, Kawakami K, Saltz LB, Blake C, et al. (2000) MethyLight: a high-throughput assay to measure DNA methylation. Nucleic Acids Res 28(8): E32.
9. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, et al. (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc Natl Acad Sci U S A 89(5): 1827-31.

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

illumina®