APPLICATION NOTE

High-Throughput DNA Methylation Analysis on the Illumina GoldenGate® Platform

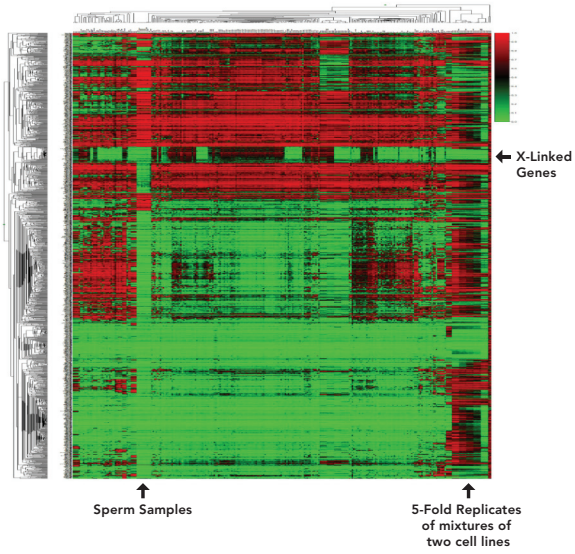
Contributed by Dr. Peter W. Laird, Associate Professor, University of Southern California

INTRODUCTION

Interest in epigenetic analysis has grown dramatically in recent years. As a consequence, a large number of DNA methylation and chromatin analysis technologies have been developed to address the needs of assay accuracy, ease of use, feature dimensionality, and sample throughput[1]. Most microarray-based DNA methylation analysis platforms rely on methylation-sensitive restriction enzyme digestion or on enrichment using methylcytosine antibodies or methyl-binding domain protein columns. These analysis methods can assay a relatively large number of features (loci) when combined with high-density microarrays, but are limited in their sample throughput capacity, and generally require large amounts of high-quality input DNA. Most other DNA methylation analysis techniques use sodium bisulfite-based conversion of unmethylated cytosines into uracils to turn epigenetic differences into sequence-based information. Many bisulfite-based techniques can accommodate large numbers of samples, but are limited in the number of loci that can be interrogated, since the increased sequence redundancy resulting from bisulfite conversion requires PCR amplification of individual loci to obtain reliable data.

So, in summary, most microarray-based methods are limited in sample throughput, while most bisulfite-based methods are limited in numbers of loci analyzed. By relying on the collective address-tagged amplification of 1,536 different CpG sites and hybridization on a BeadArray™ substrate, the bisulfite-based GoldenGate DNA methylation analysis platform recently introduced by Illumina retains the high sample throughput provided by bisulfite-based techniques, but greatly expands the number of loci that can be interrogated simultaneously. We describe here our first experiences with this platform.

FIGURE 1: CLUSTER ANALYSIS OF DNA METHYLATION DATA FOR 1,505 CPG LOCI ON 288 DNA SAMPLES

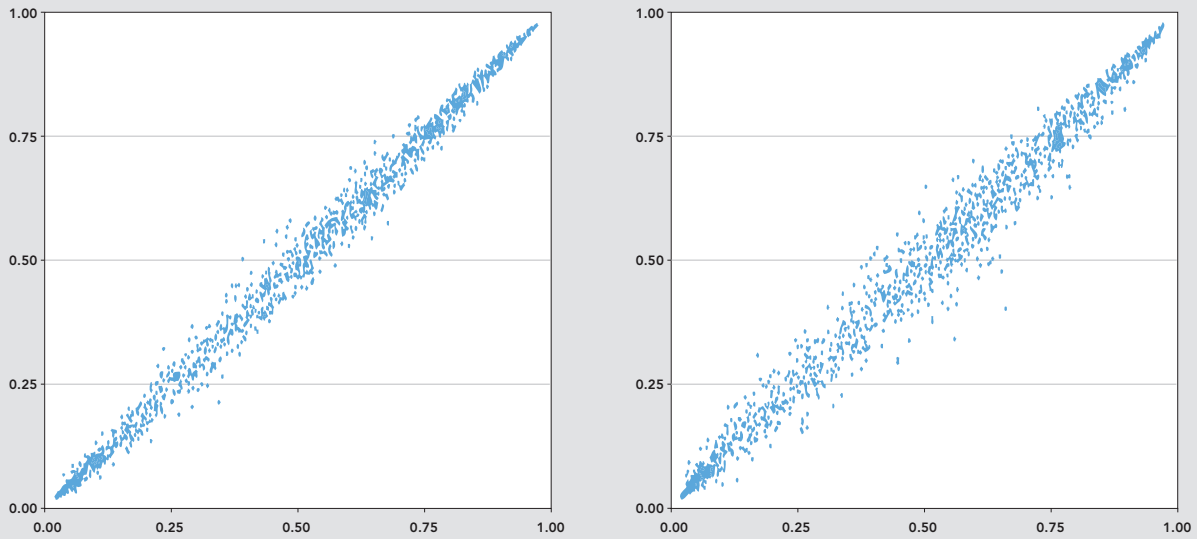


Two-dimensional cluster analysis of DNA methylation data obtained for 1,505 different reactions on 288 human DNA samples using the Illumina platform, representing 433,440 quantitative DNA methylation measurements. High levels of DNA methylation are indicated in red, while low levels are indicated in 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.

ASSAY DESCRIPTION

Each Sentrix® Universal Array Matrix (SAM) used to process the Illumina GoldenGate Assay for methylation can accommodate 96 different samples. The SAM yields up to 1,536 DNA methylation measurements per sample for a total capacity of 147,456 quantitative DNA methylation measurements per run. The platform relies on bisulfite-based conversion of unmethylated cytosines to uracil, with methylated cytosines refractory to this treatment. This creates a methylation-dependent sequence variation analogous to a SNP that can be interrogated using the Illumina BeadArray-based GoldenGate platform[2]. Extension of locus-specific oligonucleotide

FIGURE 2: REPRODUCIBILITY OF THE ILLUMINA DNA METHYLATION ANALYSIS PLATFORM

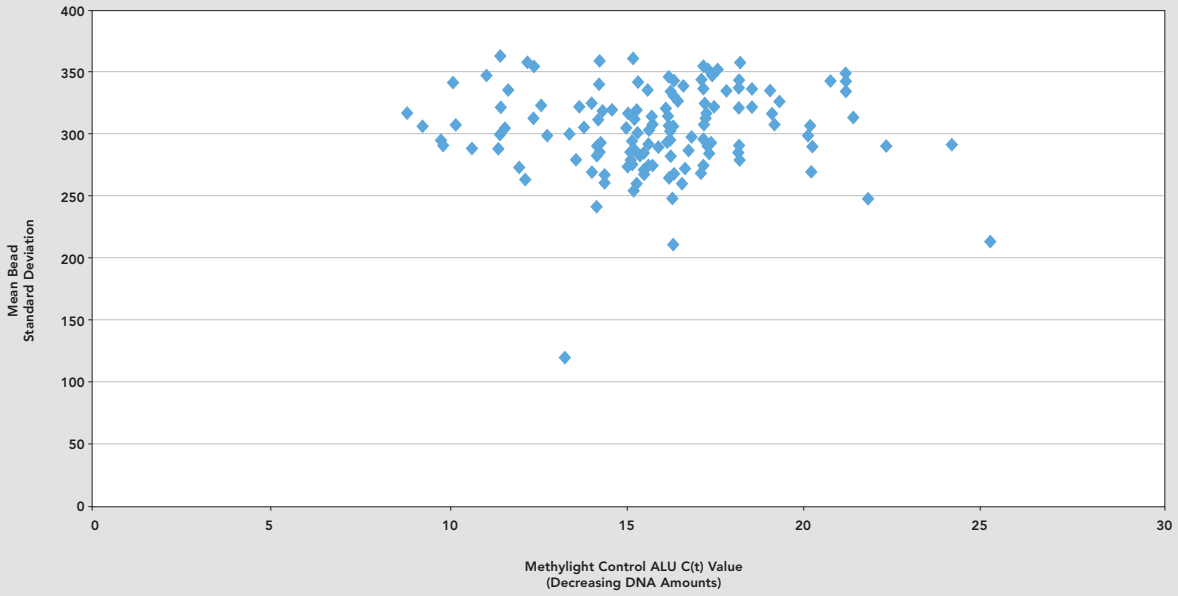


A representative result is shown for a replicate sample processed on multiple occasions. The left panel shows the correlation between beta value measurements for 1,505 CpG loci on duplicate samples processed for bisulfite conversion in the same batch. The panel on the right shows the correlation between beta value measurements for 1,505 CpG loci on duplicate samples processed for bisulfite conversion at different times in different laboratories by different laboratory personnel.

pairs complementary to this C/U site, reflecting methylated-specific or unmethylated-specific states, and ligation to downstream address-tagged locus-specific oligos, results in tagged locus-specific, methylated-specific, and unmethylated-specific templates. The primer sets for all 1,536 sites contain sequences at the 5' end of each primer that are shared among all loci with a common 5' sequence for all methylated-specific oligos, a different shared sequence for all unmethylated-specific oligos, and a third shared sequence for all address-tagged oligos. This allows for the simultaneous annealing and extension of all 1,536 loci, followed by a pooled PCR amplification with just three different shared oligos, of which two are differentially labeled with Cy5 and Cy3, reflecting the methylated and unmethylated states of each locus, respectively. The locus-specific addresses are resolved by hybridization of the PCR product to a SAM, in which each address is represented by an average of 30 different beads [3,4], allowing the relative Cy3 and Cy5 fluorescence for each bead to be determined. After hybridization and scanning of the PCR products to the bead array, Illumina BeadStudio software is used for data analysis.

We used a second generation oligonucleotide set, which differs from that described in previous publications [5,6]. This oligonucleotide set includes 31 CpG sites that did not pass final quality control and were omitted from further data analysis, leaving 1,505 different CpG sites distributed across approximately 800 different genomic loci (approximately two-fold redundancy). These 800 genomic loci have been selected by Illumina to represent loci of high epigenetic interest, including many loci known to be cancer-specifically methylated, many imprinted genes, and several X-linked genes. The selection of these loci was heavily influenced by the use of human cell lines, and may contain some loci of less relevance to the analysis of primary human tissue samples. One of the attractive features of the Illumina platform is the flexibility in assay content. As more knowledge of biological relevance of different methylation markers becomes available, we anticipate that the oligonucleotides sets will continue to evolve, improve, and ultimately perhaps specialize towards specific applications.

FIGURE 3: INFLUENCE OF DNA QUANTITY ON MEASUREMENT ERROR



The measurement error was estimated for each sample by calculating the mean of the standard deviation of the beta values (approximately 30 measurements) for each of the 1,505 CpG sites for each sample using BeadStudio software. This mean standard deviation was plotted against the MethyLight C(t) value obtained for a methylation-independent ALU control reaction 7, which is inversely related to the log of DNA quantity. No increase in mean standard deviation is observed with increasing ALU C(t) value (decreasing DNA quantity).

RESULTS AND DATA ANALYSIS

Samples Analyzed

We conducted three plate runs on the Illumina DNA methylation platform for a total of 288 DNA samples. The analysis included 48 control samples provided and processed by Illumina, nine control samples provided by Illumina but processed in our laboratory, two SssI-treated human genomic DNA controls, 120 primary human tumor samples (including colon, breast, pancreas, ovary, and lung), 50 histologically normal human tissue samples (including colon, breast, lung, buffy coat, and sperm), and 59 human cancer cell lines (including colon and ovary).

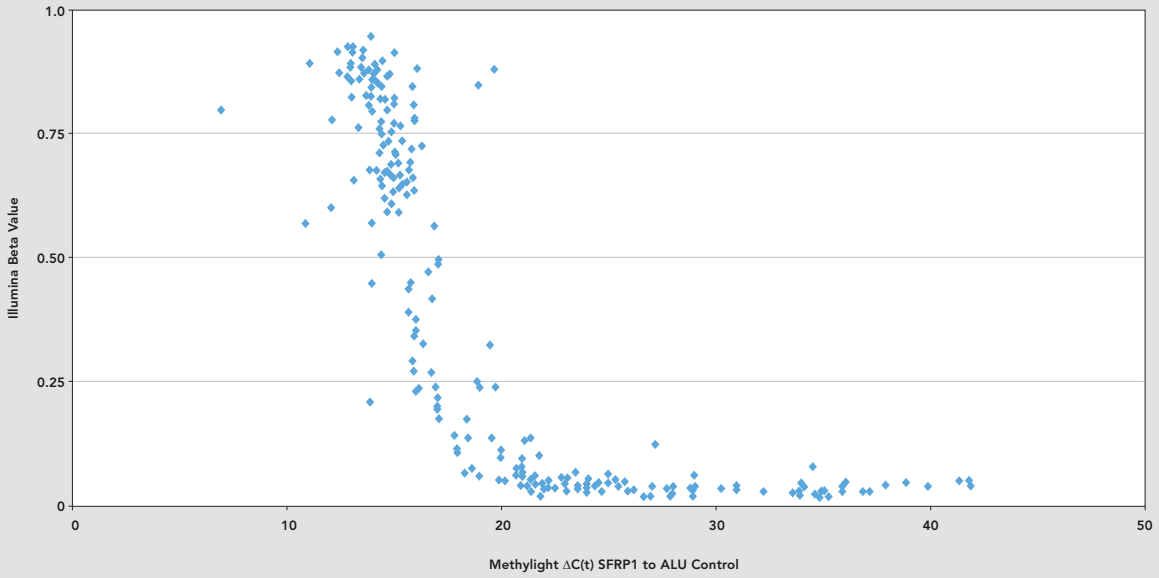
BeadStudio Software

We used the BeadStudio Methylation software to analyze the data. BeadStudio can assimilate data from multiple 96-well plates to generate heat maps and 1-D and 2-D clusters to determine methylation similarities between DNA samples. BeadStudio also allows for customizable data extraction, including average methylation (beta) values derived from approximately 30 replicate methylation measurements (beads) for each CpG site, Cy3/Cy5 ratios, as well as statistical information and CpG methylation genomic coordinates for all 1,536 CpG sites (in our case 1,505) analyzed in the platform. Histograms, line/scatter plots, bar plots of methylation beta values, Cy3/Cy5 ratios, and statistical analyses can be generated on a filtered subset of samples or on the entire data set.

Data Analysis

A heat map and dendrograms of unsupervised two-dimensional cluster analysis of all 288 samples are shown in Figure 1. It is apparent that the 1,505 CpG sites selected by Illumina contain many informative loci that display methylation heterogeneity across different sample types. X-linked genes stand out as methylated in female samples, and not in male samples. Sperm samples also stand out with heavy methylation at some loci and very low methylation at other loci. Two cell lines with highly divergent DNA methylation profiles were mixed at five different ratios (0%, 25%, 50%, 75%, and 100%). Each of these mixed samples was analyzed five

FIGURE 4: VALIDATION OF SFRP1 MEASUREMENTS BY METHYLIGHT ANALYSIS



A comparison is shown of Illumina beta values and MethyLight results obtained for the SRRP1 genes on 144 different samples. Illumina beta values represent single measurements for Target ID SFRP1_P157_F (Probe ID: 1946). MethyLight data represent the difference between the C(t) values (ΔC(t)) obtained for the SFRP1-M1 reaction (Reaction ID: HB-201) and the ALU-C4 reaction (Reaction ID: HB-313). A lower ΔC(t) indicates a higher amount of DNA methylation. The distance between the centers of the Illumina and MethyLight reactions is 53 bp.

times. The resulting 25 analyses cluster at the far right of the heat map. The concordance of the replicate measurements is apparent from the broad bands representing each five-fold replicate cluster of each of the five mixed ratios.

Reproducibility

To further explore the reproducibility of the Illumina platform, we calculated the mean correlation coefficient for all 1,505 CpG sites of the 24 replicates that were processed for bisulfite conversion at Illumina and run on two separate plates at our facility. The r^2 for this correlation was 0.990. The r^2 for replicates processed separately by Illumina and our laboratory was 0.988. *Figure 2* shows a representative example of the correlation for a sample processed in duplicate by Illumina *(left panel)* and the same sample processed separately by Illumina and in our laboratory *(right panel)*.

DNA Quantity

Our goal was to use approximately 1 µg of genomic DNA for each sample, although amounts varied since the samples were obtained from many sources and were quantitated on different occasions. We were interested to see whether samples with lower DNA quantities showed increased noise in the methylation measurements. We assessed this noise by obtaining the standard deviation of the beta values (approximately 30) for each of the 1,505 CpG sites using BeadStudio software and then calculating the mean of all 1,505 standard deviations for each sample. Estimates of DNA quantities of bisulfite-converted DNA were obtained by using a bisulfite-dependent, methylation-independent MethyLight control reaction directed against Alu repetitive elements[7]. With the ALU control reaction, we analyzed 215 of the bisulfite-converted samples and plotted the resulting C(t) value against the mean Illumina standard deviation for each sample *(Figure 3)*. There is no clear relationship between DNA quantity and observed measurement noise, suggesting that the Illumina assay may work well for the DNA quantities below 1 µg per sample.

Data Validation

We used MethyLight analysis[8] to validate the Illumina DNA methylation results. Quantitative comparison between different DNA methylation analysis technologies is often hampered by platform discrepancies. For example, Illumina amplicons generally cover a few CpG dinucleotides, while MethyLight reactions typically span more CpGs (average of 8 CpGs), since they include a methylation-specific probe in addition to the methylation-specific PCR primers. The increased specificity of MethyLight reactions reduces background and allows resolution of differences among samples with very low levels of DNA methylation. The lower number of CpG sites covered by Illumina amplicons leads to the analysis of a higher proportion of DNA molecules in the sample, and greater accuracy in resolving differences among samples that have substantial amounts of DNA methylation. This difference in the range of methylation differences resolved by the two platforms is nicely illustrated for the SFRP1 locus in *Figure 4*. Samples with substantial DNA methylation (on the left side, with low C(t) for the SFRP1-M1 reaction versus the ALU-C4 control reaction) are very well resolved by Illumina, but relatively poorly by MethyLight. Samples with low levels of DNA methylation (on the right with high MethyLight ΔC(t)) are resolved well by MethyLight, but not by Illumina. Regardless of the differing sensitivities of the two assay platforms, it is clear that a strong correlation exists between the two types of measurements.

SUMMARY

We conducted a test of the Illumina GoldenGate DNA methylation analysis platform, and found the assay to be a major advance in high-throughput DNA methylation analysis. Key features of the assay include high sample throughput capacity (batches of 96 samples), large number of informative assayed loci (up to 1,536 per assay), high reproducibility, modest DNA quantity and quality demands, good resolution of methylation differences among samples with high levels of DNA methylation, excellent technical and software support, and the promise of future content optimization and customization.

REFERENCES

(1) Laird PW (2003) The power and the promise of DNA methylation markers. Nat Rev Cancer 3: 253-66.

(2) Bickmore W, Christie S, van Heyningen V, Hastie ND, Porteous DJ (1988) Hitch-hiking from HRAS1 to the WAGR locus with CMGT markers. Nucleic Acids Res 16: 51-60.

(3) Fan JB, Chee MS, Gunderson KL (2006) Highly parallel genomic assays. Nat Rev Genet 7: 632-44.

(4) Fan JB, Gunderson K, Bibikova M, Yeakley JM, Chen J, et al. (2006) llumina universal bead arrays. Meth Enzymol 410: 57-73.

(5) Bibikova M, Lin Z, Zhou L, Chudin E, Wickman Garcia E, et al. (2006) High-throughput DNA methylation profiling using universal bead arrays. Genome Res 16: 383-393.

(6) Bibikova M, Chudin E, Wu B, Zhou L,Wickham Garcia E (2006) Human embryonic stem cells have a unique epigenetic signature. Genome Res 16: 1075-1083.

(7) Weisenberger DJ, Campan M, Long TI, Kim M, Woods C, et al. (2005) Analysis of repetitive element DNA methylation by MethyLight. Nucleic Acids Res 33: 6823-6836.

(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: E32.

ADDITIONAL INFORMATION

Visit our website or contact us at the address below to learn more about Illumina's GoldenGate Methylation Solution.

Illumina, Inc.
Customer Solutions
9885 Towne Centre Drive
San Diego, CA 92121-1975
1.800.809.4566 (toll free)
1.858.202.4566 (outside the U.S.)
techsupport@illumina.com
www.illumina.com

FOR RESEARCH ONLY

Pub. No. 270-2007-001 29Jan07