# HumanMethylation450 BeadChip Achieves Breadth of Coverage Using Two Infinium® Chemistries

## Introduction

The HumanMethylation450 BeadChip leverages the Illumina Infinium assay, the industry's most trusted, proven DNA analysis platform, offering a unique combination of comprehensive, expert-selected coverage and high-throughput compatible with large sample size, epigenome-wide association studies. By combining Infinium I and Infinium II assay chemistry technologies, the BeadChip provides coverage of 99% of RefSeq genes, 96% of CpG islands, and all of the content categories requested by an expert consortium. In addition to offering background on this unique combination of technologies, this document provides guidance and information on resources available to assist in the analysis of data generated by the BeadChip.

### Two Infinium Chemistries Enhance Breadth of Coverage

The Infinium methylation assay uses beads displaying long, target-specific probes designed to interrogate individual CpG sites within a given DNA sample. DNA methylation is measured using quantitative "genotyping" of bisulfite-converted genomic DNA. Infinium I and II assays offer complementary strengths that benefit the array's breadth of coverage.

#### Infinium I Assay

The Infinium I assay employs two probes per CpG locus: one "unmethylated" and one "methylated" query probe (Figure 1A). The 3' terminus of each probe is designed to match either the protected cytosine (methylated design) or the thymine base resulting from bisulfite conversion and whole-genome amplification (unmethylated design).

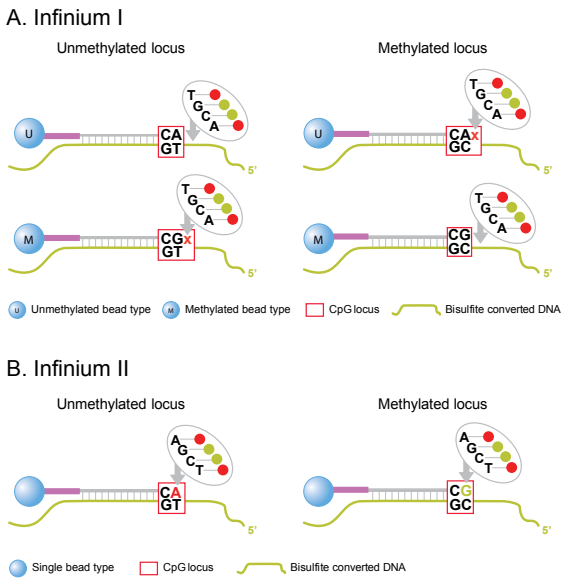
Probe designs for Infinium I assays are based on the assumption that methylation is regionally correlated within a 50 bp span and, thus, underlying CpG sites are treated as in phase with the 'methylated' (C) or 'unmethylated' (T) query sites. This co-methylation theory is supported in a study in which bisulfite sequencing of chromosomes 6, 20, and 22 showed that over 90% of CpG sites within 50 bases had the same methylation status[1]. A second study showed that, in general, methylation status at adjacent sites tends to be correlated, suggesting that correlation may depend upon the cell types or nearby polymorphic sites[2].

#### Infinium II Assay

The Infinium II assay design requires only one probe per locus (Figure 1B). The 3' terminus of the probe complements the base directly upstream of the query site while a single base extension results in the addition of a labeled G or A base, complementary to either the 'methylated' C or 'unmethylated' T.

A single, 50-mer probe is used to determine methylation state, making an "all-or-none" approach inapplicable. However, underlying CpG sites may be represented by "degenerate" R-bases. Illumina determined

**Figure 1: Broader Coverage Using Infinium I and II Assay Designs**



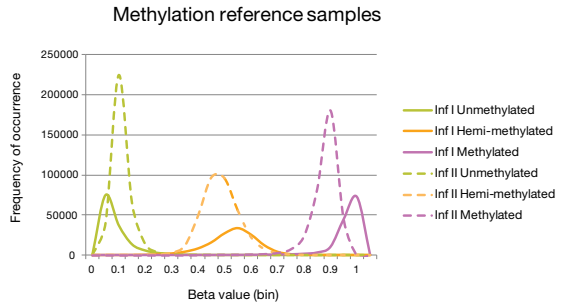

The HumanMethylation450 BeadChip employs both Infinium I and Infinium II assays, enhancing its breadth of coverage.
A) Infinium I assay design employs two bead types per CpG locus, one each for the methylated and unmethylated states.
B) The Infinium II design uses one bead type, with the methylated state determined at the single base extension step after hybridization.

that Infinium II probes can have up to three underlying CpG sites within the 50-mer probe sequence (i.e., 27 possible combinations overall) without compromising data quality. This feature enables the methylation status at a query site to be assessed independently of assumptions on the status of neighboring CpG sites. Further, the requirement for only a single bead type enables increased capacity for the number of CpG sites that can be queried. Infinium II designs are therefore applied whenever possible.

### Performance of Infinium I Versus Infinium II Assays

Due to their different chemistries, the Infinium I and II assays each have distinct advantages. Differences between the two chemistries have been observed that result in distinct beta value distributions within data sets. Figure 2 shows histograms of the beta values in bins of 0.02 and categorized by Infinium design type. In general, the peaks at the extreme ends of the beta distribution tend to be further out for Infinium I probes than for Infinium II probes, capturing the full spectrum of methylation.

**Figure 3: Infinium I and Infinium II Chemistries Cover the Full Spectrum of Methylation**




The different chemistries of Infinium I and Infinium II assays result in distinct beta value distributions[3].

The differences between the Infinium I and II assays do not affect the accuracy or reproducibility of the data generated by the Infinium HumanMethylation450 BeadChip. Importantly, the array was not designed with the intention that Infinium I and II assays, or any two assays, be compared within a single sample. Rather, the recommended comparison would be of individual sites between different samples or sample populations. Data quality metrics are assessed based on such comparisons, and technical replicates show > 98% correlation, often reaching > 99%.

Given the differences in the Infinium I and II chemistries, we sought to determine whether Infinium I and Infinium II probes vary in their performance based on relative correlation with another form of methylation assessment, whole-genome bisulfite sequencing (WGBS). DNA isolated from both lung tumor and normal tissue were assessed using both the Infinium HumanMethylation450 BeadChip and WGBS (Figure 3).

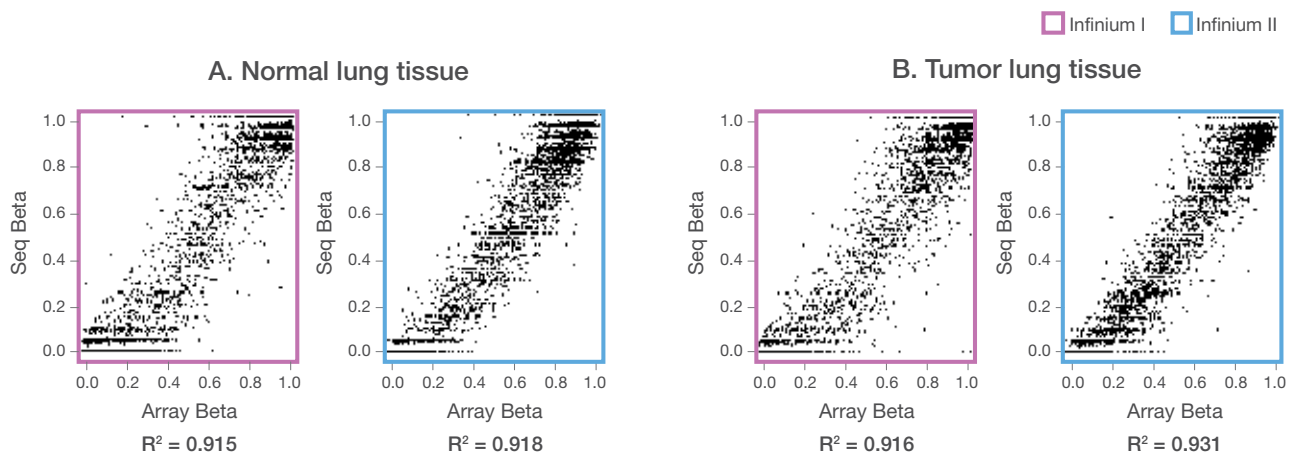
The WGBS data were filtered, requiring a minimum coverage depth of 20 reads, with the HumanMethylation450 BeadChip data minimum detection level set at $p < 0.01$. Infinium I and II probes were then separated and correlation levels measured between each subset and the corresponding sites measured by WGBS. The correlation levels for both tumor and normal samples were very similar for Infinium I and Infinium II probes. For the normal tissue sample, Infinium I probes showed an $r^2$ of 0.915, while Infinium II probes showed a value of 0.918. In the tumor sample, the results were comparable with Infinium I and II probes showing values of 0.916 and 0.931, respectively. Thus, while Infinium I and II probe chemistries are distinct, they deliver similar performance with regard to comparison against an independent control.

## Addressing Presence of SNPs Within Assay Region

The content for the Infinium HumanMethylation450 BeadChip was selected based on the recommendations of a panel of methylation experts. Prospective assays covering regions the consortium identified were filtered based on standard, Infinium design parameters. Among these filters was the avoidance of assays for which probes and query sites overlapped the positions of known DNA variants as reported in dbSNP. The potential risks associated with the presence of SNPs in the probe regions and query sites and means of addressing them in the data analysis process have been previously described for both GoldenGate® and Infinium methylation arrays[4, 5, 6].

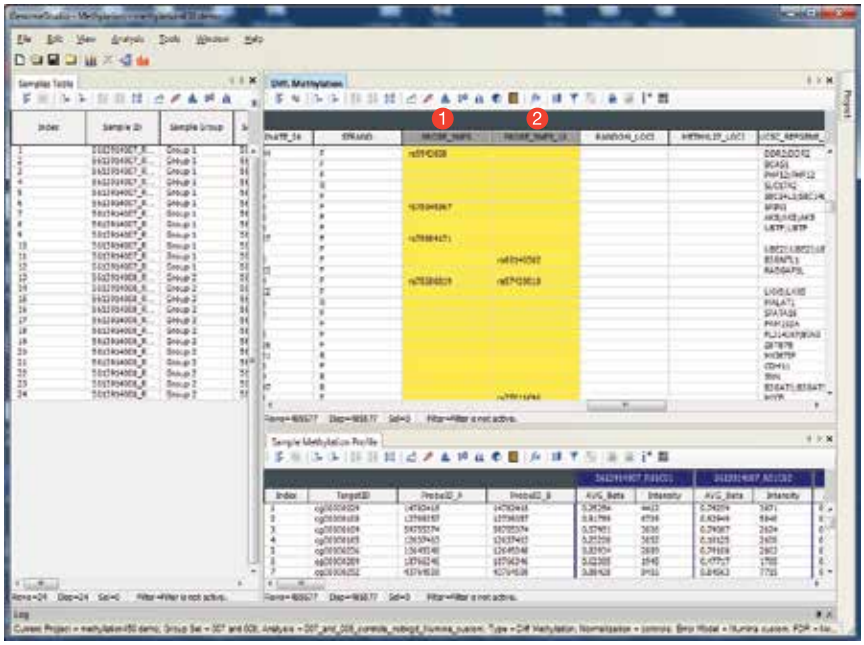
In cases in which a region identified by the consortium could only be covered by assays that included known SNP positions, the presence of the SNP locus is annotated in the manifest. This strategy places the decision of whether or not to utilize such assays in the hands of the individual investigator. Based on the make-up of their study population, the position of the SNP relative to the query site, the SNP's MAF, and other factors, researchers can assess the risk and decide for themselves whether to include these data and what validation measures to apply.

**Figure 3: Relative Correlation of Infinium I and Infinium II Probes with WGBS**




Normal lung (A) and tumor lung (B) tissue samples were assessed using the HumanMethylation450 BeadChip and WGBS, with high correlation seen between the Infinium I (purple) and Infinium II (blue) probes.

**Figure 4: HumanMethylation450 BeadChip Manifest File Data**




The HumanMethylation450 BeadChip manifest file divides data on SNP positions into two columns:
(1) whether the SNP position occurs within the first 10 bases of the 3' end of the probe.
(2) whether the SNP position occurs after the first 10 bases.

### Illumina-Provided Analysis Tools

To enable a straightforward assessment of SNP assays, Illumina provides two different sources of SNP information, the HumanMethylation450 BeadChip manifest file and Supplementary SNP list, both available from the MyIllumina customer portal. The HumanMethylation450 BeadChip manifest file includes information on the presence of known SNPs (Figure 4). Data is divided into columns based on whether the SNP position occurs within the first 10 bases of the 3' end of the probe or after the first 10 bases, due to the assumption that SNPs closer to the query site present an increased risk of impact on the data.

The more comprehensive Supplementary SNP list is updated every three months and provides a list of SNPs within either the probe sequence or, in the case of Infinium II designs, also at the query site, for each assay. Rather than categorizing based on the first 10 bases versus the next 40 bases from the 3' end of the probe, the exact position within the probe sequence is indicated. This information allows users to filter data based on their own criteria. In addition, the highest reported minor allele frequency (MAF) for each SNP is given, providing a means of identifying low-risk (based on both reported MAF and probe overlap position) cases without referring to another source. Note that while only a single MAF is indicated, a particular SNP may be reported in multiple populations at varying MAF. This information may be easily accessed directly through dbSNP. Information from the Supplementary SNP list can be imported directly into the GenomeStudio® Data Analysis Software Methylation Module (Figure 5).

## Summary

The HumanMethylation450 BeadChip allows researchers to interrogate > 485,000 methylations sites per sample at single-nucleotide resolution. Its unique design combines Infinium I and Infinium II assay chemistries, offering enhanced depth of coverage for methylation analysis. The information and resources described in this document will assist in analyzing data generated by the HumanMethylation450 BeadChip, ensuring the ability to make the most of this powerful analysis tool in their study.

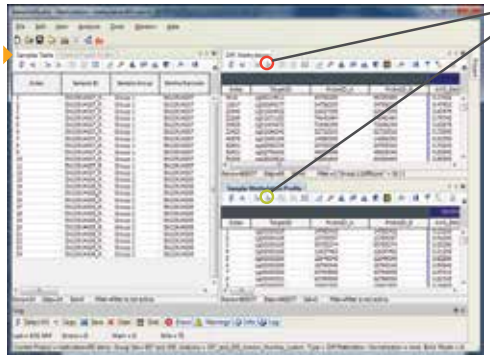
## References

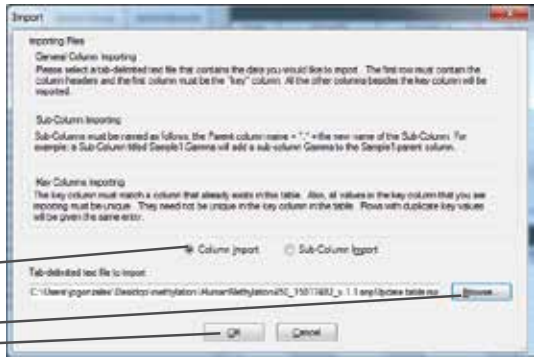
1. Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J, et.al. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22, Nat. Genet. 38: 1378–1385.
2. Shoemaker R, Deng J, Wang W, and Zhang K (2010) Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome, Genome Res. 20: 883–889.
3. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, et. al. (2011) High-density DNA methylation array with CpG resolution, Genomics 98: 288–295.
4. Hinoue T, Weisenberger DJ, Lange CP, Shen H, Byun HM, et.al. (2011) Genome-scale analysis of aberrant DNA methylation in colorectal cancer, Genome Res. June 9 [Epub ahead of print].
5. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, et.al. (2010) Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma, Cancer Cell 17:510–512.
6. Laffaire J, Everhard S, Idbaih A, Crinière E, Marie Y, et.al. (2011) Methylation profiling identifies 2 groups of gliomas according to their tumorigenesis, Neuro. Oncol. 13: 84–98.

**Figure 5: Importing the Supplemental SNP List into the GenomeStudio Methylation Module**

1. Click on to import columns into Data tables (3 choices)

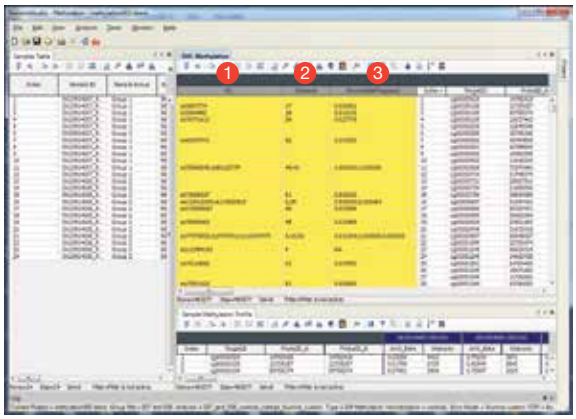
   Note: Differential Methylation Profile Table, Sample Methylation Profile Table, or Group Methylation Profile Table (not shown)




2. Select “Column Import”
3. Click “Browse” to select the SNP table file
4. Click “OK” to import SNP information into the active table

After import, three new columns will appear on the left side of the table.
In the example shown, SNP data is imported into the Differential Methylation Profile Table




The Supplementary SNP list is updated every three months and can be directly imported into GenomeStudio Methylation Module project tables by following the four steps illustrated above.

**Illumina** • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY


Pub. No. 270-2012-001 Current as of 05 March 2012

illumina®