TECHNICAL NOTE

Illumina Methylation BeadChips achieve breadth of coverage using two Infinium™ Chemistries

- Dual-probe design captures the full spectrum of DNA methylation at high resolution
- High-performance assay accommodates large sample sizes for epigenome-wide association studies
- Customizable analysis tools filter methylation data using the manifest file for simplified epigenetic analysis

illumina®

Introduction

Infinium Methylation BeadChips enable highly accurate and quantitative assessment of DNA methylation at the single-CpG-site level, offering powerful resolution for understanding epigenetic changes in health and disease. Infinium Methylation BeadChips apply the Illumina Infinium assay to detect epigenetic modifications, delivering broad coverage and high-throughput capabilities for large-scale epigenome-wide association studies (EWAS). By combining Infinium I and Infinium II probe chemistries, Infinium Methylation BeadChips provide expanded coverage of an expert-defined selection of CpG targets. This technical note provides background information and guidance on resources available for the analysis of methylation data generated by Infinium Methylation BeadChips.

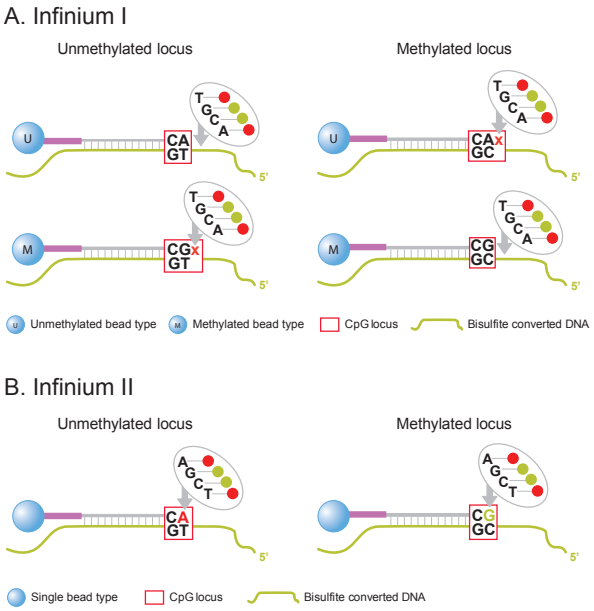

Figure 1: Infinium I and II assay designs allow for broader coverage of methylated CpG sites—(A) Infinium Type I probes use two bead types per CpG locus, one for methylated and one for unmethylated states. (B) Infinium Type II probes use a single bead type for both methylated and unmethylated DNA states.

Two Infinium chemistries enhance breadth of coverage

The Infinium methylation assay uses beads displaying target-specific probes designed to interrogate individual CpG sites within a DNA sample. Infinium I and Infinium II chemistries differ in the number of probes needed to query a single CpG locus. The Infinium I assay uses two probes per CpG, whereas the Infinium II assay requires one probe per CpG locus due to its ability to measure both unmethylated and methylated DNA states. As a result, Infinium I and II assays offer complementary strengths that enhance the breadth of coverage of the array.

Infinium I assay

The Infinium I assay employs two probes per CpG locus; one probe for unmethylated and one probe for methylated DNA states (Figure 1A). The 3′ terminus of each probe is designed to match either the protected cytosine, which indicates methylation, or the thymine base resulting from bisulfite conversion and whole-genome amplification of unmethylated cytosine. Probe designs for Infinium I assays are based on the assumption that methylation is regionally correlated within a 50 bp span. Thus, underlying CpG sites are treated as in phase with the methylated (C) or unmethylated (T) query sites. This comethylation 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 depends 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. A single base extension results in the addition of a labeled G or A base, complementary to either the methylated C or the unmethylated T bases. For the opposite strand, a labeled C or T base would be added.

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 or degenerate Y-bases for the opposite strand. Infinium II probes can have up to three underlying CpG sites within the 50-mer probe sequence 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. 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 and Infinium II assays

Due to their different chemistries, the Infinium I and II assays each have distinct advantages. Distinct beta value distributions within data sets are observed as a result of differences between the two chemistries (Figure 2). 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, thus capturing the full spectrum of methylation.

The differences between the Infinium I and II assays do not affect the accuracy or reproducibility of the data generated by Infinium Methylation BeadChips. Importantly, the microarray was not designed with the intention that Infinium I and II assays, or any two assays, be compared within a single sample. Rather, comparisons of individual sites between different samples or sample populations are recommended. Data quality metrics are assessed based on such comparisons and technical replicates exceed 98% correlation, often reaching more than 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. We looked at relative correlation with whole-genome bisulfite sequencing (WGBS), another form of methylation assessment. DNA isolated from both lung tumor and normal tissue were assessed using Infinium Methylation BeadChips and WGBS (Figure 3).

The WGBS data were filtered, requiring a minimum coverage depth of 20 reads, with the Infinium Methylation 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 similar for Infinium I and Infinium II probes. For the normal tissue sample, Infinium I probes showed an R^2 value of 0.915, whereas Infinium II probes showed an R^2 value of 0.918. In the tumor sample, the results were comparable, with Infinium I and II probes showing R^2 values of 0.916 and 0.931, respectively. Thus, although Infinium I and II probe chemistries are distinct, they deliver similar performance regarding comparison against an independent control.

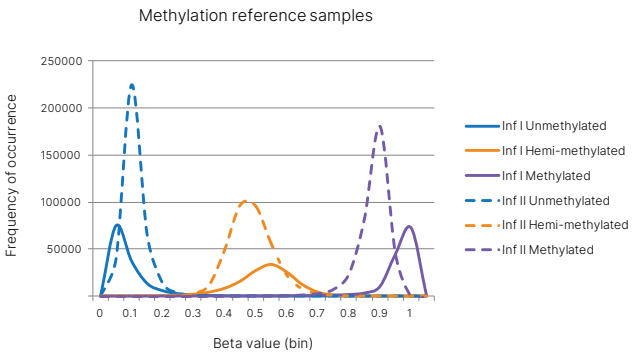

Figure 2: Infinium I and Infinium II chemistries cover the full spectrum of methylation—This figure shows histograms of the beta values in bins of 0.02 and categorized by Infinium design type. The different chemistries of Infinium I and Infinium II assays result in distinct beta value distributions.[3]

Addressing the presence of SNPs within the assay region

Content for Infinium Methylation BeadChips was selected based on the recommendations of a panel of methylation experts. Prospective assays covering regions identified by the consortium were filtered based on standard Infinium assay design parameters. Assays for which probes and query sites overlapped the positions of known DNA variants as reported in the SNP database (dbSNP) were filtered out and not included on the Infinium Methylation BeadChips. Potential risks associated with the presence of SNPs in the probe regions and query sites and considerations in data analysis have been previously described for Infinium Methylation BeadChips.[4-6]

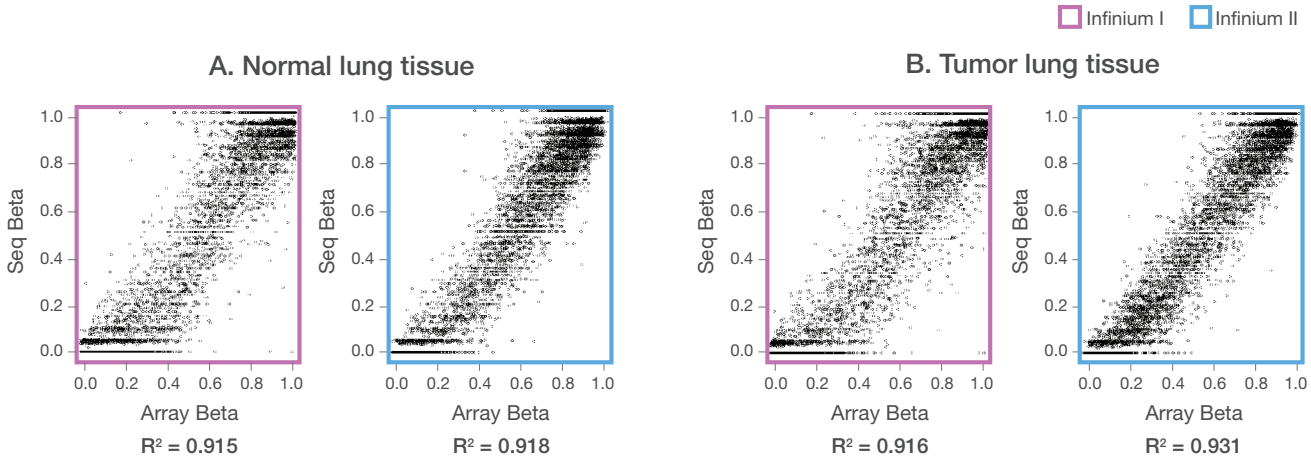

Figure 3: Relative correlation of Infinium I and Infinium II probes with WGBS—(A) Normal lung and (B) tumor lung tissue samples were assessed using an Infinium Methylation BeadChip and WGBS. Infinium I (purple) and Infinium II (blue) probes show high levels of correlation.

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 allows individual investigators to determine whether to use the specific assays in their analysis based on their study population, the position of the SNP relative to the query site, the SNP minor allele frequency (MAF), and other factors.

Resources for data analysis

Illumina provides a manifest file available on the MyIllumina customer portal to facilitate straightforward assessment of methylation assays. The manifest file includes information on the presence of known SNPs and relative positions in the probe. This information allows users to filter data based on their own criteria. In addition, the highest MAF for each SNP is listed, providing a means of identifying low-risk cases based on reported MAF and probe overlap position. Though 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 and imported directly into the GenomeStudio™ Software Methylation Module for data analysis.

In addition to GenomeStudio software, many popular third-party software packages are available on Bioconductor for downstream analysis. The two most commonly used software packages, minifi and SeSAMe, enable advanced quality control, updated normalization techniques, differential methylation analysis, and visualization capabilities.

For more information on methylation array data analysis and third-party software packages, visit illumina.com/techniques/microarrays/methylation-arrays/methylation-array-data-analysis-tips

Summary

Infinium Methylation BeadChips have a unique design that 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 help with analyzing data generated by Infinium Methylation BeadChips, offering researchers the ability to make the most of this powerful analysis tool in their studies.

References

1. Eckhardt F, Lewin J, Cortese R, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet*. 2006;38(12):1378-1385. doi:10.1038/ng1909
2. Shoemaker R, Deng J, Wang W, Zhang K. Allele-specific methylation is prevalent and is contributed by CpG-SNPs in the human genome. *Genome Res*. 2010;20(7):883-889. doi:10.1101/gr.104695.109
3. Bibikova M, Barnes B, Tsan C, et al. High density DNA methylation array with single CpG site resolution. *Genomics*. 2011;98(4):288-295. doi:10.1016/j.ygeno.2011.07.007
4. Hinoue T, Weisenberger DJ, Lange CP, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res*. 2012;22(2):271-282. doi:10.1101/gr.117523.110
5. Noushmehr H, Weisenberger DJ, Diefes K, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*. 2010;17(5):510-522. doi:10.1016/j.ccr.2010.03.017
6. Laffaire J, Everhard S, Idbaih A, et al. Methylation profiling identifies 2 groups of gliomas according to their tumorigenesis. *Neuro Oncol*. 2011;13(1):84-98. doi:10.1093/neuonc/noq110

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-00922 v1.0