Illumina Methylation BeadChips Achieve Breadth of

Introduction

Illumina Methylation BeadChips use the Illumina Infinium assay, a highly trusted, proven DNA analysis platform in the industry. They offer a unique combination of comprehensive, expert-selected coverage, and high-throughput compatible with large sample size, epigenomewide association studies. By combining Infinium I and Infinium II assay chemistry technologies, the Illumina Methylation BeadChips provide all 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 help with the analysis of data generated by Illumina Methylation BeadChips.

Coverage Using 2 Infinium[®] Chemistries

Two Infinium Chemistries Enhance Breadth of Coverage

The Infinium methylation assay uses beads displaying long, targetspecific 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 breadth of coverage of the array.

Infinium I Assay

The Infinium I assay employs 2 probes per CpG locus: 1 "unmethylated" and 1 "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. 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'. 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 sites².

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 '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 that Infinium II probes can have up to 3 underlying CpG sites within the



illumina

Single bead type
CpG locus
Bisulfite converted DNA
Figure 1: Broader Coverage Using Infinium I and II Assay Designs—Illumina
Methylation BeadChips employ both Infinium I and Infinium II assays, enhancing
the broadth of coverage ON Efficiency design employed bead types part

Methylation BeadChips employ both Infinium I and Infinium II assays, enhancing its breadth of coverage. A) Infinium I assay design employs 2 bead types per CpG locus, 1 each for the methylated and unmethylated states. B) The Infinium II design uses 1 bead type, with the methylated state determined at the single base extension step after hybridization.

50-mer probe sequence (ie, 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 2 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.

The differences between the Infinium I and II assays do not affect the accuracy or reproducibility of the data generated by the Infinium Methylation assay. Importantly, the array was not designed with the intention that Infinium I and II assays, or any 2 assays, be compared within a single sample. Rather, the recommended comparison



Methylation reference samples



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. We looked at relative correlation with wholegenome bisulfite sequencing (WGBS), another form of methylation assessment. DNA isolated from both lung tumor and normal tissue were assessed using both the Infinium HumanMethylation450 BeadChip and WGBS (Figure 3). The MethylationEPIC BeadChip has the same technology and performance specs.

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 similar for Infinium I and Infinium II probes. For the normal tissue sample, Infinium I probes showed an r^2 of 0.915, whereas 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, although Infinium I and II probe chemistries are distinct, they deliver similar

performance regarding comparison against an independent control.

Addressing Presence of SNPs Within Assay Region

The content for Infinium Methylation BeadChips 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. 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 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 to utilize such assays in the hands of the individual investigator. They can decide based on their study population, the position of the SNP relative to the query site, the SNPs MAF, and other factors.

Illumina Analysis Tools

To enable a straightforward assessment of SNP assays, Illumina provides 2 different sources of SNP information, the manifest file and Supplementary SNP list, both available from the Mylllumina customer portal. The manifest file includes information on the presence of known SNPs (Figure 4). Data are 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. This is due to the assumption that SNPs closer to the query site present an increased risk of impact on the data.

The Supplementary SNP list provides a list of SNPs within 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



A. Normal lung tissue

B. Tumor lung tissue

🔲 Infinium I

🗌 Infinium II

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.

Samples Table d D 🗙				Diff. Methylation							<	4 ⊅	
		≂, 🗙 🕊 öö ¶	<u>ج</u>	◎ ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ● ●									
Index	Sample ID	Sample Group	_	1 2									
			1	INATE_36	STRAND	PROBE_SNPS	PROBE_SNPS_10	RANDOM	LOCI M	ETHYL27_LOCI	UCSC_REFGE	NE	
	5613914007_R	Group 1	56 ×	14	F	rs9943058					DDR2:DDR2	2	
	5613914007_R	Group 1	56	8	F						BCAS1		
	5613914007_R	Group 1	56	8	F						PHF12;PHF1	12	
	5613914007_R	Group 1	56	5	R						SLC17A2		
	5613914007_R	Group 1	56	9	F						SEC14L1;SE	C1	
	5613914007_R	Group 1	56	5	F	rs75049367					SPIN1		
	5613914007_R	Group 1	50	5	F						AK5;AK5;AK	(5	
	5613914007_R	Group 1	50	5	F						UBTF;UBTF		
	5613014007 P	Group 1	50	57	F	rs75889171							
	5613914007 R	Group 1	56		F						UBE21;UBE2	21;1	
	5613914007 R	Group 1	56	2	F		rs80140302				B3GNTL1		
	5613914008 R	Group 2	56	53	F	75500040					RABGAP1L		
	5613914008 R	Group 2	56		F	12/2280813	1557439018						
	5613914008 R	Group 2	56	2	P						LHX9;LHX9		
	5613914008 R	Group 2	56	2	E						MALATI CDATA19		
	5613914008 R	Group 2	56	2	F						SPATA18		
	5613914008 R	Group 2	56	-	F						EL114107/P	TAU	
	5613914008 R	Group 2	56	6	F						707070	114.	
	5613914008_R	Group 2	56	1	P						MK1671D		
	5613914008_R	Group 2	56 E		F						CDH11		
	5613914008_R	Group 2	56	6	R						SNN		
	5613914008_R	Group 2	56	47	R						B3GAT1:B30	G۵	
	5613914008_R	Group 2	56	i i	F		rs77611696				MICB		
				4				III				- 1	
				Rows=485	i577 Disp=485577	Sel=0 Filter =Filter is r	not active.						
				Sample	Methylation Profile	A17 A			24.00		<	1	
										E6120140	1007 P01C02		
				Index	TargetID	ProbeID A	ProbeID B	AVG Beta	Intensity	AVG Beta	Intensity	4	
				- I IIII	co00000020	14792419	14702410	0.25254	4412	0.70220	2071	+	
				2	cg00000029	12709357	12709357	0.01789	6729	0.79239	5840	÷	
				3	cg00000108	59755374	59755374	0.527081	3006	0.52549	2624	÷	
				4	cg00000105	12637463	12637463	0.23228	3052	0.10125	2600	÷	
				5	cg0000036	12649348	12649348	0.83934	2889	0.79108	2863	\pm	
				6	cg00000289	18766346	18766346	0.62385	1945	0.47717	1785	t	
			-	7	cg00000292	43764508	43764508	0.88428	8416	0.84563	7715	t	
				11 ·				1 100 120		1 210 1202			

Figure 4: Infinium Methylation BeadChip Manifest File Data—Infinium Methylation BeadChip manifest file divides data on SNP positions into 2 columns: (1) whether the SNP position occurs with the first 10 bases of the 3" end of the probe.

(2) whether the SNP position occurs after the first 10 bases.

(based on both reported MAF and probe overlap position) cases without referring to another source. 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. Information from the Supplementary SNP list can be imported directly into the GenomeStudio[®] Data Analysis Software Methylation Module (Figure 5).

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, ensuring the ability to make the most of this powerful analysis tool in their study.

References

- 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.
- 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.
- 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.
- 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].
- 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.
- 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.

1. Click on Et to import columns into Data tables (3 choices) Note: Differential Methylation Profile Table, Sample Methylation Profile Table, or Group Methylation Profile Table (not shown)



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



Figure 5: Importing the Supplemental SNP List into the GenomeStudio Methylation Module — The Supplementary SNP list can be directly imported into the GenomeStudio Methylation Module project tables by following the 4 steps illustrated above.

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

For Research Use Only. Not for use in diagnostic procedures.

© 2015 Illumina, Inc. All rights reserved. Illumina, Infinium, GoldenGate, GenomeStudio, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. Pub. No. 270-2012-001 Current as of 12 October 2015

