Multiomics on the Methylation Screening Array

In addition to probes that can screen the methylation status of DNA, the Methylation Screening Array contains three types of other probes that can generate multiomics information. These include: (1) Probes that directly target SNPs (2) Dual intent probes which can interrogate a methylation as well as genetic variants and (3) Interrogation of 5- hydroxymethylcytosine (5hmC).

Section 1: Probes that directly target rsID sites

The Methylation Screening Array includes probes to interrogate 3,848 unique SNPs, the largest selection of genomic variants covered on any Infinium Methylation Array to date. Historically the presence of SNP probes on Infinium Methylation BeadChips (63 rsIDs on EPIC v2.0 and 59 rsIDs on EPIC v1.0) have been used for sample tracking. For the Methylation Screening Array, SNP selection has been expanded to include new Ancestry Informative Markers (AIM) for extended sample identification, ancestry determination, and tracking. In addition to these, probes that target common variants and low frequency pathogenic variants were carefully selected from various public databases as described below in detail.

How were these probes selected?

The following sources were used to target high-value SNPs on the Methylation Screening Array:

Content Source:	**rsID Selection Focus**
Infinium Global Screening Array-24 v3.0	Clinical variants
NHGRI-EBI GWAS Catalog	High MAF SNPs from GWAS
ClinVar (NIH)	Risk-associated SNPs for BRCA, Cystic Fibrosis, Breast Cancer, Ovarian Cancer, and Familial Hypercholesterolemia
Infinium QC Array	Ancestry Informative Markers (AIMS)
TruSight Hereditary Disease Sequencing Panel	Newborn Screening Variants
Infinium Methylation BeadChips	Legacy rsIDs
Individual studies	Variants associated w/ Alzheimer's disease

Previously existing Infinium Genotyping assays were redesigned to work with Infinium Methylation probe chemistry.

A breakdown for unique distribution of variants selected to different categories are provided in Table 1 below with additional details in the Auxiliary file (MSA-48v1-0_20102838_A1_rsIDs) provided within the Product Files. Please note some variants fall into multiple categories and hence total number will not add up to the total rsIDs on the product.

Table 1: Breakdown of unique SNP loci on the Methylation Screening Array to various categories

Clinical Category	Clinical Significance	MAF Range	# Variants
BRCA	Pathogenic/Likely pathogenic	Rare variants	167
Cystic Fibrosis	Pathogenic/Likely pathogenic	Rare variants	43
Breast and Ovarian Cancer	Pathogenic/Likely pathogenic	Rare variants	99

Familia Hypercholesterolemia	Pathogenic/Likely pathogenic	Rare variants	59
Newborn Screening	Pathogenic/Likely pathogenic	Rare variants	103
Newborn Screening	N/A	Greater than 10%	68
Ancestry Informative Markers (AIMs)	N/A	Greater than 10%	665
NHGRI GWAS	N/A	Greater than 10%	2773
NHGRI GWAS	Pathogenic/Likely pathogenic	Greater than 1%	1

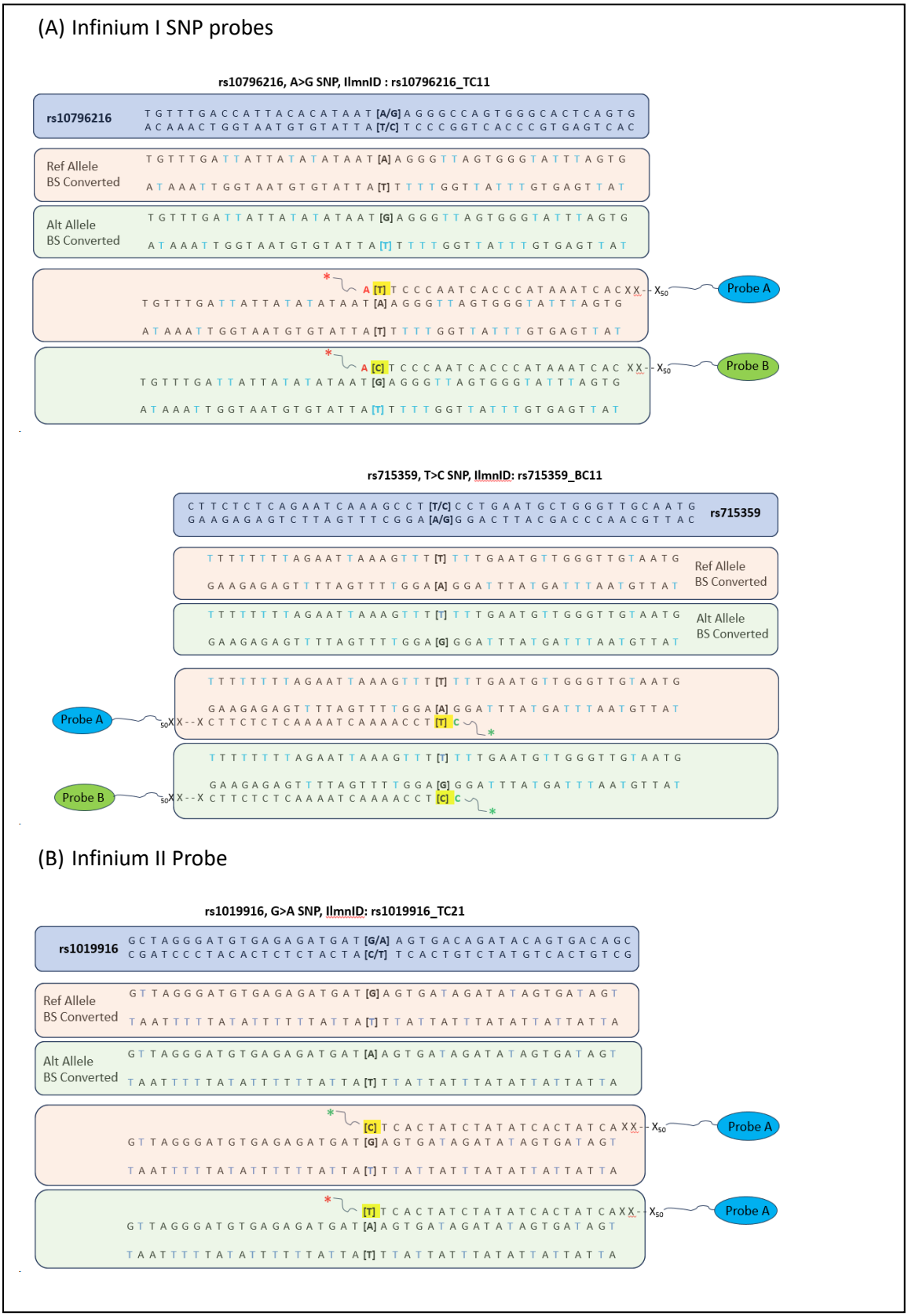
Non-Clinical Category	# Variants
Legacy Fingerprinting probes	64
Additional Fingerprinting probes	40
Alzheimer's Disease Associated from GWAS	18

How do these probes work?

SNP probes are designed as either Infinium I or II probes and query sites of known DNA variants.

The Infinium I assay contains two probes per target allele for any given locus with the base extending into the same fluorescent channel for both alleles (Figure 2A). Infinium II is a two-channel assay with one probe to assay both alleles, but the probe extends into separate fluorescent channels based on the base at the SNP position (Figure 2B).

Figure 2: Examples of Infinium I and II probes that assay SNP regions



How can these probes be used?

Within the manifest file for Methylation Screening Array (MSA-48v1-0_20102838_A1), SNP probes are designated with a prefix "rs" to the IlmnID and Name in columns A and B. Thus, customers can decide if they want to exclude these from their methylation analysis or separately assess them to infer genotyping information. Illumina currently has not verified the analytical performance of the genotyping probes on the Methylation Screening Array.

For users who wish to leverage the rsID probes, idats can be analyzed on GenomeStudio using a genotyping manifest, which contains allele and probe information with additional information including chromosome, mapinfo. If Genome Studio is used for this analysis, users will need to generate their own cluster file as described in the resources below.

- **Video: GenomeStudio™ Genotyping: Creating Custom Cluster Files for Infinium™ Arrays** https://www.illumina.com/company/video-hub/4JTrbMUbVN0.html
- **Tech Note: Infinium Genotyping Data Analysis** https://www.illumina.com/Documents/products/technotes/technote_infinium_genotyping_data_analysis.pdf

To de-identify individuals when uploading Methylation Screening Array data to public databases such as GEO, the SeSAME De-identify R package can be used.

Section 2: Dual methylation-genotyping probes

The Methylation Screening Array has a selection of approximately 11,610 unique CpGs that extend into a SNP, allowing for the interpretation of both methylation and genetic variant information. For this reason, these probes are referred to as '*dual methylation-genotyping probes'*.

Illumina is currently in-process of evaluating performance of dual methylation-genotyping probes. To find out which cg probes on the Methylation Screening Array fall under this designation, please reference the file (MSA-48v1-0_20102838_A1_dual-methyl-geno) in the Methylation Screening Array Product Files.

A detailed explanation of how these probes work can be found in the following paper:

Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res. 2017 Feb 28;45(4):e22. doi: 10.1093/nar/gkw967.

SeSAMe supports the analysis of dual methylation-genotyping probes with the Channel Inference function.

Section 3: Interrogation of 5- hydroxymethylcytosine (5hmC)

5-hydroxymethylcytosine (5hmC) is a modified form of the DNA base cytosine. It is one of the several epigenetic modifications that can occur to DNA. In this modification, a hydroxymethyl group is added to

the fifth carbon position of the cytosine ring. This modification is notable because it plays a role in the regulation of gene expression and has been identified to have significance in the development of brain diseases.

Various studies have demonstrated that 5hmC can be analyzed on Infinium Methylation Arrays using external protocols:

- Lunnon K, Hannon E, Smith RG, et al. Variation in 5-hydroxymethylcytosine across human cortex and cerebellum. Genome Biology. 2016;17(1). doi:https://doi.org/10.1186/s13059-016-0871-x
- Gross JA, Lefebvre F, Lutz PE, et al. Variations in 5-methylcytosine and 5-hydroxymethylcytosine among human brain, blood, and saliva using oxBS and the Infinium MethylationEPIC array. Biology Methods and Protocols. 2016;1(1):1-8. doi:https://doi.org/10.1093/biomethods/bpw002
- Salgado C, Oosting J, Janssen B, Kumar R, Gruis N, Doorn R. Genome-wide characterization of 5-hydoxymethylcytosine in melanoma reveals major differences with nevus. Genes, Chromosomes and Cancer. 2020;59(6):366-374. doi:https://doi.org/10.1002/gcc.22837
- Epigenome-wide analysis of multiple system atrophy and other neurodegenerative diseases Accessed February 3, 2024. https://acrobat.adobe.com/id/urn:aaid:sc:VA6C2:40a23531-1728-44e8-a08d-17707558f9e0
- Spiers H, Hannon E, Schalkwyk LC, Bray NJ, Mill J. 5-hydroxymethylcytosine is highly dynamic across human fetal brain development. BMC Genomics. 2017;18(1). doi:https://doi.org/10.1186/s12864-017-4091-x
- Zhang Z, Lee MK, Perreard L, Kelsey KT, Christensen BC, Salas LA. Navigating the hydroxymethylome: experimental biases and quality control tools for the tandem bisulfite and oxidative bisulfite Illumina microarrays. *Epigenomics*. 2022;14(3):139-152. doi:https://doi.org/10.2217/epi-2021-0490

Illumina does not provide official support for the analysis of 5hmc.