Infinium™ Methylation Screening Array GenomeStudio Manifest Column Headings

March 4, 2024

Below are detailed descriptions of the Infinium Methylation Screening Array Manifest file columns.

Column	Description
IlmnID	Unique identifier extended with the following design strand designations: Illumina strand (T/B), Assay Strand (C/O), Infinium Design Type (1/2), Rep_Num. A more detailed explanation can be found on page 3.
Name	Unique identifier from the Illumina CG database
AddressA_ID	For Infinium I beadtypes this is the Address ID for the probe specific for the A allele. For Infinium II beadtypes this is the Address ID for the probe used for both A and B alleles (in this case AddressB_ID and AlleleB_ProbeSeq columns will be empty)
AlleleA_ProbeSeq	The sequence of the probe identified in the AddressA_ID column.
AddressB_ID	For Infinium I beadtypes this is the Address ID for the probe specific for the B allele
AlleleB_ProbeSeq	For Infinium I beadtypes the sequence of the probe identified in the AddressB_ID column
Next_Base	For Infinium I probes the nucleotide immediately following the CpG. Blank for Infinium II
Color_Channel	For Infinium I probes the color channel of the Next_Base signal (Red/Green)
col	Color_Channel: For Infinium I probes the color channel of the Next_Base signal (R/G). Used for Bioconductor methylation array analysis software tools
Probe_Type	Probe type: cg=CpG, rs=dbSNP rsID, ch=Cp<nonG base>
Strand_FR	The Forward (F) or Reverse (R) designation of the Design Strand. F is equal to reference, + strand
Strand_TB	The Illumina Top (T) or Bottom (B) designation of the Design strand
Strand_CO	The assay Converted (C) or Opposite (O) designation of the Design strand
Infinium_Design	Infinium Design (1) - 2 probes or (2)- 1 probe. Used for Bioconductor methylation array analysis software tools
Infinium_Design_Type	Infinium I (2 probes/locus) or Infinium II (1 probe/locus)
Rep_Num	Reflects the replicate number if the Name has replicate designs
CHR	Chromosome containing the CpG (GRCh38)
MAPINFO	Chromosomal coordinates of the CpG (GRCh38)
Species	Species which the assays were designed for (Homo sapiens)
Genome_Build	Genome Build referenced for this manifest
Source_Seq	The original genomic sequence used for probe design prior to bisulfite conversion
Forward_Sequence	Plus (+) strand sequence (5'-3') flanking the CG
Top_Sequence	Top strand sequence (5'-3') flanking the CG
UCSC_RefGene_Group	NCBI RefSeq Gene region: 5UTR=5' untranslated region between the TSS and ATG start site, 3UTR=3' untranslated region between stop codon and poly A signal, exon_#, TSS200=1-200 bp 5' the TSS, TS1500=200-1500 bp 5' of the TSS. Intronic regions included
UCSC_RefGene_Name	NCBI RefSeq Gene Name: Target gene names from the RefSeq database. Multiple listings of the same gene name indicate splice variants
UCSC_RefGene_Accession	NCBI RefSeq Transcript ID: The RefSeq accession numbers of the target transcripts. Accession numbers are in the same order as the target gene transcripts
UCSC_CpG_Islands_Name	Chromosomal coordinates of the CpG Island from UCSC.
Relation_to_UCSC_CpG_Island	Island=within boundaries of a CpG Island, N_Shore=0-2kb 5' of Island, N_Shelf=2kb-4kb 5' of Island, S_Shore=0-2kb 3' of Island, S_Shelf=2kb-4kb 3' of Island

GencodeV41_Group	Gencode v41 Gene region: 5UTR=5' untranslated region between the TSS and ATG start site, 3UTR=3' untranslated region between stop codon and poly A signal, exon_#, TSS200=1-200 bp 5' the TSS, TS1500=200-1500 bp 5' of the TSS. Intronic regions included
GencodeV41_Name	Gencode v41 Gene Name: Target gene names from the RefSeq database. Multiple listings of the same gene name indicate splice variants
GencodeV41_Accession	Gencode v41 Transcript ID: The RefSeq accession numbers of the target transcripts. Accession numbers are in the same order as the target gene transcripts
Phantom5_Enhancers	Chromosomal coordinates from the FANTOM consortium of enhancer regions associated with FANTOM5 promoters
HMM_Island	Hidden Markov Model Islands. Chromosomal map coordinates of computationally predicted CpG islands
Regulatory_Feature_Name	Chromosomal map coordinates of the regulatory feature (informatically determined by the original ENCODE Consortium)
Regulatory_Feature_Group	Description of the regulatory feature referenced in Regulatory_Feature_Name as provided by the original ENCODE Consortium - Gene_Associated, Gene_Associated _Cell_type_specific, nonGene_Associated, Promoter_Associated_Cell_type_specific, Unclassified, Unclassified_Cell_type_specific
DNase_Hypersensitivity_NAME	Name of the DNase Hypersensitivity site as defined in ENCODE v5
Encode_CisReg_Site	Name and classification of the CisRegulatory Region as defined in ENCODE v5
Encode_CisReg_Site_Evid	Evidence for each CisReg Region/Classification based on 87 studies in ENCODE v5 based on AllData-Full classification
OpenChromatin_NAME	Classification of Open Chromatin sites based on 1600+ studies in ENCODE v5
OpenChromatin_Evidence_Count	Evidence for each Open Chromatin site classification based on 1600+ studies in ENCODE v5
EPICv2_Locus_Match	Locus Name (IlmnID) in Infinium MethylationEPICv2.0 Manifest
EPICv1_Locus_Match	Locus Name (IlmnID) in Infinium MethylationEPICv1.0 Manifest
Methyl450_Locus_Match	Locus Name (IlmnID) in Infinium Methylation450K Manifest
EPICv2_ProbeSeq_Match	Indicates if the AlleleA_ProbeSeq matches for EPIC v2
EPICv1_ProbeSeq_Match	Indicates if the AlleleA_ProbeSeq matches for EPIC v1
Methyl450_ProbeSeq_Match	Indicates if the AlleleA_ProbeSeq matches for Methylation450K
SNP_ID	rsIDs of SNPs located in the probe. Multiple listings of SNP rsIDs are allowed. dbSNP v155 used as reference
SNP_DISTANCE	Distance of SNPs from query base of the probe. Multiple listings of the distance values are associated with rsid. dbSNP v155 used as reference
SNP_MinorAlleleFrequency	Minor allele frequency of SNPs. Multiple listings of the minor allele frequencies are associated with rsid. dbSNP v155 used as reference
CHR_GRCh37	Chromosome containing the CpG (GRCh37)
MAPINFO_GRCh37	Chromosomal coordinates of the CpG (GRCh37)

The logic for how the IlmnID is determined is shown below in the leftmost table column and other relevant information on how the IlmnID is generated is given in the remaining columns:

IlmnID	Name	Strand_TB	Strand_CO	Infinium_Design_ Type	The specific synthetic oligonucleotide probe replicate
The IlmnID is a composite of multiple information fields: the name of the probe, whether the probe targets the top or bottom strand, whether the probe targets the bisulfite converted strand or complementary strand after amplification, the Infinium probe design type, and the number of times the probe was synthesized for array representation. Example of IlmnID: **cg12345678_TC13** This probe would be a CG probe with an eight digit code that relates to the probe sequence. The following "T" indicates that the probe targets the top strand. The adjacent "C" indicates that the probe targets the strand that is initially bisulfite converted. The "1" indicates that the probe has a Type I Infinium design. The "3" indicates that the probe was synthesized three times for representation on the array The multiple fields that comprise the IlmnID are provided in the table columns to the right.	The name of the probe is determined by combining the locus target identifier, which designates the function of a given probe, with an eight digit code that relates to the probe sequence. In situations where the eight digit code has not yet been generated, standard genomic coordinates are used. Example of probe name: **cg12345678** Examples of locus target identifiers: **cg**= CpG interrogating **ch**= CpH interrogating **rs** = SNP interrogating	Strand_TB relates to whether the probe is designed to target the top or bottom strands of a given locus as defined by Illumina's standard designation for ambiguous SNP probes. Examples of top/bottom strand designations: Top strand = **T** Bottom strand = **B**	Strand_CO refers to whether the probe is designed to target the originally bisulfite converted DNA strand, or the strand resulting from amplification of the originally converted DNA strand. Examples of converted/uncon-verted designations: Converted strand = **C** Opposite strand = **O**	Infinium_Design_ Type refers to whether the probe utilizes a Type I design that uses two attempted bead types to probe a single locus, or a Type II design that uses one attempted bead type to probe a single locus. Examples of Infinium design type designations: Infinium Type I design = **1** Infinium Type II design = **2**	This value depends on how many times a probe was synthesized for representation on the array. For instance, if the probe was synthesized three separate times, the probe generated by the first synthesis would be denoted as "1", while the probe generated from the second synthesis would be denoted as "2", and the probe generated from the third synthesis would be denoted as "3".