Infinium MethylationEPIC Manifest Column Headings

Oct 13 2014

Below are detailed descriptions of the Infinium MethylationEPIC Array manifest file column headings.

IlmnID: Unique identifier from the Illumina CG database. (The probe ID).
Name:The IlmnID.
Infinium_Design_Type: Infinium I (2 probes/locus) or Infinium II (1 probe/locus).
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.
Forward_Sequence: Plus (+) strand (HapMap) sequence (5'-3') flanking the CG.
Genome_Build: Genome Build referenced by the manifest.
CHR: Chromosome containing the CpG (Build 37).
MAPINFO: Chromosomal coordinates of the CpG (Build 37).
SourceSeq: The original, genomic sequence used for probe design before bisulfite conversion.
Chromosome_36: Chromosome containing the CpG (Build 36).
Coordinate_36: Chromosomal coordinates of the CpG (Build 36).
Strand: The Forward (F) or Reverse (R) designation of the Design Strand. *Note: in methylation manifest files, the Forward Strand = the genomic Plus (+) Strand and the Reverse Strand = the genomic Minus (-) Strand. In this context, Forward and Reverse ARE NOT EQUIVALENT to the Forward and Reverse Strand designations originating from dbSNP or as given in Infinium Genotyping manifests.
SNP_ID: rsid(s) of SNP(s) located in the probe. *Note: multiple listings of SNP rsid are allowed.
SNP_DISTANCE: Distance of SNP(s) from query base of the probe. *Note: multiple listings of the distance values are associated with rsid.
SNP_MinorAlleleFrequency: Minor allele frequency of SNP(s). *Note: multiple listings of the minor allele frequencies are associated with rsid.
Random_Loci: CpG loci chosen randomly by consortium members during the design process are marked "True".
Methyl27_Loci: CpG's carried over from the HumanMethylation27 array (92% carryover) are marked "True".
Methyl450_Loci: CpG's carried over from the HumanMethylation450 array (94% carryover) are marked "True".
UCSC_RefGene_Name: Target gene name(s), from the UCSC database. *Note: multiple listings of the same gene name indicate splice variants.
UCSC_RefGene_Accession: The UCSC accession number(s) of the target transcript(s). Accession numbers are given in the same order as the target gene transcripts.
UCSC_RefGene_Group: Gene region feature category describing the CpG position, from UCSC. Features listed in the same order as the target gene transcripts.
TSS200 = 0–200 bases upstream of the transcriptional start site (TSS).

TSS1500 = 200–1500 bases upstream of the TSS.

5'UTR = Within the 5' untranslated region, between the TSS and the ATG start site.

Body = Between the ATG and stop codon; irrespective of the presence of introns, exons, TSS, or promoters.

3'UTR = Between the stop codon and poly A signal.

UCSC_CpG_Islands_Name: Chromosomal coordinates of the CpG Island from UCSC.
Relation_to_UCSC_CpG_Island: The location of the CpG relative to the CpG island.
Shore = 0–2 kb from island.

Shelf = 2–4 kb from island.

N = upstream (5') of CpG island.

S = downstream (3') of CpG island.

Phantom: Classifications from the FANTOM (Functional Annotation of the Mammalian Genome) consortium as a low- or high-CpG density region associated with FANTOM 4 promoters.
DMR: Differentially methylated regions (experimentally determined).
DMR = Differentially Methylated Region.

CDMR = Cancer-specific Differentially Methylated Region.

RDMR = Reprogramming-specific Differentially Methylated Region.

450k_Enhancer: Predicted enhancer elements as annotated in the original 450K design (determined by the ENCODE Consortium using informatics and the original 450K consortia members) are marked "True".
HMM_Island:Hidden Markov Model Islands. Chromosomal map coordinates of computationally predicted CpG islands.
Regulatory_Feature_Name: Chromosomal map coordinates of the regulatory feature (determined by theENCODE Consortium using informatics).
Regulatory_Feature_Group: Description of the regulatory feature referenced in "Regulatory_Feature_Name" as provided by the Methylation Consortium.
Gene_Associated

Gene_Associated_Cell_type_specific

NonGene_Associated

Promoter_Associated

Promoter_Associated_Cell_type_specific

Unclassified

Unclassified_Cell_type_specific

DHS: DNase I Hypersensitivity Site (experimentally determined by the ENCODE project).

Any column headers omitted from this bulletin are as described in the Infinium Genotyping Manifest Column Headings bulletin. Information in the methylation manifest references Genome Build 37 (HG19) unless otherwise stated.
GencodeBasicV12_NAME: Target gene name(s), from the basic GENECODE build. *Note: multiple listings of the same gene name indicate splice variants.

GencodeBasicV12_Accession: The basic GENECODE accession number(s) of the target transcript(s). Accession numbers are given in the same order as the target gene transcripts.
GencodeBasicV12_Group: Gene region feature category describing the CpG position, from basic GENECODE. Features listed in the same order as the target gene transcripts.
TSS200 = 0–200 bases upstream of the transcriptional start site (TSS).

GencodeCompV12_NAME: Target gene name(s), from the complete GENECODE build. *Note: multiple listings of the same gene name indicate splice variants.

GencodeCompV12_Accession: The complete GENECODE accession number(s) of the target transcript(s). Accession numbers are given in the same order as the target gene transcripts.
GencodeCompV12_Group: Gene region feature category describing the CpG position, from complete GENECODE. Features listed in the same order as the target gene transcripts.
TSS200 = 0–200 bases upstream of the transcriptional start site (TSS).

DNase_Hypersensitivity_NAME: Chromosomal coordinates of the DNase hypersensitive region from ENCODE.
DNase_Hypersensitivity_Evidence_Count: Number of supporting experimental evidence for DNase hypersensitive region from ENCODE.
OpenChromatin_NAME: Chromosomal coordinates of open chromatin region from ENCODE.
OpenChromatin_Evidence_Count: Number of supporting experimental evidence for open chromatin region from ENCODE.
TFBS_NAME: Chromosomal coordinates of transcription factor binding site region from ENCODE.
TFBS_Evidence_Count: Number of supporting experimental evidence for transcription factor bind site region from ENCODE.