

The EPICv2manifest package user's guide

Peters TJ, Pidsley R

February 9, 2024

Package Contents

EPICv2manifest is an annotation package providing a data.frame object containing the complete 49 column Infinium Illumina EPIC v2.0 probe manifest ("IlmnID" constitutes the rownames), plus 31 additional columns derived from Peters et al. (2024).

```
library(EPICv2manifest)
```

Additional columns

- CpG_chrm, CpG_beg, CpG_end - EPICv2 coordinates from Sesame manifest
- MismatchPos - vector of probes where "Y" indicates discrepant genomic position between Sesame and Illumina, includes those that are missing mapping information in Illumina manifest (chr0)
- MissingPos - vector of probes where "Y" indicates probes missing mapping information in Illumina manifest
- namerep - vector of probes where "Y" indicates probes that have replicates based on probe name match (this vector can be married up with Name column to select a single probe to represent each name replicate)
- seqrep - vector of probes where "Y" indicates probes that have exact sequence matches with other probes within EPICv2
- seqrep_IlmnIDs - IlmnIDs of those probes with exact sequence matches with other probes within EPICv2 (corresponding to seqrep)
- seqrep_RepNum - replicate number of those probes with exact sequence matches with other probes within EPICv2 (corresponding to seqrep and seqrep_IlmnIDs). Can be used to e.g. filter for only 1 probe per sequence-replicate.

- **posrep** - vector of probes where “Y” indicates probes that have exact genomic position matches with other probes within EPICv2 (based on Illumina genomic positions – note, none of the affected probes are discrepant with Sesame mapping)
- **posrep_IlmnIDs** - IlmnIDs of those probes with exact genomic position matches with other probes within EPICv2 (corresponding to **posrep**)
- **posrep_RepNum** - replicate number of those probes with exact genomic position matches with other probes within EPICv2 (corresponding to **posrep** and **posrep_IlmnIDs**). Can be used to e.g. filter for only 1 probe per genomic-position-replicate.
- **EPICv1probeID** - vector of EPICv1 probe names where probe names match between EPICv1 and EPICv2
- **EPICv1seqmatch** - vector of EPICv1 probe names where probe sequences match between EPICv1 and EPICv2
- **EPICv1locmatch** - vector of EPICv1 probe names where genomic locations match between EPICv1 and EPICv2 (based on Sesame locations)
- **K450probeID** - vector of 450K probe names where probe names match between 450K and EPICv2
- **K450seqmatch** - vector of 450K probe names where probe sequences match between 450K and EPICv2
- **K450locmatch** - vector of 450K probe names where genomic locations match between 450K and EPICv2 (based on Sesame locations)
- **K450locmatch2** - vector of 450K probe names where additional 450K probes have a genomic location match between 450K and EPICv2 (based on Sesame locations)
- **K27probeID** - vector of 27K probe names where probe names match between 27K and EPICv2
- **K27seqmatch** - vector of 27K probe names where probe sequences match between 27K and EPICv2
- **K27locmatch** - vector of 27K probe names where genomic locations match between 27K and EPICv2 (based on Sesame locations)
- **K27locmatch2** - vector of 27K probe names where additional 27K probes have a genomic location match between 27K and EPICv2 (based on Sesame locations)
- **CH_BLAT** - vector of probes where “Y” indicates at least one in silico cross-hybridisation event (≥ 47 bp match) to a non-target region of the genome, predicted by BLAT (Kent 2002)

- **CH_WGBS_evidence** – subset vector of **CH_BLAT** where “Y” indicates a greater affinity for the off-target(s), via comparison to whole genome bisulphite sequencing on matched samples
- **RMSE_with_WGBS** – root mean squared error when comparing probe methylation to matched target CpG site methylation from WGBS (*M*-values)
- **Num_offtargets** – number of off-target *in silico* hybridisation events predicted by the probe sequence via BLAT (Kent 2002)
- **Suggested_offtarget** - if **CH_WGBS_evidence** == “Y”, the hg38 coordinate of the off-target cytosine conferring minimum RMSE with WGBS
- **Rep_results_by_NAME** – results of competitive comparison between replicates, with replicate probe sets defined by column **Name**
- **Rep_results_by_SEQUENCE** - results of competitive comparison between replicates, with replicate probe sets defined by column **seqrep_ilmnIDs**
- **Rep_results_by_LOCATION** - results of competitive comparison between replicates, with replicate probe sets defined by column **posrep_ilmnIDs**

```
sessionInfo()

## R Under development (unstable) (2024-01-16 r85808)
## Platform: x86_64-pc-linux-gnu
## Running under: Ubuntu 22.04.3 LTS
##
## Matrix products: default
## BLAS:   /home/biocbuild/bbs-3.19-bioc/R/lib/libRblas.so
## LAPACK: /usr/lib/x86_64-linux-gnu/lapack/liblapack.so.3.10.0
##
## locale:
##  [1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=en_GB            LC_COLLATE=C
##  [5] LC_MONETARY=en_US.UTF-8  LC_MESSAGES=en_US.UTF-8
##  [7] LC_PAPER=en_US.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C
##
## time zone: America/New_York
## tzcode source: system (glibc)
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
```

```
## [1] EPICv2manifest_0.99.0
##
## loaded via a namespace (and not attached):
## [1] rappdirs_0.3.3      utf8_1.2.4          generics_0.1.3
## [4] bitops_1.0-7        BiocVersion_3.19.1  RSQLite_2.3.5
## [7] magrittr_2.0.3      evaluate_0.23       fastmap_1.1.1
## [10] blob_1.2.4          AnnotationHub_3.11.1 AnnotationDbi_1.65.2
## [13] GenomeInfoDb_1.39.6 DBI_1.2.1           BiocManager_1.30.22
## [16] httr_1.4.7          fansi_1.0.6         Biostrings_2.71.2
## [19] cli_3.6.2           rlang_1.1.3         crayon_1.5.2
## [22] dbplyr_2.4.0        XVector_0.43.1      Biobase_2.63.0
## [25] bit64_4.0.5         cachem_1.0.8        yaml_2.3.8
## [28] tools_4.4.0         memoise_2.0.1       dplyr_1.1.4
## [31] GenomeInfoDbData_1.2.11 filelock_1.0.3      BiocGenerics_0.49.1
## [34] curl_5.2.0          vctrs_0.6.5         R6_2.5.1
## [37] png_0.1-8           stats4_4.4.0        BiocFileCache_2.11.1
## [40] lifecycle_1.0.4     zlibbioc_1.49.0     KEGGREST_1.43.0
## [43] S4Vectors_0.41.3    IRanges_2.37.1      bit_4.0.5
## [46] pkgconfig_2.0.3     pillar_1.9.0        glue_1.7.0
## [49] highr_0.10          xfun_0.42           tibble_3.2.1
## [52] tidyselect_1.2.0    knitr_1.45          compiler_4.4.0
## [55] Rcurl_1.98-1.14
```

Sources

References

Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Research*, **12**(4), 656–664.