

The DMRcatedata package user's guide

Peters TJ, Buckley MJ, Statham AL, Pidsley R, Zotenko E, Clark SJ, Molloy PL

February 13, 2024

Package Contents

DMRcatedata accompanies the DMRcate package, providing data for examples, probe filtering and transcript annotation.

```
library(DMRcatedata)
data(crosshyb)

#Proximal SNPs to EPICv1 and 450K
data(snpsall)

#Proximal SNPs to EPICv2
data(epicv2snps)

data(hg19.grt)
data(hg19.generanges)

#Betas for EPICv2
data(ALLbetas)
```

Ten objects are contained in DMRcatedata. `crosshyb` is a factor listing EPICv1 and 450K probe IDs potentially confounded by cross-hybridisation to other parts of the genome[1][2]. It is used internally by `rmSNPandCH()`.

`snpsall` is a data.frame containing probes from 450K and EPICv1 that are potentially confounded by a SNP or indel variant[1]. It lists the ID, distance (in nucleotides) to the CpG in question, and minor allele frequency for each associated variant. `epicv2snps` contains the same but for the EPICv2 array.

`XY.probes` is a vector of EPICv1 and 450K Illumina probes whose targets are on human sex chromosomes.

Objects named `.*(grt|generanges)` are annotation objects that are needed by `extractRanges()` and `DMR.plot()` respectively. `hg38` and `mm10` objects have been parsed from Release 96 of Ensembl, and `hg19` from Release 75. These are accessed within the environment of the aforementioned functions.

ALLbetas is a matrix of EPICv2 beta values from Noguera-Castells et al. (2023)[3] consisting of five B cell acute lymphoblastic leukaemia (BALL) and five T cell acute lymphoblastic leukaemia (TALL) samples for DMR calling.

Sources

- snpsall sourced from https://static-content.springer.com/esm/art\%3A10.1186\%2Fs13059-016-1066-1/MediaObjects/13059_2016_1066_MOESM4_ESM.csv, https://static-content.springer.com/esm/art\%3A10.1186\%2Fs13059-016-1066-1/MediaObjects/13059_2016_1066_MOESM5_ESM.csv, https://static-content.springer.com/esm/art\%3A10.1186\%2Fs13059-016-1066-1/MediaObjects/13059_2016_1066_MOESM6_ESM.csv, <http://www.sickkids.ca/MS-Office-Files/Research/WeksbergLab/48640-polymorphic-CpGs-Illumina450k.xlsx> (accessed October 2016)
- crosshyb sourced from https://static-content.springer.com/esm/art\%3A10.1186\%2Fs13059-016-1066-1/MediaObjects/13059_2016_1066_MOESM2_ESM.csv, https://static-content.springer.com/esm/art\%3A10.1186\%2Fs13059-016-1066-1/MediaObjects/13059_2016_1066_MOESM3_ESM.csv (accessed October 2016) and <http://www.sickkids.ca/MS-Office-Files/Research/WeksbergLab/48639-non-specific-probes-Illumina450k.xlsx>, (accessed February 2014).
- ALLbetas sourced from https://ftp.ncbi.nlm.nih.gov/geo/series/GSE222nnn/GSE222919/suppl/GSE222919_processed_data.txt.gz, (accessed February 2024).

References

- [1] Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, Van Dijk S, Muhlhausler B, Stirzaker C, Clark SJ. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biology*. 2016 17(1), 208.
- [2] Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, Gallinger S, Hudson TJ, Weksberg R. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013 Jan 11;8(2).
- [3] Noguera-Castells A, Garcia-Prieto CA, Alvarez-Errico D, Esteller M. Validation of the new EPIC DNA methylation microarray (900K EPIC v2) for high-throughput profiling of the human DNA methylome. *Epigenetics* 2023 Dec;18(1):2185742.