Package 'HarmanData'

June 6, 2023

Type Package

Title Data for the Harman package

Version 1.28.0

Date 2022-02-18

Description Datasets of accompany Harman, a PCA and constrained optimisation based technique. Contains three example datasets: IMR90, Human lung fibroblast cells exposed to nitric oxide; NPM, an experiment to test skin penetration of metal oxide nanoparticles following topical application of sunscreens in non-pregnant mice; OLF; an experiment to gauge the response of human olfactory neurosphere-derived (hONS) cells to ZnO nanoparticles.
Since version 1.24, this package also contains the Infinium5 dataset, a set of batch correction adjustments across 5 Illumina Infinium Methylation BeadChip datasets. This file does not contain methylation data, but summary statistics of 5 datasets after correction.
There is also an EpiSCOPE_sample file as exampling for the new methylation clustering functionality in Harman.

NeedsCompilation no

Suggests BiocGenerics, BiocStyle, knitr, rmarkdown, Harman (>= 1.23.3)

Depends R (>= 3.5.0)

License GPL-3

LazyData true

LazyDataCompression gzip

VignetteBuilder knitr

biocViews ExpressionData, MicroarrayData, ExperimentData, Homo_sapiens_Data, Mus_musculus_Data

URL http://www.bioinformatics.csiro.au/harman/

RoxygenNote 7.1.2

Encoding UTF-8

git_url https://git.bioconductor.org/packages/HarmanData

git_branch RELEASE_3_17

git_last_commit 62386a4

git_last_commit_date 2023-04-25

Date/Publication 2023-06-06

Author Yalchin Oytam [aut],
Josh Bowden [aut],
Jason Ross [aut, cre]

Maintainer Jason Ross <jason.ross@csiro.au>

R topics documented:

`episcope` *Infinium Methylation BeadChip batch correction reference data*

Description

A reference dataset containing beta values spanning 11 CpG probesets from the 369 arrays of the EpiSCOPE study (van Dijk, 2106). The 450K methylation data arises from neonate blood spots from children enrolled in the DOMInO (DHA to Optimise Mother Infant Outcome) cohort.

Usage

```
data(episcope)
```

Format

A list with 6 entities:

pd Phenotypic descriptors for the 369 samples

original Original uncorrected data from the study

harman Harman corrected data

combat ComBat corrected data

ref_lvr Reference log2 variance ratios for the 11 probes

ref_md Reference mean difference in beta for the 11 probes

Details

Example beta values from the EpiSCOPE study

Value

promise

References

https://doi.org/10.1186/s13148-016-0281-7/

Examples

```
library(Harman)
data(episcope)
bad_batches <- c(1, 5, 9, 17, 25)
is_bad_sample <- episcope$pd$array_num %in% bad_batches
myK <- discoverClusteredMethylation(episcope$original[, !is_bad_sample])
mykClust = kClusterMethylation(episcope$original, row_ks=myK)
res = clusterStats(pre_betas=episcope$original,
                   post_betas=episcope$harman,
                   kClusters = mykClust)
all.equal(episcope$ref_md$meandiffs_harman, res$meandiffs)
all.equal(episcope$ref_lvr$var_ratio_harman, res$log2_var_ratio)
```

IMR90 *IMR90 data, a Human lung fibroblast cell line.*

Description

Data used in the batch effect correction paper of Johnson, Li and Rabinovich. The data are from a cell-line experimental designed to reveal whether exposing mammalian cells to nitric oxide (NO) stabilizes mRNAs. The data comprises one treatment, one control and 2 time points (0 h and 7.5 h), resulting in 4 distinct (2 treatment x 2 time points) experimental conditions. There were 3 batches and a total of 12 samples, with each batch consisting of 1 replicate from each of the experimental conditions. Affymetrix HG-U133A Arrays were normalised and background adjusted as a whole using the RMA procedure in MATLAB.

Usage

```
data(IMR90)
```

Format

Two data frames. `imr90.data` has 22,223 probesets (rows) and 12 samples (columns). While `imr90.info` is a description of the samples, with two columns:

`Treatment` the treatment applied to the cells

`Batch` batch processing number

Value

promise

References

Johnson et al. Biostatistics (2007). doi: 10.1093/biostatistics/kxj037

Examples

```
data(IMR90)
```

Infinium5 *Infinium Methylation BeadChip batch correction reference data*

Description

Summary statistics after batch-effect correction per CpG probe across 5 Infinium Methylation datasets.

Usage

```
data(Infinium5)
```

Format

Four matrices each with 899255 rows and 5 columns. One row per CpG site across the 450K and EPIC designs and one column for each of the reference datasets. The matrices contain log variance ratio (LVR) statistics for ComBat and Harman and mean differences post-batch correction:

`lvr.combat` LVR for ComBat

`lvr.harman` LVR for Harman

`md.combat` mean differences for ComBat

`md.harman` mean differences for Harman

Value

promise

References

... awaiting publication

Examples

```
data(Infinium5)
```

NPM *Non-pregant mice study data*

Description

An experiment to test skin penetration of metal oxide nanoparticles following topical application of sunscreens. The data comprises three treatment groups plus a control group, with six replicates in each group, making a total of 24 Affymetrix MoGene 1.0 ST arrays. There were a total of three processing batches of eight arrays, each consisting of 2 replicates per group. Arrays were normalised and background adjusted as a whole using the RMA procedure in MATLAB.

Usage

```
data(NPM)
```

Format

Two data frames. `npm.data` has 35,512 probesets (rows) and 24 biological samples (columns). While `npm.info` is a description of the samples, with two columns:

`Treatment` the treatment applied to the mice

`Batch` batch processing number

Value

`promise`

References

Osmond-McLeod et al. Nanotoxicology (2014). doi: 10.3109/17435390.2013.855832

Examples

```
data(NPM)
```

OLF *Olfactory stem cell study data*

Description

An experiment to gauge the response of human olfactory neurosphere-derived (hONS) cells established from adult donors to ZnO nanoparticles. The data comprises six treatment groups plus a control group, each consisting of four replicates, giving a total number of 28 Affymetrix HuGene 1.0 ST arrays. The arrays were broken up into four processing batches of seven arrays each, consisting of one replicate from each of the groups. Arrays were normalised and background adjusted as a whole using the RMA procedure in MATLAB.

Usage

```
data(OLF)
```

Format

Two data frames. `olf.data` has 33,297 probesets (rows) and 28 biological samples (columns). While `olf.info` is a description of the samples, with two columns:

`Treatment` the treatment applied

`Batch` batch processing number

Value

`promise`

References

Osmond-McLeod et al. Part Fibre Toxicol. (2013). doi: 10.1186/1743-8977-10-54

Examples

```
data(OLF)
```

Index