Package 'normalize450K'

November 12, 2026

Type Package

Title Preprocessing of Illumina Infinium 450K data

Version 1.40.0

Date 2016-11-29

Author Jonathan Alexander Heiss

Maintainer Jonathan Alexander Heiss <jonathan.heiss@posteo.de>

Description

Precise measurements are important for epigenome-wide studies investigating DNA methylation in whole blood samples, where effect sizes are expected to be small in magnitude. The 450K platform is often affected by batch effects and proper preprocessing is recommended. This package provides functions to read and normalize 450K '.idat' files. The normalization corrects for dye bias and biases related to signal intensity and methylation of probes using local regression. No adjustment for probe type bias is performed to avoid the trade-off of precision for accuracy of beta-values.

Depends R (>= 3.3), Biobase, illuminaio, quadprog

Imports utils

License BSD_2_clause + file LICENSE

biocViews Normalization, DNAMethylation, Microarray, TwoChannel, Preprocessing, MethylationArray

NeedsCompilation no

git_url https://git.bioconductor.org/packages/normalize450K

git_branch RELEASE_3_23

git_last_commit 7d60aee

git_last_commit_date 2026-04-28

Repository Bioconductor 3.23

Date/Publication 2026-11-11

Contents

estimateLC 2
read_and_normalize450K 3

Index **5**

`estimateLC` *Estimation of Leukocyte composition for whole blood samples*

Description

Estimate leukocyte composition from whole blood DNA methylation

Usage

```
estimateLC(eSet)
```

Arguments

`eSet`	A Biobase eSet object as returned from a call of `normalize450K`

Details

Cell proportions are estimated using the algorithm developed by Houseman et al. (2012) by two different models. The first model was trained on a dataset of purified leukocytes (Reinius et al., 2012) and provides predictions for six cell types (granulocytes, monocytes, CD8+ T cells, CD4+ T cells, natural killer cells and CD19+ B cells), the second model was trained on whole blood samples from the LOLIPOP study as described by Heiss et al. (2016) and provides predictions for 4 cell types (neutrophils, eosinophils, lymphocytes, monocytes – ignore the prediction for basophils). Use this function only for normalized data (with `normalize450K(...,tissue='Blood')`.

Value

Returns the eSet object with cell proportions estimates added to the phenoData slot.

Author(s)

Jonathan A. Heiss

References

Houseman EA, et al. (2012) DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics, doi:10.1186/1471-2105-13-86

Reinius LE, et al. (2012) Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility. PloS ONE, doi:10.1371/journal.pone.0041361

Heiss JA, et al. (2016). Training a model for estimating leukocyte composition using whole-blood DNA methylation and cell counts as reference. Epigenomics, doi:10.2217/epi-2016-0091

read_and_normalize450K

Normalization of 450K data by LOESS method

Description

Read 450K '.idat' files and compute raw or normalized beta-values.

Usage

```
read450K(idat_files)
normalize450K(intensities,tissue='')
dont_normalize450K(intensities)
```

Arguments

idat_files	a character vector containing the paths to the .idat files stripped from the '_Grn.idat' suffix with one entry for each sample (.idat files for green and red intensities have to be in the same folder).
intensities	List object containing raw signal intensities. Result of calling read450K.
tissue	If set to 'Blood', a set of prespecified reference values are used for normalization. This is recommended if you plan to use estimateLC.

Details

Function read450K reads .idat files and returns a list object containing raw signal intensities. dont_normalize450K returns an ExpressionSet containing beta-values without normalization. normalize450K performs dye bias correction using the extension controls probes followed by normalization by local regression (Heiss and Brenner, 2015) and returns an ExpressionSet containing beta-values, too.

Value

For read450K a list containing the methylated, unmethylated and control signal intensities. For dont_normalize450K and normalize450K an ExpressionSet containing beta-values, rows corresponding to CpG sites (named) and columns to samples (in the same order as 'idat_files').

Note

A benchmark comparing the performance of this method with other normalization approaches is provided in the vignette.

Author(s)

Jonathan A. Heiss

References

Heiss JA, Brenner H (2015). Between-array normalization for 450K data. Frontiers in Genetics, doi:10.3389/fgene.2015.00092

Examples

```
## Not run:
 library(minfiData) ## this package includes some .idat files
 library(data.table)

 path <- system.file("extdata",package="minfiData")
 samples = fread(file.path(path, 'SampleSheet.csv'),integer64='character')

 samples[,file:=file.path(path,Sentrix_ID,paste0(Sentrix_ID,'_',Sentrix_Position))]
 ## samples$file is a character vector containing the location of the
 ## .idat files, but without the suffixes "_Red.idat" or "_Grn.idat"

 raw = read450K(samples$file)
 none = dont_normalize450K(raw) ## no normalization
 norm = normalize450K(raw)

## End(Not run)
```

Index

dont_normalize450K
(read_and_normalize450K), 3

estimateLC, 2

normalize450K (read_and_normalize450K),
3

read450K (read_and_normalize450K), 3
read_and_normalize450K, 3