Package 'FlowSorted.Blood.EPIC'

April 24, 2025

Type Package

Title Illumina EPIC data on immunomagnetic sorted peripheral adult blood cells

Version 2.12.0

Date 2021-10-26

Description Raw data objects to be used for blood cell proportion estimation in minfi and similar packages.
The FlowSorted.Blood.EPIC object is based in samples assayed by Brock Christensen and colleagues; for details see Salas et al. 2018.
https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110554.

License GPL-3

Depends R (>= 3.5), minfi(>= 1.21.2), ExperimentHub

Imports genefilter, nlme, quadprog, graphics, stats, utils, AnnotationHub, SummarizedExperiment, S4Vectors

Enhances FlowSorted.CordBloodCombined.450k(>= 1.11.5)

biocViews ExperimentData, Homo_sapiens_Data, Tissue, MicroarrayData, Genome, TissueMicroarrayData, MethylationArrayData

NeedsCompilation no

LazyData yes

LazyDataCompression gzip

Suggests knitr,rmarkdown, FlowSorted.CordBlood.450k, FlowSorted.Blood.450k(>= 1.0.1), EpiDISH,FlowSorted.CordBloodNorway.450k, testthat, IlluminaHumanMethylation450kmanifest(>= 0.2.0), IlluminaHumanMethylationEPICmanifest, IlluminaHumanMethylation450kanno.ilmn12.hg19(>= 0.2.1), IlluminaHumanMethylationEPICanno.ilm10b4.hg19

VignetteBuilder knitr

Encoding UTF-8

RoxygenNote 7.2.3

URL https://github.com/immunomethylomics/FlowSorted.Blood.EPIC

BugReports https://github.com/immunomethylomics/FlowSorted.Blood.EPIC/issues.

git_url https://git.bioconductor.org/packages/FlowSorted.Blood.EPIC

git_branch RELEASE_3_21

git_last_commit c6e7560

git_last_commit_date 2025-04-15

Repository Bioconductor 3.21

Date/Publication 2025-04-24

Author Lucas A. Salas [cre, aut],
Devin C. Koestler [aut],
Rondi A. Butler [ctb],
Helen M. Hansen [ctb],
John K. Wiencke [ctb],
Karl T. Kelsey [ctb],
Brock C. Christensen [ctb],
Kasper Daniel Hansen [ctb] (author and maintainer of the original minfi estimateCellCounts and internal functions),
Jean-Philippe Fortin [ctb] (contributor in the original minfi estimateCellCounts and internal functions),
Shan V. Andrews [ctb] (contributor in the original minfi estimateCellCounts and internal functions),
E. Andres Houseman [ctb] (author of the original quadratic programming function used for cell projection),
Andrew E Jaffe [ctb] (author of the original estimateCellCounts and internal functions within minfi)

Maintainer Lucas A. Salas <lucas.a.salas.diaz@dartmouth.edu>

Contents

estimateCellCounts2 . . . 3
FlowSorted.Blood.EPIC . . . 9
IDOLOptimizedCpGs . . . 10
IDOLOptimizedCpGs.compTable . . . 11
IDOLOptimizedCpGs450klegacy . . . 12
IDOLOptimizedCpGs450klegacy.compTable . . . 13
IDOLOptimizedCpGsCordBlood . . . 13
libraryDataGet . . . 14
projectCellType_CP . . . 15

Index **17**

estimateCellCounts2 *estimateCellCounts2*

Description

estimateCellCounts2 function allows the use of customized reference datasets and IDOL probes L-DMR lists

Usage

```
estimateCellCounts2(
  rgSet,
  compositeCellType = "Blood",
  processMethod = "preprocessNoob",
  probeSelect = "IDOL",
  cellTypes = c("CD8T", "CD4T", "NK", "Bcell", "Mono", "Neu"),
  referencePlatform = "IlluminaHumanMethylationEPIC",
  referenceset = NULL,
  CustomCpGs = NULL,
  returnAll = FALSE,
  meanPlot = FALSE,
  verbose = TRUE,
  lessThanOne = FALSE,
  cellcounts = NULL,
  ...
)
```

Arguments

rgSet	The input RGChannelSet or raw MethylSet for the procedure.
compositeCellType	Which composite cell type is being deconvolved. Should be one of "Blood", "CordBloodCombined", "CordBlood", "CordBloodNorway", "CordTissueAndBlood", or "DLPFC". See details for preferred approaches.
processMethod	Joint normalization/background correction for user and reference data Default input, "preprocessNoob" in minfi, you can use "auto" for preprocessQuantile for Blood and DLPFC in 450K datasets and preprocessNoob otherwise, according to existing literature. For MethylSet objects only "preprocessQuantile" is available. Set it to any minfi preprocessing function as a character if you want to override it, like "preprocessFunnorm"
probeSelect	How should probes be selected to distinguish cell types? Options include: 1) "IDOL", (default) for using a customized set of probes obtained from IDOL optimization, available for Blood and Umbilical Cord Blood 2) "both", which selects an equal number (50) of probes (with F-stat p-value < 1E-8) with the greatest magnitude of effect from the hyper- and hypo-methylated

sides, and 3) "any", which selects the 100 probes (with F-stat p-value < 1E-8) with the greatest magnitude of difference regardless of direction of effect.

This according to minfi algorithm. Default input "auto" in minfi will use "any" for cord blood and "both" otherwise. Please see references for more details.

`cellTypes` A vector of length K that contains the cell type names. Default: c("CD8T", "CD4T", "NK", "Bcell", "Mono", "Neu"). Please notice that this library use Neutrophils instead of Granulocytes. See details for your library.

`referencePlatform`
The platform for the reference dataset; options include c("IlluminaHumanMethylation450k", "IlluminaHumanMethylationEPIC" (default), "IlluminaHumanMethylation27k"). If the input rgSet belongs to another platform, it will be converted using minfi function convertArray.

`referenceset` It is NULL by default.

A custom reference RGChannelSet object (in quotes) if it is not a package installed. This option also allows the user to perform the deconvolution in closed computing clusters without internet access to ExperimentHub. For that download and save the reference and input the resulting object here. If using an installed reference package set to NULL.

`CustomCpGs` a custom vector of probe names for cell deconvolution. For custom lists it should be a vector object (no quotes).

`returnAll` Should the composition table and the normalized user supplied data be return? Default is False.

`meanPlot` Whether to plots the average DNA methylation across the cell-type discriminating probes within the mixed and sorted samples.

`verbose` Should the function be verbose?

`lessThanOne` Should the predictions be constrained to exactly one, in minfi default is FALSE, now you can select the option

`cellcounts` If cell counts are available (CBC, of flow sort) add a vector of lenght equal to the samples being deconvolved

`...` Other arguments for preprocessquantile or other normalizations

Value

This function will return a list containing matrix of cell counts (counts), if returnAll=FALSE, or a list containing the counts, mean methylation per cellType, and the normalized betas (if returnAll is set to TRUE). These objects are important if you decide to use a different deconvolution algorithm such as CIBERSORT or robust partial correlation (RPC).

References

LA Salas et al. (2018). *An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray*. Genome Biology 19, 64. doi: 10.1186/s13059-018-1448-7

LA Salas et al. (2022). *Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling*. Nat Comm 13, 761 (2022). doi: 10.1038/s41467-021-27864-7

DC Koestler et al. (2016). *Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL)*. BMC bioinformatics. 17, 120. doi: 10.1186/s13059-018-1448-7.

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

AE Jaffe and RA Irizarry.(2014) *Accounting for cellular heterogeneity is critical in epigenome-wide association studies*. Genome Biology 15:R31. doi: 10.1186/gb-2014-15-2-r31.

K Gervin, LA Salas et al. (2019) *Systematic evaluation and validation of references and library selection methods for deconvolution of cord blood DNA methylation data*. Clin Epigenetics 11,125. doi: 10.1186/s13148-019-0717-y.

KM Bakulski, et al. (2016) *DNA methylation of cord blood cell types: Applications for mixed cell birth studies*. Epigenetics 11:5. doi:10.1080/15592294.2016.1161875.

AJ Titus, et al. (2017). *Cell-type deconvolution from DNA methylation: a review of recent applications*. Hum Mol Genet 26: R216-R224. doi:10.1093/hmg/ddx275

AE Teschendorff, et al. (2017). *A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies*. BMC Bioinformatics 18: 105. doi: 10.1186/s12859-017-1511-5

Examples

```
# FlowSorted.Blood.EPIC
# Step 1: Load the reference library to extract the artificial mixtures
# Note: If your machine does not allow access to internet you can download
# and save the file. Then load it manually into the R environment
# library(FlowSorted.Blood.EPIC)
FlowSorted.Blood.EPIC <- libraryDataGet("FlowSorted.Blood.EPIC")

# Step 2 separate the reference from the testing dataset if you want to run
# examples for estimations for this function example
RGsetTargets <- FlowSorted.Blood.EPIC[,
    FlowSorted.Blood.EPIC$CellType == "MIX"
]

sampleNames(RGsetTargets) <- paste(RGsetTargets$CellType,
    seq_len(dim(RGsetTargets)[2]),
    sep = "_"
)
RGsetTargets
# Step 3: use your favorite package for deconvolution.
# Deconvolve a target data set consisting of EPIC DNA methylation
# data profiled in blood, using your prefered method.

# You can use our IDOL optimized DMR library for the EPIC array.  This object
# contains a vector of length 450 consisting of the IDs of the IDOL optimized
# CpG probes.  These CpGs are used as the backbone for deconvolution and were
# selected because their methylation signature differs across the six normal
# leukocyte subtypes. Use the option "IDOL"

head(IDOLOptimizedCpGs)
# If you need to deconvolve a 450k legacy dataset use
```

```
# IDOLOptimizedCpGs450klegacy instead

# We recommend using Noob processMethod = "preprocessNoob" in minfi for the
# target and reference datasets.
# Cell types included are "CD8T", "CD4T", "NK", "Bcell", "Mono", "Neu"

# To use the IDOL optimized list of CpGs (IDOLOptimizedCpGs) use
# estimateCellCounts2 an adaptation of the popular estimateCellCounts in
# minfi. This function also allows including customized reference arrays.

# Do not run with limited RAM the normalization step requires a big amount
# of memory resources

    propEPIC <- estimateCellCounts2(RGsetTargets,
        compositeCellType = "Blood",
        processMethod = "preprocessNoob",
        probeSelect = "IDOL",
        cellTypes = c(
            "CD8T", "CD4T", "NK", "Bcell",
            "Mono", "Neu"
        )
    )

    head(propEPIC$prop)
    percEPIC <- round(propEPIC$prop * 100, 1)

# #Advanced deconvolution CP/QP, CIBERSORT and/or RPC deconvolution
# noobset<- preprocessNoob(RGsetTargets)
#  propEPIC<-projectCellType_CP (
#  getBeta(noobset)[IDOLOptimizedCpGs,],
#  IDOLOptimizedCpGs.compTable, contrastWBC=NULL, nonnegative=TRUE,
#  lessThanOne=FALSE)
#
# head(propEPIC)
# percEPIC<-round(propEPIC*100,1)
#
# #If you prefer CIBERSORT or RPC deconvolution use EpiDISH or similar
# #Example not to run
# library(EpiDISH)
# RPC <- epidish(getBeta(noobset)[IDOLOptimizedCpGs,],
# IDOLOptimizedCpGs.compTable), method = "RPC")
# RPC$estF#RPC proportion estimates
# percEPICRPC<-round(RPC$estF*100,1)#percentages

# CBS <- epidish(getBeta(noobset)[IDOLOptimizedCpGs,],
# IDOLOptimizedCpGs.compTable), method = "CBS")
# CBS$estF#CBS proportion estimates
# percEPICCBS<-round(CBS$estF*100,1)#percentages

# UMBILICAL CORD BLOOD DECONVOLUTION


#library(FlowSorted.CordBloodCombined.450k)
```

```
## Step 1: Load the reference library to extract the umbilical cord samples
#FlowSorted.CordBloodCombined.450k <-
#    libraryDataGet("FlowSorted.CordBloodCombined.450k")
#FlowSorted.CordBloodCombined.450k

## Step 2 separate the reference from the testing dataset if you want to run
## examples for estimations for this function example

#RGsetTargets <- FlowSorted.CordBloodCombined.450k[,
#    FlowSorted.CordBloodCombined.450k$CellType == "WBC"
#]
#sampleNames(RGsetTargets) <- paste(RGsetTargets$CellType,
#    seq_len(dim(RGsetTargets)[2]),
#    sep = "_"
#)
#RGsetTargets

## Step 3: use your favorite package for deconvolution.
## Deconvolute a target data set consisting of 450K DNA methylation
## data profiled in blood, using your prefered method.
## You can use our IDOL optimized DMR library for the Cord Blood,  This object
## contains a vector of length 517 consisting of the IDs of the IDOL optimized
## CpG probes.  These CpGs are used as the backbone for deconvolution and were
## selected because their methylation signature differs across the six normal
## leukocyte subtypes plus the nucleated red blood cells.
#data("IDOLOptimizedCpGsCordBlood")
#head(IDOLOptimizedCpGsCordBlood)
## We recommend using Noob processMethod = "preprocessNoob" in minfi for the
## target and reference datasets.
## Cell types included are "CD8T", "CD4T", "NK", "Bcell", "Mono", "Gran",
## "nRBC"
## To use the IDOL optimized list of CpGs (IDOLOptimizedCpGsCordBlood) use
## estimateCellCounts2 from FlowSorted.Blood.EPIC.
## Do not run with limited RAM the normalization step requires a big amount
## of memory resources. Use the parameters as specified below for
## reproducibility.
#
#    propUCB <- estimateCellCounts2(RGsetTargets,
#        compositeCellType = "CordBloodCombined",
#        processMethod = "preprocessNoob",
#        probeSelect = "IDOL",
#        cellTypes = c(
#            "CD8T", "CD4T", "NK",
#            "Bcell", "Mono", "Gran", "nRBC"
#        )
#    )
#
#    head(propUCB$prop)
#    percUCB <- round(propUCB$prop * 100, 1)

## Using cell counts instead of proportions
#library(FlowSorted.Blood.450k)
#RGsetTargets2 <- FlowSorted.Blood.450k[,
```

```
#    FlowSorted.Blood.450k$CellType == "WBC"
#]
#sampleNames(RGsetTargets2) <- paste(RGsetTargets2$CellType,
#    seq_len(dim(RGsetTargets2)[2]),
#    sep = "_"
#)
#RGsetTargets2
#propEPIC2 <- estimateCellCounts2(RGsetTargets2,
#    compositeCellType = "Blood",
#    processMethod = "preprocessNoob",
#    probeSelect = "IDOL",
#    cellTypes = c(
#        "CD8T", "CD4T", "NK", "Bcell",
#        "Mono", "Neu"
#    ), cellcounts = rep(10000, 6)
#)
#head(propEPIC2$prop)
#head(propEPIC2$counts)
#percEPIC2 <- round(propEPIC2$prop * 100, 1)

## Blood Extended deconvolution or any external reference
## please contact \email{Technology.Transfer@dartmouth.edu}

# library (FlowSorted.BloodExtended.EPIC)

## Step 1: Extract the mix samples
# library(FlowSorted.Blood.EPIC)
# library(ExperimentHub)
# FlowSorted.Blood.EPIC <- libraryDataGet('FlowSorted.Blood.EPIC')
## Step 2 separate the reference from the testing dataset if you want to run
## examples for estimations for this function example

# RGsetTargets <- FlowSorted.Blood.EPIC[,
# FlowSorted.Blood.EPIC$CellType == "MIX"]
# sampleNames(RGsetTargets) <- paste(RGsetTargets$CellType,
#                                seq_len(dim(RGsetTargets)[2]), sep = "_")
# RGsetTargets

## Step 3: use your favorite package for deconvolution.
## Deconvolve the target data set 450K or EPIC blood DNA methylation.
## We recommend ONLY the IDOL method, the automatic method can lead to severe
## biases.

## We recommend using Noob processMethod = "preprocessNoob" in minfi for the
## target and reference datasets.
## Cell types included are "Bas", "Bmem", "Bnv", "CD4mem", "CD4nv",
## "CD8mem", "CD8nv", "Eos", "Mono", "Neu", "NK", and "Treg"
## Use estimateCellCounts2 from FlowSorted.Blood.EPIC.
## Do not run with limited RAM the normalization step requires a big amount
## of memory resources. Use the parameters as specified below for
## reproducibility.
#
#    prop_ext <- estimateCellCounts2(RGsetTargets,
```

```
#        compositeCellType = "BloodExtended",
#        processMethod = "preprocessNoob",
#        probeSelect = "IDOL",
#        cellTypes = c(
#            "Bas", "Bmem", "Bnv",
#            "CD4mem", "CD4nv",
#            "CD8mem", "CD8nv", "Eos",
#            "Mono", "Neu", "NK", "Treg"),
#CustomCpGs =if(RGsetTargets@annotation[1]=="IlluminaHumanMethylationEPIC"){
#IDOLOptimizedCpGsBloodExtended}else{IDOLOptimizedCpGsBloodExtended450k})
#    perc_ext<-round(prop_ext$prop*100,1)
#    head(perc_ext)
#
## End of example
```

`FlowSorted.Blood.EPIC` *FlowSorted.Blood.EPIC*

Description

Illumina Human Methylation data from EPIC on immunomagnetic sorted adult blood cell populations. The FlowSorted.Blood.EPIC package contains Illumina HumanMethylationEPIC ("EPIC")) DNA methylation microarray data from the immunomethylomics group (Salas et al. 2018), consisting of 37 magnetic sorted blood cell references and 12 samples, formatted as an RGChannelSet object for integration and normalization using most of the existing Bioconductor packages.

This package contains data similar to the FlowSorted.Blood.450k package consisting of data from peripheral blood samples generated from adult men and women. However, when using the newer EPIC microarray minfi estimates of cell type composition using the FlowSorted.Blood.450k package are less precise compared to actual cell counts. Hence, this package consists of appropriate data for deconvolution of adult blood samples used in for example EWAS relying in the newer EPIC technology.

Researchers may find this package useful as these samples represent different cellular populations (T lymphocytes (CD4+ and CD8+), B cells (CD19+), monocytes (CD14+), NK cells (CD56+) and Neutrophils of cell sorted blood generated with high purity estimates. As a test of accuracy 12 experimental mixtures were reconstructed using fixed amounts of DNA from purified cells. We offer the function estimateCellCounts2 a modification of the popular estimatesCellCounts function in minfi. This function allows estimating cellular composition in users' whole blood Illumina EPIC samples using a modified version of the algorithm constrained projection/quadratic programming described in Houseman et al. 2012. For a slightly more accurate estimations we also offered an IDOL optimized CpG selection for cell deconvolution as the object IDOLOptimizedCpGs, and the IDOLOptimizedCpGs450klegacy object for legacy 450K datasets. See the objects help for details.

Usage

```
FlowSorted.Blood.EPIC
#See ?estimateCellCounts2 for cell deconvolution guidelines
```

Format

A class: RGChannelSet, dimensions: 1051815 49

Value

RGChannelSet 49 samples

Source

The FlowSorted.Blood.EPIC object is based in samples assayed by Brock Christensen and colleagues; Salas et al. 2018. GSE110554

See Also

References

1. LA Salas et al. (2018). *An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray*. Genome Biology 19, 64. doi: 10.1186/s13059-018-1448-7.
2. DC Koestler et al. (2016). *Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL)*. BMC bioinformatics. 17, 120. doi:10.1186/s12859-016-0943-7.
3. EA Houseman et al. (2012) *DNA methylation arrays as surrogate measures of cell mixture distribution*. BMC Bioinformatics 13, 86. doi:10.1186/1471-2105-13-86.
4. **minfi** package, tools for analyzing DNA methylation microarrays

Examples

```
# FlowSorted.Blood.EPIC<-
# libraryDataGet('FlowSorted.Blood.EPIC')
# FlowSorted.Blood.EPIC
# table(FlowSorted.Blood.EPIC$CellType)
```

`IDOLOptimizedCpGs` *IDOL Optimized CpGs for adult blood DNA methylation deconvolution EPIC*

Description

This object is a vector of length 450 consisting of the names of the IDOL optimized CpGs. These CpGs are used as the backbone for deconvolution and were selected because their methylation signature differs across the six normal leukocyte subtypes.

Usage

```
IDOLOptimizedCpGs
```

Format

An object of class "character" of length 450.

The format is: chr [1:450] "cg08769189" "cg07661835" "cg00219921" "cg13468685" ...

References

LA Salas et al. (2018). *An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray*. Genome Biology 19, 64. doi: 10.1186/s13059-018-1448-7

DC Koestler et al. (2016). *Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL)*. BMC bioinformatics. 17, 120. doi: 10.1186/s13059-018-1448-7.

Examples

```
# Do not run
# head(IDOLOptimizedCpGs)
```

IDOLOptimizedCpGs.compTable

IDOL Optimized CpGs matrix for adult blood DNA methylation deconvolution EPIC

Description

This object is a matrix of dimensions 450 x 6 consisting of the average DNA methylation values fo the probes included in the IDOL optimized CpGs per each of the six cell types available. These CpGs are used as the backbone for deconvolution and were selected because their methylation signature differs across the six normal leukocyte subtypes.

Usage

```
IDOLOptimizedCpGs.compTable
```

Format

An object of class "matrix" of dimensions 450 x 6.

The format is: num [1:450, 1:6] 0.197 0.105 0.135 0.654 0.246 ...

References

LA Salas et al. (2018). *An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray*. Genome Biology 19, 64. doi: 10.1186/s13059-018-1448-7

DC Koestler et al. (2016). *Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL)*. BMC bioinformatics. 17, 120. doi: 10.1186/s13059-018-1448-7.

Examples

```
# data ("IDOLOptimizedCpGs.compTable")
# head(IDOLOptimizedCpGs.compTable)
# See ?estimateCellCounts2 for deconvolution examples (advanced user option)
```

`IDOLOptimizedCpGs450klegacy` *IDOL Optimized CpGs for adult blood DNA methylation deconvolution EPIC*

Description

This object is a vector of length 350 consisting of the names of the IDOL optimized CpGs to use in the older 450K platform. These CpGs are used as the backbone for deconvolution and were selected because their methylation signature differs across the six normal leukocyte subtypes.

Usage

```
IDOLOptimizedCpGs450klegacy
```

Format

An object of class "character" of length 350.

The format is: chr [1:350] "cg14232368" "cg15087459" "cg20538211" "cg11944101" ...

References

LA Salas et al. (2018). *An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray*. Genome Biology 19, 64. doi: 10.1186/s13059-018-1448-7

DC Koestler et al. (2016). *Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL)*. BMC bioinformatics. 17, 120. doi: 10.1186/s13059-018-1448-7.

Examples

```
# data ("IDOLOptimizedCpGs450klegacy")
# head(IDOLOptimizedCpGs450klegacy)
# See ?estimateCellCounts2 for deconvolution examples
```

IDOLOptimizedCpGs450klegacy.compTable

IDOL Optimized CpGs matrix for adult blood DNA methylation deconvolution EPIC

Description

This object is a matrix of dimensions 350 x 6 consisting of the average DNA methylation values fo the probes included in the IDOL optimized CpGs per each of the six cell types available to use in the older 450K platform. These CpGs are used as the backbone for deconvolution and were selected because their methylation signature differs across the six normal leukocyte subtypes.

Usage

```
IDOLOptimizedCpGs450klegacy.compTable
```

Format

An object of class "matrix" of dimensions 450 x 6.

The format is: num [1:350, 1:6] 0.904 0.122 0.633 0.841 0.135 ...

References

LA Salas et al. (2018). *An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray*. Genome Biology 19, 64. doi: 10.1186/s13059-018-1448-7

DC Koestler et al. (2016). *Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL)*. BMC bioinformatics. 17, 120. doi: 10.1186/s13059-018-1448-7.

Examples

```
# data("IDOLOptimizedCpGs450klegacy.compTable")
# head(IDOLOptimizedCpGs450klegacy.compTable)
# See ?estimateCellCounts2 for deconvolution examples (advanced user)
```

IDOLOptimizedCpGsCordBlood

IDOL Optimized CpGs for umbilical cord blood DNA methylation deconvolution

Description

This object is a vector of length 517 consisting of the names of the IDOL optimized CpGs. These CpGs are used as the backbone for deconvolution and were selected because their methylation signature differs across the six normal leukocyte subtypes and the nucleated red blood cells.

Usage

```
#data ("IDOLOptimizedCpGsCordBlood")
#head(IDOLOptimizedCpGsCordBlood)
#See ?estimateCellCounts2 for deconvolution examples
```

Format

An object of class "character" of length 517.

The format is: chr [1:517] "cg12603453" "cg24765783" "cg06975018" "cg19708055" ...

References

K Gervin, LA Salas et al. (2019) *Systematic evaluation and validation of references and library selection methods for deconvolution of cord blood DNA methylation data*. Clin Epigenetics 11,125. doi: 10.1186/s13148-019-0717-y

LA Salas et al. (2018). *An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray*. Genome Biology 19, 64. doi: 10.1186/s13059-018-1448-7.

DC Koestler et al. (2016). *Improving cell mixture deconvolution by identifying optimal DNA methylation libraries (IDOL)*. BMC bioinformatics. 17, 120. doi: 10.1186/s12859-016-0943-7.

Examples

```
#data ("IDOLOptimizedCpGsCordBlood")
#head(IDOLOptimizedCpGsCordBlood)
#See ?estimateCellCounts2 for deconvolution examples
```

---

libraryDataGet *libraryDataGet*

Description

Function to load the library data from ExperimentHub

Usage

```
libraryDataGet(title)
```

Arguments

title	title of the data, e.g., 'FlowSorted.Blood.EPIC'

Value

The function will look for the dataset in ExperimentHub and load the object

This function will return an object matching the title of the ExperimenHub

Examples

```
FlowSorted.Blood.EPIC <-
    libraryDataGet("FlowSorted.Blood.EPIC")
FlowSorted.Blood.EPIC
```

`projectCellType_CP` *projectCellType_CP*

Description

This function predicts the underlying cellular composition of heterogeneous tissue types (i.e., WB) using the constrained projection procedure described by Houseman et al. 2012. This is equivalent to the internal projectCellType function in minfi, Jaffe et al. 2014. We recommend this function only for advanced users. Please preprocess your dataset filtering potential bad quality samples.

Usage

```
projectCellType_CP(
  Y,
  coefWBC,
  contrastWBC = NULL,
  nonnegative = TRUE,
  lessThanOne = FALSE
)
```

Arguments

`Y`	A J x N matrix of methylation beta-values collected from mixed/ heterogeneous biospecimen (i.e., Whole Blood). Target set.
`coefWBC`	A J x K projection matrix;, i.e., within-cell type mean methylation matrix across J L-DMRs and K number of cell types
`contrastWBC`	Contrast for cell composition predictions set to NULL by default. The user should not modify this
`nonnegative`	Should cell predictions be nonnegative. Defaults to TRUE
`lessThanOne`	Should the predictions sum less than one. Default is FALSE

Value

A N x K matrix of cell proportion estimates across the K cell types for each of the N subjects contained in the Target Set.

Examples

```
# Step 1: Load the reference library to extract the artificial mixtures
FlowSorted.Blood.EPIC <- libraryDataGet("FlowSorted.Blood.EPIC")
FlowSorted.Blood.EPIC

# Step 2 separate the reference from the testing dataset if you want to run
# examples for estimations for this function example

RGsetTargets <- FlowSorted.Blood.EPIC[
    ,
    FlowSorted.Blood.EPIC$CellType == "MIX"
]

sampleNames(RGsetTargets) <- paste(RGsetTargets$CellType,
    seq_len(dim(RGsetTargets)[2]),
    sep = "_"
)
RGsetTargets

# Step 3: apply the CP approach using the preloaded matrix of IDOL.
# Deconvolute a target data set consisting of EPIC DNA methylation
# data profiled in blood, using your prefered method.

# You can use our IDOL optimized DMR library for the EPIC array.  This object
# contains a matrix of dimensions 450 X 6 consisting on average methylation
# values obtained from the IDs of the IDOL optimized CpG probes.  These
# CpGs are used as the backbone for deconvolution and were selected because
# their methylation signature differs across the six normal leukocyte
# subtypes.

head(IDOLOptimizedCpGs.compTable)
# If you need to deconvolute a 450k legacy dataset use
# IDOLOptimizedCpGs450klegacy.compTable instead

# We recommend using Noob processMethod = "preprocessNoob" in minfi for the
# target and reference datasets.
# Cell types included are "CD8T", "CD4T", "NK", "Bcell", "Mono", "Neu"

# If you need to normalize your data do not run with limited RAM. The
# normalization step requires a big amount of memory resources

if (memory.limit() > 8000) {
    propEPIC <- projectCellType_CP(
        getBeta(preprocessNoob(RGsetTargets))[IDOLOptimizedCpGs, ],
        IDOLOptimizedCpGs.compTable,
        contrastWBC = NULL, nonnegative = TRUE,
        lessThanOne = FALSE
    )

    head(propEPIC)
    percEPIC <- round(propEPIC * 100, 1)
}
```

Index

∗ **datasets**
FlowSorted.Blood.EPIC, 9
IDOLOptimizedCpGs, 10
IDOLOptimizedCpGs.compTable, 11
IDOLOptimizedCpGs450klegacy, 12
IDOLOptimizedCpGs450klegacy.compTable, 13
IDOLOptimizedCpGsCordBlood, 13

estimateCellCounts2, 3

FlowSorted.Blood.EPIC, 9

IDOLOptimizedCpGs, 10
IDOLOptimizedCpGs.compTable, 11
IDOLOptimizedCpGs450klegacy, 12
IDOLOptimizedCpGs450klegacy.compTable, 13
IDOLOptimizedCpGsCordBlood, 13

libraryDataGet, 14

projectCellType_CP, 15