

Package ‘FlowSorted.Blood.EPIC’

November 19, 2019

Type Package

Title Illumina EPIC data on immunomagnetic sorted peripheral adult blood cells

Version 1.4.1

Date 2019-10-24

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), SummarizedExperiment, genefilter, quadprog, nlme, S4Vectors, graphics, stats, utils, IlluminaHumanMethylationEPICanno.ilm10b4.hg19, ExperimentHub

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

NeedsCompilation no

LazyData yes

Suggests knitr, rmarkdown, EpiDISH, FlowSorted.CordBloodCombined.450k, FlowSorted.Blood.450k(>= 1.0.1), FlowSorted.CordBlood.450k, FlowSorted.CordBloodNorway.450k, testthat, IlluminaHumanMethylation450kmanifest(>= 0.2.0), IlluminaHumanMethylation450kanno.ilmn12.hg19(>= 0.2.1), IlluminaHumanMethylationEPICanno.ilm10b2.hg19

VignetteBuilder knitr

RoxygenNote 6.1.1

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_10

git_last_commit 64be62a

git_last_commit_date 2019-11-06

Date/Publication 2019-11-19

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>

R topics documented:

estimateCellCounts2	2
FlowSorted.Blood.EPIC	7
IDOLOptimizedCpGs	8
IDOLOptimizedCpGs450klegacy	9

Index **10**

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 = c("auto", "any",
  "IDOL"), cellTypes = c("CD8T", "CD4T", "NK", "Bcell", "Mono", "Neu"),
  referencePlatform = c("IlluminaHumanMethylation450k",
  "IlluminaHumanMethylationEPIC", "IlluminaHumanMethylation27k"),
  referenceset = NULL, IDOLOptimizedCpGs = NULL, returnAll = FALSE,
  meanPlot = FALSE, verbose = TRUE, lessThanOne = FALSE, ...)
```

Arguments

rgSet The input RGChannelSet or raw MethylSet for the procedure.

compositeCellType	Which composite cell type is being deconvoluted. Should be one of "Blood", "CordBloodCombined", "CordBlood", "CordBloodNorway", "CordTissueAnd-Blood", 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 "preprocess-Funnorm"
probeSelect	How should probes be selected to distinguish cell types? Options include: 1) "IDOL", 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; 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.
IDOLOptimizedCpGs	a vector of probe names for cell deconvolution. For EPIC datasets it should be equal to IDOLOptimizedCpGs (no quotes). For legacy 450K datasets you can use IDOLOptimizedCpGs450klegacy (no quotes) if you want to use Flow-Sorted.EPIC.Blood for the deconvolution. For umbilical cord blood use IDOLOptimizedCpGsCordBlood in the FlowSorted.CordBloodCombined.450k package for both 450k and EPIC arrays. See details in help for each IDOL list.
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
...	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.
- 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.
- 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.
- 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.

Examples

```
# Step 1: Load the reference library to extract the artificial mixtures

library(ExperimentHub)
hub <- ExperimentHub()
query(hub, "FlowSorted.Blood.EPIC")
FlowSorted.Blood.EPIC <- hub[["EH1136"]]
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.
# Deconvolute a target data set consisting of EPIC DNA methylation
# data profiled in blood, using your preferred 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.

data (IDOLOptimizedCpGs)
head (IDOLOptimizedCpGs)
# If you need to deconvolute 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

if (memory.limit()>8000){
  countsEPIC<-estimateCellCounts2(RGsetTargets, compositeCellType = "Blood",
    processMethod = "preprocessNoob",
    probeSelect = "IDOL",
    cellTypes = c("CD8T", "CD4T", "NK", "Bcell",
"Mono", "Neu"),
    referencePlatform =
"IlluminaHumanMethylationEPIC",
    referenceset = NULL,
    IDOLOptimizedCpGs =IDOLOptimizedCpGs,
    returnAll = FALSE)

  head(countsEPIC$counts)
}

# CIBERSORT and/or RPC deconvolution
# If you prefer CIBERSORT or RPC deconvolution use EpiDISH or similar

# Example not to run

# countsEPIC<-estimateCellCounts2(RGsetTargets, compositeCellType = "Blood",
#
#                               processMethod = "preprocessNoob",
#                               probeSelect = "IDOL",
#                               cellTypes = c("CD8T", "CD4T", "NK",
#                               "Bcell", "Mono", "Neu"),
#                               referencePlatform =
#                               "IlluminaHumanMethylationEPIC",
#                               referenceset = NULL,
#                               IDOLOptimizedCpGs =IDOLOptimizedCpGs,
#                               returnAll = TRUE)

# library(EpiDISH)
# RPC <- epidish(getBeta(countsEPIC2$normalizedData),
# # as.matrix(countsEPIC2$compTable[IDOLOptimizedCpGs, 3:8]), method = "RPC")
# RPC$estF#RPC count estimates

```

```

# CBS <- epidish(getBeta(countsEPIC2$normalizedData),
# as.matrix(countsEPIC2$compTable[IDOLOptimizedCpGs, 3:8]), method = "CBS")
# CBS$estF#CBS count estimates

# UMBILICAL CORD BLOOD DECONVOLUTION

# Do not run
# library (FlowSorted.Blood.EPIC)
# data("IDOLOptimizedCpGsCordBlood")
# Step 1: Load the reference library to extract the umbilical cord samples
# library(ExperimentHub)
# hub <- ExperimentHub()
# myfiles <- query(hub, "FlowSorted.CordBloodCombined.450k")
# FlowSorted.CordBloodCombined.450k <- myfiles[[1]]
# 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 preferred 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.
#
# if (memory.limit())>8000){
#   countsUCB<-estimateCellCounts2(RGsetTargets,
#                                   compositeCellType =
#                                     "CordBloodCombined",
#                                   processMethod = "preprocessNoob",
#                                   probeSelect = "IDOL",
#                                   cellTypes = c("CD8T", "CD4T", "NK",
#                                                 "Bcell", "Mono", "Gran", "nRBC"),
#                                   referencePlatform =
#                                     "IlluminaHumanMethylation450k",
#                                   referenceset =

```

```

#                               "FlowSorted.CordBloodCombined.450k",
#                               IDOLOptimizedCpGs =
#                               IDOLOptimizedCpGsCordBlood,
#                               returnAll = FALSE)
#
#   head(countsUCB$counts)
# }

```

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 (manuscript submitted), 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
```

Format

A class: RGChannelSet, dimensions: 1051815 49

Source

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110554> The FlowSorted.Blood.EPIC object is based in samples assayed by Brock Christensen and colleagues; manuscript submitted.

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 for tools for estimating cell type composition in blood using these data

Examples

```
# Explore the reference library
library(ExperimentHub)
hub <- ExperimentHub()
query(hub, "FlowSorted.Blood.EPIC")
FlowSorted.Blood.EPIC <- hub[["EH1136"]]
FlowSorted.Blood.EPIC
```

IDOLOptimizedCpGs	<i>IDOL Optimized CpGs for adult blood DNA methylation deconvolution EPIC</i>
-------------------	---

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/s12859-016-0943-7.

Examples

```
# Do not run
# data("IDOLOptimizedCpGs")
```

```
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.

D 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

```
# Do not run
# data("IDOLOptimizedCpGs450klegacy")
```

Index

*Topic **datasets**

FlowSorted.Blood.EPIC, [7](#)

IDLOptimizedCpGs, [8](#)

IDLOptimizedCpGs450klegacy, [9](#)

estimateCellCounts2, [2](#)

FlowSorted.Blood.EPIC, [7](#)

IDLOptimizedCpGs, [8](#)

IDLOptimizedCpGs450klegacy, [9](#)