

Package 'FlowSorted.CordBloodCombined.450k'

October 17, 2019

Type Package

Title Illumina 450k/EPIC data on FACS and MACS umbilical blood cells

Version 1.0.0

Date 2019-01-18

Description Raw data objects to be used for umbilical cord blood cell proportion estimation in minfi and similar packages.
The FlowSorted.CordBloodCombined.450k object is based in samples assayed by Bakulski et al, Gervin et al., de Goede et al., and Lin et al.

License GPL-3

Depends R (>= 3.6), minfi (>= 1.21.2), SummarizedExperiment, IlluminaHumanMethylation450kanno.ilmn12.hg19(>= 0.2.1), IlluminaHumanMethylationEPICanno.ilm10b4.hg19, ExperimentHub(>= 1.9.1), FlowSorted.Blood.EPIC, utils

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

NeedsCompilation no

LazyData no

Suggests knitr, rmarkdown, testthat, IlluminaHumanMethylation450kmanifest(>= 0.2.0), IlluminaHumanMethylationEPICanno.ilm10b2.hg19

VignetteBuilder knitr

RoxygenNote 6.1.1

URL <https://github.com/immunomethylomics/FlowSorted.CordBloodCombined.450k>

BugReports <https://github.com/immunomethylomics/FlowSorted.CordBloodCombined.450k/issues>.

git_url <https://git.bioconductor.org/packages/FlowSorted.CordBloodCombined.450k>

git_branch RELEASE_3_9

git_last_commit 7c4280e

git_last_commit_date 2019-05-02

Date/Publication 2019-10-17

Author Lucas A. Salas [cre, aut],
 Kristina Gervin [aut],
 Meaghan C. Jones [aut],
 Kelly M. Bakulski [ctb],
 Devin C. Koestler [ctb],
 John K. Wiencke [ctb],
 Karl T. Kelsey [ctb],
 Robert Lyle [ctb],
 Brock C. Christensen [ctb],
 Janine Felix [ctb]

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

R topics documented:

FlowSorted.CordBloodCombined.450k	2
FlowSorted.CordBloodCombined.450k.compTable	4
IDOLOptimizedCpGsCordBlood	5

Index	7
--------------	----------

FlowSorted.CordBloodCombined.450k

FlowSorted.CordBloodCombined.450k

Description

The FlowSorted.CordBloodCombined.450k package contains data derived from Illumina HumanMethylation450K and Illumina HumanMethylationEPIC DNA methylation microarrays (Gervin K, Salas LA et al. under review), consisting of 263 blood cell references and 26 umbilical cord blood samples, formatted as an RGChannelSet object for integration and normalization using most of the existing Bioconductor packages.

This package contains cleaned data from four different umbilical cord blood references similar to the FlowSorted.CordBlood.450K package consisting of data from umbilical cord blood samples generated from healthy newborns. However, when using the cleaned dataset (eliminating potential cell cross contamination) and using the IDOL procedure compared to minfi estimates the cell type composition obtained through FlowSorted.CordBlood.450k package were less precise and biased compared to actual cell counts. Hence, this package consists of appropriate data for deconvolution of umbilical cord blood samples used in for example EWAS relying in both 450K and 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+), Granulocytes, and nucleated red blood cells of cell sorted umbilicalcord blood. The estimates were contrasted versus FACS proportions in 22 umbilical samples, and validated using 197 umbilical cord blood samples.

These data can be integrated with the minfi Bioconductor package to estimate cellular composition in users' umbilical cord blood Illumina 450K and EPIC samples using a modified version of the algorithm constrained projection/quadratic programming described in Houseman et al. 2012. However, for more accurate estimations we suggests that the user prefers IDOL over minfi automatic estimations, using the function estimateCellCounts2 from the package FlowSorted.Blood.EPIC which allows using customized sets of probes from IDOL (see IDOLOptimizedCpGsCordBlood for an example).

Usage

```
FlowSorted.CordBloodCombined.450k
```

Format

A class: RGChannelSet, dimensions: 575130 289

Value

RGChannelSet 289 samples

See Also

References

1. 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.* (Under review)
2. 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.
3. K Gervin, et al. (2016) *Cell type specific DNA methylation in cord blood: A 450K-reference data set and cell count-based validation of estimated cell type composition.* Epigenetics 11:690-8. doi:10.1080/15592294.2016.1214782.
4. OM de Goede, et al. (2015) *Nucleated red blood cells impact DNA methylation and expression analyses of cord blood hematopoietic cells.* Clin Epigenetics. 7:95. doi:10.1186/s13148-015-0129-6.
5. X Lin, et al. (2018) *Cell type-specific DNA methylation in neonatal cord tissue and cord blood: A 850K-reference panel and comparison of cell-types.* Epigenetics. 13:941-58. doi:10.1080/15592294.2018.152297.
6. 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.
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.
8. 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.
9. **minfi** package for tools for estimating cell type composition in blood using these data

Examples

```
library(ExperimentHub)
hub <- ExperimentHub()
myfiles <- query(hub, "FlowSorted.CordBloodCombined.450k")
myfiles$ah_id=="EH2256"#ExperimenHub ID
# Do not run unless you have enough memory resources
# FlowSorted.CordBloodCombined.450k <- myfiles[[1]]
# FlowSorted.CordBloodCombined.450k
```

FlowSorted.CordBloodCombined.450k.compTable

FlowSorted.CordBloodCombined.450k.compTable

Description

The FlowSorted.CordBloodCombined.450k.compTable contains the average DNA methylation values used for IDOL deconvolution (Gervin K, Salas LA et al. under review), these data are derived from 263 umbilical blood cell references available in ExperimentHub (FlowSorted.CordBloodCombined.450k).

Researchers who want to project directly their estimates can use this matrix of different cellular populations (T lymphocytes (CD4+ and CD8+), B cells (CD19+), monocytes (CD14+), NK cells (CD56+), Granulocytes, and nucleated red blood cells of cell sorted umbilical cord blood. The estimates were contrasted versus FACS proportions in 22 umbilical samples, and validated using 197 umbilical cord blood samples.

Usage

FlowSorted.CordBloodCombined.450k.compTable

Format

A class: matrix, dimensions: 517 7 The format is: num [1:517, 1:7] 0.0568 0.214 0.908 0.839 ...

Value

numeric matrix 517 rows 7 columns

See Also

References

1. 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.* (Under review)
2. 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.
3. K Gervin, et al. (2016) *Cell type specific DNA methylation in cord blood: A 450K-reference data set and cell count-based validation of estimated cell type composition.* Epigenetics 11:690-8. doi:10.1080/15592294.2016.1214782.
4. OM de Goede, et al. (2015) *Nucleated red blood cells impact DNA methylation and expression analyses of cord blood hematopoietic cells.* Clin Epigenetics. 7:95. doi:10.1186/s13148-015-0129-6.
5. X Lin, et al. (2018) *Cell type-specific DNA methylation in neonatal cord tissue and cord blood: A 850K-reference panel and comparison of cell-types.* Epigenetics. 13:941-58. doi:10.1080/15592294.2018.152292.
6. 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.
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.

8. 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.
9. **minfi** package for tools for estimating cell type composition in blood using these data

Examples

```
# Explore the reference library
```

```
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

```
IDOLOptimizedCpGsCordBlood
```

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*. (Under review).
- 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
# 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){
#   countsEPIC<-estimateCellCounts2(RGsetTargets,
#                                   compositeCellType = "Blood",
#                                   processMethod = "preprocessNoob",
#                                   probeSelect = "IDOL",
#                                   cellTypes = c("CD8T", "CD4T", "NK",
#                                                "Bcell", "Mono", "Gran", "nRBC"),
#                                   referencePlatform =
#                                     "IlluminaHumanMethylationEPIC",
#                                   referenceset =
#                                     "FlowSorted.CordBloodCombined.450k",
#                                   IDOLOptimizedCpGs =
#                                     IDOLOptimizedCpGsCordBlood,
#                                   returnAll = FALSE)
#   head(countsEPIC$counts)
# }

```

Index

*Topic **datasets**

FlowSorted.CordBloodCombined.450k.compTable,
[4](#)

IDOLOptimizedCpGsCordBlood, [5](#)

FlowSorted.CordBloodCombined.450k, [2](#)

FlowSorted.CordBloodCombined.450k.compTable,
[4](#)

IDOLOptimizedCpGsCordBlood, [5](#)