Package 'granulator'

May 1, 2024

Type Package

Title Rapid benchmarking of methods for *in silico* deconvolution of bulk RNA-seq data

Version 1.13.0

Description granulator is an R package for the cell type deconvolution of heterogeneous tissues based on bulk RNA-seq data or single cell RNA-seq expression profiles. The package provides a unified testing interface to rapidly run and benchmark multiple state-of-the-art deconvolution methods. Data for the deconvolution of peripheral blood mononuclear cells (PBMCs) into individual immune cell types is provided as well.

URL https://github.com/xanibas/granulator

BugReports https://github.com/xanibas/granulator/issues

Depends R (>= 4.1)

Suggests BiocStyle, knitr, rmarkdown, testthat

VignetteBuilder knitr

License GPL-3

Encoding UTF-8

LazyData FALSE

RoxygenNote 7.1.1

biocViews RNASeq, GeneExpression, DifferentialExpression, Transcriptomics, SingleCell, StatisticalMethod, Regression

Imports cowplot, e1071, epiR, dplyr, dtangle, ggplot2, ggplotify, grDevices, limSolve, magrittr, MASS, nnls, parallel, pheatmap, purrr, rlang, stats, tibble, tidyr, utils

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

git_branch devel

git_last_commit e912228

git_last_commit_date 2024-04-30

Repository Bioconductor 3.20

Date/Publication 2024-05-01

Author Sabina Pfister [aut, cre],
Vincent Kuettel [aut],
Enrico Ferrero [aut]

Maintainer Sabina Pfister <sabina.pfister@novartis.com>

Contents

`benchmark` *Regress estimated cell type proportions against the ground truth*

Description

`regress` computes regression between estimated cell type proportions and the measured cell type proportions (ground truth).

Usage

```
benchmark(deconvoluted, ground_truth)
```

Arguments

`deconvoluted`	Output object of the function `deconvolute`.
`ground_truth`	A matrix containing measured cell type proportions in percentages. Samples names are inlcuded in rownames.

Value

Returns a list containing thres elements:

- data: a list of data frames with celltype matched estimated and predicted proportions
- stats: a list of data frames with regression statistics comprising Pearson Correlation Coefficient ('pcc'), Concordance Correlation Coefficient ('ccc'), Coefficient of Determination ('adj.r2') and Root Mean Square Error ('rmse')
- summary: a data frame with summary statistics by cell type
- rank: ranking of deconvolution alghoritms by highest all-to-all correlation of coefficients
- summay: summary statistics of regression coefficients by method, signature and cell type
- rank: ranking of methods and signatures by highest average regression coefficient
- combinations: combination of methods and signatures tested

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# load demo PBMCS data
load_ABIS()

# deconvolute
decon <- deconvolute(m = bulkRNAseq_ABIS,
sigMatrix = sigMatrix_ABIS_S0)

# bechmark
bench <- benchmark(deconvoluted = decon,
ground_truth = groundTruth_ABIS)
```

---

bulkRNAseq_ABIS *PBMCs expression profiles (ABIS dataset)*

Description

Public dataset (GSE107011) containing the TPM-normalized gene expression values from bulk RNAseq of PBMCs of 12 healthy individuals. We include here only genes selected in the signature matrices.

Usage

```
data(bulkRNAseq_ABIS)
```

Format

A matrix with 1296 rows (genes) and 12 variables (samples)

Source

GEO

References

Monaco et al. (2019) Cell Reports 26, 1627–1640 (Cell Reports)

`correlate` *Pearson correlation of cell type proportions across cell types and methods*

Description

`correlate` computes Pearson correlations between estimated cell type proportions generated by different methods.

Usage

```
correlate(deconvoluted, scale = TRUE)
```

Arguments

`deconvoluted`	A list: output object from `deconvolute`
`scale`	Boolean: indicate whether the coefficients should be transformed to standard scores (default: scale = TRUE).

Details

`correlation_analysis` is particularly useful to assess the performance of the different methods when no ground truth is available. If several methods agree on similar relative abundances of cell types across samples, the results are more likely to reflect true differences in cell type composition.

Value

Returns a list encompassing two data frames:

- the pearson correlation of coefficients with all other coefficients
- summay: summary statistics of all-to-all correlation of coefficients by cell type
- rank: ranking of deconvolution alghoritms by highest all-to-all correlation of coefficients
- rank: ranking of deconvolution alghoritms by highest average regression all-to-all correlation of coefficients
- combinations: combination of methods and signatures tested

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# load data
load_ABIS()

# deconvolute
decon <- deconvolute(m = bulkRNAseq_ABIS,
sigMatrix = sigMatrix_ABIS_S0)

# correlate
correl <- correlate(deconvoluted = decon)
```

`deconvolute` *Deconvolution from bulk RNAseq*

Description

`deconvolute` predicts cell type proportions from bulk RNAseq data by applying multiple deconvolution methods.

Usage

```
deconvolute(m, sigMatrix, methods = get_decon_methods(), use_cores = 1)
```

Arguments

`m`	Bulk RNAseq: a genes (rows) by samples (columns) matrix containing transcript-per-million (TPM)-normalized gene expression values.
`sigMatrix`	Reference profile: a matrix or a named list of matrices. Each signature matrix should be a genes (rows) by cell types (columns) data frame containing TPM-normalized gene expression values of signature genes.
`methods`	Deconvolution methods: a character vector containing the names of the deconvolution methods to be applied. By default, all methods are run. Functions are either reimplementations of published methods or wrapper functions for published packages: • ols: ordinary least squares • nnls: non negative least squares regression model. Adapted from Abas et al. (2009) • qprog: quadratic programming without constraints • qprogwc: quadratic programming non-negative and sum-to-one constraints. Adapted from Gong et al. (2015) • dtangle: wrapper for the cell deconvolution function `dtangle` form the package **dtangle** • rls: robust linear regression. Adapted from Monaco et al. (2019) • svr: support vector regression. Adapted from Newman et al. (2015)
`use_cores`	Number of cores to use for parallel processing

Value

Returns a list containing two elements:

- coefficients: estimated cell type coefficients
- proportions: estimated cell type proportions in percentage
- combinations: combination of methods and signatures tested

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# load demo PBMCS data
load_ABIS()

# generate list of reference profiles to be tested
sigMatrix <- list(
sig1 = sigMatrix_ABIS_S0,
sig2 = sigMatrix_ABIS_S1)

# deconvolute
decon <- deconvolute(m = bulkRNAseq_ABIS,
sigMatrix = sigMatrix)
```

---

get_decon_methods *Deconvolution methods acronyms*

Description

`get_decon_methods` returns supported deconvolution methods acronyms.

Usage

```
get_decon_methods()
```

Value

vector containing the acronyms of deconvolution methods.

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# get available deconvolution methods
get_decon_methods()
```

get_TPM *Convert raw counts to TPM*

Description

get_TPM is used to convert raw counts to TPMs, which is the most suitable normalization for deconvolution.

Usage

```
get_TPM(counts, effLen)
```

Arguments

counts	Bulk RNAseq: a genes (rows) by samples (columns) matrix containing gene raw counts.
effLen	Vector of gene lengths.

Value

Returns a transcript-per-million (TPM)-normalized matrix.

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# get TPMs from raw counts and gene lengths.
mat <- round(matrix(rexp(200, rate=.01), ncol=20))
len <- round(matrix(rexp(10, rate=.001), ncol=1))+10
tpm <- get_TPM(mat,as.vector(len))
```

groundTruth_ABIS *PBMCS true cell type proprotions (ABIS dataset)*

Description

Public dataset (GSE107011)containing the true proportions for all combinations of cell types (PBMCs) for 12 individuals.

Usage

```
data(groundTruth_ABIS)
```

Format

A matrix with 12 rows (samples) and 24 variables (cell types)

Source

Github

References

Monaco et al. (2019) Cell Reports 26, 1627–1640 (Cell Reports)

---

load_ABIS *Load demo PBMCs deconvolution data*

Description

`load_ABIS` is used to load a demo dataset for the deconvolution of PBMCs samples from published data under the accession number GSE107011. The dataset consists of the following datasets:

- bulkRNAseq_ABIS: PBMCs expression profiles
- sigMatrix_ABIS_S0: Signature matrix for deconvolution of PBMCs in 17 cell types
- sigMatrix_ABIS_S1: Signature matrix for deconvolution of PBMCs in 13 cell types
- sigMatrix_ABIS_S2: Signature matrix for deconvolution of PBMCs in 11 cell types
- sigMatrix_ABIS_S3: Signature matrix for deconvolution of PBMCs in 9 cell types
- groundTruth_ABIS: PBMCS true cell type proprotions

Usage

```
load_ABIS()
```

Value

Returns string confirming successfull loading of the data.

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# load data
load_ABIS()
```

plot_benchmark *Plot benchmarking analysis scores*

Description

plot_benchmark plots the median correlation scores between estimated and measured cell types across methods and cell types.

Usage

```
plot_benchmark(benchmarked, metric = "pcc")
```

Arguments

benchmarked	List: output object from function benchmarked.
metric	Character: the metric of evaluation. Options include Pearson Correlation Coefficient ('pcc'), Concordance Correlation Coefficient ('ccc'), Coefficient of Determination ('adj.r2') and Root Mean Square Error ('rmse') of the linear regression model.

Value

Plot showing correlations across algorithms and cell types.

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# load demo PBMCS data
load_ABIS()

# deconvolute
decon <- deconvolute(m = bulkRNAseq_ABIS,
sigMatrix = sigMatrix_ABIS_S0)

# bechmark
bench<- benchmark(deconvoluted = decon,
ground_truth = groundTruth_ABIS)

# plot bechmark
plot_benchmark(benchmarked = bench,
metric = 'pcc')
```

plot_correlate *Plot of correlations between deconvolution methods*

Description

plot_correlate is used to visualize the results obtained by correlation_analysis.

Usage

```
plot_correlate(correlated, method = "heatmap", legend = TRUE)
```

Arguments

correlated	output object from correlate
method	plot type ("heatmap" or "boxplot")
legend	boolean to display color legend

Details

plot_correlate plots the correlation of cell type proportions across methods in form of a heatmap or a violin plot. If methods agree, cell type proportions of the same cell type should by strongly correlated. For cell types with weak correlation across methods, corresploding estimated cell type proportions should be interpreted with caution.

Value

Returns a heatmap or violin plot showing the correlation distribution of by different methods/signature matrices for each cell type

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# load demo PBMCS data
load_ABIS()

# deconvolute
decon <- deconvolute(m = bulkRNAseq_ABIS,
sigMatrix = sigMatrix_ABIS_S0)

# correlate
correl <- correlate(deconvoluted = decon)

# plot correlate
plot_correlate(correlated = correl,
method="heatmap")
```

plot_deconvolute *Estimated cell types across methods*

Description

plot_deconvolute allows to compare methods across cell types, where the different methods show a high level of agreement or potentially generate diverging proportion estimates.

Usage

```
plot_deconvolute(
  deconvoluted = deconvoluted,
  scale = TRUE,
  labels = TRUE,
  markers = TRUE
)
```

Arguments

deconvoluted	output object from function deconvolute.
scale	Boolean: indicate whether the coefficients should be transformed to standard scores (default: scale = TRUE).
labels	Boolean: indicate if x axis labels should be included (default: labels = TRUE).
markers	Boolean: indicate if data points markers should be drawn (default: markers = TRUE).

Details

Plots the estimated cell types generated by different deconvolution methods/signature matrices across samples. Scaling is used to directly compare deconvolution outputs across methods.

Value

line plot

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# load demo PBMCS data
load_ABIS()

# deconvolute
decon <- deconvolute(m = bulkRNAseq_ABIS,
sigMatrix = sigMatrix_ABIS_S0)
```

```
# plot deconvolute
plot_deconvolute(deconvoluted = decon)
```

`plot_proportions` *Plot estimated cell type proportions*

Description

`plot_proportions` plots the estimated cell type proportions as computed by a given method and signature matrix.

Usage

```
plot_proportions(deconvoluted, method = "svr", signature = "sig1")
```

Arguments

`deconvoluted`	Output object from function `deconvolute`.
`method`	Character string with name of method to be regressed.
`signature`	Character string with name of signature to be regressed.

Value

Plot showing regression of estimated versus measured cell type coefficients.

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# load demo PBMCS data
load_ABIS()

# deconvolute
decon <- deconvolute(m = bulkRNAseq_ABIS,
sigMatrix = sigMatrix_ABIS_S0)

# plot cell type proportions
plot_proportions(deconvoluted = decon,
method = 'svr', signature = 'sig1')
```

plot_regress *Plot estimated cell type coefficients against the ground truth*

Description

plot_regress depicts the measured cell type proportions (x-axis) vs. the estimated proportions (y-axis).

Usage

```
plot_regress(benchmarked, method = "svr", signature = "sig1")
```

Arguments

benchmarked	List: output object from function benchmarked.
method	Character string with name of method to be regressed.
signature	Character string with name of signature to be regressed.

Value

Plot showing regression of estimated versus measured cell type coefficients.

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# load demo PBMCS data
load_ABIS()

# deconvolute
decon <- deconvolute(m = bulkRNAseq_ABIS,
sigMatrix = sigMatrix_ABIS_S0)

# bechmark
bench <- benchmark(deconvoluted = decon,
ground_truth = groundTruth_ABIS)

# plot regress
plot_regress(benchmarked = bench,
method = 'svr', signature = 'sig1')
```

plot_similarity *Plot reference profile similariy matrix*

Description

plot_similarity plots cell type similarity matrix by computing the Kendall rank correlations between cell type expression profiles. Kendall rank correlation is used to test the similarities in the ordering of data when it is ranked by quantities, and provides a less inflated measure of accuracy than Pearson correlation by accounting for ties in the data.

Usage

```
plot_similarity(sigMatrix)
```

Arguments

sigMatrix	Signature matrix: a data frame or a named list of data frames. Each signature matrix should be a genes (rows) by cell types (columns) data frame containing TPM-normalized gene expression values of signature genes.

Value

Plot showing the Kendall rank correlations similariy matrix.

Author(s)

Vincent Kuettel, Sabina Pfister

Examples

```
# load demo PBMCS data
load_ABIS()

# generate list of reference profiles to be tested
sigMatrix <- list(sig1 = sigMatrix_ABIS_S0,
sig2 = sigMatrix_ABIS_S2)

# plot similarity
plot_similarity(sigMatrix = sigMatrix)
```

sigMatrix_ABIS_S0 *Signature matrix for deconvolution of PBMCs in 17 cell types*

Description

A dataset containing the TPM-normalized RNA-seq gene expression values for signature genes of 17 PBMCs.

Usage

```
data(sigMatrix_ABIS_S0)
```

Format

A matrix with 1296 rows (genes) and 17 variables (cell types)

Source

Github

References

Monaco et al. (2019) Cell Reports 26, 1627–1640 (Cell Reports)

sigMatrix_ABIS_S1 *Signature matrix for deconvolution of PBMCs in 13 cell types*

Description

A dataset containing the TPM-normalized RNA-seq gene expression values for signature genes of 17 PBMCs.

Usage

```
data(sigMatrix_ABIS_S1)
```

Format

A matrix with 1296 rows (genes) and 13 variables (cell types)

References

Monaco et al. (2019) Cell Reports 26, 1627–1640 (Cell Reports)

sigMatrix_ABIS_S2 *Signature matrix for deconvolution of PBMCs in 11 cell types*

Description

A dataset containing the TPM-normalized RNA-seq gene expression values for signature genes of 17 PBMCs.

Usage

```
data(sigMatrix_ABIS_S2)
```

Format

A matrix with 1296 rows (genes) and 11 variables (cell types)

References

Monaco et al. (2019) Cell Reports 26, 1627–1640 (Cell Reports)

sigMatrix_ABIS_S3 *Signature matrix for deconvolution of PBMCs in 9 cell types*

Description

A dataset containing the TPM-normalized RNA-seq gene expression values for signature genes of 17 PBMCs.

Usage

```
data(sigMatrix_ABIS_S3)
```

Format

A matrix with 1296 rows (genes) and 9 variables (cell types)

References

Monaco et al. (2019) Cell Reports 26, 1627–1640 (Cell Reports)

Index