Package 'methylDeConv'

June 19, 2021

Title Computational pipeline for deconvolution of methylation data.

Version 0.1.1

Description An integrated computational pipeline for methylaion data of different platforms and tissues, with a variety of feature selection methods, deconvolution algorithms and extensions of reference library.

License `use_gpl3_license()`

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

biocViews

Imports caret, dplyr, doParallel, EpiDISH, ExperimentHub, foreach, FlowSorted.Blood.450k, FlowSorted.DLPFC.450k, FlowSorted.Blood.EPIC, FlowSorted.CordBlood.450k, glmnet, GEOquery, GSVA, GSEABase, genefilter, limma, minfi, matrixStats, MCMCpack, MASS, MKmisc, Metrics, plyr, pracma, parallel, quadprog, recipes, randomForest, robustbase, stats, varhandle

R topics documented:

betaSim_mixtureProfiles

Beta Mixture Simulation

Description

Beta Mixture Simulation to generate the mixture profiles with true/known proportions.

Usage

```
betaSim_mixtureProfiles(nonimmune_level = 1, purified_datasets_sim, n = 30)
```

Arguments

nonimmune_level

> The levels are 1,2,3,4,5. Default value is 1. 1: No non-immune component; level 2: the non-immune proportion is 0.1-0.2; level 3: the non-immune proportion is 0.2-0.5; level 4: the non-immune proportion is 0.5-0.8; level 5: the non-immune proportion is 0.8-0.9.

purified_datasets_sim

> The reference matrix ref_betamatrix.

n   The number of simulated proportions. Default value: 30.

Value

The simulated mixture profiles.

betaSim_purifiedProfiles

Purified profiles sampling in Beta Mixture Simulation.

Description

Purified profiles sampling in Beta Mixture Simulation.

Usage

```
betaSim_purifiedProfiles(ref_betamatrix, ref_phenotype, n = 20)
```

Arguments

ref_betamatrix	The reference matrix ref_betamatrix.
ref_phenotype	The cell type information for the reference matrix.
n	The number of simulated purfied cell profiles. Default value: 20.

Value

The simulated purified profiles.

---

build_reference_450k   *Build the reference library for 450k arrays*

Description

Build the reference library for 450k to include "CD8T", "CD4T", "NK", "Bcell", "Mono", "Gran", And extend the reference library for 450k to include "Epithelial", "Fibroblast".

Usage

```
build_reference_450k(extend = TRUE)
```

Arguments

extend	If TRUE, generate the extended reference library; otherwise, generate the reference library of only 6 immune cell types. Default value is TRUE.

Value

A list of beta value reference matrix (ref_betamatrix) and reference cell types (ref_phenotype).

```
build_reference_450k_neuron
```
Build the reference library for 450k arrays of brain tissue

Description

Build the reference library for 450k arrays of brain tissues (DLPFC) to include "NeuN_neg", "NeuN_pos".

Usage

```
build_reference_450k_neuron()
```

Value

A list of beta value reference matrix (ref_betamatrix) and reference cell types (ref_phenotype).

```
build_reference_EPIC
```      *Build the reference library for EPIC arrays*

Description

Build the reference library for EPIC arrays to include "CD8T", "CD4T", "NK", "Bcell", "Mono", "Neu", And extend the reference library for 450k to include "Epithelial" or "cfDNA".

Usage

```
build_reference_EPIC(extend = TRUE, include = "Epithelial")
```

Arguments

extend	If TRUE, generate the extended reference library; otherwise, generate the reference library of only 6 immune cell types. Default value is TRUE.
include	If include = "Epithelial", extend the reference library by only adding epithelial cells; if include = "cfDNA", extend the reference library by only adding cell free DNAs; if include = "Epithelial and cfDNA", extend the reference library by add both epithelial and cell free DNAs. Default value is "Epithelial".

Value

A list of beta value reference matrix (ref_betamatrix) and reference cell types (ref_phenotype).

build_reference_EPIC_neuron

Build the reference library for EPIC arrays of brain tissue

Description

Build the reference library for EPIC arrays of brain tissues to include "NeuN_neg", "NeuN_pos".

Usage

```
build_reference_EPIC_neuron()
```

Value

A list of beta value reference matrix (ref_betamatrix) and reference cell types (ref_phenotype).

CBS *CBS algorithm*

Description

CBS deconvolution algorithm

Usage

```
CBS(betamatrix, compTable, probes_select)
```

Arguments

betamatrix	Beta value matrix of methylation array for mixture samples.
compTable	Average reference profiles over each cell type.
probes_select	Probes selected.

Value

The matrix of deconvolution results with rows as mixture samples and columns as cell types.

---

enrichment_score   *Enrichment score-based algorithm*

Description

CBS deconvolution algorithm

Usage

```
enrichment_score(
  betamatrix,
  ref_betamatrix,
  ref_phenotype,
  method = "MCP-counter",
  pv = 1e-08,
  MaxDMRs = 100
)
```

Arguments

betamatrix	Beta value matrix of methylation array for mixture samples.
ref_betamatrix	The reference matrix ref_betamatrix.
ref_phenotype	The cell type information for the reference matrix.
method	The enrichment score-based deconvolution algorithm such as "MCP-counter", "ssGSEA" and "ESTIMATE", which are adapted from gene-expression based algorithm. Default value: "MCP-counter".
pv	The p-value threshold with default value as 1e-8
MaxDMRs	The number of probes selected with default value as 100.

Value

The matrix of deconvolution results with rows as mixture samples and columns as cell types.

GaussianSim_mixtureProfiles
                          Gaussian Mixture Simulation

Description

Gaussian Mixture Simulation to generate the mixture profiles with true/known proportions.

Usage

```
GaussianSim_mixtureProfiles(nonimmune_level = 1, purified_datasets_sim, n = 30)
```

Arguments

nonimmune_level

> The levels are 1,2,3,4,5. Default value is 1. 1: No non-immune component; level 2: the non-immune proportion is 0.1-0.2; level 3: the non-immune proportion is 0.2-0.5; level 4: the non-immune proportion is 0.5-0.8; level 5: the non-immune proportion is 0.8-0.9.

purified_datasets_sim

> The reference matrix ref_betamatrix.

n   The number of simulated proportions. Default value: 30.

Value

The simulated mixture profiles.

---

gaussianSim_purifiedProfiles

> *Purified profiles sampling in Gaussian Mixture Simulation.*

Description

Purified profiles sampling in Gaussian Mixture Simulation.

Usage

```
gaussianSim_purifiedProfiles(ref_betamatrix, ref_phenotype, n = 20)
```

Arguments

ref_betamatrix   The reference matrix ref_betamatrix.

ref_phenotype   The cell type information for the reference matrix.

n   The number of simulated purfied cell profiles. Default value: 20.

Value

The simulated purified profiles.

Houseman_project   *Houseman algorithm*

Description

Houseman deconvolution algorithm

Usage

```
Houseman_project(betamatrix, compTable, probes_select)
```

Arguments

`betamatrix`	Beta value matrix of methylation array for mixture samples.
`compTable`	Average reference profiles over each cell type.
`probes_select`	Probes selected.

Value

The matrix of deconvolution results with rows as mixture samples and columns as cell types.

`MethylDeconv_pipeline`   *Deconvolution Pipeline*

Description

Deconvolution Pipeline

Usage

```
MethylDeconv_pipeline(
  input_methyl,
  normalized = FALSE,
  array_type = "450k",
  feature_selection = "oneVsAllttest",
  deconv_algorithm = "Houseman",
  tissue = "general",
  extend_reference = TRUE,
  custom_probes = NULL
)
```

Arguments

`input_methyl`	Methylation array for mixture samples. The rows correpsond to the probes/CpGs; the columns correspond to the samples.
`normalized`	If TRUE, the methylation array is the beta value matrix. Default value: FALSE.
`array_type`	The array type can be 450k or EPIC. Default value: "450k".
`feature_selection`	
	Different feature selection methods such as "oneVsAllttest" (Default value), "oneVsAllLimma", "pairwiseLimma", "pairwiseGlmnet", "multiGlmnet", "glmnetpreselect", "RFpreselect", "OptVariables".
`deconv_algorithm`	
	Different deconvolution algorithms such as "Houseman" (default value), "RPC", "CBS", "MethylResolver", "MCP-counter", "ssGSEA", "ESTIMATE".
`tissue`	Different tissue types such as "general" (default value) which correspond to blood tissue and general epithelial tissue; "brain" correpsonds to brain tissues.
`extend_reference`	
	Default: TRUE. If TRUE, extend the reference library.
`custom_probes`	Default: NULL. The user-defined probe lists.

Value

The deconvolution results (cell type proportions) with rows as the samples and the columns as cell types.

MethylResolver *MethylResolver algorithm*

Description

MethylResolver deconvolution algorithm

Usage

```
MethylResolver(
  methylMix,
  methylSig,
  probes_select,
  alpha = seq(0.5, 0.9, by = 0.05)
)
```

Arguments

methylMix	Beta value matrix of methylation array for mixture samples.
methylSig	Average reference profiles over each cell type.
probes_select	Probes selected.
alpha	Tuning parameters. Default range 0.5–0.9, which corresponds to fitting a trimmed least square regression to 50–90%, of the cpgs. Get the best alpha value based on RMSE between original and reconstructed mixture, and select best alpha for each mixture sample separately

Value

The matrix of deconvolution results with rows as mixture samples and columns as cell types.

---

ref_compTable *Average reference profiles over each cell type*

Description

Compute the average reference profiles for each cell type, using the reference matrix of purified samples from different cell types (phenotype).

Usage

```
ref_compTable(ref_betamatrix, ref_phenotype)
```

Arguments

`ref_betamatrix`	The reference matrix ref_betamatrix.
`ref_phenotype`	The cell type information for the reference matrix.

Value

The data maxtirx of average reference profiles (each column corresponds to each cell type), except for the first column (Fstat) and the last three columns (low, high, range) which are the summary statistics.

`ref_probe_selection_HighVar`
          Feature selection based on variability

Description

Select the most variable probes across all samples in the reference matrix

Usage

```
ref_probe_selection_HighVar(ref_betamatrix, ranks = 601:1200)
```

Arguments

`ref_betamatrix`	The reference matrix ref_betamatrix.
`ranks`	The top number of most variable probes. Default value is 601:1200, which corresponds the second 600 highest variable probes.

Value

A vector of the selected probes.

`ref_probe_selection_multiclassGlmnet_cv`
          Multi-class elastic net modeling feature selection

Description

Select the non-zero features from the multi-class elastic net modeling on the reference matrix.

Usage

```
ref_probe_selection_multiclassGlmnet_cv(
  ref_betamatrix,
  ref_phenotype,
  nCores = 4,
  reps.resamp = 5,
  reps.repeats = 3
)
```

Arguments

`ref_betamatrix`	The reference matrix ref_betamatrix.
`ref_phenotype`	The cell type information for the reference matrix.
`nCores`	Number of parallel cores. Default value is 4.
`reps.resamp`	The number of folds in k-fold cross-validation. Default value is 5.
`reps.repeats`	The number of complete sets of folds to compute. Default value is 3.

Value

A vector of the selected probes.

`ref_probe_selection_multiclassRF_cv`

Multi-class RF modeling feature selection

Description

Select features based on multi-class Random Forest modeling on the reference matrix of all probes.

Usage

```
ref_probe_selection_multiclassRF_cv(
  ref_betamatrix,
  ref_phenotype,
  nCores = 4,
  reps.resamp = 5,
  reps.repeats = 3,
  tune_grid = 1:30
)
```

Arguments

`ref_betamatrix`	The reference matrix ref_betamatrix.
`ref_phenotype`	The cell type information for the reference matrix.

Value

A vector of the selected probes.

---

ref_probe_selection_oneVsAllLimma

One-vs-All moderated t test feature selection

Description

The moderated One-vs-All t test feature selection based on the reference matrix.

Usage

```
ref_probe_selection_oneVsAllLimma(
  ref_betamatrix,
  ref_phenotype,
  probeSelect = "both",
  FDR = 1e-06,
  MaxDMRs = 100
)
```

Arguments

ref_betamatrix	The reference matrix ref_betamatrix.
ref_phenotype	The cell type information for the reference matrix.
probeSelect	The selection can be "any" or "both". If "any", the function selects top probes regardless of up-regulation or down-regulation. If "both", half of top probes are picked up from the up-regulated probes while the other half of the top probes are picked up from the down-regulated probes. Default value is "both".
FDR	The fdr threshold with default value as 1e-6
MaxDMRs	The number of probes selected with default value as 100.

Value

A vector of the selected probes.

---

ref_probe_selection_oneVsAllttest

One-vs-All t test feature selection

Description

The One-vs-All t test feature selection based on the reference matrix.

Usage

```
ref_probe_selection_oneVsAllttest(
  ref_betamatrix,
  ref_phenotype,
  probeSelect = "both",
  pv = 1e-08,
  MaxDMRs = 100
)
```

Arguments

`ref_betamatrix`	The reference matrix ref_betamatrix.
`ref_phenotype`	The cell type information for the reference matrix.
`probeSelect`	The selection can be "any" or "both". If "any", the function selects top probes regardless of up-regulation or down-regulation. If "both", half of top probes are picked up from the up-regulated probes while the other half of the top probes are picked up from the down-regulated probes. Default value is "both".
`pv`	The p-value threshold with default value as 1e-8
`MaxDMRs`	The number of probes selected with default value as 100.

Value

A vector of the selected probes.

`ref_probe_selection_pairwiseGlmnet_cv`
Pairwise elastic net modeling feature selection

Description

Select the non-zero features from the pairwise elastic net modeling on the reference matrix.

Usage

```
ref_probe_selection_pairwiseGlmnet_cv(
  ref_betamatrix,
  ref_phenotype,
  nCores = 4,
  reps.resamp = 5,
  reps.repeats = 3
)
```

Arguments

`ref_betamatrix`	The reference matrix ref_betamatrix.
`ref_phenotype`	The cell type information for the reference matrix.
`nCores`	Number of parallel cores. Default value is 4.
`reps.resamp`	The number of folds in k-fold cross-validation. Default value is 5.
`reps.repeats`	The number of complete sets of folds to compute. Default value is 3.

Value

A vector of the selected probes.

---

ref_probe_selection_pairwiseLimma

Pairwise moderated t test feature selection

Description

The pairwise moderated t test feature selection based on the reference matrix.

Usage

```
ref_probe_selection_pairwiseLimma(
  ref_betamatrix,
  ref_phenotype,
  FDR = 0.01,
  deltaBeta = 0.2,
  MaxDMRs = 100
)
```

Arguments

ref_betamatrix	The reference matrix ref_betamatrix.
ref_phenotype	The cell type information for the reference matrix.
FDR	The fdr threshold with default value as 0.01
deltaBeta	The threshold for the absolute mean difference. Default value is 0.2.
MaxDMRs	The number of probes selected with default value as 100.

Value

A vector of the selected probes.

---

ref_probe_selection_twoStage

Two-stage feature selection

Description

Select features in two stages: firstly, select top features from one-vs-all t test; secondly, select the features with machine learning modeling.

Usage

```
ref_probe_selection_twoStage(
  ref_betamatrix,
  ref_phenotype,
  preselect = 300,
  ml_model = "elastic net"
)
```

Arguments

`ref_betamatrix`	The reference matrix ref_betamatrix.
`ref_phenotype`	The cell type information for the reference matrix.
`preselect`	The number of top features per cell type selected from one-vs-all t tests. The default value is 300.
`ml_model`	The machine learning model for feature selection in the second stage. The default value is "elastic net", which correpsonds to selecting the non-zero features from multi-class elastic net modeling on the reference matrix. Otherwise, if the parameter value is "RF", the model selection is based on the important variables learnt from multi-class Random forest modeling; if the parameter value is "rfe", it selects features based on recursive feature elimination algorithm and a Random Forest algorithm is used on each iteration to evaluate the model.

Value

Model class.

RPC	*RPC algorithm*

Description

RPC deconvolution algorithm

Usage

```
RPC(betamatrix, compTable, probes_select)
```

Arguments

`betamatrix`	Beta value matrix of methylation array for mixture samples.
`compTable`	Average reference profiles over each cell type.
`probes_select`	Probes selected.

Value

The matrix of deconvolution results with rows as mixture samples and columns as cell types.

---

up_probes_oneVsAllttest_celltype

One-vs-All t test feature selection (up-regulated)

Description

The One-vs-All t test feature selection (up-regulated) based on the reference matrix.

Usage

```
up_probes_oneVsAllttest_celltype(
  ref_betamatrix,
  ref_phenotype,
  pv = 1e-08,
  MaxDMRs = 100
)
```

Arguments

ref_betamatrix	The reference matrix ref_betamatrix.
ref_phenotype	The cell type information for the reference matrix.
pv	The p-value threshold with default value as 1e-8
MaxDMRs	The number of probes selected with default value as 100.

Value

A vector of the selected probes.

Index