Package 'RESOLVE'

April 10, 2023

Version 1.0.0

Date 2022-10-19

Title RESOLVE: An R package for the efficient analysis of mutational signatures from cancer genomes

Depends NMF

Imports Biostrings, BSgenome, BSgenome.Hsapiens.1000genomes.hs37d5, cluster, data.table, GenomeInfoDb, GenomicRanges, glmnet, ggplot2, gridExtra, IRanges, lsa, nnls, parallel, reshape2

Suggests BiocGenerics, BiocStyle, testthat, knitr

Description Cancer is a genetic disease caused by somatic mutations in genes controlling key biological functions such as cellular growth and division. Such mutations may arise both through cell-intrinsic and exogenous processes, generating characteristic mutational patterns over the genome named mutational signatures. The study of mutational signatures have become a standard component of modern genomics studies, since it can reveal which (environmental and endogenous) mutagenic processes are active in a tumor, and may highlight markers for therapeutic response. Mutational signatures computational analysis presents many pitfalls. First, the task of determining the number of signatures is very complex and depends on heuristics. Second, several signatures have no clear etiology, casting doubt on them being computational artifacts rather than due to mutagenic processes. Last, approaches for signatures assignment are greatly influenced by the set of signatures used for the analysis. To overcome these limitations, we developed RESOLVE (Robust EStimation Of mutationaL signatures Via rEgularization), a framework that allows the efficient extraction and assignment of mutational signatures. RESOLVE implements a novel algorithm that enables (i) the efficient extraction, (ii) exposure estimation, and (iii) confidence assessment during the computational inference of mutational signatures.

Encoding UTF-8

License file LICENSE

URL https://github.com/danro9685/RESOLVE

BugReports https://github.com/danro9685/RESOLVE/issues

biocViews BiomedicalInformatics, SomaticMutation

RoxygenNote 7.2.1

VignetteBuilder knitr

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

git_branch RELEASE_3_16

git_last_commit 5223f8c

git_last_commit_date 2022-1-01

Date/Publication 2023-04-10

Author Daniele Ramazzotti [aut] (<https://orcid.org/0000-0002-6087-2666>),
Luca De Sano [cre, aut] (<https://orcid.org/0000-0002-9618-3774>)

Maintainer Luca De Sano <luca.desano@gmail.com>

R topics documented:

RESOLVE-package *An R package for the efficient analysis of mutational signatures from cancer genomes.*

Description

This package provides a framework that allows for the efficient extraction and assignment of mutational signatures. It implements a novel algorithm that enables: (i) the efficient extraction, (ii) exposure estimation, and (iii) confidence assessment during the computational inference of mutational signatures. RESOLVE performs de novo signatures extraction by regularized Non-Negative Matrix Factorization. The method incorporates a background signature during the inference step and adopts elastic net regression to reduce the impact of overfitting. The estimation of the optimal number of signatures is performed by bi-cross-validation.

background *germline replication error*

Description

germline replication error estimated in Rahbari, Raheleh, et al. (2016).

Usage

```
data(background)
```

Format

vector of rates

Value

vector of rates for the 96 trinucleotides

Source

Nat Genet. 2016 Feb;48(2):126-133 (https://www.nature.com/articles/ng.3469).

background2 *COSMIC replication error*

Description

background replication error signature derived from COSMIC SBS5.

Usage

```
data(background2)
```

Format

vector of rates

Value

vector of rates for the 96 trinucleotides

Source

COSMIC database (https://cancer.sanger.ac.uk/cosmic/signatures) v3.

getMNVCounts *getMNVCounts*

Description

Create Multi-Nucleotide Variants (MNVs) counts matrix from input data.

Usage

```
getMNVCounts(data)
```

Arguments

data	a data.frame with variants having 6 columns: sample name, chromosome, start position, end position, ref, alt.

Value

A matrix with Multi-Nucleotide Variants (MNVs) counts per patient.

Examples

```
data(ssm560_reduced)
res <- getMNVCounts(data = ssm560_reduced)
```

getSBSCounts *getSBSCounts*

Description

Create Single Base Substitutions (SBS) counts matrix from input data for a provided reference genome.

Usage

```
getSBSCounts(data, reference = NULL)
```

Arguments

data	a data.frame with variants having 6 columns: sample name, chromosome, start position, end position, ref, alt.
reference	a BSgenome object with the reference genome to be used to retrieve flanking bases.

Value

A matrix with Single Base Substitutions (SBS) counts per patient.

Examples

```
library('BSgenome.Hsapiens.1000genomes.hs37d5')
data(ssm560_reduced)
res <- getSBSCounts(data = ssm560_reduced, reference = BSgenome.Hsapiens.1000genomes.hs37d5)
```

groupsCNPlot *groupsCNPlot*

Description

Plot observed Copy Number (CN) counts for different groups of patients.

Usage

```
groupsCNPlot(counts, groups, normalize = TRUE, xlabels = FALSE)
```

Arguments

counts	matrix with Copy Number (CN) counts data.
groups	list where names are groups labels and elements are patients labels corresponding to rownames in counts.
normalize	boolean value; shall I normalize observed counts?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
counts <- plot_data_examples[['groups.CN.plot']][['counts']]
groups <- plot_data_examples[['groups.CN.plot']][['groups']]
groupsCNPlot(counts=counts,groups=groups)
```

---

groupsCXPlot *groupsCXPlot*

Description

Plot observed Copy Number (Reduced, CX) counts for different groups of patients.

Usage

```
groupsCXPlot(counts, groups, normalize = TRUE, xlabels = FALSE)
```

Arguments

counts	matrix with Copy Number (Reduced, CX) counts data.
groups	list where names are groups labels and elements are patients labels corresponding to rownames in counts.
normalize	boolean value; shall I normalize observed counts?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
counts <- plot_data_examples[['groups.CX.plot']][['counts']]
groups <- plot_data_examples[['groups.CX.plot']][['groups']]
groupsCXPlot(counts=counts,groups=groups)
```

groupsMNVPlot *groupsMNVPlot*

Description

Plot observed Multi-Nucleotide Variants (MNVs) counts for different groups of patients.

Usage

```
groupsMNVPlot(counts, groups, normalize = TRUE, xlabels = FALSE)
```

Arguments

counts	matrix with Multi-Nucleotide Variants (MNVs) counts data.
groups	list where names are groups labels and elements are patients labels corresponding to rownames in counts.
normalize	boolean value; shall I normalize observed counts?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
counts <- plot_data_examples[['groups.MNV.plot']][['counts']]
groups <- plot_data_examples[['groups.MNV.plot']][['groups']]
groupsMNVPlot(counts=counts,groups=groups)
```

groupsSBSPlot *groupsSBSPlot*

Description

Plot observed Single Base Substitutions (SBS) counts for different groups of patients.

Usage

```
groupsSBSPlot(counts, groups, normalize = TRUE, xlabels = FALSE)
```

Arguments

counts	matrix with Single Base Substitutions (SBS) counts data.
groups	list where names are groups labels and elements are patients labels corresponding to rownames in counts.
normalize	boolean value; shall I normalize observed counts?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
counts <- plot_data_examples[['groups.SBS.plot']][['counts']]
groups <- plot_data_examples[['groups.SBS.plot']][['groups']]
groupsSBSPlot(counts=counts,groups=groups)
```

---

patients *point mutations for 560 breast tumors*

Description

dataset of counts of the point mutations detected in 560 breast tumors published in Nik-Zainal, Serena, et al. (2016).

Usage

```
data(patients)
```

Format

counts of the point mutations

Value

counts of point mutations for 560 tumors and 96 trinucleotides

Source

Nature. 2016 Jun 2;534(7605):47-54 (https://www.nature.com/articles/nature17676).

patientsCNPlot *patientsCNPlot*

Description

Plot Copy Number (CN) counts for a set of given patients.

Usage

```
patientsCNPlot(
  cn_data_counts,
  samples = rownames(cn_data_counts),
  freq = FALSE,
  xlabels = FALSE
)
```

Arguments

cn_data_counts	Copy Number counts matrix.
samples	name of the samples. This should match a rownames in cn_data_counts
freq	boolean value; shall I display rates instead of counts?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
counts <- plot_data_examples[['patients.CN.plot']][['counts']]
patientsCNPlot(cn_data_counts=counts,samples=rownames(counts)[seq_len(2)])
```

patientsCXPlot *patientsCXPlot*

Description

Plot Copy Number (Reduced, CX) counts for a set of given patients.

Usage

```
patientsCXPlot(
  cn_data_counts,
  samples = rownames(cn_data_counts),
  freq = FALSE,
  xlabels = FALSE
)
```

Arguments

cn_data_counts	Copy Number counts matrix.
samples	name of the samples. This should match a rownames in cn_data_counts
freq	boolean value; shall I display rates instead of counts?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
counts <- plot_data_examples[['patients.CX.plot']][['counts']]
patientsCXPlot(cn_data_counts=counts,samples=rownames(counts)[seq_len(2)])
```

`patientsMNVPlot` *patientsMNVPlot*

Description

Plot Multi-Nucleotide Variants (MNVs) counts for a set of given patients.

Usage

```
patientsMNVPlot(
  multi_nucleotides_counts,
  samples = rownames(multi_nucleotides_counts),
  freq = FALSE,
  xlabels = FALSE
)
```

Arguments

multi_nucleotides_counts	Multi-Nucleotide counts matrix.
samples	name of the samples. This should match a rownames in multi_nucleotides_counts
freq	boolean value; shall I display rates instead of counts?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
counts <- plot_data_examples[['patients.MNV.plot']][['counts']]
patientsMNVPlot(multi_nucleotides_counts=counts,samples=rownames(counts)[seq_len(2)])
```

---

patientsSBSPlot *patientsSBSPlot*

Description

Plot Single Base Substitutions (SBS) counts for a set of given patients.

Usage

```
patientsSBSPlot(
  trinucleotides_counts,
  samples = rownames(trinucleotides_counts),
  freq = FALSE,
  xlabels = FALSE
)
```

Arguments

trinucleotides_counts	trinucleotides counts matrix.
samples	name of the samples. This should match a rownames in trinucleotides_counts.
freq	boolean value; shall I display rates instead of counts?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
counts <- plot_data_examples[['patients.SBS.plot']][['counts']]
patientsSBSPlot(trinucleotides_counts=counts,samples=rownames(counts)[seq_len(2)])
```

`plot_data_examples` *list data structure to run examples*

Description

list data structure to run examples.

Usage

```
data(plot_data_examples)
```

Format

list data structure to run examples

Value

list data structure to run examples

Source

list data structure to run examples.

`signaturesAssignment` *signaturesAssignment*

Description

Perform the assignment of K somatic mutational signatures provided as input to samples given a set of observed counts x. This function can be used to estimate different types of mutational signatures such as: SBS (single base substitutions) and MNV (multi-nucleotide variant) (see Degasperi, Andrea, et al. 'Substitution mutational signatures in whole-genome–sequenced cancers in the UK population.' Science 376.6591 (2022): abl9283), CX (chromosomal instability) (see Drews, Ruben M., et al. 'A pan-cancer compendium of chromosomal instability.' Nature 606.7916 (2022): 976-983) and CN (copy number) signatures (see Steele, Christopher D., et al. 'Signatures of copy number alterations in human cancer.' Nature 606.7916 (2022): 984-991).

Usage

```
signaturesAssignment(
  x,
  beta,
  normalize_counts = FALSE,
  sparsify = TRUE,
  verbose = TRUE
)
```

Arguments

x	counts matrix for a set of n patients and m categories. These can be, e.g., SBS, MNV, CN or CN counts; in the case of SBS it would be an n patients x 96 trinucleotides matrix.
beta	matrix of the discovered signatures to be used for the assignment.
normalize_counts	if true, the input counts matrix x is normalized such that the patients have the same number of mutation.
sparsify	boolean; Shall I perform regularization using LASSO?
verbose	boolean; Shall I print information messages?

Value

A list with the discovered signatures. It includes 2 elements: alpha: matrix of the discovered exposure values. beta: matrix of the discovered signatures.

Examples

```
data(background)
data(patients)
set.seed(12345)
beta <- signaturesDecomposition(x = patients[seq_len(3),],
                                K = 3,
                                background_signature = background,
                                nmf_runs = 2,
                                sparsify = FALSE,
                                num_processes = 1)
set.seed(12345)
res <- signaturesAssignment(x = patients[seq_len(3),], beta = beta$beta[[1]], sparsify = FALSE)
```

---

signaturesCNPlot *signaturesCNPlot*

Description

Plot the inferred Copy Number (CN) mutational signatures.

Usage

```
signaturesCNPlot(beta, useRowNames = FALSE, xlabels = FALSE)
```

Arguments

beta	matrix with the inferred mutational signatures.
useRowNames	boolean value; shall I use the rownames from beta as names for the signatures?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
beta <- plot_data_examples[['signatures.CN.plot']][['beta']]
signaturesCNPlot(beta=beta)
```

---

signaturesCV *signaturesCV*

Description

Perform the assessment of different signaturesDecomposition solutions by cross validation for K (beta, as estimated by signaturesDecomposition) somatic mutational signatures given a set of observations x and discovered signatures beta.

Usage

```
signaturesCV(
  x,
  beta,
  normalize_counts = FALSE,
  cross_validation_entries = 0.01,
  cross_validation_iterations = 5,
  cross_validation_repetitions = 100,
  num_processes = Inf,
  verbose = TRUE
)
```

Arguments

x	counts matrix for a set of n patients and m categories. These can be, e.g., trinucleotides counts for n patients and 96 trinucleotides.
beta	a set of inferred signatures as returned by signaturesDecomposition function.
normalize_counts	if true, the input counts matrix x is normalized such that the patients have the same number of mutation.
cross_validation_entries	Percentage of cells in the counts matrix to be replaced by 0s during cross validation.
cross_validation_iterations	For each configuration, the first time the signatures are fitted form a matrix with a percentage of values replaced by 0s. This may result in poor fit/results. Then, we perform predictions of these entries and replace them with such predicted values. This parameter is the number of restarts to be performed to improve this estimate and obtain more stable solutions.
cross_validation_repetitions	Number of time cross-validation should be repeated. Higher values result in better estimate, but are computationally more expensive.
num_processes	Number of processes to be used during parallel execution. To execute in single process mode, this parameter needs to be set to either NA or NULL.
verbose	boolean; Shall I print information messages?

Value

A list of 2 elements: estimates and summary. Here, cv_estimates reports the mean squared error for each configuration of performed cross validation; rank_estimates reports mean and median values for each value of K.

Examples

```
data(background)
data(patients)
set.seed(12345)
sigs <- signaturesDecomposition(x = patients[seq_len(3),],
                                K = 3:4,
                                background_signature = background,
                                nmf_runs = 2,
                                sparsify = FALSE,
                                num_processes = 1)
set.seed(12345)
res <- signaturesCV(x = patients[seq_len(3),],
                    beta = sigs$beta,
                    cross_validation_iterations = 2,
                    cross_validation_repetitions = 2,
                    num_processes = 1)
```

signaturesCXPlot *signaturesCXPlot*

Description

Plot the inferred Copy Number (Reduced, CX) mutational signatures.

Usage

```
signaturesCXPlot(beta, useRowNames = FALSE, xlabels = FALSE)
```

Arguments

beta	matrix with the inferred mutational signatures.
useRowNames	boolean value; shall I use the rownames from beta as names for the signatures?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
beta <- plot_data_examples[['signatures.CX.plot']][['beta']]
signaturesCXPlot(beta=beta)
```

signaturesDecomposition
signaturesDecomposition

Description

Perform signatures discovery and rank estimation for a range of K somatic mutational signatures given a set of observed counts x. This function can be used to estimate different types of mutational signatures such as: SBS (single base substitutions) and MNV (multi-nucleotide variant) (see Degasperi, Andrea, et al. 'Substitution mutational signatures in whole-genome–sequenced cancers in the UK population.' Science 376.6591 (2022): abl9283), CX (chromosomal instability) (see Drews, Ruben M., et al. 'A pan-cancer compendium of chromosomal instability.' Nature 606.7916 (2022): 976-983) and CN (copy number) signatures (see Steele, Christopher D., et al. 'Signatures of copy number alterations in human cancer.' Nature 606.7916 (2022): 984-991).

Usage

```
signaturesDecomposition(
  x,
  K,
  background_signature = NULL,
  normalize_counts = FALSE,
  nmf_runs = 50,
  sparsify = TRUE,
  num_processes = Inf,
  verbose = TRUE
)
```

Arguments

x	counts matrix for a set of n patients and m categories. These can be, e.g., SBS, MNV, CN or CN counts; in the case of SBS it would be an n patients x 96 trinucleotides matrix.
K	either one value or a range of numeric values (each of them greater than 0) indicating the number of signatures to be considered.
background_signature	background signature to be used.
normalize_counts	if true, the input counts matrix x is normalized such that the patients have the same number of mutation.
nmf_runs	number of iteration (minimum 1) of NMF to be performed for a robust estimation of beta.
sparsify	boolean; Shall I perform regularization using LASSO?
num_processes	Number of processes to be used during parallel execution. To execute in single process mode, this parameter needs to be set to either NA or NULL.
verbose	boolean; Shall I print information messages?

Value

A list with the discovered signatures and related rank measures. It includes 3 elements: alpha: list of matrices of the discovered exposure values for each possible rank in the range K. beta: list of matrices of the discovered signatures for each possible rank in the range K. measures: a data.frame containing the quality measures for each possible rank in the range K.

Examples

```
data(background)
data(patients)
set.seed(12345)
res <- signaturesDecomposition(x = patients[seq_len(3),],
                               K = 3:4,
                               background_signature = background,
                               nmf_runs = 2,
```

```
                    sparsify = FALSE,
                    num_processes = 1)
```

---

signaturesMNVPlot *signaturesMNVPlot*

Description

Plot the inferred Multi-Nucleotide Variants (MNVs) mutational signatures.

Usage

```
signaturesMNVPlot(beta, useRowNames = FALSE, xlabels = FALSE)
```

Arguments

beta	matrix with the inferred mutational signatures.
useRowNames	boolean value; shall I use the rownames from beta as names for the signatures?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
beta <- plot_data_examples[['signatures.MNV.plot']][['beta']]
signaturesMNVPlot(beta=beta)
```

---

signaturesSBSPlot *signaturesSBSPlot*

Description

Plot the inferred Single Base Substitutions (SBS) mutational signatures.

Usage

```
signaturesSBSPlot(beta, useRowNames = FALSE, xlabels = FALSE)
```

Arguments

beta	matrix with the inferred mutational signatures.
useRowNames	boolean value; shall I use the rownames from beta as names for the signatures?
xlabels	boolean value; shall I display x labels?

Value

A ggplot2 object.

Examples

```
data(plot_data_examples)
beta <- plot_data_examples[['signatures.SBS.plot']][['beta']]
signaturesSBSPlot(beta=beta)
```

---

signaturesSignificance *signaturesSignificance*

Description

Perform a robust estimation by bootstrap of alpha coefficients until a given level of cosine similarity is reached given a set of observed counts x and discovered signatures beta.

Usage

```
signaturesSignificance(
  x,
  beta,
  cosine_thr = 0.95,
  min_contribution = 0.05,
  pvalue_thr = 0.05,
  sparsify = TRUE,
  nboot = 100,
  num_processes = Inf,
  verbose = TRUE
)
```

Arguments

x	counts matrix for a set of n patients and m categories. These can be, e.g., trinucleotides counts for n patients and 96 trinucleotides.
beta	discovered signatures to be used for the fit of alpha.
cosine_thr	Level of cosine similarity to be reached for the fit of alpha.

min_contribution	Minimum contribution of a signature to be considered significant.
pvalue_thr	Pvalue level to be used to assess significance.
sparsify	boolean; Shall I perform regularization using LASSO?
nboot	Number of bootstrap iterations to be performed.
num_processes	Number of processes to be used during parallel execution. To execute in single process mode, this parameter needs to be set to either NA or NULL.
verbose	boolean; Shall I print information messages?

Value

A list with the bootstrap estimates. It includes 4 elements: alpha: matrix of the discovered exposure values considering significant signatures as estimated by bootstrap. beta: matrix of the discovered signatures. goodness_fit: vector reporting cosine similarities between predictions and observations. bootstrap_estimates: list of matrices reporting results by bootstrap estimates.

Examples

```
data(background)
data(patients)
set.seed(12345)
beta <- signaturesDecomposition(x = patients[seq_len(3),],
                                K = 3:4,
                                background_signature = background,
                                nmf_runs = 2,
                                sparsify = FALSE,
                                num_processes = 1)
set.seed(12345)
res <- signaturesSignificance(x = patients[seq_len(3),],
                              beta = beta$beta[[1]],
                              cosine_thr = 0.95,
                              min_contribution = 0.05,
                              pvalue_thr = 0.05,
                              sparsify = FALSE,
                              nboot = 2,
                              num_processes = 1)
```

---

ssm560_reduced *a reduced version of the point mutations for 560 breast tumors in the format compatible with the import function*

Description

reduced versione of the dataset of counts of the point mutations detected in 560 breast tumors published in Nik-Zainal, Serena, et al. (2016).

Usage

```
data(ssm560_reduced)
```

Format

reduced versione of the counts of the point mutations in the format compatible with the import function

Value

reduced versione of the counts of point mutations for 560 tumors and 96 trinucleotides in the format compatible with the import function

Source

Nature. 2016 Jun 2;534(7605):47-54 (https://www.nature.com/articles/nature17676).

Index