Package 'indelsig.tools.lib'

October 11, 2022

Type Package

Title Indel signature tools R package

Version 0.1.0

Author Xueqing Zou and Scott Nanda

Maintainer Xueqing Zou <xz388@cam.ac.uk>

Description indelsig.tools.lib is an R package for analyzing indel signatures from sequencing data.
Use four spaces when indenting paragraphs within the Description.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

Depends R (>= 3.5)

Imports ggplot2, plyr, reshape2, gridExtra, scales, stringr,
factoextra, VariantAnnotation, GenomicFeatures,
BSgenome.Hsapiens.UCSC.hg19, BSgenome.Hsapiens.UCSC.hg38,
Rtsne, ggbeeswarm, umap, signature.tools.lib, rlang, grDevices,
utils, Rcpp, Biostrings, dendextend, gplots

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

LinkingTo Rcpp

R topics documented:

assign_channels_m5	*Assign each indel a channel*

Description

Assign each indel a channel

Usage

```
assign_channels_m5(indel.df)
```

Arguments

indel.df   A classified indel list generated by function indel_classifier89()

Value

An indel list with channel information

---

assign_channels_mf *Assign each indel a channel from 464 channels*

Description

Assign each indel a channel from 464 channels

Usage

```
assign_channels_mf(indel.df)
```

Arguments

indel.df A classified indel list generated by function indel_classifier_full()

Value

An indel list with channel information

---

cal_cossim_2sigsets *Calculate cosine similarity between two signature/profile sets*

Description

Calculate cosine similarity between two signature/profile sets

Usage

```
cal_cossim_2sigsets(sigset1, sigset2, h, w, text_size, ifplot = T, outputname)
```

Arguments

sigset1	The first signature set
sigset2	The second signature set
h	Hight of the plot
w	Width of the plot
text_size	Size of text
ifplot	Plot the output or not, default is True
outputname	Output file name of the plot

Value

A text file including all the cosine similarities between signatures from two data sets

---

cos_similarity   *Calculate cosine similarity between two vectors*

Description

Calculate cosine similarity between two vectors

Usage

```
cos_similarity(v1, v2)
```

Arguments

v1   The first vector

v2   The second vector

Value

The cosine similarity between two vectors

---

divid_indelcat   *Separate indel catlogue according to a indel number cutoff*

Description

Separate indel catlogue according to a indel number cutoff

Usage

```
divid_indelcat(mut_cat, cutoff = 500, outputname)
```

Arguments

mut_cat   An indel catalogue of multiple samples

cutoff   Threshold of indel number to separate samples according to indel burden

outputname   file names

Value

An list of two data.frame

---

fix_load_indels   *Fix load indels*

Description

Fix load indels

Usage

```
fix_load_indels(name)
```

Arguments

name   TO BE ADDED

Value

function TO BE ADDED

---

fix_prepare_indels   *Fix prepare indels*

Description

Fix prepare indels

Usage

```
fix_prepare_indels(name)
```

Arguments

name   TO BE ADDED

Value

function TO BE ADDED

---

gen_catalogue21   *Generate indel catalogue in 21 channels*

Description

Generate indel catalogue in 21 channels

Usage

```
gen_catalogue21(muts_list, sample_col)
```

Arguments

muts_list   A indel list

sample_col   Sample column name

Value

A 21 channel indel catalogue

---

gen_catalogue89   *Generate indel catalogue in 89 channels*

Description

Generate indel catalogue in 89 channels

Usage

```
gen_catalogue89(muts_list, sample_col)
```

Arguments

muts_list   A indel list

sample_col   Sample column name

Value

A 89 channel indel catalogue

gen_fullcatalogue   *Generate indel catalogue in full channels*

Description

Generate indel catalogue in full channels

Usage

```
gen_fullcatalogue(muts_list, sample_col)
```

Arguments

muts_list	A indel list
sample_col	Sample column name

Value

A full channel indel catalogue

gen_plot_catalouge21_single
                    Plot indel profile in a 21-channel bar plot for single sample

Description

Plot indel profile in a 21-channel bar plot for single sample

Usage

```
gen_plot_catalouge21_single(muts_basis, text_size, plot_title)
```

Arguments

muts_basis	A indel catalogue of a single sample
text_size	Size of text
plot_title	Title of the plot

Value

A 21-channel indel profile plot

---

gen_plot_catalouge89_single
Plot indel profile in a 89-channel bar plot for single sample Show all x labels

---

Description

Plot indel profile in a 89-channel bar plot for single sample Show all x labels

Usage

```
gen_plot_catalouge89_single(muts_basis, text_size, plot_title)
```

Arguments

muts_basis	A indel catalogue of a single sample
text_size	Size of text
plot_title	Title of the plot

Value

A 89-channel indel profile plot

---

gen_plot_catalouge89_single_noXlabel
Plot indel profile in a 89-channel bar plot for single sample Not show X labels

---

Description

Plot indel profile in a 89-channel bar plot for single sample Not show X labels

Usage

```
gen_plot_catalouge89_single_noXlabel(muts_basis, text_size, plot_title)
```

Arguments

muts_basis	A indel catalogue of a single sample
text_size	Size of text
plot_title	Title of the plot

Value

A 89-channel indel profile plot

gen_plot_catalouge89_single_old

Old function to plot indel profile/signature in a 89-channel bar plot for single sample

Description

Old function to plot indel profile/signature in a 89-channel bar plot for single sample

Usage

```
gen_plot_catalouge89_single_old(muts_basis, text_size, plot_title)
```

Arguments

muts_basis	A indel catalogue of a single sample
text_size	Size of text
plot_title	Title of the plot

Value

A 89-channel indel profile plot

gen_plot_catalouge89_single_percentage

Plot indel signature in a 89-channel bar plot for single sample

Description

Plot indel signature in a 89-channel bar plot for single sample

Usage

```
gen_plot_catalouge89_single_percentage(muts_basis, text_size, plot_title)
```

Arguments

muts_basis	A indel catalogue of a single sample
text_size	Size of text
plot_title	Title of the plot

Value

A 89-channel indel signature plot

---

gen_plot_catalougefull_single

Plot indel profile in a extended channel bar plot for single sample Show all x labels

---

Description

Plot indel profile in a extended channel bar plot for single sample Show all x labels

Usage

```
gen_plot_catalougefull_single(muts_basis, text_size, plot_title)
```

Arguments

muts_basis	A indel catalogue of a single sample
text_size	Size of text
plot_title	Title of the plot

Value

A extended channel indel profile plot

---

gen_plot_catalougefull_single_noXlabel

Plot indel profile in a extended channel bar plot for single sample Not show X labels

---

Description

Plot indel profile in a extended channel bar plot for single sample Not show X labels

Usage

```
gen_plot_catalougefull_single_noXlabel(muts_basis, text_size, plot_title)
```

Arguments

muts_basis	A indel catalogue of a single sample
text_size	Size of text
plot_title	Title of the plot

Value

A extended channel indel profile plot

---

hclust_indel   *Hierarchical Clustering to identify high-burden clusters*

Description

Hierarchical Clustering to identify high-burden clusters

Usage

```
hclust_indel(
  mut_cat,
  hclust_method = "complete",
  tissue_type,
  ncluster = 6,
  cutoff_height = NULL
)
```

Arguments

mut_cat	An indel catalogue of multiple samples
hclust_method	This should be (an unambiguous abbreviation of) one of "ward.D", "ward.D2", "single", "complete", "average" (= UPGMA), "mcquitty" (= WPGMA), "median" (= WPGMC) or "centroid" (= UPGMC).
tissue_type	Tissue type
ncluster	An integer scalar or vector with the desired number of clusters
cutoff_height	Numeric scalar or vector with heights where the tree should be cut

Value

An object of class hclust which describes the tree produced by the clustering process

---

indel_classifier89   *Classify indels into different types*

Description

Classify indels into different types

Usage

```
indel_classifier89(indels, genome.v)
```

Arguments

indels	A indel list. Columns include: "Sample","chr", "position", "REF", "ALT"
genome.v	: "hg19", "hg38"

Value

Classified indel list

---

indel_classifier_full *expand 89 channels*

Description

expand 89 channels

Usage

```
indel_classifier_full(indels, genome.v)
```

Arguments

indels	A indel list. Columns include: "Sample","chr", "position", "REF", "ALT"
genome.v	: "hg19", "hg38"

Value

Classified indel list

---

indel_highspecific *For signature analysis, indels have to be as clean as possible. This function will: 1) remove single indels that have original_reps >=10 2) remove nMer indels that have original_reps >=10 3) small indels: length <=100*

Description

For signature analysis, indels have to be as clean as possible. This function will: 1) remove single indels that have original_reps >=10 2) remove nMer indels that have original_reps >=10 3) small indels: length <=100

Usage

```
indel_highspecific(indel.classified)
```

Arguments

indel.classified
                A classified indel list.

Value

A filtered indel list

indel_template_type_4 *indel_template_type_4*

Description

Template for 89 channel indel profile

Usage

```
indel_template_type_4
```

Format

A data frame with three variables:

IndelType Indel type in 89 channels

Indel Indel type in 8 channels

indel_template_type_4_full

indel_template_type_4_full

Description

Template for extended channel indel profile

Usage

```
indel_template_type_4_full
```

Format

A data frame with three variables:

IndelType Indel type in extended channels

Indel Indel type in 8 channels

---

indel_template_type_4_full_figurelabel

indel_template_type_4_full_figurelabel

Description

Template for extended channel indel profile

Usage

    indel_template_type_4_full_figurelabel

Format

A data frame with three variables:

IndelType Indel type in extended channels

Indel Indel type in 8 channels

Indel3 Indel type in 3 channels

Figlabel Lable for Figure

---

indel_type_4_figurelabel

indel_type_4_figurelabel

Description

Template for 89 channel indel profile

Usage

    indel_type_4_figurelabel

Format

A data frame with three variables:

IndelType Indel type in 89 channels

Indel Indel type in 8 channels

Indel3 Indel type in 3 channels

Figlabel Lable for Figure

---

normalize_euclidean *Normalize a vector by euclidean length*

Description

Normalize a vector by euclidean length

Usage

```
normalize_euclidean(v)
```

Arguments

v   The target vector

Value

normalized vector

---

normalize_manhattan *Normalize a vector by manhattan length*

Description

Normalize a vector by manhattan length

Usage

```
normalize_manhattan(v)
```

Arguments

v   The target vector

Value

normalized vector

plots_indelprofile_21ch

Plot indel profile in a 21-channel bar plot, original plots_type_4_m4_89 function

Description

Plot indel profile in a 21-channel bar plot, original plots_type_4_m4_89 function

Usage

```
plots_indelprofile_21ch(muts_basis, colnum, h, w, text_size, outputname)
```

Arguments

muts_basis	A indel catalogue of multiple samples
colnum	Number of columns
h	Hight of the plot
w	Width of the plot
text_size	Size of text
outputname	Output file name of the plot

Value

A plot including 89-channel indel profile of multiple samples

plots_indelprofile_89ch

Plot indel profile in a 89-channel bar plot, original plots_type_4_m4_89 function

Description

Plot indel profile in a 89-channel bar plot, original plots_type_4_m4_89 function

Usage

```
plots_indelprofile_89ch(
  muts_basis,
  colnum,
  h,
  w,
  text_size,
  print_Xlabel = T,
  outputname
)
```

Arguments

`muts_basis`	A indel catalogue of multiple samples
`colnum`	Number of columns
`h`	Hight of the plot
`w`	Width of the plot
`text_size`	Size of text
`print_Xlabel`	print X axis label or not
`outputname`	Output file name of the plot

Value

A plot including 89-channel indel profile of multiple samples

---

`plots_indelprofile_89ch_noXlabel`

Plot indel profile in a 89-channel bar plot, original plots_type_4_m4_89 function Not show X labels

---

Description

Plot indel profile in a 89-channel bar plot, original plots_type_4_m4_89 function Not show X labels

Usage

```
plots_indelprofile_89ch_noXlabel(
  muts_basis,
  colnum,
  h,
  w,
  text_size,
  outputname
)
```

Arguments

`muts_basis`	A indel catalogue of multiple samples
`colnum`	Number of columns
`h`	Hight of the plot
`w`	Width of the plot
`text_size`	Size of text
`outputname`	Output file name of the plot

Value

A plot including 89-channel indel profile of multiple samples

---

plots_indelprofile_full

> *Plot indel profile in a extended channel bar plot, original plots_type_4_m4_89 function*

Description

Plot indel profile in a extended channel bar plot, original plots_type_4_m4_89 function

Usage

```
plots_indelprofile_full(
  muts_basis,
  colnum,
  h,
  w,
  text_size,
  print_Xlabel = T,
  outputname
)
```

Arguments

muts_basis	A indel catalogue of multiple samples
colnum	Number of columns
h	Hight of the plot
w	Width of the plot
text_size	Size of text
print_Xlabel	print X axis label or not
outputname	Output file name of the plot

Value

A plot including extended channel indel profile of multiple samples

---

plots_indelprofile_fullch_noXlabel

> *Plot indel profile in a extended channel bar plot, original plots_type_4_m4_89 function Not show X labels*

Description

Plot indel profile in a extended channel bar plot, original plots_type_4_m4_89 function Not show X labels

Usage

```
plots_indelprofile_fullch_noXlabel(
  muts_basis,
  colnum,
  h,
  w,
  text_size,
  outputname
)
```

Arguments

muts_basis	A indel catalogue of multiple samples
colnum	Number of columns
h	Hight of the plot
w	Width of the plot
text_size	Size of text
outputname	Output file name of the plot

Value

A plot including extended channel indel profile of multiple samples

---

plots_indelsig_89ch	*Fixed size plot indel signatures in a 89-channel bar plot, original plots_type_4_m4_89 function*

Description

Fixed size plot indel signatures in a 89-channel bar plot, original plots_type_4_m4_89 function

Usage

```
plots_indelsig_89ch(muts_basis, rownum = 5, h = 15, w = 40, outputname)
```

Arguments

muts_basis	A indel catalogue of multiple samples
rownum	Number of rowumns
h	Hight of the plot
w	Width of the plot
outputname	Output file name of the plot

Value

A plot including 89-channel indel signatures of multiple signatures

---

plot_sig_count   *Plot Count of signature in each sample for a given cohort*

Description

Plot Count of signature in each sample for a given cohort

Usage

```
plot_sig_count(sample_sig_count, SampleCol, h, w, text_size, outputname)
```

Arguments

sample_sig_count

                 A data frame shows the number of mutations of signatures in each sample

SampleCol	Sample column name
h	Hight of the plot
w	Width of the plot
text_size	Size of text
outputname	Output file name of the plot

Value

A bar plot of signature distribution (Count) in each sample

---

plot_sig_percentage   *Plot percentage of signature in each sample for a given cohort*

Description

Plot percentage of signature in each sample for a given cohort

Usage

```
plot_sig_percentage(sample_sig_count, SampleCol, h, w, text_size, outputname)
```

Arguments

sample_sig_count

                 A data frame shows the number of mutations of signatures in each sample

SampleCol	Sample column name
h	Hight of the plot
w	Width of the plot
text_size	Size of text
outputname	Output file name of the plot

Value

A bar plot of signature distribution (percentage) in each sample

---

prepare_indels   *Prepare indels sequence context information for classification*

Description

Prepare indels sequence context information for classification

Usage

```
prepare_indels(indels, sampleID, genome.v)
```

Arguments

indels	A indel list. Columns include: "Sample","chr", "position", "REF", "ALT"
sampleID	Cohort name
genome.v	: "hg19", "hg38"

Value

Extended indel list with 5' and 3' sequence information

---

preprocess.sample   *Preprocess samples before signature extraction*

Description

Preprocess samples before signature extraction

Usage

```
preprocess.sample(mut_cat, ss_threshold = 500, rescale.type = "none")
```

Arguments

mut_cat	An indel catalogue of multiple samples
ss_threshold	Threshold of indel number (default: 500)
rescale.type	types of rescale: "manhattan_hb", "euclidean". "manhattan_hb" is to normalize mutation burden of hypermutators to between 2,000-5,000. "euclidean" is 2 norm distance.

Value

An indel catalogue of multiple samples

rescale.sample	*Plot indel profile in a 89-channel bar plot, original plots_type_4_m4_89 function*

Description

Plot indel profile in a 89-channel bar plot, original plots_type_4_m4_89 function

Usage

```
rescale.sample(mut_cat, rescale.type)
```

Arguments

mut_cat	An indel catalogue of multiple samples
rescale.type	types of rescale: "hypermutator", "manhattan", "euclidean". "hypermutator" is to normalize mutation burden of hypermutators to between 2,000-5,000. "manhattan" is 1 norm distance. "euclidean" is 2 norm distance.

Value

Rescaled indel catalogue to total indel number = 1000

segment	*segmentation function*

Description

segmentation function

Usage

```
segment(string, context)
```

Arguments

string	A string
context	Context of a string

---

segment_indels   *Segment indels types for classification*

Description

Segment indels types for classification

Usage

```
segment_indels(df)
```

Arguments

df   An extended indel list generated by prepare_indels().

Value

Extended indel list with 5' and 3' sequence information

---

ss.sample   *Combine low indel burden samples to make synthetic sample with in-*
                        del number more than a given threshold

---

Description

Combine low indel burden samples to make synthetic sample with indel number more than a given threshold

Usage

```
ss.sample(mut_cat, ss_threshold = 500)
```

Arguments

mut_cat   An indel catalogue of multiple samples
ss_threshold   Threshold of indel number (default: 500)

Value

An indel catalogue of multiple samples

Index