Package: mutaliskRutils (via r-universe)

September 10, 2024

Title What the Package Does (One Line, Title Case)

Version 0.0.0.9000

Description What the package does (one paragraph).

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Imports assertions (>= 0.0.0.9000), checkmate, data.table, dplyr,
forcats, ggplot2, ggthemes, gtools, magrittr, pals, purrr,
rlang, tidyr,

Suggests covr, stats, testthat (>= 3.0.0), vdiffr

Config/testthat/edition 3

Remotes selkamand/assertions

Repository https://selkamand.r-universe.dev

RemoteUrl https://github.com/selkamand/mutaliskRutils

RemoteRef HEAD

RemoteSha c2f360c51c1def1f729d6ef0b23831ccfac92dd1

Contents

Index   **10**

---

extract_sample_names_from_mutalisk_filenames
Sample Names From Mutalisk Output

Description

Sample Names From Mutalisk Output

Usage

```
extract_sample_names_from_mutalisk_filenames(mutalisk_filenames)
```

Arguments

mutalisk_filenames
              names of mutalisk output files (character)

Value

sample name (string)

---

extract_sample_names_from_mutalisk_files
Sample Names From Mutalisk File Contents

Description

Sample Names From Mutalisk File Contents

Usage

```
extract_sample_names_from_mutalisk_files(mutalisk_filenames)
```

Arguments

mutalisk_filenames
              names of mutalisk output files (character)

Value

sample name (string)

---

mutalisk_best_signature_directory_to_dataframe

Mutalisk directory to dataframe

Description

Mutalisk directory to dataframe

Usage

```
mutalisk_best_signature_directory_to_dataframe(directory, metadata = NA)
```

Arguments

directory
: path to **mutalisk_best_fit** folder. To obtain, run your VCFs through mutalisk. Select Mutational Signature (Best only) and click 'Get the selected result for all samples at once'. Then unzip the file, and youre ready to go

metadata
: Either a path to csv file OR a dataframe. Must contain a header line which contains a SampleID column that matches that of mutalisk_dataframe (string)

Value

tibble

---

mutalisk_dataframe_add_metadata

Add sample metadata

Description

Add sample metadata from dataframe to mutalisk data.frame

Usage

```
mutalisk_dataframe_add_metadata(mutalisk_dataframe, sample_metadata)
```

Arguments

mutalisk_dataframe

: a dataframe that can be produced using mutalisk_best_signature_directory_to_dataframe. Can also just make it yourself, if you want to visualise non-mutalisk data. Dataframe just needs 3 columns:

  1. **SampleID**: a sample identifier.
  2. **Signatures**: an identifier for a particular signature.

3. **Contributions**: the percentage contribution of the signature to the patients genetic profile (0.1 = 10%).

sample_metadata

a dataframe containing a SampleID column and additional columns for each property you want to add as metadata (data.frame)

Value

mutalisk dataframe with additional metadata columns (data.frame)

---

mutalisk_dataframe_expand

Expand mutalisk_dataframe

Description

In normal mutalisk dataframe, each sample has data for ONLY the 1-7 signatures that comprise the 'best fit'. Basically, we end up with a dataframe where not all signatures have entries for all samples. We call this 'implicit' missing values. This means that a signature level jitterplot won't show the samples where it was not included in this 'best fit' set, when we'd actually want to know that it contributed 0% to that sample. This function fixes the issue by adding entries for ALL signature - sample pairs, with Contributions set to 0% where relevant.

Usage

```
mutalisk_dataframe_expand(mutalisk_dataframe)
```

Arguments

mutalisk_dataframe

a dataframe that can be produced using mutalisk_best_signature_directory_to_dataframe. Can also just make it yourself, if you want to visualise non-mutalisk data. Dataframe just needs 3 columns:

1. **SampleID**: a sample identifier.
2. **Signatures**: an identifier for a particular signature.
3. **Contributions**: the percentage contribution of the signature to the patients genetic profile (0.1 = 10%).

Value

dataframe containing all combinations of Sample ID and Signatures. For cases where a signature was not included in the 'best fit' subset, Contribution is set to 0%.

---

mutalisk_dataframe_inform_user_of_metadata
Mutalisk Dataframe

Description

Adds metadata from a file to the mutalisk dataframe

Usage

```
mutalisk_dataframe_inform_user_of_metadata(mutalisk_dataframe, metadata)
```

Arguments

mutalisk_dataframe

a dataframe that can be produced using mutalisk_best_signature_directory_to_dataframe. Can also just make it yourself, if you want to visualise non-mutalisk data. Dataframe just needs 3 columns:

1. **SampleID**: a sample identifier.
2. **Signatures**: an identifier for a particular signature.
3. **Contributions**: the percentage contribution of the signature to the patients genetic profile (0.1 = 10%).

metadata

Either a path to csv file OR a dataframe. Must contain a header line which contains a SampleID column that matches that of mutalisk_dataframe (string)

Value

mutalisk dataframe with metadata columns (data.frame)

---

mutalisk_dataframe_metadata_column_names
Mutalisk Dataframe

Description

Get a vector of metadata columns from a mutalisk_dataframe.

Usage

```
mutalisk_dataframe_metadata_column_names(mutalisk_dataframe)
```

Arguments

mutalisk_dataframe

> a dataframe that can be produced using mutalisk_best_signature_directory_to_dataframe. Can also just make it yourself, if you want to visualise non-mutalisk data. Dataframe just needs 3 columns:
>
> 1. **SampleID**: a sample identifier.
> 2. **Signatures**: an identifier for a particular signature.
> 3. **Contributions**: the percentage contribution of the signature to the patients genetic profile (0.1 = 10%).

Value

a character vector containing names of metadata columns. If no metadata columns have been added, returns a zero length character vector. (character)

---

mutalisk_to_dataframe   *Mutalisk files to dataframe*

Description

Mutalisk files to dataframe

Usage

```
mutalisk_to_dataframe(mutalisk_files, sample_names_from_file_contents = FALSE)
```

Arguments

mutalisk_files   a vector of filepaths, each leading to the report.txt files output when downloading best_signature results for all vcfs in cohort

sample_names_from_file_contents

> guess sample names from filecontents instead of filenames (flag)

Value

a dataframe containing three columns:

1. **SampleID**: a sample identifier.
2. **Signatures**: an identifier for a particular signature.
3. **Contributions**: the percentage contribution of the signature to the patients genetic profile (0.1 = 10 percent).

---

mutalisk_to_dataframe_single_sample

Mutalisk to dataframe

Description

You probably want to use tags$strongmutalisk_to_dataframe instead. See ?mutalisk_to_dataframe

Usage

```
mutalisk_to_dataframe_single_sample(
  mutalisk_file,
  sample_names_from_file_contents = FALSE
)
```

Arguments

mutalisk_file   a vector of filepaths, each leading to the report.txt files output when downloading
                  best_signature results for all vcfs in cohort

sample_names_from_file_contents
                  guess sample names from filecontents instead of filenames (flag)

Value

tibble

---

pallette_cols23_customised

Custom Palette

Description

Custom Palette

Usage

```
pallette_cols23_customised(n)
```

Arguments

n   number of colors in pallete to return

Value

character vector of colors

Examples

```
pallette_cols23_customised(2)
```

```
plot_signature_contribution_jitterplot
```
Plot Signature-Level Dotplot Plots a signature-Level dotplot

Description

Plot Signature-Level Dotplot Plots a signature-Level dotplot

Usage

```
plot_signature_contribution_jitterplot(mutalisk_dataframe)
```

Arguments

mutalisk_dataframe

> a dataframe that can be produced using mutalisk_best_signature_directory_to_dataframe. Can also just make it yourself, if you want to visualise non-mutalisk data. Dataframe just needs 3 columns:
>
> 1. **SampleID**: a sample identifier.
> 2. **Signatures**: an identifier for a particular signature.
> 3. **Contributions**: the percentage contribution of the signature to the patients genetic profile (0.1 = 10%).

Value

a ggplot object

---

plot_stacked_bar   *Cohort-Level Mutational Signature Visualisation*

Description

A note of warning: for different mutalisk runs, this function will not enforce uniform colours for a single mutational signature. Better to get all data in at once and add a facet_wrap call

Usage

```
plot_stacked_bar(
  mutalisk_dataframe,
  lump_type = "min_prop",
  lump_min = 0.1,
  topn = 5,
  legend = "right",
  legend_direction = NA,
  pal = pals::kovesi.diverging_rainbow_bgymr_45_85_c67,
  color_of_other = "grey60",
  facet_column = NA,
  fontsize_strip = 18,
  fontsize_axis_title = 18
)
```

Arguments

mutalisk_dataframe

a dataframe that can be produced using mutalisk_best_signature_directory_to_dataframe. Can also just make it yourself, if you want to visualise non-mutalisk data. Dataframe just needs 3 columns:

1. **SampleID**: a sample identifier.
2. **Signatures**: an identifier for a particular signature.
3. **Contributions**: the percentage contribution of the signature to the patients genetic profile (0.1 = 10%).

lump_type   one of "min_prop", "topn", or "none".

min_prop will allow lump together all signatures that contribute less than **lump_min**. **topn** will keep the topn contributing signatures distinct, and lump the rest together (string)

lump_min   see **lump_type**

topn   see **lump_type**

legend   Where should the legend be placed? ("top", "left", "bottom", "right")

legend_direction

How should the legend be oriented. By defualt will guess based on position of legend ("vertical", "horizontal")

pal   Palette to use for generating colours.

color_of_other colour of signatures lumped into 'other' (string)

facet_column   name of column to use for faceting (string)

fontsize_strip fontsize of facet titles (number)

fontsize_axis_title

fontsize of axis titles (number)

Value

a ggplot (gg)

Index