Package 'errRt'

January 6, 2020

Type Package

Title Examine RT error rates through sequencing.

Version 1.0

Date 2018-03-01

Author Ashton Trey Belew, Keith Hughitt

Maintainer Ashton Trey Belew <abelew@gmail.com>

Description This is the R component of my current method for analyzing error rates of RT via sequencing. In its current form, it is intended to be used with errrt: github.com/abelew/errrt.

License GPL-2 | file LICENSE

Depends dplyr, tidyr

VignetteBuilder knitr

ByteCompile true

Encoding UTF-8

RoxygenNote 7.0.2

Collate '01_errRt.r'
'error_rate.r'
'plots.r'
'quant.r'

R topics documented:

barplot_matrices *Make a bar plot for every type of data returned by create_matrices.*

Description

The function create_matrices() creates two lists containing the various categorizations of the data. This takes that information and makes a bar plot for every element in those lists.

Usage

```
barplot_matrices(summary)
```

Arguments

summary	Result from create_matrices()

Value

List full of bar plots!

create_matrices *Given a samples sheet with some metadata, create a big pile of matrices describing the data.*

Description

Given a samples sheet with some metadata, create a big pile of matrices describing the data.

Usage

```
create_matrices(
  sample_sheet = "sample_sheets/all_samples.xlsx",
  ident_column = "identtable",
  mut_column = "mutationtable",
  min_reads = NULL,
  min_indexes = NULL,
  min_sequencer = 10,
  min_position = NULL,
  prune_n = TRUE,
  verbose = TRUE
)
```

Arguments

sample_sheet	xlsx/csv/whatever file of metadata.
ident_column	Column containing the files of reads identical to the template.
mut_column	Column containing the files of reads not identical to the template.
min_reads	Filter for the minimum number of reads / index.
min_indexes	Filter for the minimum numer of indexes / mutation.

min_sequencer	Filter defining the minimum number of reads when looking for sequencer-based mutations.
min_position	Filter against weird data at the beginning of the template.
prune_n	Remove mutations which are something to N.
verbose	Print information about the running of this function?

Value

List with lots of matrices of the resulting data.

---

errRt *errRt: Some functions for analyzing sequencing data intended to quantify RT error rates.*

Description

errRt: Some functions for analyzing sequencing data intended to quantify RT error rates.

---

expand_string *Combine the columns describing a mutation into a single column and categorize.*

Description

Given a table with columns including the position, mutation type, reference nt at that position, and product nt at that position; create a single column from it, standardize it, and make some categories describing each mutation. These columns currently include: 'mt': X_Y telling that this is a mutation from X to Y, 'transition_transversion': this is a transition or transversion mismatch, 'strong_weak': this mismatch went from strong->weak, weak->strong, weak->weak, or strong->strong. undef is used in the case of indels.

Usage

```
expand_string(t)
```

Arguments

t	Table from readr.

Value

The same table with some new columns describing the mutations therein.

`prune_indexes` *Summarize the mutant/identical data with respect to the number of reads/index.*

Description

This should provide a table of how many reads/index are (a)identical, (b)contain mutations, and the sum of (a + b). If a minimum number of reads is requested (e.g. min_reads is a number), return the list of indexes which have at least that many reads. This set of indexes may be used in other contexts to limit the data.

Usage

```
prune_indexes(chng, ident, min_reads = 3, verbose = TRUE)
```

Arguments

`chng`	The result of read_tsv() on a file containing a mutation table.
`ident`	The result of read_tsv() on a file containing identical reads.
`min_reads`	Minimum read / index filter.
`verbose`	Print information about this while it runs?

Value

List with the summary of the numbers of reads observed and the indexes kept. The list of indexes kept may just be 'all'.

`quantify_parsed` *Quantify the tables of reads identical/different to/from the template sequence.*

Description

The heavy lifting of locating reads which are/not identical to the template was performed by errrt.pm and returns a series of compressed tables of identical read ids and non-identical insertions, deletions, and mismatches. This function is intended to read those two files, gather the resulting data, and make some sense of it.

Usage

```
quantify_parsed(
  changed = NULL,
  identical = NULL,
  min_reads = NULL,
  min_indexes = NULL,
  min_sequencer = 10,
  prune_n = TRUE,
  min_position = 24,
  verbose = TRUE
)
```

Arguments

changed	File containing the set of changed reads/indexes, by default named 'step4.txt.xz' by errrt.pm.
identical	File containing the set of identical reads/indexes, by default named 'step2_identical_reads.txt.xz' by errrt.pm.
min_reads	Minimum number of reads for each index required to include each index in the final result.
min_indexes	Minimum number of indexes required to include a given mutation in the final result.
min_sequencer	Minimum number of total reads to consider an index as associated with a sequencer-based error.
prune_n	Remove mutations of a base to 'N'?
min_position	Filter mutations before this position?
verbose	Print information describing what is happening while this runs.

Value

List containing a bunch of summary information about the data.

---

sequencer_error *Calculate a lower limit error rate for the sequencer.*

Description

I say 'lower limit' because I assume my set of sequencer-based errors did not detect the full set of error actually from the sequencer. I chose to take only the set with >= 10 reads / index for which only n (1) read(s) is a mutant. My hope is that this will avoid false positives, but it will also limit the perceived sequencer error rate.

Usage

```
sequencer_error(data_summary)
```

Arguments

summary	Result from create_matrices().

Value

Estimate of errors/nucleotide deemed to originate from the sequencer.

Index