Package 'scCNAutils'

November 3, 2018

Title Functions to analyze copy number aberrations in single-cell data

Version 0.0.0.9000

Description Functions to analyze copy number aberrations in single-cell data. A bunch of scripts and workflows to read and analyze scRNA-seq data and look at CNA-oriented signal.

Depends R (¿= 3.4.4)

License MIT License + file LICENSE

Encoding UTF-8

LazyData true

Imports Matrix,
dplyr,
magrittr,
tidyr,
rlang,
parallel,
data.table

Suggests testthat

RoxygenNote 6.1.0

R topics documented:

bin_genes *Merge consecutive genes into expressed bins*

Description

Merge consecutive genes into expressed bins

Usage

```
bin_genes(ge_df, mean_exp = 3, nb_cores = 1)
```

Arguments

ge_df	the input gene expression with coordinate columns (chr, start, end) and then one column per cell.
mean_exp	the desired minimum mean expression in the bin.
nb_cores	the number of processors to use.

Value

a data.frame with bin expression.

Author(s)

Jean Monlong

convert_to_coord *Convert gene symbols to coordinates*

Description

If genes_coord is a filename, the file is expected to be a tab-delimited file with four columns: 'chr', 'start', 'end', 'symbol'. The order of the columns is not important.

Usage

```
convert_to_coord(ge_df, genes_coord, chrs = c(1:22, "X", "Y"),
  rm_dup = TRUE)
```

Arguments

ge_df	the data.frame with gene expression and one column 'symbol' with gene names.
genes_coord	either a file name or a data.frame with coordinates and gene names.
chrs	the chromosome names to keep. NULL to include all the chromosomes.
rm_dup	remove duplicated coordinates? Default is TRUE.

Details

The gene names in column 'symbol' should match the gene names in the input ge_df.

Value

a data.frame with columns 'chr', 'start', 'end' columns with genes coordinates (and still one column per barcode).

Author(s)

Jean Monlong

`norm_ge` *Normalize gene expression*

Description

Normalize gene expression

Usage

```
norm_ge(ge_df, method = c("tmm", "total"), nb_cores = 1)
```

Arguments

`ge_df`	the input gene expression with a 'symbol' column and then one column per cell.
`method`	the normalization method
`nb_cores`	the number of processors to use.

Value

a data.frame with the normalized expression.

Author(s)

Jean Monlong

qc_cells *Compute quality control metrics for each cell*

Description

If cell_cycle is provided it should be a data.frame (or a tsv file) with two columns: 'symbol' with gene names, and 'phase' with the cell cycle phase (e.g. either 'G1.S' or 'G2.M').

Usage

```
qc_cells(ge_df, cell_cycle = NULL)
```

Arguments

ge_df	the input gene expression with a 'symbol' column and then one column per cell.
cell_cycle	if non-null, either a file or data.frame to compute cell cycle scores. See details.

Value

a data.frame with qc metrics per cell.

Author(s)

Jean Monlong

read_mtx *Read a trio of genes, barcodes and mtx files.*

Description

Read a trio of genes, barcodes and mtx files.

Usage

```
read_mtx(mtx_file = "matrix.mtx", genes_file = "genes.tsv",
  barcodes_file = "barcodes.tsv", path = ".", rm_dup = TRUE,
  genes_col = 2)
```

Arguments

mtx_file	the path to the mtx file
genes_file	the path to the genes file.
barcodes_file	the path to the barcodes file
path	the path to the folder containing the files
rm_dup	remove duplicated gene names? Default is TRUE.
genes_col	the column to use in genes_file. Default is 2.

Value

a data.frame with a 'symbol' column with gene names and one column per barcode.

Author(s)

Jean Monlong

Index