Package 'scStratify'

January 21, 2020

Version 0.1.21

Author Jie Hao, Xin Zou, Jiawei Zou, Zeguang Han

Maintainer Jie Hao <j.hao@sjtu.edu.cn>

Title scStratify

Description This algorithm can estimate a gene filtering thresholding curve optimized for the given data. The filtering makes a balance between technical noise reduction and biological information preservation. Please cite: J. Hao, et al, Optimal Gene Filtering for Single-Cell data (OGFSC) <a8>C a gene filtering algorithm for single-cell RNA-seq data, Bioinformatics, 2019.

Depends reticulate, RColorBrewer, pls, ggplot2, gplots, FNN, MASS, qvalue

License GPL-2

LazyLoad yes

NeedsCompilation no

R topics documented:

`DEgeneId` *DEgeneId*

Description

Perform OGFSC gene filtering and identify noise robust DE genes.

Usage

```
DEgeneId(Ctr, Case, nBins = 20, alpha = 0.5, plot_option = 0)
```

Arguments

Ctr	The scRNA-seq data from the control group, with genes versus cells
Case	The scRNA-seq data from the case group, with genes versus cells
Case	The scRNA-seq data from the case group, with genes versus cells
nBins	A OGFSC parameter, defining the size of each gene bin. The default value is 20, but the users can try 5-30.
alpha	A OGFSC parameter, defining the candidate gene filtering curves. The default value is 0.5.
plot_option	A OGFSC parameter, defining the plot option. 1 - the OGFSC figures are shown, 0 - otherwise.

Value

Ctr_filtered	The control data with OGFSC selected genes only.
Case_filtered	The case data with OGFSC selected genes only.
idx_OGFSC	The index of genes selected by OGFSC.
idx_DEgenes	The index of DE genes.

---

findAnchors *findAnchors*

Description

To find anchor cells for PLS1 model construction using the KNN method.

Usage

```
findAnchors(Ctr, Case, k = 3, dim = 5)
```

Arguments

Ctr	The OGFSC filtered control data.
Case	The OGFSC filtered case data.
k	The number of neighbours for KNN. The default value is 3.
dim	The number of principle components used for data normalization. The default value is 5.

Value

anchorCells	The list of anchor cells, paired with control (column 1) vs. case (column 2).

GSEAanalysis *GSEAanalysis*

Description

The GSEA analysis of the DE genes identified by scStratify.

Usage

```
GSEAanalysis(stratifiedDEgenes, species = "1", geneset = 1, qvalue = 0.05,
  outdir = getwd(), driverpath = paste(system.file("extdata",
    package = "scStratify"), "/chromedriver.exe", sep = ""),
  outputfileName = "stratified_GSEA.csv")
```

Arguments

stratifiedDEgenes	The output of scStratify function.
species	Select the species. Input one of 'Human', 'Mouse' and 'Rat'.
geneset	Type 'hallmark' or 'Hallmark' represent hallmark, Type 'KEGG' or 'kegg' represent KEGG, Type 'BP' or 'GO biological process' represent GO biological process, Type 'CC' or 'GO cellular component' represent GO cellular component, Type 'MF' or 'GO molecular function' represent GO molecular function.
qvalue	The FDR q-value cutoff for enriched GSEA terms identification.
outdir	The directory of output files. By default, it is the current working directory.
driverpath	The webdriver path. For windows OS, Google Chrome or Firefox is need. Download driver from chrome: http://npm.taobao.org/mirrors/chromedriver. Type about://version or chrome://version in the chrome address bar to get your chrome version.
outputfileName	The enriched GSEA terms for different cell bins.

scStratify *scStratify*

Description

The main function to profile the amount of variations related to the biological quesion of interest. The function will sort the cells from the case group according to the amount of signal they contain. The cells from the case group are either categorized into S-resp and W-resp subgroups, or a cell transition trajectory is reconstructed by stratifying the cells into up to 4 bins.

Usage

```
scStratify(Ctr, Case, anchorCells, idx_DEgenes, trajectory, geneList,
  legendPosition1 = "topright", legendPosition2 = "topleft")
```

Arguments

`Ctr, Case`	The OGFSC filtered data from the control and the case group, respectively.
`anchorCells`	The paired anchor cells identified using the findAnchors function.
`trajectory`	The option to perform trajectogory analysis. If "1", the case cells will be stratified into up to 4 bins and the cell trancition trajectory construction analysis will be performed. "0" means the case cells will be categorized into S-resp and W-resp sub-groups.
`geneList`	The list of gene symbles.
`legendPosition1`	Specify the location of figure legend for the SS_PLS2 distribution plot. The argument is "topright" or "topleft".
`legendPosition2`	Specify the location of figure legend for the transition velocty scatter plot. The argument is "topright" or "topleft".

Value

`scStratify_idx`	The case cell categories. For trajectory analysis,
`Ctr_corrected`	The control cells with batch effect corrected.
`Case_corrected`	The case cells with batch effect corrected.
`scStratify_metric`	The value of SS_PLS2 of case cells.
`Hartigans`	The results of Hartigans bimodality test.
`stratifiedDEgenes`	The DE genes identified from each bin.
`P`	The correlation p values of transition velocity.

Index