Package 'STged'

November 27, 2023

Type Package

Title STged: Gene expression deconvolution for spatial transcriptomic data

Version 0.1.0

Author Jian-Juan Tu

Maintainer Jian-Juan Tu<tujiajuan@163.com>

Description STged is a gene expression deconvolution mathod. STged employs graph-guided spatial correlations and prior-guided gene expression similarities within a non-negative least-squares regression framework.

Depends R (>= 4.2.1)

Imports MASS, Matrix, nnls, reticulate

Suggests knitr, rmarkdown

VignetteBuilder knitr

License GPL(>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 7.2.3

R topics documented:

`create_group_exp` *This function focuses on cleaning scRNA-seq and SRT datasets.*

Description

This function focuses on cleaning scRNA-seq and SRT datasets.

Usage

```
create_group_exp(sc_exp, sc_label)
```

Arguments

sc_exp	scRNA-seq matrix, genes * cells. The format should be raw-counts. The matrix need include gene names and cell names.
sc_label	cell type information. The cells are need be divided into multiple category.

Examples

```
data("PDAC")
refmu = create_group_exp(sc_exp,sc_label)
```

---

data_process *This function focuses on cleaning scRNA-seq and SRT datasets.*

Description

This function focuses on cleaning scRNA-seq and SRT datasets.

Usage

```
data_process(
  sc_exp,
  sc_label,
  spot_exp,
  spot_loc,
  gene_det_in_min_cells_per = 0.01,
  expression_threshold = 0,
  nUMI = 100,
  verbose = FALSE,
  depthscale = 1,
  clean.only = FALSE
)
```

Arguments

sc_exp	scRNA-seq matrix, genes * cells. The format should be raw-counts. The matrix need include gene names and cell names.
sc_label	cell type information. The cells are need be divided into multiple category.
spot_exp	SRT gene expression matrix, genes * spots. The format should be raw counts. The matrix need include gene names and spot names.
spot_loc	coordinate matrix, spots * coordinates. The matrix need include spot names and coordinate name (x, y).
gene_det_in_min_cells_per	a floor variable. minimum percent of genes that need to be detected in a cell.
expression_threshold	a floor variable. Threshold to consider a gene expressed.
nUMI	a floor variable. minimum of read count that need to be detected in a cell or spot.
verbose	a logical variable that defines whether to print the processing flow of data process.

Value

a list includes processed scRNA-seq matrix, cell type, stRNA-seq matrix.

Examples

```
data("PDAC")
datax = data_process(sc_exp = sc_exp, sc_label = sc_label, spot_exp = spot_exp, spot_loc = spot_loc)
```

`dis_weight` *This function focuses on cleaning scRNA-seq and SRT datasets.*

Description

This function focuses on cleaning scRNA-seq and SRT datasets.

Usage

```
dis_weight(
  spot_loc = spot_loc,
  spot_exp = spot_exp,
  k = 6,
  quantile_prob_bandwidth = 1/3,
  method = "Hex",
  coord_type = "grid"
)
```

Arguments

`spot_loc`	coordinate matrix, spots * coordinates. The matrix need include spot names and coordinate name (x, y).
`spot_exp`	stRNA-seq matrix, genes * spots. The format should be raw-counts. The matrix need include gene names and spot names.
`k`	number of neighbor spots for construct spatial neighboring graph, details refer to Squidpy.
`quantile_prob_bandwidth`	selection of bandwidth of the spatial kernel of each spot.
`method`	the method used to construct spatial neighboring graph, details refer to Squidpy.
`coord_type`	grid coordinates, details refer to Squidpy.

PDAC *Input data of the STged*

Description

Input data of the STged

Usage

```
data(PDAC)
```

Format

a list of matrices contains SRT gene expression matrix with coordinates information and corresponding cell type proportion information, and annotated scRNA-seq data as cell type reference (bottom).

Examples

```
data(PDAC)
```

STged *STged: Gene expression deconvolution for spatial transcriptomic data*

Description

STged is a gene expression deconvolution mathod. STged employs graph-guided spatial correlations and prior-guided gene expression similarities within a non-negative least-squares regression framework.

Usage

```
model.est = STged(sc_exp = sc_exp, sc_label = sc_label, spot_exp = spot_exp, spot_loc = spot_loc
```

Arguments

sc_exp	scRNA-seq matrix, genes * cells. The format should be raw-counts. The matrix need include gene names and cell names.
sc_label	cell type information. The cell need be divided into multiple categories.
spot_exp	stRNA-seq matrix, genes * spots. The format should be raw-counts. The matrix need include gene names and spot names.
spot_loc	coordinate matrix, spots * coordinates. The matrix need include spot names and coordinate name (x, y).
beta	cell type proportion matrix, spots * cell types.
gene_det_in_min_cells_per	a floor variable. minimum percent # of genes that need to be detected in a cell.
expression_threshold	a floor variable. Threshold to consider a gene expressed.

nUMI	a floor variable. minimum # of read count that need to be detected in a cell or spot.
verbose	a logical variable that defines whether to print the processing flow of data process.
clean.only	a logical variable that defines whether to normalize data with log transform.
python_env	the path of python environment.
truncate	a logical variable. that defines whether to truncate the gene expression data for both scRNA-seq and SRT.
qt	Winsorize expression values to prevent outliers. Values below this quantile and above 1-this quantile will be set to the quantile value.
knei	number of neighbor spots for construct spatial neighboring graph, details refer to Squidpy.
methodL	the method used to construct spatial neighboring graph, details refer to Squidpy.
coord_type	grid coordinates, details refer to Squidpy.
quantile_prob_bandwidth	selection of bandwidth of the spatial kernel of each spot.
lambda1	tuning parameter to balance the graph regularization term. If the tuning parameter is set to NULL, then, we will adopt the value in our algorithm.
lambda2	tuning parameter to balance the prior regularization term. If the tuning parameter is set to NULL, then, we will adopt the value in our algorithm.
cutoff	a cutoff value of cell type proportion.
rho	the penalty parameter in the ADMM algorithm.
rho.incr	the increase step parameter for varying penalty parameter rho.
rho.max	the maximum value of rho.
maxiter	a positive integer represents the maximum number of updating algorithm. Default setting is 100.
epsilon	a parameter represents the stop criterion.

Index