Package 'STged'

June 28, 2024

Type Package

Title Precise gene expression deconvolution in spatial transcriptomics with STged

Version 0.1.0

Author Jian-Juan Tu

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

Description STged uses graph-driven spatial structure correlations and reference gene signature-driven 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` *Create Grouped Expression Matrix by Cell Type*

Description

This function computes the mean expression for each cell type from scRNA-seq data. It groups cells by their specified type and calculates the mean of all cells within each type to create a summarized expression profile per cell type.

Usage

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

Arguments

`sc_exp`	A matrix of single-cell RNA-seq data (genes x cells), containing expression counts.
`sc_label`	A vector of labels corresponding to the cells in the sc_exp matrix.

Value

A matrix of mean expression values for each cell type (genes x cell types). # Assuming 'sc_data' is a matrix of gene expression data and 'cell_types' is a vector of cell type labels

Examples

```
data(Fishplus)
refmu = create_group_exp(sc_exp,sc_label)
```

`data_process` *Data Processing for Spatial and Single-cell Transcriptomics. Processes and cleans spatial transcriptomics (SRT) and single-cell RNA-seq (scRNA-seq) data by filtering genes and cells/spots based on specified criteria. It aligns the gene expression data between scRNA-seq and SRT datasets by retaining only the common genes.*

Description

Data Processing for Spatial and Single-cell Transcriptomics. Processes and cleans spatial transcriptomics (SRT) and single-cell RNA-seq (scRNA-seq) data by filtering genes and cells/spots based on specified criteria. It aligns the gene expression data between scRNA-seq and SRT datasets by retaining only the common genes.

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 = 1e+06,
  clean.only = TRUE
)
```

Arguments

sc_exp	A matrix of scRNA-seq data (genes x cells) containing expression counts.
sc_label	A vector of labels corresponding to the cells in the sc_exp matrix.
spot_exp	A matrix of spatial transcriptomics data (genes x spots) containing expression counts.
spot_loc	A matrix of coordinates for each spot in the spot_exp matrix.
gene_det_in_min_cells_per	Minimum percentage of cells in which a gene must be detected.
expression_threshold	The minimum expression level a gene must exhibit to be considered expressed.
nUMI	Minimum number of UMIs (unique molecular identifiers) required for a cell/spot to be included.
verbose	Logical; if TRUE, detailed processing information will be printed.
depthscale	A scaling factor used for normalization (default is 100,000).
clean.only	Logical; if TRUE, only performs data cleaning without additional normalization.

Value

A list containing filtered and processed scRNA-seq and SRT data: - ‘sc_exp‘: Filtered scRNA-seq expression matrix. - ‘sc_label‘: Corresponding cell labels for the filtered scRNA-seq data. - ‘spot_exp‘: Filtered SRT expression matrix.

Examples

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

dis_weight *Calculate Weighted Adjacency Matrix for Spatial Data*

Description

This function computes the weighted adjacency matrix for spatial data based on the connectivity and distances between spots. It supports only the 'Hex' method for defining spatial neighbors in a hexagonal grid system.

Usage

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

Arguments

spot_loc	A matrix of coordinates for each spot, typically named (x, y).
spot_exp	A matrix of spatial transcriptomics data (genes x spots) containing expression counts.
k	Integer; number of neighboring spots to consider when constructing the spatial neighboring graph.
quantile_prob_bandwidth	The bandwidth quantile for weighting the adjacency matrix.
method	Method used for constructing the spatial neighboring graph. Currently supports only 'Hex'.
coord_type	Type of coordinates system used, typically 'grid' for standard spatial transcriptomics data.

Value

A list containing: - 'dis_weight': The Gaussian-weighted adjacency matrix based on spatial distances and connectivity. - 'weight_adj': The unweighted adjacency matrix derived from spatial connectivity alone.

Examples

```
# Assuming `coordinates` is a matrix of spatial coordinates and `expression` is the corresponding expression da
data(Fishplus)
weights <- dis_weight(spot_loc = spot_loc, spot_exp = spot_exp, k = 6, method = "Hex", coord_type = "grid")
```

Fishplus *Input data of the STged*

Description

Input data of the STged

Usage

```
data(Fishplus)
```

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(Fishplus)
```

MUR.STged *Multiplicative update rule (MUR) for Spatial Transcriptomics Gene Expression Deconvolution (MUR.STged) This function performs the gene expression deconvolution using a multilevel perceptron approach for spatial transcriptomics data.*

Description

Multiplicative update rule (MUR) for Spatial Transcriptomics Gene Expression Deconvolution (MUR.STged) This function performs the gene expression deconvolution using a multilevel perceptron approach for spatial transcriptomics data.

Usage

```
MUR.STged(
  srt_exp = srt_exp,
  ref_exp = ref_exp,
  beta.type = beta.type,
  W = W,
  lambda1 = lambda1,
  lambda2 = lambda2,
  cutoff = 0.05,
  epsilon = 1e-05,
  maxiter = 100
)
```

Arguments

srt_exp	A matrix of spatial transcriptomics RNA-seq gene expression data (genes x spots), containing raw counts.
ref_exp	A matrix of reference single-cell RNA-seq gene expression data (genes x cell types).
beta.type	A matrix describing the initial cell type proportions at each spot (spots x cell types).
W	A spatial weight matrix (spots x spots) describing the spatial correlation between spots.
lambda1	Regularization parameter for the graph regularization term. Automatically determined if set to NULL.
lambda2	Regularization parameter for the prior regularization term. Automatically determined if set to NULL.
cutoff	Cutoff value for cell type proportion to filter insignificant cell type contributions. Default is 0.05.
epsilon	Convergence threshold for stopping the algorithm. Default is 1e-5.
maxiter	Maximum number of iterations allowed in the algorithm. Default is 100.

Value

A list containing the following elements:

V.hat	Estimated gene expression matrix (genes x spots).
F_list	List of estimated cell type-specific gene expression matrices.
lambda1	Selected value of lambda1.
lambda2	Selected value of lambda2.
beta	Final estimated cell type proportions matrix (spots x cell types).
obj.loss	Objective loss value of the final model.

Examples

```
## Not run:
# Example usage:
result <- MUR.STged(srt_exp = spatial_exp, ref_exp = reference_exp, beta.type = beta,
                    W = spatial_weights, lambda1 = NULL, lambda2 = NULL, cutoff = 0.05,
                    epsilon = 1e-5, maxiter = 100)

## End(Not run)
```

STged *Spatial Transcriptomics Gene Expression Deconvolution (STged)*

Description

Spatial Transcriptomics Gene Expression Deconvolution (STged)

Usage

```
model_results <- STged(sc_exp = sc_exp, sc_label = sc_label,
                       spot_exp = spot_exp, spot_loc = spot_loc, beta = beta)
```

Arguments

sc_exp	A matrix of single-cell RNA-seq gene expression data (genes x cells), containing raw counts.
sc_label	A vector or factor of cell type labels corresponding to each cell in the sc_exp data.
spot_exp	A matrix of spatial transcriptomics RNA-seq gene expression data (genes x spots), containing raw counts.
spot_loc	A matrix specifying the coordinates of each spot (spots x coordinates), typically named (x, y).
beta	A matrix describing the cell type proportions at each spot (spots x cell types).
gene_det_in_min_cells_per	Minimum percentage of genes that must be detected in a cell, used as a filtering criterion.
expression_threshold	Minimum expression level to consider a gene as expressed, used for data cleaning.
nUMI	Minimum number of read counts required to consider a cell or spot, used for data cleaning.
verbose	Logical; if TRUE, prints details of the data processing steps.
clean.only	Logical; if TRUE, applies only data cleaning steps without normalization.
depthscale	A scaling factor used for normalization (default is 100,000).
python_env	Path to the Python environment, used for calling Python functions.
truncate	Logical; if TRUE, truncates gene expression data to reduce the impact of extreme values.
qt	Quantile used for Winsorizing expression values to control the effects of outliers.
knei	Integer; number of neighboring spots to consider when constructing the spatial neighboring graph.
methodL	Method used for constructing the spatial neighboring graph, details provided by the Squidpy package.
coord_type	Type of coordinates system used, e.g., 'grid' as described by Squidpy.
quantile_prob_bandwidth	Bandwidth for the spatial kernel used in constructing the neighboring graph.
lambda1	Tuning parameter for balancing the graph regularization term. Automatically determined if set to NULL.
lambda2	Tuning parameter for balancing the prior regularization term. Automatically determined if set to NULL.
cutoff	Cutoff value for cell type proportion to filter insignificant cell type contributions.
maxiter	Maximum number of iterations allowed in the algorithm.
epsilon	Convergence threshold for stopping the algorithm.
rho	Penalty parameter in the ADMM algorithm, with adaptive adjustments.
rho.incr	Increment step for varying the penalty parameter rho.
rho.max	Maximum allowable value for rho.

Value

A list containing model estimation results, including estimated parameters and diagnostics.

Index