Package: SpaCCI (via r-universe)

October 25, 2024

Type Package

Title Spatially Aware Cell-Cell Interaction Analysis

Version 1.0.2

Date 2024-10-23

Maintainer Li-Ting Ku <lku@mdanderson.org>

Description Provides tools for analyzing spatial cell-cell interactions based on ligand-receptor pairs, including functions for local, regional, and global analysis using spatial transcriptomics data. Integrates with databases like 'CellChat' <http://www.cellchat.org/>, 'CellPhoneDB' <https://www.cellphonedb.org/>, 'Cellinker' <https://www.rna-society.org/cellinker/>, 'ICELLNET' <https://github.com/soumelis-lab/ICELLNET>, and 'ConnectomeDB' <https://humanconnectome.org/software/connectomedb/> to identify ligand-receptor pairs, visualize interactions through heatmaps, chord diagrams, and infer interactions on different spatial scales.

License GPL (>= 2)

Encoding UTF-8

Depends R (>= 3.5.0)

Imports Rcpp (>= 1.0.13), Seurat (>= 4.0.0), nnls, ggrepel, pheatmap, circlize (>= 0.4.12), Matrix, dplyr, patchwork, grDevices, reshape2, graphics, ggplot2

LinkingTo Rcpp, RcppArmadillo

RoxygenNote 7.3.2

NeedsCompilation yes

Author Li-Ting Ku [aut, cre]

Repository CRAN

Date/Publication 2024-10-24 10:30:51 UTC

Contents

`Find_regional_IDs` *Find the spatial neighborhood Spot IDs*

Description

This function identifies the spatial neighborhood Spot IDs around a given center Spot ID within a specified radius.

Usage

```
Find_regional_IDs(
  object,
  spatial_coord,
  centerID,
  enhanced = FALSE,
  radius,
  avern = 5
)
```

Arguments

object	An object that could be either 1. A Seurat object, or 2. A data frame where the columns are Spot_IDs (i.e., the gene*spot expression matrix).
spatial_coord	A data frame of the spatial coordinates. The column names should include 'c("Spot_ID", "imagerow", "imagecol")', and the row names must be the Spot_ID, which is the same as the row names in the cell type proportion data frame or the column names of the gene*spot expression data frame.
centerID	A vector of length 1, representing a single Spot_ID that serves as the center for the neighborhood.
enhanced	Logical; if 'TRUE', enhances the Seurat object by marking the neighborhood in a special way. Defaults to 'FALSE'.
radius	The radius of the spatial neighborhood, specified as a numeric value.
avern	Numeric; the number of samples to average over when determining the unit distance. Defaults to 5.

Value

A list containing:

centerID The input center Spot_ID.

closeID The Spot_IDs of the neighboring spots within the specified radius.

unit The unit distance used to determine the neighborhood.

GetShuffledCT *Perform a Deranged Shuffle of Cell Types*

Description

This function takes a vector of cell types and returns a shuffled version where no element remains in its original position.

Usage

```
GetShuffledCT(CellType)
```

Arguments

CellType	A character vector representing the cell types to be shuffled.

Value

A character vector of the same length as 'CellType', with elements shuffled such that no element remains in its original position.

Examples

```
original <- c("B_cell", "T_cell", "NK_cell", "Macrophage")
shuffled <- GetShuffledCT(original)
print(shuffled)
```

`Global_Permutations` *Perform Global Permutations*

Description

This function performs global permutations on the spatial transcriptomics data.

Usage

```
Global_Permutations(
  permutationMatrix,
  permut_null_regionMatrix,
  permut_col,
  cellPropMatrix,
  spotGeneMatrix,
  LigandVectorIndex,
  ReceptorVectorIndex,
  null_expression,
  nBoot
)
```

Arguments

`permutationMatrix`	A matrix containing permutations.
`permut_null_regionMatrix`	A matrix of null region permutations.
`permut_col`	A column matrix of permutations.
`cellPropMatrix`	A matrix of cell type proportions.
`spotGeneMatrix`	A matrix of gene expressions at spots.
`LigandVectorIndex`	A vector of ligand indices.
`ReceptorVectorIndex`	A vector of receptor indices.
`null_expression`	A matrix of null expression values.
`nBoot`	Number of bootstrap iterations.

Value

A matrix with the results of the global permutations.

`Local_Regional_Permutations` *Perform Local and Regional Permutations*

Description

This function performs local and regional permutations on the spatial transcriptomics data.

Usage

```
Local_Regional_Permutations(
  permutationMatrix,
  permut_col,
  cellPropMatrix,
  spotGeneMatrix,
  LigandVectorIndex,
  ReceptorVectorIndex,
  null_expression,
  nBoot
)
```

Arguments

`permutationMatrix`	A matrix containing permutations.
`permut_col`	A column matrix of permutations.
`cellPropMatrix`	A matrix of cell type proportions.
`spotGeneMatrix`	A matrix of gene expressions at spots.
`LigandVectorIndex`	A vector of ligand indices.
`ReceptorVectorIndex`	A vector of receptor indices.
`null_expression`	A matrix of null expression values.
`nBoot`	Number of bootstrap iterations.

Value

A matrix with the results of the local and regional permutations.

LR_database *Identify Possible Ligand-Receptor Pairs for Cell-Cell Communication*

Description

This function identifies possible ligand-receptor (L-R) pairs for cell-cell communication analysis using a selected database. It checks for the presence of all genes involved in each L-R pair within the provided gene expression matrix, filtering based on a specified expression percentage threshold. The function supports multiple databases including CellChat, CellPhoneDB, Cellinker, ICELLNET, and ConnectomeDB.

Usage

```
LR_database(
  species,
  database_name,
  gene_spot_expression_dataframe,
  percentage = 10
)
```

Arguments

species	A string specifying the species ("Human" or "Mouse").
database_name	A string specifying the L-R database to use. Options include "CellChat", "CellPhoneDB", "Cellinker", "ICELLNET", and "ConnectomeDB".
gene_spot_expression_dataframe	A gene expression data frame with genes as row names and Spot IDs as column names. This data frame is used to verify the presence of all genes involved in the L-R pairs.
percentage	A numeric value specifying the minimum percentage of spots in which a gene must be expressed to be considered. The default is 10, meaning the gene express over 10% of spots.

Value

A list containing:

possible_LR_pairs A data frame of L-R pairs where all genes are present in the gene_spot_expression_dataframe and meet the expression threshold. The data frame includes the ligand and receptor vectors, and the combined gene vectors.

possible_LR_pairs_info A data frame with detailed information about the identified L-R pairs, including their original annotations from the selected database.

Examples

```
library(SpaCCI)
#Load the example data
data(test_data)
gene_spot_df <- test_data$gene_spot_df
result <- LR_database(species = "Human",
            database_name = "CellChat",
            gene_spot_expression_dataframe = gene_spot_df)
```

`plot_localized` *Plot Localized Hotspot Pattern*

Description

Visualize the inferred cell-cell interaction localized pattern if NOT using Seurat Object

Usage

```
plot_localized(
  spatial_coord,
  resultdf_list,
  RegionIDs_matrix,
  celltype_ligand,
  celltype_receptor,
  plot_size,
  L_R_pair_name = NULL,
  alpha = 0.05
)
```

Arguments

`spatial_coord`	A data frame of the spatial coordinates. The columns should include `"Spot_ID"`, `"imagerow"`, and `"imagecol"`. And the row names must be the names of `"Spot_ID"`, which is the same as the rownames in cell type proportion data frame or the colnames of the gene* spot expression data frame
`resultdf_list`	A result of data frame list from the output of `run_SpaCCI(..., analysis_scale = "local",...)` `` `dataframelist` ``
`RegionIDs_matrix`	A result of matrix list from the output of `run_SpaCCI(..., analysis_scale = "local",...)` `` `RegionIDs_matrix` ``
`celltype_ligand`	Ligand cell type string inputted by user, the name of the cell type should match the names in the `` `spot_cell_proportion_dataframe` `` during the `run_SpaCCI` analysis.

`celltype_receptor`
: Receptor cell type string inputted by user, the name of the cell type should match the names in the `` `spot_cell_proportion_dataframe` `` during the `run_SpaCCI` analysis.

`plot_size`
: As this function incorporate with `Seurat`'s `` `SpatialFeaturePlot` ``, this parameter could control the plotting size of the each spot.

`L_R_pair_name`
: Initially this is set to `NULL`, if one is interested in a specific Ligand-Receptor pair, then one could specify the L_R_pair_name here. Note: the input name should match the L-R pair name exists in the dataframe in the output of SpaCCI_local "dataframelist".

`alpha`
: This is the significant cutoff for the adjusted-p-value of thr permutation test. Initially this is set to `0.05`, one could adjust the cutoff.

Value

The localized plot from the inferred cell-cell interaction on the local scale.

Examples

```
# Run localized hotspot plot
Result <- run_SpaCCI(..., analysis_scale = "local",...)
local_plot <- plot_localized(spatial_coord = spatial_coords_df,
                             resultdf_list = Result$dataframelist,
                             RegionIDs_matrix = Result$RegionIDs_matrix,
                             celltype_ligand = "Beta_cells",
                             celltype_receptor = "T_ells",
                             plot_size = 3)
```

---

plot_localized_Seurat *Plot Localized Hotspot Pattern on Seurat Object*

Description

Visualize the inferred cell-cell interaction localized pattern on the tissue image with Seurat_object

Usage

```
plot_localized_Seurat(
  Seurat_object,
  resultdf_list,
  RegionIDs_matrix,
  celltype_ligand,
  celltype_receptor,
  plot_size,
  L_R_pair_name = NULL,
  alpha = 0.05
)
```

Arguments

Seurat_object	A Seurat object
resultdf_list	A result of data frame list from the output of run_SpaCCI(..., analysis_scale = "local",...) `dataframelist`
RegionIDs_matrix	A result of matrix list from the output of run_SpaCCI(..., analysis_scale = "local",...) `RegionIDs_matrix`
celltype_ligand	Ligand cell type string inputted by user, the name of the cell type should match the names in the `spot_cell_proportion_dataframe` during the run_SpaCCI analysis.
celltype_receptor	Receptor cell type string inputted by user, the name of the cell type should match the names in the `spot_cell_proportion_dataframe` during the run_SpaCCI analysis.
plot_size	As this function incorporate with Seurat's `SpatialFeaturePlot`, this parameter could control the plotting size of the each spot.
L_R_pair_name	Initially this is set to NULL, if one is interested in a specific Ligand-Receptor pair, then one could specify the L_R_pair_name here. Note: the input name should match the L-R pair name exists in the dataframe in the output of SpaCCI_local "dataframelist".
alpha	This is the significant cutoff for the adjusted-p-value of thr permutation test. Initially this is set to 0.05, one could adjust the cutoff.

Value

The localized plot from the inferred cell-cell interaction on the local scale.

Examples

```
# Not Run

# Run localized hotspot plot
Result <- run_SpaCCI(..., analysis_scale = "local",...)
local_plot <- plot_localized_Seurat(Seurat_object = gene_spot_df,
                                    resultdf_list = Result$dataframelist,
                                    RegionIDs_matrix = Result$RegionIDs_matrix,
                                    celltype_ligand = "Beta_cells",
                                    celltype_receptor = "T_ells",
                                    plot_size = 3)
```

plot_SpaCCI_chordDiagram *Plot SpaCCI Results on Chord Diagram*

Description

This function generates a chord diagram to visualize cell-cell interactions based on ligand-receptor pairs. The interactions can be filtered by specific cell types, pathways, or interaction names.

Usage

```
plot_SpaCCI_chordDiagram(
  SpaCCI_Result_List,
  specific_celltypes = NULL,
  pathway_name = NULL,
  L_R_pair_name = NULL,
  color = NULL,
  alpha = 0.05
)
```

Arguments

SpaCCI_Result_List	A list containing the results from a SpaCCI "regional" or "global" analysis. This list should include pvalue_df, which are the outputs from run_SpaCCI(..., analysis_scale = "regional",...) or run_SpaCCI(..., analysis_scale = "global",...).
specific_celltypes	A character vector specifying the cell types to include in the plot, RECOMMEND using colnames of cell type proportion matrix to include all cell types. If NULL, cell types that involved in significant interactions are included.
pathway_name	A single character string specifying the pathway name to filter the interactions. If NULL, all pathways are included.
L_R_pair_name	A character vector specifying the ligand-receptor pair names to include in the plot. If NULL, all interactions are included.
color	A named vector of colors to use for the cell types. If NULL, a default color palette is used.
alpha	A numeric value specifying the significance threshold for adjusted P-values. Initially, set to 0.05.

Value

A chord diagram plot visualizing the significant cell-cell interactions.

Examples

```
library(SpaCCI)
library(dplyr)
library(circlize)
data(result_global)
celltypes <- c("beta" , "delta" , "ductal","macrophage",
               "activated_stellate", "quiescent_stellate")
# Run the result chordDiagram for global analysis
plot_SpaCCI_chordDiagram(SpaCCI_Result_List = result_global,
                         specific_celltypes = c(celltypes),
                         L_R_pair_name  = "AREG_EGFR")
```

`plot_SpaCCI_heatmap` *Plot SpaCCI Results on the heatmap Visualize inferred significant cell-cell interactions using a heatmap*

Description

Plot SpaCCI Results on the heatmap Visualize inferred significant cell-cell interactions using a heatmap

Usage

```
plot_SpaCCI_heatmap(
  SpaCCI_Result_List,
  specific_celltypes = NULL,
  pathways = NULL,
  interaction = NULL,
  log1p_transform = FALSE,
  show_rownames = TRUE,
  show_colnames = TRUE,
  scale = "none",
  cluster_cols = TRUE,
  cluster_rows = TRUE,
  border_color = "white",
  fontsize_row = 11,
  fontsize_col = 11,
  family = "Arial",
  main = "",
  treeheight_col = 0,
  treeheight_row = 0,
  low_col = "dodgerblue4",
  mid_col = "peachpuff",
  high_col = "deeppink4",
  alpha = 0.05,
  return_tables = FALSE,
```

```
  symmetrical = FALSE,
  ...
)
```

Arguments

SpaCCI_Result_List	A list containing the results from a SpaCCI "regional" or "global" analysis. This list should include pvalue_df, which are the outputs from run_SpaCCI(..., analysis_scale = "regional",...) or run_SpaCCI(..., analysis_scale = "global",...).
specific_celltypes	A vector of cell types to include in the heatmap, i.e c("Celltype_A","Celltype_B"). NOTE: the cell type names should match the names input in the SpaCCI analysis.
pathways	A vector of pathways to filter the interactions. Initially set to NULL, if not, then it will aggregate the results of the selected pathways.
interaction	A vector of interactions to filter. Initially set to NULL, if not, then it will aggregate the results of the selected interactions.
log1p_transform	Logical; whether to apply a log(1 + x) transformation to the count matrix.
show_rownames	Logical; whether to show row names in the heatmap.
show_colnames	Logical; whether to show column names in the heatmap.
scale	Character; whether to scale the data ("row", "column", "none").
cluster_cols	Logical; whether to cluster columns.
cluster_rows	Logical; whether to cluster rows.
border_color	Character; color of the heatmap borders.
fontsize_row	Numeric; font size for row names.
fontsize_col	Numeric; font size for column names.
family	Character; font family for text in the heatmap.
main	Character; title of the heatmap.
treeheight_col	Numeric; height of the column dendrogram.
treeheight_row	Numeric; height of the row dendrogram.
low_col	Character; color for low values in the heatmap.
mid_col	Character; color for mid values in the heatmap.
high_col	Character; color for high values in the heatmap.
alpha	Numeric; significance threshold for p-values, initailly set to 0.05.
return_tables	Logical; whether to return the count matrix and summary tables.
symmetrical	Logical; whether to make the heatmap symmetrical.
...	Additional arguments passed to ‘pheatmap‘.

Value

If 'return_tables' is FALSE (default), the function returns a heatmap object created by pheatmap, showing the count of significant cell-cell interactions. If 'return_tables' is TRUE, the function returns a list containing:

heatmap The heatmap object showing the significant cell-cell interactions.

heatmap_countmatrix The matrix used to generate the heatmap, with cell types as rows and columns, and counts of significant interactions as values.

table A data frame summarizing the counts of significant interactions between each ligand and receptor cell type combination.

Examples

```
library(SpaCCI)
library(dplyr)
library(reshape2)
library(grDevices)
library(pheatmap)
data(result_global)
celltypes <- c("beta" , "delta" , "ductal","macrophage",
               "activated_stellate", "quiescent_stellate")
plot_SpaCCI_heatmap(SpaCCI_Result_List = result_global,
                    symmetrical = FALSE, cluster_cols = FALSE, return_tables = FALSE,
                    cluster_rows = FALSE, #cellheight = 10, cellwidth = 10,
                    specific_celltypes = c(celltypes),
                    main= "Cell-Cell Interaction Count")
```

---

plot_SpaCCI_local *Plot SpaCCI Localized Interaction Results*

Description

This function provides a unified interface to visualize the localized cell-cell interaction patterns inferred by SpaCCI, either using a Seurat object with a spatial image or a spatial coordinates data frame.

Usage

```
plot_SpaCCI_local(
  Seurat_Object = NULL,
  spatial_coordinates_dataframe = NULL,
  SpaCCI_local_Result_List,
  Ligand_cell_type,
  Receptor_cell_type,
  spot_plot_size,
  specific_LR_pair_name = NULL,
  significant_cutoff = 0.05
)
```

Arguments

`Seurat_Object` Optional. A Seurat object containing spatial data. If provided, the function will plot the interaction patterns on the tissue image.

`spatial_coordinates_dataframe`
Optional. A data frame containing the spatial coordinates of the spots. The columns should include `"Spot_ID"`, `"imagerow"`, and `"imagecol"`. The row names must be the names of `"Spot_ID"`, matching those in the cell type proportion data frame or the gene expression data frame.

`SpaCCI_local_Result_List`
A list containing the results from a SpaCCI local analysis. This list should include `dataframelist` and `RegionIDs_matrix`, which are the outputs from `run_SpaCCI(..., analysis_scale = "local",...)`.

`Ligand_cell_type`
The name of the ligand cell type to plot. This should match the cell type names used in the `run_SpaCCI` analysis.

`Receptor_cell_type`
The name of the receptor cell type to plot. This should match the cell type names used in the `run_SpaCCI` analysis.

`spot_plot_size` A numeric value controlling the size of the spots in the plot.

`specific_LR_pair_name`
Optional. The name of a specific ligand-receptor pair to plot. If provided, the plot will focus on this interaction. The name should match those in the `SpaCCI_local_Result_List$dataframelist`.

`significant_cutoff`
A numeric value specifying the significance cutoff for the adjusted P-values from the permutation test. Default is `0.05`.

Value

A plot object showing the localized interaction patterns. The plot will be generated using either the Seurat object or the spatial coordinates data frame, depending on the input provided.

Examples

```
# Plot localized SpaCCI results using Seurat object
library(SpaCCI)
data(result_local)
data(test_data)
spatial_coords_df <- test_data$spatial_coords_df
#plot_SpaCCI_local(Seurat_Object = seurat_object,.....)

# Plot localized SpaCCI results using spatial coordinates
plot_SpaCCI_local(spatial_coordinates_dataframe = spatial_coords_df,
                  SpaCCI_local_Result_List = result_local,
                  Ligand_cell_type = "beta",
                  Receptor_cell_type = "delta",
                  spot_plot_size = 3)
```

possible_L_R_pairs_cellchat *CellChat Database: Identify Possible Ligand-Receptor Pairs for Cell-Cell Communication*

Description

This function identifies possible ligand-receptor (L-R) pairs for cell-cell communication analysis using a subset of the CellChat database. It checks for the presence of all genes involved in each L-R pair within the provided gene expression matrix.

Usage

```
possible_L_R_pairs_cellchat(
  species,
  gene_spot_expression_dataframe,
  percentage
)
```

Arguments

species	A string specifying the species ("Human" or "Mouse"). The function selects the appropriate CellChatDB object, typically 'CellChatDB.human' or 'CellChatDB.mouse', which contains information on ligand-receptor interactions.
gene_spot_expression_dataframe	A gene expression data frame with genes as row names and Spot IDs as column names. This data frame is used to verify the presence of all genes involved in the L-R pairs.
percentage	A numeric value specifying the minimum percentage of spots in which a gene must be expressed to be considered. The default is 10.

Value

A list containing:

possible_L_R_pairs A data frame of L-R pairs where all genes are present in the 'gene_spot_expression_dataframe'. The data frame includes the ligand and receptor vectors, and the combined gene vectors.

possible_L_R_pairs_details A data frame with detailed information about the L-R pairs, including the original annotations from the CellChatDB.

Examples

```
library(SpaCCI)
#Load the example data
load(system.file("extdata", "Tutorial_example_data.rda", package = "SpaCCI"))
Example_Seurat <- NormalizeData(Example_Seurat)
```

```
gene_spot_df <- as.data.frame(Example_Seurat@assays$Spatial@data)
result <- possible_L_R_pairs_cellchat(CellChatDB.human,
      gene_spot_expression_dataframe = gene_spot_df)
```

`possible_L_R_pairs_Cellinker` *Identify Possible Ligand-Receptor Pairs for Cell-Cell Communication*

Description

Cellinker Database: This function identifies possible ligand-receptor (L-R) pairs for cell-cell communication analysis using data from the Cellinker database. It checks for the presence of all genes involved in each L-R pair within the provided gene expression matrix, filtering based on a specified expression percentage threshold.

Usage

```
possible_L_R_pairs_Cellinker(
  species,
  gene_spot_expression_dataframe,
  percentage
)
```

Arguments

`species`	A string specifying the species ("Human" or "Mouse"). The function selects the appropriate Cellinker interaction file based on this input.
`gene_spot_expression_dataframe`	A gene expression data frame with genes as row names and Spot IDs as column names. This data frame is used to verify the presence of all genes involved in the L-R pairs.
`percentage`	A numeric value specifying the minimum percentage of spots in which a gene must be expressed to be considered. The default is 10.

Value

A list containing:

possible_L_R_pairs A data frame of L-R pairs where all genes are present in the 'gene_spot_expression_dataframe' and meet the expression threshold. The data frame includes the ligand and receptor vectors, and the combined gene vectors.

possible_L_R_pairs_details A data frame with detailed information about the identified L-R pairs, including their original annotations from the Cellinker dataset.

possible_L_R_pairs_cellphoneDB *Identify Possible Ligand-Receptor Pairs for Cell-Cell Communication*

Description

CellPhone Database: This function identifies possible ligand-receptor (L-R) pairs for cell-cell communication analysis using data from a CellPhoneDB dataset. It checks for the presence of all genes involved in each L-R pair within the provided gene expression matrix and filters based on a specified expression percentage threshold.

Usage

```
possible_L_R_pairs_cellphoneDB(gene_spot_expression_dataframe, percentage)
```

Arguments

gene_spot_expression_dataframe	A gene expression data frame with genes as row names and Spot IDs as column names. This data frame is used to verify the presence of all genes involved in the L-R pairs.
percentage	A numeric value specifying the minimum percentage of spots in which a gene must be expressed to be considered. The default is 10.

Value

A list containing:

possible_L_R_pairs A data frame of L-R pairs where all genes are present in the 'gene_spot_expression_dataframe' and meet the expression threshold. The data frame includes the ligand and receptor vectors, and the combined gene vectors.

possible_L_R_pairs_details A data frame with detailed information about the identified L-R pairs, including their original annotations from the CellPhoneDB dataset.

---

possible_L_R_pairs_connectome *Identify Possible Ligand-Receptor Pairs for Cell-Cell Communication*

Description

ConnectomeDB 2020 Database: This function identifies possible ligand-receptor (L-R) pairs based on gene expression data.

Usage

```
possible_L_R_pairs_connectome(gene_spot_expression_dataframe, percentage)
```

Arguments

`gene_spot_expression_dataframe`
: A gene expression data frame with genes as row names and Spot IDs as column names. This data frame is used to verify the presence of all genes involved in the L-R pairs.

`percentage`
: A numeric value specifying the minimum percentage of spots in which a gene must be expressed to be considered. The default is 10.

Value

A list containing:

possible_L_R_pairs A data frame of L-R pairs where all genes are present in the 'gene_spot_expression_dataframe' and meet the expression threshold. The data frame includes the ligand and receptor vectors, and the combined gene vectors.

possible_L_R_pairs_details A data frame with detailed information about the identified L-R pairs, including their original annotations from the ConnectomeDB 2020 dataset.

`possible_L_R_pairs_ICELLNET` *Identify Possible Ligand-Receptor Pairs for Cell-Cell Communication*

Description

ICELLENT Database: This function identifies possible ligand-receptor (L-R) pairs for cell-cell communication analysis using data from the ICELLNET database. It checks for the presence of all genes involved in each L-R pair within the provided gene expression matrix, filtering based on a specified expression percentage threshold.

Usage

```
possible_L_R_pairs_ICELLNET(gene_spot_expression_dataframe, percentage)
```

Arguments

`gene_spot_expression_dataframe`
: A gene expression data frame with genes as row names and Spot IDs as column names. This data frame is used to verify the presence of all genes involved in the L-R pairs.

`percentage`
: A numeric value specifying the minimum percentage of spots in which a gene must be expressed to be considered. The default is 10.

Value

A list containing:

possible_L_R_pairs A data frame of L-R pairs where all genes are present in the 'gene_spot_expression_dataframe' and meet the expression threshold. The data frame includes the ligand and receptor vectors, and the combined gene vectors.

possible_L_R_pairs_details A data frame with detailed information about the identified L-R pairs, including their original annotations from the ICELLNET dataset.

`random_region` *Select Closest Spatial IDs to a Center Point: this is used for permutation*

Description

This function identifies and returns the IDs of the closest spatial points to a specified center point based on Euclidean distance.

Usage

```
random_region(spatial_coord, center_id, n_ids)
```

Arguments

`spatial_coord`	A data frame of the spatial coordinates. The column names should include 'c("Spot_ID", "imagerow", "imagecol")', and the row names must be the Spot_IDs, which is the same as the row names in the cell type proportion data frame or the column names of the gene*spot expression data frame.
`center_id`	A character string specifying the ID of the center spot from which distances are calculated.
`n_ids`	An integer specifying the number of closest IDs to select.

Value

A character vector of the 'n_ids' closest IDs to the specified center ID.

Examples

```
spatial_coord <- data.frame(
  imagecol = c(1, 2, 3, 4, 5),
  imagerow = c(5, 4, 3, 2, 1),
  row.names = c("Spot1", "Spot2", "Spot3", "Spot4", "Spot5")
)
center_id <- "Spot3"
closest_ids <- random_region(spatial_coord, center_id, 3)
print(closest_ids)
```

result_global *result_global data for SpaCCI*

Description

This example dataset is the result of running the `run_SpaCCI` function. It contains the inferred cell-cell interactions across the global scale.

Usage

```
data(result_global)
```

Format

A list containing:

pvalue_df A data frame of p-values and adjusted p-values for cell-cell interactions.

Details

These objects can be used for testing and running example analyses with the SpaCCI package.

Examples

```
data(result_global)
print(result_global)
```

result_local *result_local data for SpaCCI*

Description

This example dataset is the result of running the `run_SpaCCI` function. It contains the inferred cell-cell interactions across the global scale.

Usage

```
data(result_local)
```

Format

A list containing:

dataframelist A list of data frame of the p-value reults of each spatial neighborhood.

RegionIDs_matrix A list of matrix contains the spot IDs of each spatial neighborhood.

Details

These objects can be used for testing and running example analyses with the SpaCCI package.

Examples

```
data(result_local)
print(result_local)
```

`run_SpaCCI` *Run SpaCCI Analysis*

Description

This function runs the SpaCCI analysis to infer cell-cell interactions based on ligand-receptor pairs at global, regional, or local spatial scales. It integrates gene expression data, cell type proportions, and spatial coordinates with a user-specified ligand-receptor database.

Usage

```
run_SpaCCI(
  gene_spot_expression_dataframe,
  spot_cell_proportion_dataframe,
  spatial_coordinates_dataframe,
  LR_database_list,
  specific_LR_pair = NULL,
  analysis_scale,
  region_spot_IDs = NULL,
  local_scale_proportion = 1,
  neighborhood_radius = 2.5
)
```

Arguments

`gene_spot_expression_dataframe`
: A data frame of gene expression values, where row names are genes and column names are spot IDs.

`spot_cell_proportion_dataframe`
: A data frame of cell type proportions, where row names are spot IDs and column names are cell types.

`spatial_coordinates_dataframe`
: A data frame containing the spatial coordinates of the spots. The columns should include `"Spot_ID"`, `"imagerow"`, and `"imagecol"`. And the row names must be the names of `"Spot_ID"`.

`LR_database_list`
: A list containing ligand-receptor pairs and additional information, generated by functions using `` `LR_database()` ``.

`specific_LR_pair`
: Required if `analysis_scale` is `"local"`. A vector of ligand-receptor pair names for localized analysis. The names should match those in row names in `` `LR_database_list$possible_LR ``

`analysis_scale` A string specifying the scale of analysis: `"global"`, `"regional"`, or `"local"`.

`region_spot_IDs`
: Required if `analysis_scale` is "regional". A vector of spot IDs defining the region for regional analysis.

`local_scale_proportion`
: Optional. A numeric value ranging from 0 to 1, (0,1] specifying the proportion of spots to use for localized analysis. Default is `1`, meaning using `100%` proportion of spots. One could modified if want to reducing computing time.

`neighborhood_radius`
: Optional. A numeric value specifying the radius of the neighborhood for localized analysis. Default is `2.5`, according to the 10X Visium ST data accounting for 200-250 μm interacting distance.

Details

The function supports three scales of analysis:

`global` Analyzes interactions across the entire dataset.

`regional` Analyzes interactions within a specified region of spots. Requires `region_spot_IDs`.

`local` Analyzes localized hotspot of interactions for specific ligand-receptor pairs on the entire slides. Requires `specific_LR_pair`.

Value

A list containing:

If `analysis_scale` **is** `"local"`**:** A list containing:

- `dataframelist` A list of data frames, each representing the inferred interactions for a specific center spot. Each data frame includes information on ligand and receptor cell types, P-values, and adjusted P-values.
- `RegionIDs_matrix` A list of matrices, each containing the IDs of the spots within the specified radius of each center spot.

If `analysis_scale` **is** `"regional"` **or** `"global"`**:** A list containing:

- `pvalue_df` A data frame of inferred interactions within the specified region or globally, including information on ligand and receptor cell types, P-values, and adjusted P-values.

Examples

```
library(SpaCCI)
library(nnls)
#Load the example data
data(test_data)
gene_spot_df <- test_data$gene_spot_df
cell_prop_df <- test_data$cell_prop_df
spatial_coords_df <- test_data$spatial_coords_df
```

```
result <- LR_database(species = "Human",
                      database_name = "CellChat",
                      gene_spot_expression_dataframe = gene_spot_df)
# global
result_global <- run_SpaCCI(gene_spot_expression_dataframe = gene_spot_df,
                            spot_cell_proportion_dataframe = cell_prop_df,
                            spatial_coordinates_dataframe = spatial_coords_df,
                            LR_database_list = result,
                            analysis_scale = "global")

# local
result_local <- run_SpaCCI(gene_spot_expression_dataframe = gene_spot_df,
                           spot_cell_proportion_dataframe = cell_prop_df,
                           spatial_coordinates_dataframe = spatial_coords_df,
                           LR_database_list = result,
                           specific_LR_pair = "EDN2_EDNRA",
                           analysis_scale = "local",
                           local_scale_proportion = 0.1,
                           neighborhood_radius = 2.5)
```

---

scPalette *Generate a Color Palette*

Description

This function generates a color palette. It selects colors from a predefined color space, and if more colors are needed than are available in the predefined set, it generates a palette using color interpolation.

Usage

```
scPalette(n)
```

Arguments

n	An integer specifying the number of colors needed.

Value

A character vector of colors in hexadecimal format.

Examples

```
# Generate a palette with 5 colors
palette <- scPalette(5)
print(palette)

# Generate a palette with 30 colors
```

```
large_palette <- scPalette(30)
print(large_palette)
```

SpaCCI_global *Infer Cell-Cell Interactions on a Global Scale*

Description

This function infers cell-cell interactions on a global scale using spatial transcriptomics data. It applies permutation testing to identify significant ligand-receptor interactions across all spots.

Usage

```
SpaCCI_global(
  gene_spot_df,
  spot_cell_prop_df,
  matching_L_R_pairs,
  matching_L_R_pairs_info
)
```

Arguments

`gene_spot_df`	A data frame where the rows are genes and the columns are spots (Spot_IDs), representing gene expression levels across spatial spots.
`spot_cell_prop_df`	A data frame of cell type proportions for each spot. The rows represent spots (Spot_IDs), and the columns represent different cell types.
`matching_L_R_pairs`	A data frame containing matching ligand-receptor pairs. Each row corresponds to a ligand-receptor pair, with columns for `ligand_vector` and `receptor_vector`.
`matching_L_R_pairs_info`	A data frame providing additional information for each ligand-receptor pair, such as pathway information.

Value

A list containing:

pvalue_df A data frame of inferred interactions across the global scale, including information on ligand and receptor cell types, interaction strength, P-values, and adjusted P-values.

SpaCCI_local *Infer Cell-Cell Interactions on a Local Scale*

Description

This function infers cell-cell interactions on a local scale using spatial transcriptomics data. It utilizes permutation testing to identify significant ligand-receptor interactions within specified neighborhoods around randomly selected center spots.

Usage

```
SpaCCI_local(
  gene_spot_df,
  spot_cell_prop_df,
  spatial_coord,
  prop,
  radius,
  matching_L_R_pairs,
  matching_L_R_pairs_info
)
```

Arguments

gene_spot_df	A data frame where the rows are genes and the columns are spots (Spot_IDs), representing gene expression levels across spatial spots.
spot_cell_prop_df	A data frame of cell type proportions for each spot. The rows represent spots (Spot_IDs), and the columns represent different cell types.
spatial_coord	A data frame of the spatial coordinates. The column names should include ‘c("Spot_ID", "imagerow", "imagecol")‘, and the row names must be the Spot_IDs, which is the same as the row names in the cell type proportion data frame or the column names of the gene*spot expression data frame.
prop	A numeric value representing the proportion of spots to randomly sample as center spots for local neighborhood analysis.
radius	A numeric value specifying the radius of the spatial neighborhood around each center spot.
matching_L_R_pairs	A data frame containing matching ligand-receptor pairs. Each row corresponds to a ligand-receptor pair, with columns for `ligand_vector` and `receptor_vector`.
matching_L_R_pairs_info	A data frame providing additional information for each ligand-receptor pair, such as pathway information.

Value

A list containing:

dataframelist A list of data frames, each representing the inferred interactions for a specific center spot. Each data frame includes information on ligand and receptor cell types, P-values, and adjusted P-values.

RegionIDs_matrix A list of matrices, each containing the IDs of the spots within the specified radius of each center spot.

`SpaCCI_region` *Infer Cell-Cell Interactions in a Specified Region*

Description

This function infers cell-cell interactions within a specified region using spatial transcriptomics data. It applies permutation testing to identify significant ligand-receptor interactions in the region.

Usage

```
SpaCCI_region(
  gene_spot_df,
  spot_cell_prop_df,
  region_spot_IDs,
  matching_L_R_pairs,
  matching_L_R_pairs_info
)
```

Arguments

`gene_spot_df` A data frame where the rows are genes and the columns are spots (Spot_IDs), representing gene expression levels across spatial spots.

`spot_cell_prop_df`
A data frame of cell type proportions for each spot. The rows represent spots (Spot_IDs), and the columns represent different cell types.

`region_spot_IDs`
A vector of Spot_IDs representing the spots included in the region of interest.

`matching_L_R_pairs`
A data frame containing matching ligand-receptor pairs. Each row corresponds to a ligand-receptor pair, with columns for `ligand_vector` and `receptor_vector`.

`matching_L_R_pairs_info`
A data frame providing additional information for each ligand-receptor pair, such as pathway information.

Value

A list containing:

pvalue_df A data frame of inferred interactions within the specified region, including information on ligand and receptor cell types, P-values, and adjusted P-values.

`test_data` *Test data for SpaCCI*

Description

This dataset includes:

- gene_spt_df: A data frame of spot-level gene expression.
- cell_prop_df: A data frame with cell type proportions.
- spatial_coords_df: A data frame of spatial coordinates.

Usage

```
data(test_data)
```

Format

An object of class `list` of length 3.

Examples

```
library(SpaCCI)
data(test_data)
```

Index