# Package 'sccca'

March 13, 2024

**Type** Package

**Title** Single-Cell Correlation Based Cell Type Annotation

**Version** 0.1.1

**Maintainer** Mohamed Soudy <Mohmedsoudy2009@gmail.com>

**Description** Performing cell type annotation based on cell markers from a unified database. The approach utilizes correlation-based approach combined with association analysis using Fisher-exact and phyper statistical tests (Upton, Graham JG. (1992) <DOI:10.2307/2982890>).

**License** GPL (>= 3)

**Encoding** UTF-8

**Imports** Seurat, dplyr, plyr, scales, HGNChelper, openxlsx

**RoxygenNote** 7.3.0

**NeedsCompilation** no

**Author** Mohamed Soudy [aut, cre],
Sophie LE BARS [aut],
Enrico Glaab [aut]

**Repository** CRAN

**Date/Publication** 2024-03-13 11:40:02 UTC

## R topics documented:

**Index**          **12**

---

| calculate_cor_mat | *Performs aggregation based on cell clusters and condition. Then, it calculates correlation matrix of genes* |
|---|---|

---

### Description

This Function is used to perform cell aggregation by averaging the expression of scRNA-seq matrix and then perform correlation matrix

### Usage

```
calculate_cor_mat(expression_mat, condition = NULL, clusters, assay = "RNA")
```

### Arguments

expression_mat  Seurat object that contains the expression matrix.

condition        column name of the condition in th meta data of the Seurat object.

clusters          column name of the cluster numbers in the meta data of the Seurat object.

assay             the assay to be used default is set to RNA

### Value

correlation matrix of genes.

### Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

---

calculate_normalized_ratio

*Calculate cell scores based on number of genes*

---

### Description

This Function is used to calculate cell scores based on number of genes

### Usage

```
calculate_normalized_ratio(vec)
```

## Arguments

vec list of genes of cell types.

## Value

vector of cell scores based on the number of overlapped genes with the input matrix.

## Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

---

correct_gene_symbols *Process the cell markers names*

---

## Description

This Function is used to return the cell markers names processed for the sctype approach

## Usage

```
correct_gene_symbols(markers)
```

## Arguments

markers list of unique cell markers.

## Value

vector of genes names which overlap with the correlation matrix.

## Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

---

enrich_genes                    *Performs parallel function on two lists*

---

### Description

This Function is used to perform parallel function on two lists

### Usage

```
enrich_genes(ref_list, overlap_list, func)
```

### Arguments

| | |
|---|---|
| ref_list | reference list. |
| overlap_list | overlap list. |
| func | function to be applied. |

### Value

list where each element is the result of applying the function 'func' to the corresponding elements of 'ref_list' and 'overlap_list'.

### Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

---

filter_correlation              *Filter the genes based on specific correlation threshold*

---

### Description

This Function is used to filter the gene correlation matrix based on user-defined threshold

### Usage

```
filter_correlation(cor_mat, gene_list, threshold = 0.7)
```

### Arguments

| | |
|---|---|
| cor_mat | correlation matrix generated from calculate_cor_mat function. |
| gene_list | cell markers that passed threshold. |
| threshold | absolute correlation threshold. |

## Value

vector of gene names that pass user-defined correlation threshold.

## Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

---

| filter_list | *Process the cell markers that overlap between the cell markers and scRNA matrix* |
| --- | --- |

---

## Description

This Function is used to return the cell markers that overlap between the cell markers and scRNA matrix

## Usage

```
filter_list(gene_list, passed_cells)
```

## Arguments

gene_list       list of unique genes of cell types.

passed_cells    cells types that pass the specified threshold.

## Value

list of cell types which genes are found in the input matrix.

## Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

---

fisher_test                    *Performs fisher exact test to get the significant overlap between genes for cell type assignment*

---

### Description

This Function is used to perform fisher exact test to get cell types

### Usage

```
fisher_test(ref, gene_overlap)
```

### Arguments

ref                 reference gene set.

gene_overlap        genes that pass the correlation threshold.

### Value

vector of p-value and overlap.

### Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

### Examples

```
fisher_test(c("PAX8","PAX6","TP53","AOC3","LIPF"), c("LIPF","PAX8","PAX6","TP53","TSHB","AOC3"))
```

---

match_characters               *Process the cell markers that pass specific threshold in the gene correlation matrix*

---

### Description

This Function is used to return the cell markers that pass specific threshold in the gene correlation matrix

### Usage

```
match_characters(genes, gene_mat)
```

## Arguments

| | |
|---|---|
| genes | list of unique genes of cell types. |
| gene_mat | correlation matrix of genes. |

## Value

vector of genes names which overlap with the correlation matrix.

## Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

---

| phyper_test | *Performs phyper test to get the significant overlap between genes for cell type assignment* |
|---|---|

---

## Description

This Function is used to perform phyper test to get cell types

## Usage

```
phyper_test(ref, overlap)
```

## Arguments

| | |
|---|---|
| ref | reference gene set. |
| overlap | genes that pass the correlation threshold. |

## Value

vector of p-value and overlap.

## Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

## Examples

```
phyper_test(c("PAX8","PAX6","TP53","AOC3","LIPF"), c("LIPF","PAX8","PAX6","TP53","TSHB","AOC3"))
```

---

process_clus                    *Gets the associated cell types using correlation-based approach*

---

### Description

This Function is used to get the associated cell clusters using correlation-based approach

### Usage

```
process_clus(cluster,sobj,assay="RNA",clus,markers,cor_m,m_t=0.9,c_t=0.7,test="p")
```

### Arguments

| | |
|---|---|
| cluster | associated cluster name. |
| sobj | Seurat object. |
| assay | assay to be used default is set to RNA. |
| clus | cell clusters. |
| markers | cell markers database. |
| cor_m | gene correlation matrix. |
| m_t | overlap threshold between cell markers and expression matrix. |
| c_t | correlation threshold between genes. |
| test | statistical test that check if overlap is significant could be "p" for phyper or "f" for fisher. |

### Value

data frame of proposed cell types.

### Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

---

process_database                *Process the database for the sctype approach*

---

### Description

This Function is used to process the database that will be used for sctype approach

### Usage

```
process_database(database_name = "sctype", org = 'a', tissue, tissue_type = 'n')
```

## Arguments

| | |
|---|---|
| database_name | name of the database to be used that can be 'sctype' or 'UMD'. |
| org | name of organism to be used that can be 'h' for human, 'm' for mouse, and 'a' for all markers. |
| tissue | specified tissue from which the data comes. |
| tissue_type | tissue type whether 'a' for all types 'n' for normal tissues only or "c" for cancer tissues. |

## Value

vector of genes names which overlap with the correlation matrix.

## Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

---

| process_markers | *Process the cell markers database and return the processed list* |
|---|---|

---

## Description

This Function is used to process the cell markers database and return the processed list

## Usage

```
process_markers(markers_df)
```

## Arguments

| | |
|---|---|
| markers_df | data frame with markers named as gene_original and cell names as cell type. |

## Value

list of lists of the processed markers

## Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

---

sccca                    *Run the pipeline for the cell type assignment*

---

### Description

This Function is used to run the main pipeline that does the cell type assignment

### Usage

```
sccca(sobj,assay="RNA",cluster,marker,tissue,tt="a",cond,m_t=0.9,c_t=0.7,test="p",org="a")
```

### Arguments

| | |
|---|---|
| sobj | Seurat object. |
| assay | assay to be used default is set to RNA. |
| cluster | colname in the mata.data that have the cell cluster numbers. |
| marker | cell markers database path. |
| tissue | specified tissue from which the data comes. |
| tt | tissue type whether 'a' for all types 'n' for normal tissues only or "c" for cancer tissues. |
| cond | colname in the meta.data that have the condition names. |
| m_t | overlap threshold between cell markers and expression matrix. |
| c_t | correlation threshold between genes. |
| test | statistical test that check if overlap is significant could be "p" for phyper or "f" for fisher. |
| org | organism to be used that can be 'h' for human, 'm' for mouse, and 'a' for all markers. |

### Value

list of Seurat object that have the assigned clusters, and top 3 proposed cell types.

### Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

---

| sctype | *Run the sctype approach as it's implemented by Ianevski, A., Giri, A.K. and Aittokallio, T.* |

---

### Description

This Function is used to run the sctype approach with faster implementation

### Usage

```
sctype(sobj,assay="RNA",tissue,tt="a",clus,org="a",scaled=T,database="sctype")
```

### Arguments

| | |
|---|---|
| sobj | Seurat object. |
| assay | assay to be used default is set to RNA. |
| tissue | specified tissue from which the data comes. |
| tt | tissue type whether 'a' for all types 'n' for normal tissues only or "c" for cancer tissues. |
| clus | colname in the mata.data that have the cell cluster numbers. |
| org | organism to be used that can be 'h' for human, 'm' for mouse, and 'a' for all markers. |
| scaled | indicates whether the matrix is scaled (TRUE by default) |
| database | name of the database to be used that can be 'sctype' or 'UMD' |

### Value

vector of genes names which overlap with the correlation matrix.

### Author(s)

Mohmed Soudy <Mohamed.soudy@uni.lu> and Sohpie LE BARS <sophie.lebars@uni.lu> and Enrico Glaab <enrico.glaab@uni.lu>

# Index