Package 'CACIMAR'

October 12, 2022

Title Cross-Species Analysis of Cell Identities, Markers and Regulations

Version 1.0.0

Maintainer Junyao Jiang <jiangjunyao789@163.com>

Author Junyao Jiang <jiangjunyao789@163.com>

Description A toolkit to perform cross-species analysis based on scRNA-seq data. This package contains 5 main features. (1) identify Markers in each cluster. (2) Cell type annotation (3) identify conserved markers. (4) identify conserved cell types. (5) identify conserved modules of regulatory networks.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

Imports pheatmap, pbapply, psych, ROCR, reshape2, dplyr, grDevices, stats, methods, viridisLite

Depends R (>= 4.0), Seurat

Suggests knitr, testthat (>= 3.0.0)

Config/testthat/edition 3

URL https://github.com/jiang-junyao/CACIMAR

BugReports https://github.com/jiang-junyao/CACIMAR/issues

NeedsCompilation no

Repository CRAN

Date/Publication 2022-05-18 08:20:02 UTC

R topics documented:

CACIMAR_cols 2
Format_Markers_Frac 3
Heatmap_Cor 3

Identify_CellType 4
Identify_ConservedCellTypes 5
Identify_ConservedMarkers 6
Identify_ConservedNetworks 7
Identify_Markers 8
OrthG_Hs_Ch 9
OrthG_Hs_Mm 9
OrthG_Hs_Zf 9
OrthG_Mm_Ch 10
OrthG_Mm_Zf 10
OrthG_Zf_Ch 10
Plot_MarkersHeatmap 11

Index **13**

`CACIMAR_cols` *CACIMAR colors palette*

Description

CACIMAR colors palette

Usage

```
CACIMAR_cols(color_number)
```

Arguments

`color_number`	numeric, indicating used colors number

Value

vector of colors

Examples

```
CACIMAR_cols(10)
CACIMAR_cols(20)
```

Format_Markers_Frac *Format marker genes for plotting*

Description

Order the gene expression in each cluster to make the heatmap look better

Usage

```
Format_Markers_Frac(Marker_genes)
```

Arguments

Marker_genes	data.frame, generated by Identify_Markers

Value

Markers corresponding to certain cluster

Examples

```
data("pbmc_small")
all.markers <- Identify_Markers(pbmc_small)
all.markers2 <- Format_Markers_Frac(all.markers)
```

Heatmap_Cor *plot the heatmap of marker genes across different species*

Description

plot the heatmap of marker genes across different species

Usage

```
Heatmap_Cor(
  RNA1,
  RowType1 = "",
  ColType1 = "",
  cluster_cols = TRUE,
  cluster_rows = FALSE,
  Color1 = NULL,
  ...
)
```

Arguments

RNA1	correlation of expression in each cell type
RowType1	character, indicating the cell types that you want to show on the row in heatmap. RowType1=" means show all cell types
ColType1	character, indicating the cell types that you want to show on the column in heatmap. RowType1=" means show all cell types
cluster_cols	boolean values determining if columns should be clustered or hclust object
cluster_rows	boolean values determining if rows should be clustered or hclust object
Color1	vector of colors used in heatmap
...	parameter in pheatmap

Value

pheatmap object

Examples

```
load(system.file("extdata", "network_example.rda", package = "CACIMAR"))
n1 <- Identify_ConservedNetworks(OrthG_Mm_Zf,mmNetwork,zfNetwork,'mm','zf')
Heatmap_Cor(n1[[2]],cluster_cols=TRUE, cluster_rows=FALSE)
```

Identify_CellType *Identify cell type of each cluster*

Description

This function has three steps to identify cell type of each cluster. (1) Calculate the power of each known marker based on AUC (area under the receiver operating characteristic curve of gene expression) which indicates the capability of marker i from cell type m to distinguish cluster j and the other clusters. (2) Calculate the united power (UP) for cell type m across each cluster j. (3) For each cluster j we determine the cell type according to UP. Generally, the cluster beongs to the cell type which have the highest united power or higher than the threshold of the united power (for example > 0.9 power).

Usage

```
Identify_CellType(seurat_object, Marker_gene_table)
```

Arguments

seurat_object	seurat object
Marker_gene_table	data.frame, indicating marker gene and its corresponding cell type. Marker_gene_table should contain two columns: 'CellType' represent corresponding cell types of each marker and 'Marker' represent Markers

Value

Cell type with the highest power in each cluster

Examples

```
KnownMarker=data.frame(c('AIF1','BID','CCL5','CD79A','CD79B','MS4A6A'),c('a','a','a','b','b','b'))
data("pbmc_small")
colnames(KnownMarker)=c('Marker','CellType')
CT <- Identify_CellType(pbmc_small,KnownMarker)
```

Identify_ConservedCellTypes

Identify conserved cell types based on power of genes and orthologs database

---

Description

Identify conserved cell types based on power of genes and orthologs database

Usage

```
Identify_ConservedCellTypes(
  OrthG,
  Species1_Marker_table,
  Species2_Marker_table,
  Species_name1,
  Species_name2
)
```

Arguments

OrthG	ortholog genes database
Species1_Marker_table	data.frame of species 1, should contain three column: 'gene', 'cluster' and 'power'
Species2_Marker_table	data.frame of species 2, should contain three column: 'gene', 'cluster' and 'power'
Species_name1	character, indicating the species names of Species1_Marker_table
Species_name2	character, indicating the species names of Species2_Marker_table

Value

list contains two elements: first one is details of conserved cell types, second one is matrix of cell types conserved score

Examples

```
load(system.file("extdata", "CellTypeAllMarkers.rda", package = "CACIMAR"))
expression <- Identify_ConservedCellTypes(OrthG_Mm_Zf,mm_Marker[1:30,],zf_Marker[1:30,],'mm','zf')
```

`Identify_ConservedMarkers` *Identify orthologs marker genes for two species*

Description

Identify orthologs marker genes for two species based on orthologs database

Usage

```
Identify_ConservedMarkers(
  OrthG,
  Species1_Marker_table,
  Species2_Marker_table,
  Species_name1,
  Species_name2,
  match_cell_name = NULL
)
```

Arguments

`OrthG`	ortholog genes database
`Species1_Marker_table`	data.frame of species 1, first column should be gene name, second column should be Clusters corresponding to marker gene
`Species2_Marker_table`	data.frame of species 2, first column should be gene name, second column should be Clusters corresponding to marker gene of marker genes.
`Species_name1`	character, indicating the species names of Species1_Marker_table.
`Species_name2`	character, indicating the species names of Species2_Marker_table
`match_cell_name`	characters contained in both cell names to match similar cell types

Value

Data frame of conserved markers

Examples

```
load(system.file("extdata", "CellMarkers.rda", package = "CACIMAR"))
o1 <- Identify_ConservedMarkers(OrthG_Mm_Zf,Mm_marker_cell_type,
Zf_marker_cell_type,Species_name1 = 'mm',Species_name2 = 'zf')
o2 <- Identify_ConservedMarkers(OrthG_Zf_Ch,Ch_marker_cell_type,
Zf_marker_cell_type,Species_name1 = 'ch',Species_name2 = 'zf')
```

`Identify_ConservedNetworks` *Identify conserved regulatory networks*

Description

Use Score of Conserved network to identify conserved regulatory network modules based on homologous genes databased and topology of networks

Usage

```
Identify_ConservedNetworks(
  OrthG,
  Species1_GRN,
  Species2_GRN,
  Species_name1,
  Species_name2
)
```

Arguments

`OrthG`	ortholog genes database
`Species1_GRN`	gene regulatory network of species 1
`Species2_GRN`	gene regulatory network of species 2
`Species_name1`	character, indicating the species names of Species1_GRN
`Species_name2`	character, indicating the species names of Species2_GRN

Value

list contains two df. First df contains details of conserved regulatory network, second df contains NCS between module pairs

Examples

```
load(system.file("extdata", "gene_network.rda", package = "CACIMAR"))
n1 <- Identify_ConservedNetworks(OrthG_Mm_Zf,mm_gene_network,zf_gene_network,'mm','zf')
```

`Identify_Markers` *Identify markers of each cluster*

Description

This function first identify marker genes in each cluster with Roc threshold > RocThr. Then, based on marker genes identified above, this function calculates the difference and power of marker genes in each cluster, and marker genes with Difference threshold > DiffThr will be retained. Next, gene with the largest power in which cluster will be the marker gene in this cluster. Eventually, make fisher test for power of each cluster, cluster with p.value < 0.05 will be retained as the final cluster for marker gene

Usage

```
Identify_Markers(
  Seurat_object,
  PowerCutoff = 0.4,
  DifferenceCutoff = 0,
  PvalueCutoff = 0.05
)
```

Arguments

`Seurat_object`	Seurat object, should contain cluster information
`PowerCutoff`	numeric, indicating the cutoff of gene power to refine marker genes
`DifferenceCutoff`	numeric, indicating the cutoff of difference in marker genes between clusters to refine marker genes
`PvalueCutoff`	numeric, indicating the p.value cutoff of chi-square test to refine marker genes

Value

Data frame of conserved markers

Examples

```
data("pbmc_small")
all.markers <- Identify_Markers(pbmc_small)
```

OrthG_Hs_Ch — *Orthologs genes database for homo sapiens and zebrafish*

Description

Orthologs genes database for homo sapiens and zebrafish

Usage

```
OrthG_Hs_Ch
```

Format

An object of class `data.frame` with 12219 rows and 5 columns.

OrthG_Hs_Mm — *Orthologs genes database for homo sapiens and mus musculus*

Description

Orthologs genes database for homo sapiens and mus musculus

Usage

```
OrthG_Hs_Mm
```

Format

An object of class `data.frame` with 16754 rows and 5 columns.

OrthG_Hs_Zf — *Orthologs genes database for homo sapiens and zebrafish*

Description

Orthologs genes database for homo sapiens and zebrafish

Usage

```
OrthG_Hs_Zf
```

Format

An object of class `data.frame` with 12017 rows and 5 columns.

OrthG_Mm_Ch *Orthologs genes database for mus musculus and chicken*

Description

Orthologs genes database for mus musculus and chicken

Usage

```
OrthG_Mm_Ch
```

Format

An object of class `data.frame` with 62661 rows and 5 columns.

OrthG_Mm_Zf *Orthologs genes database for mus musculus and zebrafish*

Description

Orthologs genes database for mus musculus and zebrafish

Usage

```
OrthG_Mm_Zf
```

Format

An object of class `data.frame` with 65631 rows and 5 columns.

OrthG_Zf_Ch *Orthologs genes database for mus zebrafish and chicken*

Description

Orthologs genes database for mus zebrafish and chicken

Usage

```
OrthG_Zf_Ch
```

Format

An object of class `data.frame` with 38394 rows and 5 columns.

`Plot_MarkersHeatmap` *Plot Markers in each cell type*

Description

This function integrate R package pheatmap to plot markers in each cell type

Usage

```
Plot_MarkersHeatmap(
  ConservedMarker,
  start_col = 2,
  module_colors = NA,
  heatmap_colors = NA,
  cluster_rows = FALSE,
  cluster_cols = FALSE,
  show_rownames = FALSE,
  show_colnames = FALSE,
  cellwidth = NA,
  cellheight = NA,
  legend = FALSE,
  annotation_legend = FALSE,
  annotation_names_row = FALSE,
  ...
)
```

Arguments

`ConservedMarker`	Markers table
`start_col`	numeric, indicating the start column of marker power in each cell type
`module_colors`	vector, indicating colors of modules (annotation_colors)
`heatmap_colors`	vector, indicating colors used in heatmap
`cluster_rows`	boolean values determining if rows should be clustered or hclust object
`cluster_cols`	boolean values determining if columns should be clustered or hclust object
`show_rownames`	boolean specifying if column names are be shown
`show_colnames`	boolean specifying if column names are be shown
`cellwidth`	individual cell width in points. If left as NA, then the values depend on the size of plotting window
`cellheight`	individual cell height in points. If left as NA, then the values depend on the size of plotting window
`legend`	logical to determine if legend should be drawn or not
`annotation_legend`	boolean value showing if the legend for annotation tracks should be drawn

annotation_names_row
: boolean value showing if the names for row annotation tracks should be drawn

... : parameter in pheatmap

Value

pheatmap object

Examples

```
data("pbmc_small")
all.markers <- Identify_Markers(pbmc_small)
all.markers <- Format_Markers_Frac(all.markers)
Plot_MarkersHeatmap(all.markers[,c(2,6,7,8)])
```

Index

∗ **datasets**
OrthG_Hs_Ch, 9
OrthG_Hs_Mm, 9
OrthG_Hs_Zf, 9
OrthG_Mm_Ch, 10
OrthG_Mm_Zf, 10
OrthG_Zf_Ch, 10

CACIMAR_cols, 2

Format_Markers_Frac, 3

Heatmap_Cor, 3

Identify_CellType, 4
Identify_ConservedCellTypes, 5
Identify_ConservedMarkers, 6
Identify_ConservedNetworks, 7
Identify_Markers, *3*, 8

OrthG_Hs_Ch, 9
OrthG_Hs_Mm, 9
OrthG_Hs_Zf, 9
OrthG_Mm_Ch, 10
OrthG_Mm_Zf, 10
OrthG_Zf_Ch, 10

Plot_MarkersHeatmap, 11