Package ‘EMTscore’

April 17, 2025

Type Package

Title Calculate 'EMT' Scores Based on 'Omics' Data

Version 0.1.1

Date 2025-04-11

Maintainer Haimei Wen <hudie.luoluo@gmail.com>

Description Epithelial-Mesenchymal transition ('EMT') is an important form of cellular plasticity that is fully or partially activated in several biological scenarios including development and disease progression. 'EMT' involves altered expression of hundreds of protein-coding and non-protein-coding genes. Recent studies showed the prevalence of partial 'EMT' in multiple processes such as various cancers and organ fibrosis, which necessitates rigorous quantification of the degree of 'EMT'. While traditional gene set scoring methods such as gene set variation analysis have been used to generate 'EMT' scores from omics data, multiple 'EMT' scoring algorithms and 'EMT' gene sets have been used by different groups without standardization. Furthermore, comparisons of 'EMT' scores computed from different methods and/or different EMT gene sets are generally difficult due to both the context dependent nature of 'EMT' and the lack of tools that comprehensively integrate varying components for 'EMT' scoring. To address this problem, we have built a toolbox named 'EMTscore' that enables users to select scoring methods from a list of previously used algorithms and 'EMT' gene sets from a list of gene sets produced from different experiments. We provided several visualization methods for making publication quality plots of 'EMT' scores from 'omics' data. Furthermore, we showed a unique utility of a method based on principal component analysis for scoring divergent 'EMT' processes from a single dataset. Overall, 'EMTscore' provides an integrated solution for assessing the degree and complexity of 'EMT' from 'omics' data, and it paves the way for standardizing the comparison of EMT programs across multiple contexts.

License GPL-3

Encoding UTF-8

Imports ggplot2, dplyr, AUCell, GSVA, ggpubr, ComplexHeatmap, circlize, stats, gridExtra, magrittr, GSA, nsprcomp, stringr, foreach, doParallel, Seurat, grid, pheatmap, paletteer, ggthemes, curl

Suggests testthat, knitr, rmarkdown

RoxygenNote 7.3.2

LazyData false

NeedsCompilation no

Depends R (>= 3.5.0)

Author Haimei Wen [aut, cre],
Daniel Lopez [aut],
Tian Hong [aut]

Repository CRAN

Date/Publication 2025-04-17 07:40:05 UTC

Contents

`AUCellfunc` *AUCell EMT Score Calculation*

Description

Calculate AUCell scores for a given gene set across single-cell gene expression data.

Usage

```
AUCellfunc(geneExp, genes)
```

Arguments

`geneExp`	A matrix or data frame of gene expression data, where rows are genes and columns are cells/samples.
`genes`	A vector of gene names that defines the gene set for the AUCell analysis.

Value

A data frame with two columns: ‘SampleID‘ and ‘AUCell‘, representing the AUCell score for each cell/sample.

Examples

```
library(curl)
url <- "https://zenodo.org/record/15213845/files/geneExp.rda"
destfile <- tempfile(fileext = ".rda")
download.file(url, destfile, mode = "wb")
load(destfile)
data(Panchy_et_al_E_signature)
geneList = Panchy_et_al_E_signature
genes = unlist(geneList$GeneName)
AUCellMethod = AUCellfunc(t(geneExp),genes)
```

`cell_annotation_file` *cell_annotation_file*

Description

This is a cell type annotation file

Usage

```
data(cell_annotation_file)
```

Format

example data
    celltype annotation

Examples

```
data(cell_annotation_file)
```

`Execute_AUCell` *AUCell methods to calculate EMT score*

Description

AUCell methods to calculate EMT score

Usage

```
Execute_AUCell(geneExp, geneList, colnames)
```

Arguments

`geneExp`	A numeric matrix of gene expression values
`geneList`	A list of signature gene sets
`colnames`	A character string specifying the name to assign to the column containing the scores in the output data frame.

Value

A data frame containing sample/cell ID and EMT scores

Examples

```
library(curl)
url <- "https://zenodo.org/record/15213845/files/geneExp.rda"
destfile <- tempfile(fileext = ".rda")
download.file(url, destfile, mode = "wb")
load(destfile)
data(Panchy_et_al_E_signature)
Execute_AUCell(geneExp, geneList = Panchy_et_al_E_signature, colnames = "Escore")
```

`Execute_E_M_plot` *present EMT score result*

Description

present EMT score result

Usage

```
Execute_E_M_plot(
  cell_annotation_file,
  Methods,
  geneExp,
  M_signature_gene_sets,
  E_signature_gene_sets
)
```

Arguments

`cell_annotation_file`	cell annotation file in data frame, The column name for cell type is "celltype_annotation" and the column name for cell sampleID is "name".
`Methods`	can select nnPCA, ssGSEA, AUCell, SCSE
`geneExp`	gene expression matrix
`M_signature_gene_sets`	M signature gene list in dataframe, we supply example list such as geneList_M, M_signature_for_cancer, M_signature_for_cell
`E_signature_gene_sets`	E signature gene list in dataframe, we supply example list such as geneList_E, E_signature_for_cancer, E_signature_for_cell

Value

Figure represent EMT scores and cell type

Examples

```
library(nsprcomp)
library(ggplot2)
library(dplyr)
library(gridExtra)
library(curl)
Methods = "nnPCA"
data(cell_annotation_file)
url <- "https://zenodo.org/record/15213845/files/geneExp.rda"
destfile <- tempfile(fileext = ".rda")
download.file(url, destfile, mode = "wb")
load(destfile)
data(Panchy_et_al_E_signature)
data(Panchy_et_al_M_signature)


Execute_E_M_plot(cell_annotation_file,
Methods,
geneExp,
M_signature_gene_sets = Panchy_et_al_M_signature,
E_signature_gene_sets = Panchy_et_al_E_signature
)
```

`Execute_M_dimension_plot`

present M dimension score result

Description

present M dimension score result

Usage

```
Execute_M_dimension_plot(cell_annotation_file, geneExp, M_signature_gene_sets)
```

Arguments

`cell_annotation_file`	cell annotation file in data frame, The column name for cell type is "celltype_annotation" and the column name for cell sampleID is "name".
`geneExp`	gene expression matrix
`M_signature_gene_sets`	M signature gene list in dataframe, we supply example list such as geneList_M, M_signature_for_cancer, M_signature_for_cell

Value

Figure represent M dimension 1 and dimension 2 scores and cell type

Examples

```
library(nsprcomp)
library(ggplot2)
library(dplyr)
library(gridExtra)
library(curl)
data(cell_annotation_file)
url <- "https://zenodo.org/record/15213845/files/geneExp.rda"
destfile <- tempfile(fileext = ".rda")
download.file(url, destfile, mode = "wb")
load(destfile)
data(Panchy_et_al_M_signature)


Execute_M_dimension_plot(
cell_annotation_file ,
geneExp,
M_signature_gene_sets = Panchy_et_al_M_signature
)  # only for nnPCA method
```

`Execute_nnPCA` *nnPCA methods to calculate EMT score*

Description

nnPCA methods to calculate EMT score

Usage

```
Execute_nnPCA(geneExp, geneList, dimension, score_names)
```

Arguments

`geneExp`	A numeric matrix of gene expression values where rows represent genes and columns represent samples/cells.
`geneList`	A list of signature gene sets. It should contain a column named ‘GeneName‘ which lists the genes to be used in the analysis.
`dimension`	An integer specifying the number of principal components to compute (default is 1).
`score_names`	A character vector specifying the names to assign to the columns of the returned data frame containing the scores (e.g., c(’M1_score’, ’M2_score’)).

Value

A data frame containing sample/cell IDs and their respective EMT scores.

Examples

```
library(curl)
url <- "https://zenodo.org/record/15213845/files/geneExp.rda”
destfile <- tempfile(fileext = ".rda”)
download.file(url, destfile, mode = "wb”)
load(destfile)
data(Panchy_et_al_E_signature)
library(nsprcomp)
Execute_nnPCA(geneExp, Panchy_et_al_E_signature, dimension=1, score_names='E_score')
nnPCA_Mscore <- Execute_nnPCA(
geneExp,
Panchy_et_al_E_signature,
dimension=2,
score_names=c('M1_score','M2_score')
)
```

`Execute_nnPCA_plot` *present EM score and M dimension score result*

Description

present EM score and M dimension score result

Usage

```
Execute_nnPCA_plot(cell_annotation_file, geneExp, gene_lists, fig_title)
```

Arguments

`cell_annotation_file`	A data frame containing cell annotation, with columns ‘celltype_annotation‘ and ‘name‘.
`geneExp`	A gene expression matrix.
`gene_lists`	A list of M and E signature gene sets to be used for the nnPCA analysis.
`fig_title`	The title of the figure

Value

Figure represent two plots: 1): EM score 2): M dimension 1 and dimension 2 scores and cell type

Examples

```
library(ggpubr)
library(ggplot2)
library(nsprcomp)
library(dplyr)
library(gridExtra)
data(cell_annotation_file)
library(curl)
url <- "https://zenodo.org/record/15213845/files/geneExp.rda"
destfile <- tempfile(fileext = ".rda")
download.file(url, destfile, mode = "wb")
load(destfile)
data(Panchy_et_al_M_signature)
data(Panchy_et_al_E_signature)


gene_lists <- list(Panchy_et_al = list(Panchy_et_al_M_signature, Panchy_et_al_E_signature))

Execute_nnPCA_plot(
cell_annotation_file,
geneExp,
gene_lists,
fig_title = 'RPM'
)  # only for nnPCA method
```

`Execute_SCSE` *SCSE methods to calculate EMT score*

Description

SCSE methods to calculate EMT score

Usage

```
Execute_SCSE(geneExp, geneList, colnames)
```

Arguments

`geneExp`	A numeric matrix of gene expression values, where rows represent genes and columns represent samples/cells.
`geneList`	A list of signature gene sets, with each element containing gene names for the respective signature.
`colnames`	A character string representing the name to assign to the column containing the EMT scores in the output data frame.

Value

A data frame containing sample/cell ID and their respective EMT scores.

Examples

```
library(curl)
url <- "https://zenodo.org/record/15213845/files/geneExp.rda"
destfile <- tempfile(fileext = ".rda")
download.file(url, destfile, mode = "wb")
load(destfile)
data(Panchy_et_al_E_signature)
Execute_SCSE(geneExp, geneList = Panchy_et_al_E_signature,colnames = "Escore")
```

`Execute_ssGSEA` *ssGSEA methods to calculate EMT score*

Description

ssGSEA methods to calculate EMT score

Usage

```
Execute_ssGSEA(geneExp, geneList, colnames)
```

Arguments

`geneExp`	A numeric matrix of gene expression values where rows represent genes and columns represent samples/cells.
`geneList`	A list of signature gene sets to be used for ssGSEA analysis.
`colnames`	A character string specifying the name to assign to the column containing the scores in the output data frame.

Value

A data frame containing sample/cell ID and EMT scores

Examples

```
library(GSVA)
library(curl)
url <- "https://zenodo.org/record/15213845/files/geneExp.rda"
destfile <- tempfile(fileext = ".rda")
download.file(url, destfile, mode = "wb")
load(destfile)
data(Panchy_et_al_E_signature)
Execute_ssGSEA(geneExp, geneList = Panchy_et_al_E_signature, colnames = "Escore")
```

`Panchy_et_al_E_signature`

Panchy_et_al_E_signature

Description

This is an example dataset.

Usage

```
data(Panchy_et_al_E_signature)
```

Format

E gene list from Panchy_et_al

E gene

Examples

```
data(Panchy_et_al_E_signature)
```

`Panchy_et_al_M_signature`

Panchy_et_al_M_signature

Description

This is an example dataset.

Usage

```
data(Panchy_et_al_M_signature)
```

Format

M gene list from Panchy_et_al

M gene

Examples

```
data(Panchy_et_al_M_signature)
```

`plot_heatmap_function` *present result of heat map of genes that contribute to M score of dimension 1 and disension 2*

Description

present result of heat map of genes that contribute to M score of dimension 1 and disension 2

Usage

```
plot_heatmap_function(geneExp, geneList_M)
```

Arguments

`geneExp`	gene expression matrix
`geneList_M`	M signature gene list in dataframe, we supply example list such as geneList_M, M_signature_for_cancer, M_signature_for_cell

Value

Figure represent EMT score for different methods and gene sets

Examples

```
library(ggpubr)
library(ggplot2)
library(dplyr)
library(nsprcomp)
library(pheatmap)
library(grid)
library(gridExtra)
library(circlize)
library(paletteer)
library(ggthemes)
library(ComplexHeatmap)
library(curl)
url <- "https://zenodo.org/record/15213845/files/geneExp.rda"
destfile <- tempfile(fileext = ".rda")
download.file(url, destfile, mode = "wb")
```

```
load(destfile)
data(Panchy_et_al_M_signature)
geneList_M = Panchy_et_al_M_signature


plot_heatmap_function( geneExp, geneList_M)
```

`SingleCellSigExplorer` *SingleCellSigExplorer*

Description

A function to print the words "New function!"

Usage

```
SingleCellSigExplorer(data, genes)
```

Arguments

`data`	A numeric matrix where rows are genes and columns are cells or samples.
`genes`	A character vector of gene names to include in the signature.

Value

A character vector

Index