

# Package ‘SEPA’

April 23, 2016

**Type** Package

**Title** SEPA

**Version** 1.0.0

**Date** 2015-04-17

**Author** Zhicheng Ji, Hongkai Ji

**Maintainer** Zhicheng Ji <zji4@jhu.edu>

**Description** Given single-cell RNA-seq data and true experiment time of cells or pseudo-time cell ordering, SEPA provides convenient functions for users to assign genes into different gene expression patterns such as constant, monotone increasing and increasing then decreasing. SEPA then performs GO enrichment analysis to analysis the functional roles of genes with same or similar patterns.

**License** GPL(>=2)

**Imports** ggplot2, shiny, topGO, segmented, reshape2, org.Hs.eg.db, org.Mm.eg.db

**VignetteBuilder** knitr

**Suggests** knitr

**Depends** R(>= 2.10.0)

**biocViews** GeneExpression, Visualization, GUI, GO

**NeedsCompilation** no

## R topics documented:

HSMMdata . . . . .	2
patternGOanalysis . . . . .	2
patternsummary . . . . .	3
pseudotime . . . . .	4
pseudotimepattern . . . . .	5
pseudotimevisualize . . . . .	6
SEPA . . . . .	7
SEPAui . . . . .	7
truetime . . . . .	8

truetimepattern . . . . .	8
truetimevisualize . . . . .	9
windowGOanalysis . . . . .	10
windowGOvisualize . . . . .	11

<b>Index</b>	<b>12</b>
--------------	-----------

---

HSMMdata	<i>Single-cell RNA-seq data for HSMM (Human Skeletal Muscle Myoblast)</i>
----------	---

---

### Description

The R object is a log2 transformed expression matrix (HSMMdata) with 518 genes and 271 cells. This dataset is a subset of original gene expression profiles, which is available in the HSMMSingleCell package.

### Format

A matrix

### Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52529>

### References

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., ... & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*.

---

patternGOanalysis	<i>patternGOanalysis</i>
-------------------	--------------------------

---

### Description

Performs GO analysis on genes for each pattern

### Usage

```
patternGOanalysis(pattern, type = NULL, termnum = 20,
  identifier = "ENSEMBL", species = "Human")
```

**Arguments**

pattern	The direct output of <code>truetimepattern</code> or <code>pseudotimepattern</code> function.
type	Character value of specific pattern to perform the GO analysis. If NULL GO analysis will be performed for all patterns.
termnum	Number of top GO terms to be displayed.
identifier	The identifier of the genes. It should be one of the following: "Entrez", "GenBank", "Alias", "Ensembl", "Gene", "Symbol", "GeneName" and "UniGene"
species	The species of the genes. Currently only "Human" and "Mouse" are supported

**Details**

This function is basically a wrap up of functions in the `topGO` package. It takes as input the output of `truetimepattern` or `pseudotimepattern` function. For each pattern, the GO enrichment analysis is performed where input genes are genes with specific patterns and background genes are all genes in the expression profile. Users should correctly select identifier and species, otherwise the function may breakdown.

**Value**

A list where each element is a `data.frame` containing the results of GO analysis.

**Author(s)**

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

**Examples**

```
library(topGO)
data(HSMMdata)
pattern <- truetimepattern(HSMMdata, truetime, removeconstant=TRUE)
patternGOanalysis(pattern, termnum=20, identifier="ENSEMBL", species="Human")
```

---

patternsummary

*patternsummary*

---

**Description**

Count table for pattern

**Usage**

```
patternsummary(pattern)
```

**Arguments**

pattern	The direct output of <code>truetimepattern</code> or <code>pseudotimepattern</code> function.
---------	---

**Details**

This function generates a count table of number of genes with each pattern.

**Value**

A data.frame object. First column: pattern; Second column: number of genes

**Author(s)**

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

**Examples**

```
data(HSMMdata)
pattern <- truetimepattern(HSMMdata, truetime, removeconstant=TRUE)
patternsummary(pattern)
```

---

pseudotime

*Pseudo-time Cell Ordering for HSMM Single-cell RNA-seq data*

---

**Description**

The dataset contains a data.frame with two columns specifying pseudo-time calculated by Monocle. Note that only cells on the main path are retained. Cells belonging to the contaminated path are excluded. Please check HSMMSingleCell package for details.

**Format**

A data.frame with two columns

**Source**

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52529>

**References**

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., ... & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*.

---

pseudotimepattern      *pseudotimepattern*

---

## Description

Identify Pattern for Pseudo Temporal Cell Ordering

## Usage

```
pseudotimepattern(expr, pseudotime, simplify = T, removeconstant = F,  
plot = F, gap = 10)
```

## Arguments

expr	The matrix of gene expression profile.
pseudotime	A character data.frame or matrix of pseudo-time. First column: Cell name; Second column: pseudo-time.
simplify	Whether to simplify pattern so that same neighboring patterns will be reduced to one. For example "up_up_constant" will be simplified to "up_constant".
removeconstant	Whether to remove all constant patterns. For example "up_up_constant" will be simplified to "up_up". This step will be performed before simplify.
plot	Whether to generate plot for genes with transition points.
gap	Number of first and last gap cells that will be excluded when considering transition points.

## Details

Identify the gene expression patterns for true experiment time. For the expressions of each gene, the function performs t-tests for cells from neighboring time points. The expression pattern for cells from neighboring time points could be increasing, decreasing or constant. All patterns are concatenated using "-" to form the final pattern.

## Value

A list. expr: original expression matrix; pseudotime: original pseudotime; pattern: a list containing results of different patterns. For single patterns, it is a named vector where values are the p-values of the t-test of the simple linear regression slope coefficient. The vector is ordered according to the p-values. For transition patterns, a data.frame containing the mean and confidence interval of the transition point. It is ordered according to the transition points; fitexpr: the fitted expression matrix

## Author(s)

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

## Examples

```
data(HSMMdata)
pseudotimepattern(HSMMdata,pseudotime)
```

---

pseudotimevisualize *pseudotimevisualize*

---

## Description

Visualize Gene Expression Pattern for Pseudo Temporal Cell Ordering

## Usage

```
pseudotimevisualize(pattern, gene, showtrue = F)
```

## Arguments

pattern	The exact output of the pseudotimepattern function.
gene	A character value or vector of gene names. Should be included in the gene expression matrix.
showtrue	For the heatmap of multiple gene, whether to display true gene expression values or fitted gene expression values.

## Details

Visualize gene expression pattern of one or multiple genes for pseudo temporal cell ordering. For one gene, a scatterplot with fitted lines will be generated. For multiple genes, a heatmap with fitted or true expression values will be generated.

## Value

A ggplot2 object

## Author(s)

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

## Examples

```
data(HSMMdata)
pattern <- pseudotimepattern(HSMMdata,pseudotime)
pseudotimevisualize(pattern,"ENSG00000108821.9")
pseudotimevisualize(pattern,c("ENSG00000108821.9","ENSG00000187193.8"))
```

---

SEPA

*SEPA: Single-Cell Gene Expression Pattern Analysis*

---

### Description

SEPA provides useful functions to analysis gene expression patterns and perform GO analysis for single-cell RNA-seq data.

### Details

Given single-cell RNA-seq data and true experiment time of cells or pseudo-time cell ordering, SEPA provides convenient functions for users to assign genes into different gene expression patterns such as constant, monotone increasing and increasing then decreasing. SEPA then performs GO enrichment analysis to analysis the functional roles of genes with same or similar patterns. SEPA comes with a user-friendly Graphical User Interface written in shiny.

---

SEPAui

*SEPAui*

---

### Description

Launch the SEPA user interface in local machine

### Usage

```
SEPAui()
```

### Details

This function will automatically launch the SEPA user interface in a web browser. It provides a much easier and more convenient way to analysis gene expression patterns and perform GO analysis for single-cell RNA-seq data. The user interface can also be accessed by <http://zhiji.shinyapps.io/SEPA>. Neither R nor any packages are required in this online version. However, it is highly recommended that the user interface be launched locally for faster running speed.

### Author(s)

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

### Examples

```
## Not run:  
SEPAui()  
  
## End(Not run)
```

---

truetime

*True Experiment Time Points for HSMM Single-cell RNA-seq data*


---

### Description

The dataset contains a data.frame with two columns specifying true experiment time points. Note that only cells on the main path are retained. Cells belonging to the contaminated path are excluded. Please check HSMMSingleCell package for details.

### Format

A data.frame with two columns

### Source

<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52529>

### References

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., ... & Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology*.

---

truetimepattern

*truetimepattern*


---

### Description

Identify Pattern for True Experiment Time

### Usage

```
truetimepattern(expr, truetime, simplify = T, removeconstant = F)
```

### Arguments

expr	The matrix of gene expression profile.
truetime	A character data.frame or matrix of true experimental time. First column: Cell name; Second column: experiment time.
simplify	Whether to simplify pattern so that same neighboring patterns will be reduced to one. For example "up_up_constant" will be simplified to "up_constant".
removeconstant	Whether to remove all constant patterns. For example "up_up_constant" will be simplified to "up_up". This step will be performed before simplify.



**Details**

Identify the gene expression patterns for true experiment time. For the expressions of each gene, the function performs t-tests for cells from neighboring time points. The expression pattern for cells from neighboring time points could be increasing, decreasing or constant. All patterns are concatenated using "-" to form the final pattern.

**Value**

A named vector or patterns. The names corresponds to gene names.

**Author(s)**

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

**Examples**

```
data(HSMMdata)
truetimepattern(HSMMdata, truetime)
```

---

truetimevisualize      *truetimevisualize*

---

**Description**

Visualize Gene Expression Pattern for True Experiment Time

**Usage**

```
truetimevisualize(expr, truetime, gene, mode = c("mean", "median"))
```

**Arguments**

<code>expr</code>	The matrix of gene expression profile.
<code>truetime</code>	A character data.frame or matrix of true experimental time. First column: Cell name; Second column: experiment time.
<code>gene</code>	A vector of gene names to be plotted.
<code>mode</code>	A character value specifying mean or median to be displayed

**Details**

Identify the gene expression patterns for true experiment time. For the expressions of each gene, the function performs t-tests for cells from neighboring time points. The expression pattern for cells from neighboring time points could be increasing, decreasing or constant. All patterns are concatenated using "-" to form the final pattern.

**Value**

A ggplot2 object.

**Author(s)**

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

**Examples**

```
data(HSMMdata)
truetimevisualize(HSMMdata,truetime,c("ENSG00000122180.4", "ENSG00000125968.7"))
```

---

windowGOanalysis      *windowGOanalysis*

---

**Description**

Performs GO analysis with moving window for transition patterns

**Usage**

```
windowGOanalysis(pattern, type = "constant_up", windowsize = NULL,
  movesize = NULL, termnum = 20, identifier = "ENSEMBL",
  species = "Human")
```

**Arguments**

pattern	The direct output of pseudotimepattern function.
type	The type of transition pattern.
windowsize	The number of genes in each group.
movesize	How many genes to move forward each time.
termnum	Number of top GO terms to be displayed.
identifier	The identifier of the genes. It should be one of the following: "Entrez", "GenBank", "Alias", "Ensembl", "Gene", "Symbol", "GeneName" and "UniGene"
species	The species of the genes. Currently only "Human" and "Mouse" are supported

**Details**

This function is specifically designed for GO analysis of genes with a specific transition pattern. GO analyses are performed iteratively on a group of genes with similar transition points. Users can define the windowsize (the number of genes in each group) and the movesize (how many genes to move forward each time).

**Value**

A list where each element is a data.frame containing the results of GO analysis. The name of the list specifies the group of genes.

**Author(s)**

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

**Examples**

```
library(topGO)
data(HSMMdata)
pattern <- pseudotimepattern(HSMMdata,pseudotime)
windowGOanalysis(pattern,type="constant_up")
```

---

windowGOvisualize      *windowGOvisualize*

---

**Description**

Visualize results of GO analysis with moving window for transition patterns

**Usage**

```
windowGOvisualize(GOres, GOTerm = NULL, topterm = 2, mode = "Heatmap")
```

**Arguments**

GOres	The direct output of windowGOanalysis function.
GOTerm	The name of GO term to be displayed. If NULL, top GO terms will be displayed instead
topterm	The number of top GO terms to be displayed. This argument only works when GOTerm is NULL.
mode	To plot in heatmap or line graph. Either "Heatmap" or "Line".

**Details**

This function is specifically designed to visualize the results obtained from windowGOanalysis function. Users can choose to visualize specific GO terms or top GO terms in each time window

**Value**

A ggplot2 object.

**Author(s)**

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

**Examples**

```
library(topGO)
data(HSMMdata)
pattern <- pseudotimepattern(HSMMdata,pseudotime)
windowGOvisualize(windowGOanalysis(pattern,type="constant_up"))
```

# Index

HSMMdata, [2](#)

patternGOanalysis, [2](#)

patternsummary, [3](#)

pseudotime, [4](#)

pseudotimepattern, [5](#)

pseudotimevisualize, [6](#)

SEPA, [7](#)

SEPA-package (SEPA), [7](#)

SEPAui, [7](#)

truetime, [8](#)

truetimepattern, [8](#)

truetimevisualize, [9](#)

windowGOanalysis, [10](#)

windowGOvisualize, [11](#)