Package 'LEAP'

September 13, 2016

Type Package

Title Constructing Gene Co-Expression Networks for Single-Cell RNA-Sequencing Data Using Pseudotime Ordering

Version 0.2

Date 2016-09-09

Author Alicia T. Specht and Jun Li

Maintainer Alicia T. Specht <aspecht2@nd.edu>

Suggests ggplot2

Description Advances in sequencing technology now allow researchers to capture the expression profiles of individual cells. Several algorithms have been developed to attempt to account for these effects by determining a cell's so-called `pseudotime', or relative biological state of transition. By applying these algorithms to single-cell sequencing data, we can sort cells into their pseudotemporal ordering based on gene expression. LEAP (Lag-based Expression Association for Pseudotime-series) then applies a time-series inspired lag-based correlation analysis to reveal linearly dependent genetic associations.

License GPL-2

LazyData yes

NeedsCompilation no

Repository CRAN

Date/Publication 2016-09-13 21:19:05

R topics documented:

LEAP-package *Constructing Gene Co-Expression Networks for Single-Cell RNA-Sequencing Data Using Pseudotime Ordering*

Description

Advances in sequencing technology now allow researchers to capture the expression profiles of individual cells. Several algorithms have been developed to attempt to account for these effects by determining a cell's so-called 'pseudotime', or relative biological state of transition. By applying these algorithms to single-cell sequencing data, we can sort cells into their pseudotemporal ordering based on gene expression. LEAP (Lag-based Expression Association for Pseudotime-series) then applies a time-series inspired lag-based correlation analysis to reveal linearly dependent genetic associations.

Details

The DESCRIPTION file:

Package:	LEAP
Type:	Package
Title:	Constructing Gene Co-Expression Networks for Single-Cell RNA-Sequencing Data Using Pseudotime Orderin
Version:	0.2
Date:	2016-09-09
Author:	Alicia T. Specht and Jun Li
Maintainer:	Alicia T. Specht <aspecht2@nd.edu>
Suggests:	ggplot2
Description:	Advances in sequencing technology now allow researchers to capture the expression profiles of individual cells
License:	GPL-2
LazyData:	yes

Index of help topics:

LEAP-package	Constructing Gene Co-Expression Networks for Single-Cell RNA-Sequencing Data Using Pseudotime Ordering
MAC_counter	Function to perform lag-based correlation anaylsis of single-cell sequencing data, sorted by pseudotime.
MAC_example	Numerical matrix
MAC_lags	Internal function used by MAC_counter and MAC_perm
MAC_perm	Function to perform a permutation analysis to determine a cutoff for significant MAC values.
MAC_symmetric	Numeric data frame
example_data	Numeric data frame of example data

lag_example	Integer data frame
perm_example	The resulting data ouptut from applying MAC_perm() to example_data

Further information is available in the following vignettes:

LEAP_Vignette	LEAP_Vignette (source, pdf)

~~ An overview of how to use the package, including the most important functions ~~

Author(s)

Alicia T. Specht and Jun Li

Maintainer: Alicia T. Specht <aspecht2@nd.edu>

References

Shalek AK, Satija R., Shuga J., Trombetta J.J. et al. (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature, 510(7505), 363-369. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4193940/

Examples

```
## Not run: MAC_results = MAC_counter(data=example_data)

## Not run: MAC_perm(data=example_data, MACs_observ=MAC_example)
```

---

example_data *Numeric data frame of example data*

Description

Contains 20 genes across 564 single-cell sequencing experiments, sorted by pseudotime using the package Monocle.

Usage

```
data("example_data")
```

Format

A data frame with 20 observations on 564 variables.

Details

See vignette for more details.

Source

Shalek AK, Satija R., Shuga J., Trombetta J.J. et al. (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature, 510(7505), 363-369. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4193940/

Examples

```
data(example_data)
```

---

lag_example *Integer data frame*

Description

Contains the resulting lag matrix from applying MAC_counter() to example_data

Usage

```
data("lag_example")
```

Format

A data frame with 20 observations on 20 variables.

Details

See vignette for more details.

Examples

```
data(lag_example)
```

MAC_counter *Function to perform lag-based correlation anaylsis of single-cell sequencing data, sorted by pseudotime.*

Description

See vignette for more details.

Usage

```
MAC_counter(data, max_lag_prop = 1/3, MAC_cutoff = 0.2,
file_name = F, lag_matrix = T, symmetric = F)
```

Arguments

data	A data matrix for which the rows are genes and the columns are experiments, sorted by their pseudotime.
max_lag_prop	The largest proportion of your experiments that you want the size of the lag to be. Recommended not to go beyond 1/3. Default value is 1/3.
MAC_cutoff	The lowest MAC desired to be shown in results. Default value is 0.2.
file_name	The name to be used for resulting .csv files. I.e., using file_name="mine" would create the file "MAC_mine.csv" and "lag_mine.csv"
lag_matrix	Logical. TRUE indicates that the resulting matrix of lags should be saved to a csv. Default value is TRUE.
symmetric	Logical. TRUE indicates that a symmetric matrix is required for output MAC .csv file. Results in the absolute maximum value for each pair (i,j) and (j,i). Default value is FALSE.

Details

See vignette for more details.

Value

Returns a dataset with four columns: Row gene index and Column gene index correspond to the indeces for the gene pair (i,j), Correlation is the maximum absolute correlation (MAC) achieved for the pair, and Lag is the lag at which the MAC occurred.

Examples

```
x <- matrix(rnorm(6),2,3)
y <- MAC_counter(x)

## Not run: MAC_results = MAC_counter(data=example_data)
```

MAC_example *Numerical matrix*

Description

Contains the matrix of MACs that result from applying MAC_counter() to example_data

Usage

```
data("MAC_example")
```

Format

The format is: num [1:20, 1:20] NA 0.141 0.418 0.253 0.164 ...

Details

See vignette for more details.

Examples

```
data(MAC_example)
```

`MAC_lags` *Internal function used by MAC_counter and MAC_perm*

Description

Performs the lag-based correlation analysis

Usage

```
MAC_lags(data, max_lag_prop = 1/3, symmetric = F)
```

Arguments

`data`	Data being passed to the function
`max_lag_prop`	The largest proportion of your experiments that you want the size of the lag to be. Recommended not to go beyond 1/3. Default value is 1/3.
`symmetric`	Logical. TRUE indicates that a symmetric matrix is required for output MAC .csv file. Results in the absolute maximum value for each pair (i,j) and (j,i). Default value is FALSE.

Details

See vignette for more details.

Value

Returns MAC and associated lag matrices.

Examples

```
x <- matrix(rnorm(6),2,3)
y <- MAC_lags(x)

## Not run: MAC_results = MAC_lags(data=example_data, max_lag_prop=1/3, symmetric=F)
```

MAC_perm *Function to perform a permutation analysis to determine a cutoff for significant MAC values.*

Description

See vignette for more details.

Usage

```
MAC_perm(data, MACs_observ, num_perms = 100, max_lag_prop = 1/3,
FDR_cutoffs = 101, perm_file_name = F)
```

Arguments

data	A data matrix for which the rows are genes and the columns are experiments, sorted by their pseudotime.
MACs_observ	The resulting matrix of MACs from running MAC_counter on the dataset
num_perms	The number of permutations to be performed. Default is 100.
max_lag_prop	The largest proportion of your experiments that you want the size of the lag to be. Recommended not to go beyond 1/3. Default value is 1/3.
FDR_cutoffs	The number of cutoffs between 0 and 1 to use for FDR analysis. Default value is 101, resulting in 0,0.01,0.02,...,0.98,0.99,1.
perm_file_name	The name to be used for resulting .csv file. I.e., using perm_file_name="mine" would create the file "perm_mine.csv"

Details

See vignette for more details.

Value

Returns a dataset with four columns: cors are the correlation cutoffs, MACs_observed are the number of observed correlations at that cutoff, MACs_ave_perm are the average number observed in the permuted datasets at that cutoff, and fdr is the false discovery rate (FDR) observed at that cutoff.

Examples

```
x <- matrix(rnorm(6),2,3)
cor <- cor(x)
y <- MAC_perm(x, cor)

## Not run: MAC_perm(data=example_data, MACs_observ=MAC_example)
```

MAC_symmetric *Numeric data frame*

Description

Contains the results of applying MAC_counter(symmetric=T) to example_data

Usage

```
data("MAC_symmetric")
```

Format

A data frame with 20 observations on 20 variables.

Examples

```
data(MAC_symmetric)
```

perm_example *The resulting data ouptut from applying MAC_perm() to example_data*

Description

101 observations at correlation cutoffs, from 0 to 1, of the four variables Cors, MACs_observed, MACs_ave_perm, and fdr.

Usage

```
data("perm_example")
```

Format

A data frame with 101 observations on 4 variables.

Details

See vignette for more details

Examples

```
data(perm_example)
```

Index