Package 'timeSeq'

October 14, 2022

Type Package

Title Detecting Differentially Expressed Genes in Time Course RNA-Seq Data

Version 1.0.4

Date 2019-2-08

Author Fan Gao, Xiaoxiao Sun

Maintainer Fan Gao <fangaohz@gmail.com>

Description A negative binomial mixed-effects (NBME) model to detect nonparallel differential expression(NPDE) genes and parallel differential expression(PDE) genes in the time course RNA-seq data.

Depends R (>= 3.0.1)

Imports gss, mgcv, lattice, pheatmap, reshape, grDevices, graphics

License GPL (>= 2)

Suggests rmarkdown, knitr

VignetteBuilder knitr

Repository CRAN

NeedsCompilation no

RoxygenNote 6.1.1.9000

Date/Publication 2019-02-09 05:50:03 UTC

R topics documented:

`timeSeq_1.0.4-package` *Statistical Inference for Time Course RNA-Seq Data using a Negative Binomial Mixed-Effects Model*

Description

In this package, we propose a negative binomial mixed-effects (NBME) model to identify differentially expressed (DE) genes, including nonparallel differentially expressed (NPDE) and parallel differentially expressed (PDE) genes, in time course RNA-seq data.

Details

Package:	timeSeq
Type:	Package
Version:	1.0.4
Date:	2019-2-08
License:	GPL-2 | GPL-3

Author(s)

Fan Gao and Xiaoxiao Sun

Maintainer: Fan Gao <fangaohz@gmail.com>

References

Sun, Xiaoxiao, David Dalpiaz, Di Wu, Jun S. Liu, Wenxuan Zhong, and Ping Ma. "Statistical inference for time course RNA-Seq data using a negative binomial mixed-effect model." BMC Bioinformatics, 17(1):324, 2016.

Chong Gu. Model diagnostics for smoothing spline ANOVA models. Canadian Journal of Statistics, 32(4):347-358, 2004.

Chong Gu. Smoothing spline ANOVA models. Springer, second edition, 2013.

Chong Gu and Ping Ma. Optimal smoothing in nonparametric mixed-effect models. Annals of Statistics, pages 1357-1379, 2005a.

Wood (2001) mgcv:GAMs and Generalized Ridge Regression for R. R News 1(2):20-25

pAbp *Example Data in Vignette*

Description

The data 'pAbp'(list) has 4 arguments, including 'data.count', 'exon.length', 'gene.names', and 'group.label'. Please refer to 'timeSeq.Rd' for more information.

Usage

```
data(pAbp)
```

Format

A list with 4 arguments:

Examples

```
data(pAbp)
str(pAbp)
```

simulate.dt *Example Data in Vignette*

Description

The data 'simulate.dt'(list) has 3 arguments, including 'data.count', 'group.label', and 'gene.names'. Please refer to 'timeSeq.Rd' for more information.

Usage

```
data(simulate.dt)
```

Format

A list with 3 arguments:

Examples

```
data(simulate.dt)
str(simulate.dt)
```

timeSeq *Statistical Inference for Time Course RNA-Seq Data using a Negative Binomial Mixed-Effects Model*

Description

Accurately identifying differentially expressed (DE) genes from time course RNA-seq data has been of tremendous significance in creating a global picture of cellular function. DE genes from the time course RNA-seq data can be classified into two types, parallel DE genes (PDE) and non-parallel DE (NPDE) genes. The former are often biologically irrelevant, whereas the latter are often biologically interesting. In this package, we propose a negative binomial mixed-effects (NBME) model to identify both PDE and NPDE genes in time course RNA-seq data.

Usage

```
timeSeq(data.count,group.label,gene.names,exon.length=NULL,exon.level=FALSE,pvalue=TRUE)
```

Arguments

data.count	a n by p matrix of expression values. Data should be appropriately normalized beforehand.
group.label	a vector indicating the experimental conditions of each time point.
gene.names	a vector containing all the gene names.
exon.length	a vector containing the length of exons, only used in exon level data.
exon.level	logical:indicating if this is an exon level dataset. Default is FALSE.
pvalue	logical:indicating if p-values are returned. Default is TRUE.

Details

Nonparallel differential expression(NPDE) genes and parallel differential expression(PDE) genes detection.

Value

A list with components

sorted	an object returned by timeSeq.sort function. It contains sorted Kullback Leibler Ratios(KLRs) or p-values for identifying DE genes.
count	the number of exons or replicates for each gene.
NPDE	the NPDE ratios or p-values.
PDE	the PDE ratios or p-values.
genenames	gene names.
table	gene expression values.
data	a n by p matrix of expression values.

gene.names	a vector including all the gene names.
group.label	a vector indicating the experimental conditions of each time point.
group.length	the total number of time points.
group1.length	the number of time points of condition one.
group2.length	the number of time points of condition two.
exon.level	logical:indicating if this is an exon level dataset. Default is FALSE.
pvalue	logical:indicating if p-values are returned. Default is TRUE.

Author(s)

Fan Gao and Xiaoxiao Sun

References

Sun, Xiaoxiao, David Dalpiaz, Di Wu, Jun S. Liu, Wenxuan Zhong, and Ping Ma. "Statistical inference for time course RNA-Seq data using a negative binomial mixed-effect model." BMC Bioinformatics, 17(1):324, 2016.

Chong Gu. Model diagnostics for smoothing spline ANOVA models. Canadian Journal of Statistics, 32(4):347-358, 2004.

Chong Gu. Smoothing spline ANOVA models. Springer, second edition, 2013.

Chong Gu and Ping Ma. Optimal smoothing in nonparametric mixed-effect models. Annals of Statistics, 1357-1379, 2005.

Wood (2001) mgcv:GAMs and Generalized Ridge Regression for R. R News 1(2):20-25

Examples

```
####Data should be appropriately normalized beforehand####

##Exon level data (The p-values calculation is not supported)
data(pAbp)
attach(pAbp)
model.fit <- timeSeq(data.count,group.label,gene.names,exon.length,exon.level=TRUE,pvalue=FALSE)
#NPDE genes have large KLRs
model.fit$NPDE
detach(pAbp)

##Gene level data (three replicates)
data(simulate.dt)
attach(simulate.dt)
model.fit <- timeSeq(data.count,group.label,gene.names,exon.level=FALSE,pvalue=TRUE)
#p-values
model.fit$NPDE
```

`timeSeq.heatmap` *Heatmap of the Most Significant NDPE Genes*

Description

Heatmap for the most significant NDPE genes.

Usage

```
timeSeq.heatmap(timeSeq.obj,n)
```

Arguments

`timeSeq.obj`	an object returned by timeSeq function
`n`	the number of the most significant NPDE genes. It must be a positive integer.

Author(s)

Fan Gao and Xiaoxiao Sun

References

Sun, Xiaoxiao, David Dalpiaz, Di Wu, Jun S. Liu, Wenxuan Zhong, and Ping Ma. "Statistical inference for time course RNA-Seq data using a negative binomial mixed-effect model." BMC Bioinformatics, 17(1):324, 2016.

Chong Gu. Model diagnostics for smoothing spline ANOVA models. Canadian Journal of Statistics, 32(4):347-358, 2004.

Chong Gu. Smoothing spline ANOVA models. Springer, second edition, 2013.

Chong Gu and Ping Ma. Optimal smoothing in nonparametric mixed-effect models. Annals of Statistics, pages 1357-1379, 2005a.

Wood (2001) mgcv:GAMs and Generalized Ridge Regression for R. R News 1(2):20-25

Examples

```
data(simulate.dt)
attach(simulate.dt)
model.fit <- timeSeq(data.count,group.label,gene.names,exon.level=FALSE,pvalue=TRUE)
timeSeq.heatmap(model.fit,n=10)
```

`timeSeq.screeplot` *Scree Plot of Kullback Leibler Distance Ratios*

Description

Plot the scree plot for all genes in the dataset.

Usage

```
timeSeq.screeplot(timeSeq.obj,type=c("barplot","lines"))
```

Arguments

`timeSeq.obj`	an object returned by timeSeq function.
`type`	type of plot: "barplot" for the bar plot and "lines" for the line graph.

Author(s)

Fan Gao and Xiaoxiao Sun

References

Sun, Xiaoxiao, David Dalpiaz, Di Wu, Jun S. Liu, Wenxuan Zhong, and Ping Ma. "Statistical inference for time course RNA-Seq data using a negative binomial mixed-effect model." BMC Bioinformatics, 17(1):324, 2016.

Chong Gu. Model diagnostics for smoothing spline ANOVA models. Canadian Journal of Statistics, 32(4):347-358, 2004.

Chong Gu. Smoothing spline ANOVA models. Springer, second edition, 2013.

Chong Gu and Ping Ma. Optimal smoothing in nonparametric mixed-effect models. Annals of Statistics, pages 1357-1379, 2005a.

Wood (2001) mgcv:GAMs and Generalized Ridge Regression for R. R News 1(2):20-25

Examples

```
data(simulate.dt)
attach(simulate.dt)
model.fit <- timeSeq(data.count,group.label,gene.names,exon.level=FALSE,pvalue=FALSE)
timeSeq.screeplot(model.fit,"lines")
```

`timeSeq.sort` *Sort NDPE Genes by Kullback Leibler Distance Ratios or P-values*

Description

Sort all genes in the dataset by their Kullback Leibler distance ratios or p-values.

Usage

```
timeSeq.sort(genenames,npde,pde,table,count,pvalue)
```

Arguments

`genenames`	a vector of gene names.
`npde`	a vector of Kullback Leibler distance ratios or p-values for NPDE genes.
`pde`	a vector of Kullback Leibler distance ratios or p-values for PDE genes.
`table`	gene expression values.
`count`	the number of exons or replicates for each gene.
`pvalue`	logical:indicating if p-values are returned.

Value

A list with components

`npde.list`	dataframe of NPDE genes sorted by KLRs or p-values.
`pde.list`	dataframe of PDE genes sorted by KLRs or p-values.
`table1`	gene expression values for each gene, corresponding to npde.list.
`table2`	gene expression values for each gene, corresponding to pde.list.

Author(s)

Fan Gao and Xiaoxiao Sun

References

Sun, Xiaoxiao, David Dalpiaz, Di Wu, Jun S. Liu, Wenxuan Zhong, and Ping Ma. "Statistical inference for time course RNA-Seq data using a negative binomial mixed-effect model." BMC Bioinformatics, 17(1):324, 2016.

Chong Gu. Model diagnostics for smoothing spline ANOVA models. Canadian Journal of Statistics, 32(4):347-358, 2004.

Chong Gu. Smoothing spline ANOVA models. Springer, second edition, 2013.

Chong Gu and Ping Ma. Optimal smoothing in nonparametric mixed-effect models. Annals of Statistics, pages 1357-1379, 2005a.

Wood (2001) mgcv:GAMs and Generalized Ridge Regression for R. R News 1(2):20-25

Index