

Package ‘DREAMSeq’

December 25, 2017

Type Package

Title Differentially expressed gene detection based on double Poisson model, negative binomial model or the mixture of them for RNA-seq data

Version 1.0

Date 2017-10-31

Author Zhihua Gao, Wenqiang Tang

Maintainer Zhihua Gao <gzhheuet@sina.com>

Description Using double Poisson model, negative binomial model or the mixture of them detect differentially expressed genes for RNA-seq data.

License GPL (>= 3)

Imports matrixStats (>= 0.50.1), edgeR (>= 3.10.5), DESeq (>= 1.20.0)

Depends R (>= 3.3.2), Biobase (>= 2.28.0), methods, limma (>= 3.24.15), BiocParallel (>= 1.2.22), locfit (>= 1.5-9.1), pbapply (>= 1.3-2)

NeedsCompilation no

LazyLoad yes

R topics documented:

DREAMSeq-package	2
comMeanVariance	3
DEGCountSet-class	3
DP	4
dpWaldTest	5
estCommonDispersion	5
estNormFactor	6
estShrinkDispersion	6
exCounts	7
exDispLNMean	7
exDispLNSd	8
exEstDispersionMethod	8
exEstNormFactorMethod	9
exFitMethod	9
exModel	10
exNormFactor	10

exResult	11
foxtail	11
headWaldTest	12
main	12
modelPipeline	13
nbWaldTest	14
scanMeans	15

Index	16
--------------	-----------

DREAMSeq-package	<i>Introduction to the DREAMSeq package</i>
------------------	---

Description

DREAMSeq is developed based on double Poisson model, negative binomial model or the mixture of them to identify differentially expressed genes for RNA-seq data.

Details

Package: DREAMSeq
 Type: Package
 Version: 1.0
 Date: 2017-10-31
 License: GPL (>= 3)

For more details, please refer to the publication below.

Author(s)

Zhihua Gao, gzhheuet@sina.com

References

Zhihua Gao, Zhiying Zhao and Wenqiang Tang (2017) DREAMSeq: an improved method for analyzing differential expression in RNA-seq data.

Examples

```
data(foxtail)
counts <- foxtail
groups <- gl(2, 3)
cnt <- DEGCountSet(counts, groups)
cnt <- DREAMSeq(cnt, model = "DP", parallel = TRUE, fc = 1.5)
res <- exResult(cnt)
```

comMeanVariance	<i>Compute the means and variances of samples</i>
-----------------	---

Description

The means and variances of samples are computed based on across-groups, pooling or pregroup.

Usage

```
comCommonMeanVariance( normCounts )
comPooledMeanVariance( normCounts, groups )
comGroupMeanVariance( normCounts, groups )
```

Arguments

normCounts	normalized read counts.
groups	treatment groups in a RNA-seq experiment.

Author(s)

Zhihua Gao, gzhheuet@sina.com

Examples

```
## Not run:
data(foxtail)
counts <- foxtail
groups <- gl(2, 3)
cnt <- DEGCountSet(counts, groups)
normCnts <- exCounts(cnt, normalized=TRUE)
mv1 <- comCommonMeanVariance(normCnts)
mv2 <- comPooledMeanVariance(normCnts, groups)
mv3 <- comGroupMeanVariance(normCnts, groups)

## End(Not run)
```

DEGCountSet-class	<i>DEGCountSet object and constructors</i>
-------------------	--

Description

The DEGCountSet is a container for read count data from RNA-seq experiments, and is the main class for the present package.

Usage

```
DEGCountSet( counts, groups, normFactor = NULL, phenoData = NULL, featureData = NULL )
```

Arguments

counts	the read count data of non-negative integers.
groups	treatment groups in a RNA-seq experiment.
normFactor	the normalization factors.
phenoData	a slot used to store variables describing sample phenotypes (i.e., columns in counts).
featureData	a slot used to store variables describing features (i.e., rows in counts).

Author(s)

Zhihua Gao, gzhheuet@sina.com

Examples

```
## Not run:
data(foxtail)
counts <- foxtail
groups <- gl(2,3)
cnt <- DEGCountSet(counts,groups)

## End(Not run)
```

DP

Double Poisson distribution

Description

Compute the probability density of double Poisson distribution with parameters (i.e. μ , θ) estimated from given RNA-seq data.

Usage

```
ddp(x, mu, theta, log = F, tol = 1e6)
```

Arguments

x	the read counts of non-negative integers.
mu	the average read counts.
theta	the dispersion of read count data in a RNA-seq experiment.
log	logical indicating whether or not to calculate logarithmic probability density of double Poisson distribution.
tol	the computational precision.

Author(s)

Zhihua Gao, gzhheuet@sina.com

dpWaldTest	<i>Wald test using double Poisson distribution</i>
------------	--

Description

Compute the significance levels of different genes from given RNA-seq data using Wald test based on double Poisson distribution.

Usage

```
dpWaldTest( cntsA, cntsB, muA, muB, thetaA, thetaB, normFactorA, normFactorB )
```

Arguments

cntsA	the read counts matrix of non-negative integers in group A.
cntsB	the read counts matrix of non-negative integers in group B.
muA	the average read counts of samples in group A.
muB	the average read counts of samples in group B.
thetaA	the dispersion of read count data in group A.
thetaB	the dispersion of read count data in group B.
normFactorA	the normalization factors in group A.
normFactorB	the normalization factors in group B.

Author(s)

Zhihua Gao, gzhheuet@sina.com

estCommonDispersion	<i>Estimate common dispersion</i>
---------------------	-----------------------------------

Description

Estimate common dispersion for read count data.

Usage

```
estCommonDispersion( object, estCommonDispMethod, fitMethod, model, maxLdisp, minLdisp, parallel
```

Arguments

object	a DEGCountSet.
estCommonDispMethod	an estimation method for common dispersion.
fitMethod	a method fitting the relationships between the means and the dispersions.
model	a model used to fit the read count data.
maxLdisp	a maximum value of logarithmic dispersion.
minLdisp	a minimum value of logarithmic dispersion.
parallel	logical indicating whether or not to use the parallel computing algorithm.

Author(s)

Zhihua Gao, gzhheuet@sina.com

estNormFactor	<i>Estimate normalization factors</i>
---------------	---------------------------------------

Description

Estimate normalization factors for read count data.

Usage

```
estNormFactor( object, estNormFactorMethod )
```

Arguments

`object` a DEGCountSet.
`estNormFactorMethod` an estimation method for normalization factors.

Author(s)

Zhihua Gao, gzhheuet@sina.com

Examples

```
## Not run:
data(foxtail)
counts <- foxtail
groups <- gl(2, 3)
cnt <- DEGCountSet(counts, groups)
cnt <- estNormFactor(cnt)
nf <- exNormFactor(cnt)

## End(Not run)
```

estShrinkDispersion	<i>Estimate shrink dispersion</i>
---------------------	-----------------------------------

Description

Estimate shrink dispersion for read count data.

Usage

```
estShrinkDispersion( object, model, step, parallel )
```

Arguments

object	a DEGCountSet.
model	a model used to fit the read count data.
step	the step length of using "window scan" approach to scan the average read counts.
parallel	logical indicating whether or not to use the parallel computing algorithm.

Author(s)

Zhihua Gao, gzhheuet@sina.com

exCounts	<i>Extract counts</i>
----------	-----------------------

Description

Extract counts for read count data.

Usage

```
## S4 method for signature 'DEGCountSet'
exCounts(object, normalized=FALSE)
## S4 replacement method for signature 'DEGCountSet,matrix'
exCounts(object) <- value
```

Arguments

object	a DEGCountSet.
normalized	logical indicating whether or not to divide the counts by the size factors before returning.
value	an integer matrix.

Author(s)

Zhihua Gao, gzhheuet@sina.com

exDispLNMean	<i>Extract a mean of logarithmic dispersions</i>
--------------	--

Description

Extract a mean of logarithmic dispersions for read count data.

Usage

```
## S4 method for signature 'DEGCountSet'
exDispLNMean(object)
## S4 replacement method for signature 'DEGCountSet,numeric'
exDispLNMean(object) <- value
```

Arguments

object a DEGCountSet.
 value a numeric.

Author(s)

Zhihua Gao, gzhheuet@sina.com

exDisplNSd

Extract a standard deviation of logarithmic dispersions

Description

Extract a standard deviation of logarithmic dispersions for read count data.

Usage

```
## S4 method for signature 'DEGCountSet'
exDisplNSd(object)
## S4 replacement method for signature 'DEGCountSet,numeric'
exDisplNSd(object) <- value
```

Arguments

object a DEGCountSet.
 value a numeric.

Author(s)

Zhihua Gao, gzhheuet@sina.com

exEstDispersionMethod *Extract an estimated method of dispersion*

Description

Extract an estimated method of dispersion for read count data.

Usage

```
## S4 method for signature 'DEGCountSet'
exEstDispersionMethod(object)
## S4 replacement method for signature 'DEGCountSet,character'
exEstDispersionMethod(object) <- value
```

Arguments

object a DEGCountSet.
 value a character that is "pooled" or "common".

Author(s)

Zhihua Gao, gzhheuet@sina.com

exEstNormFactorMethod *Extract an estimated method of normalization factors*

Description

Extract an estimated method of normalization factors for read count data.

Usage

```
## S4 method for signature 'DEGCountSet'
exEstNormFactorMethod(object)
## S4 replacement method for signature 'DEGCountSet,character'
exEstNormFactorMethod(object) <- value
```

Arguments

object	a DEGCountSet.
value	a character that is "TMM", "TC", "UQ", "M", "Q" or "RLE". TMM, trimmed mean of M-values; TC, total counts; UQ, upper quartile; M, median; Q, quantile; RLE, relative logarithmic expression (DESeq normalization).

Author(s)

Zhihua Gao, gzhheuet@sina.com

exFitMethod *Extract a fitted method*

Description

Extract a method to fit the relationships between the means and the dispersions.

Usage

```
## S4 method for signature 'DEGCountSet'
exFitMethod(object)
## S4 replacement method for signature 'DEGCountSet,character'
exFitMethod(object) <- value
```

Arguments

object	a DEGCountSet.
value	a character that is "weightedLowess", "locfit", "spline" or "loess".

Author(s)

Zhihua Gao, gzhheuet@sina.com

exModel	<i>Extract a model</i>
---------	------------------------

Description

Extract a model to fit read count data.

Usage

```
## S4 method for signature 'DEGCountSet'
exModel(object)
## S4 replacement method for signature 'DEGCountSet,character'
exModel(object) <- value
```

Arguments

object	a DEGCountSet.
value	a character that is "DP", "NB" or "Mix". DP, double Poisson model; NB, negative binomial model; Mix, the mixture of DP and NB models.

Author(s)

Zhihua Gao, gzhheuet@sina.com

exNormFactor	<i>Extract normalization factors</i>
--------------	--------------------------------------

Description

Extract normalization factors for read count data.

Usage

```
## S4 method for signature 'DEGCountSet'
exNormFactor(object)
## S4 replacement method for signature 'DEGCountSet,vector'
exNormFactor(object) <- value
```

Arguments

object	a DEGCountSet.
value	a numeric vector.

Author(s)

Zhihua Gao, gzhheuet@sina.com

exResult	<i>Extract a differential expression analysis result</i>
----------	--

Description

Extract a differential expression analysis result for read count data.

Usage

```
## S4 method for signature 'DEGCountSet'
exResult(object)
## S4 replacement method for signature 'DEGCountSet,data.frame'
exResult(object) <- value
```

Arguments

object	a DEGCountSet.
value	a data frame.

Author(s)

Zhihua Gao, gzhheuet@sina.com

foxtail	<i>Foxtail dataset</i>
---------	------------------------

Description

A Foxtail dataset includes three biological replicates, in which roots from 1-week-old Foxtail millet seedlings (Yugu-1 cultivar) were treated with or without 0.001mM epi-Brassinolide (eBL).

Usage

```
data(foxtail)
```

Author(s)

Zhihua Gao, gzhheuet@sina.com

headWaldTest	<i>Head of Wald test</i>
--------------	--------------------------

Description

Provide all the parameters for the following WaldTest.

Usage

```
headWaldTest( object, model, parallel )
```

Arguments

object	a DEGCountSet.
model	a model used to fit the read count data.
parallel	logical indicating whether or not to use the parallel computing algorithm.

Author(s)

Zhihua Gao, gzhheuet@sina.com

main	<i>The main function of DREAMSeq</i>
------	--------------------------------------

Description

This is a main function of DREAMSeq.

Usage

```
DREAMSeq( object,
  estNormFactorMethod = c( "TMM", "TC", "UQ", "M", "Q", "RLE" ),
  estCommonDispMethod = c( "pooled", "common" ),
  fitMethod = c( "weightedLowess", "locfit", "spline", "loess" ),
  model = c( "DP", "NB", "Mix" ),
  maxLdisp = 5,
  minLdisp = -18.42068,
  step = 1,
  dispLNMean = NA_real_,
  dispLNSd = NA_real_,
  parallel = FALSE,
  fc = NULL,
  quiet = FALSE
)
```

Arguments

object	a DEGCountSet.
estNormFactorMethod	an estimation method for normalization factors.
estCommonDispMethod	an estimation method for common dispersion.
fitMethod	a method fitting the relationships between the means and dispersions.
model	a model used to fit the read count data.
maxLdisp	a maximum value of logarithmic dispersion.
minLdisp	a minimum value of logarithmic dispersion.
step	the step length of using "Window scan" approach to scan the average read counts.
dispLNMean	the mean of the log-normal prior fitted dispersion.
dispLNSd	the standard deviation of the log-normal prior fitted dispersion.
parallel	logical indicating whether or not to use the parallel computing algorithm.
fc	a fold change based on the ratio of the average read counts between group B and A.
quiet	logical indicating whether or not to print messages at each step.

Author(s)

Zhihua Gao, gzhheuet@sina.com

Examples

```
## Not run:
data(foxtail)
counts <- foxtail
groups <- gl(2, 3)
cnt <- DEGCountSet(counts, groups)
cnt <- DREAMSeq(cnt, model = "DP", parallel = TRUE, fc = 1.5)
res <- exResult(cnt)

## End(Not run)
```

modelPipeline

Differential expression analysis pipeline for read count data

Description

This is a pipeline to identify differentially expressed genes for read count data using double Poisson model, negative binomial model or the mixture of them.

Usage

```
modelPipeline( object,
  estNormFactorMethod = estNormFactorMethod,
  estCommonDispMethod = estCommonDispMethod,
  fitMethod = fitMethod,
  model = model,
  maxLdisp = maxLdisp,
  minLdisp = minLdisp,
  step = step,
  dispLNMean = dispLNMean,
  dispLNSd = dispLNSd,
  parallel = parallel,
  quiet = quiet
)
```

Arguments

object	a DEGCountSet.
estNormFactorMethod	an estimation method for normalization factors.
estCommonDispMethod	an estimation method for common dispersion.
fitMethod	a method fitting the relationships between the means and dispersions.
model	a model used to fit the read count data.
maxLdisp	a maximum value of logarithmic dispersion.
minLdisp	a minimum value of logarithmic dispersion.
step	the step length of using "Window scan" approach to scan the average read counts.
dispLNMean	the mean of the log-normal prior fitted dispersion.
dispLNSd	the standard deviation of the log-normal prior fitted dispersion.
parallel	logical indicating whether or not to use the parallel computing algorithm.
quiet	logical indicating whether or not to print messages at each step.

Author(s)

Zhihua Gao, gzhheuet@sina.com

nbWaldTest

Wald test using negative binomial distribution

Description

Compute the significance levels of different genes from given RNA-seq data using Wald test based on negative binomial distribution.

Usage

```
nbWaldTest( cntsA, cntsB, muA, muB, dspA, dspB, normFactorA, normFactorB )
```

Arguments

cntsA	the read counts matrix of non-negative integers in group A.
cntsB	the read counts matrix of non-negative integers in group B.
muA	the average read counts of samples in group A.
muB	the average read counts of samples in group B.
dspA	the dispersion of read count data in group A.
dspB	the dispersion of read count data in group B.
normFactorA	the normalization factors in group A.
normFactorB	the normalization factors in group B.

Author(s)

Zhihua Gao, gzhheuet@sina.com

scanMeans

Scan means by a window

Description

Scan means by shifting a window of setted counts range.

Usage

```
scanMeans( mean, step )
```

Arguments

mean	the average read counts.
step	the step length of using "window scan" approach to scan the average read counts.

Author(s)

Zhihua Gao, gzhheuet@sina.com

Index

comCommonMeanVariance
 (comMeanVariance), 3
comGroupMeanVariance (comMeanVariance),
 3
comMeanVariance, 3
comPooledMeanVariance
 (comMeanVariance), 3

ddp (DP), 4
DEGCountSet (DEGCountSet-class), 3
DEGCountSet-class, 3
DP, 4
dpWaldTest, 5
DREAMSeq (main), 12
DREAMSeq-package, 2

estCommonDispersion, 5
estCommonDispersion, DEGCountSet-method
 (estCommonDispersion), 5
estNormFactor, 6
estNormFactor, DEGCountSet-method
 (estNormFactor), 6
estShrinkDispersion, 6
estShrinkDispersion, DEGCountSet-method
 (estShrinkDispersion), 6
exCounts, 7
exCounts, DEGCountSet-method (exCounts),
 7
exCounts<- (exCounts), 7
exCounts<-, DEGCountSet, matrix-method
 (exCounts), 7
exDispLNMean, 7
exDispLNMean, DEGCountSet-method
 (exDispLNMean), 7
exDispLNMean<- (exDispLNMean), 7
exDispLNMean<-, DEGCountSet, numeric-method
 (exDispLNMean), 7
exDispLNSd, 8
exDispLNSd, DEGCountSet-method
 (exDispLNSd), 8
exDispLNSd<- (exDispLNSd), 8
exDispLNSd<-, DEGCountSet, numeric-method
 (exDispLNSd), 8
exEstDispersionMethod, 8
exEstDispersionMethod, DEGCountSet-method
 (exEstDispersionMethod), 8
exEstDispersionMethod<-
 (exEstDispersionMethod), 8
exEstDispersionMethod<-, DEGCountSet, character-method
 (exEstDispersionMethod), 8
exEstNormFactorMethod, 9
exEstNormFactorMethod, DEGCountSet-method
 (exEstNormFactorMethod), 9
exEstNormFactorMethod<-
 (exEstNormFactorMethod), 9
exEstNormFactorMethod<-, DEGCountSet, character-method
 (exEstNormFactorMethod), 9
exFitMethod, 9
exFitMethod, DEGCountSet-method
 (exFitMethod), 9
exFitMethod<- (exFitMethod), 9
exFitMethod<-, DEGCountSet, character-method
 (exFitMethod), 9
exModel, 10
exModel, DEGCountSet-method (exModel), 10
exModel<- (exModel), 10
exModel<-, DEGCountSet, character-method
 (exModel), 10
exNormFactor, 10
exNormFactor, DEGCountSet-method
 (exNormFactor), 10
exNormFactor<- (exNormFactor), 10
exNormFactor<-, DEGCountSet, vector-method
 (exNormFactor), 10
exResult, 11
exResult, DEGCountSet-method (exResult),
 11
exResult<- (exResult), 11
exResult<-, DEGCountSet, data.frame-method
 (exResult), 11

foxtail, 11

headWaldTest, 12
headWaldTest, DEGCountSet-method
 (headWaldTest), 12

main, 12

modelPipeline, [13](#)

nbWaldTest, [14](#)

scanMeans, [15](#)