Package 'RoMA'

December 5, 2017

Type Package

Title RNA-seq analyses of Molecular Abundance (RoMA)

Version 1.0

Date 2019-09-17

Author Guoshuai Cai

Maintainer Guoshuai Cai <GCAI@mailbox.sc.edu>

Description An efficient method to detect RNA-seq differential expression, by modeling mRNA abundance with precision weights from raw counts, provides an accurate quantification of mRNA abundance and a data adjustment-tolerant analysis and therefore integrative with various well-established methods of mRNA abundance.

License GPL (>=2)

Depends edgeR, limma

R topics documented:

RoMA-package 1
calcLibSizes 2
calcNormRPKMs 3
DEGList 3
length.gene.hg19 4
length.isoform.hg19 4
plotMA 5
roma 5
x 6

Index **7**

`RoMA-package` *RNA-seq analyses of Molecular Abundance (RoMA)*

Description

An efficient method to detect RNA-seq differential expression, by modeling mRNA abundance with precision weights from raw counts, provides an accurate quantification of mRNA abundance and a data adjustment-tolerant analysis and therefore integrative with various well-established methods of mRNA abundance.

Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.

Author(s)

Guoshuai Cai

Maintainer: Guoshuai Cai <GCAI@mailbox.sc.edu>

Examples

```
data(x)

e<-DEGList(counts=x)

e <- calcNormFactors(e,method="TMM")

e<-calcLibSizes(e)
e<-calcNormRPKMs(e)

des<-factor(c(rep("BR",7),rep("UHR",7)))

des<-as.matrix(cbind(1,as.numeric(des)))

fit<-roma(e, design = des, lib.size = NULL, normalize.method = "none",
span = 0.5, plot = TRUE)

plotMA(fit,coef=2,h=1,n=3000,sort="p",p=0.05)

res<-topTable(fit,coef=2,n=Inf,sort = "p", p = 0.05)

res<-topTable(fit,coef=2,n=Inf)
```

---

calcLibSizes *Calculation of Library Sizes*

Description

The effective library size is estimated by norm factor * total counts, where norm factor can be estimated by TMM.

Usage

```
calcLibSizes(e)
```

Arguments

e	The DGEList object with estimated normalization factors.

Value

The DGEList object with estimated effective library sizes in e$lib.size.

`calcNormRPKMs` *Calculation of Normalized RPKMs*

Description

Although RPKM/FPKM normalizes the RNA-seq data, it does not account for composition bias. Composition bias can be accounted for using TMM normalization by equaling the median of log-ratios of read counts between samples. RoMA use this function to normalize RPKMs by the TMM method.

Usage

```
calcNormRPKMs(e)
```

Arguments

`e`	The DGEList object with estimated library sizes and transcript lengths.

Value

The DGEList object with calculated RPKMs in e$rpkms.

`DEGList` *DEGList Maker*

Description

Creates a DGEList object with various information including library sizes, gene length and others.

Usage

```
DEGList(counts = matrix(0, 0, 0), lib.size = colSums(counts, na.rm = TRUE),
norm.factors = rep(1, ncol(counts)), group = rep(1, ncol(counts)),
length = NULL, length.gene="length.gene.hg19",
length.isoform="length.isoform.hg19", rpkms = NULL, genes = NULL,
remove.zeros = FALSE, level = "gene")
```

Arguments

`counts`	numeric matrix of read counts.
`lib.size`	numeric vector giving sequence depth for each library, which is the total count by default.
`norm.factors`	numeric vector of normalization factors that modify the library sizes, which is 1 by default.
`group`	vector or factor giving the experimental group/condition for each sample/library.

`length`	numeric matrix of gene length, which could be gene expression effective length estimated by RSEM.
`length.gene`	data frame containing annotation information for each gene, which is from the hg19 human assembly UCSC annotation by default.
`length.isoform`	frame containing annotation of gene length for each gene isoform, which is from the hg19 human assembly UCSC annotation by default.
`rpkms`	numeric matrix of rpkms, which is NULL by default.
`genes`	data frame containing annotation information for each gene.
`remove.zeros`	logical, whether to remove rows that have 0 total count.
`level`	indicate the data is on gene level or isoform level.

Value

DGEList object

`length.gene.hg19` *Collapsed Gene Length from UCSC HG19 annotation*

Description

Collapsed Gene Length from UCSC HG19 annotation

Usage

```
data("length.gene.hg19")
```

Format

The format is: int [1:47352, 1] 230 320 1774 2134 9544 2907 4689 2190 5400 454 ... - attr(*, "dimnames")=List of 2 ..$: chr [1:47352] "5S_rRNA" "7SK" "A1BG" "A1BG-AS1" ... ..$: chr "Length"

`length.isoform.hg19`
Transcript Length from UCSC HG19 annotation

Description

Transcript Length from UCSC HG19 annotation

Usage

```
data("length.isoform.hg19")
```

Format

The format is: int [1:82960, 1] 1652 1746 2412 1769 1577 1130 918 32 62 4370 ... - attr(*, "dimnames")=List of 2 ..$: chr [1:82960] "uc001aaa.3" "uc001aac.4" "uc001aae.4" "uc001aah.4" ... ..$: chr "Length"

plotMA *MA-plot Maker*

Description

Visualize the difference between two measurements by log ratio and mean average of data.

Usage

```
plotMA(fit, coef = 2, h = 1, n = 3000, sort = "p", p = 0.05, colors = NULL,
legend = NULL)
```

Arguments

fit	the fit object from RoMA.
coef	column number or column name of coefficient matrix which is of interest.
h	threshold line on log ratio
n	maximum number of significant genes will be shown
sort	values which will be ranked to select signficiant genes.
p	shreshold value
colors	colors will be display
legend	legend will be display

roma *The Main Function of RoMA*

Description

RoMA incorporates information from both mRNA abundance and raw counts by modeling RPKM (reads per kilobase per million), which represents the relative abundance of mRNA transcripts, and borrowing mean-variance dependency from CPM (counts per million) as a precision weight accounting for the variability in sequencing depth. LOWESS curve is fitted to capture the CPM mean-variance dependency and the inverse variance is taken into the modeling as the precision weight. Moderate-t statistics is performed by the eBayes function in limma.

Usage

```
roma(counts, rpkms = NULL, design = NULL, lib.size = NULL,
normalize.method = "none", span = 0.5, plot = FALSE, save.plot = FALSE, ...)
```

Arguments

counts	numeric matrix containing raw counts or a DGEList object.
rpkms	numeric vector containing RPKMs for each sample. If NULL and counts is a DGEList then it can be obtained from counts$rpkms.
design	design matrix with rows corresponding to samples and columns to coefficients of study condition and covariates to be estimated.
lib.size	numeric vector containing total library sizes for each sample. If NULL and counts is a DGEList then library sizes are calculated from the columnwise counts totals.
normalize.method	normalization method to be applied to the logCPM values. Choices are as for the method argument of normalizeBetweenArrays of limma. As normalization is performed ahead in RoMA standard analysis, this argument is suggested to set as "none".
span	width of the lowess smoothing window as a proportion.
plot	logical, should a plot of the mean-variance trend be displayed?
save.plot	logical, should the coordinates and line of the plot be saved in the output?
...	

---

x *Counts Matrix*

Description

Counts Matrix

Usage

```
data("x")
```

Format

A data frame with 15998 observations on the following 14 variables.

`X127_SRR037452` a numeric vector
`X127_SRR037453` a numeric vector
`X127_SRR037454` a numeric vector
`X127_SRR037455` a numeric vector
`X127_SRR037456` a numeric vector
`X127_SRR037457` a numeric vector
`X127_SRR037458` a numeric vector
`X128_SRR037459` a numeric vector
`X128_SRR037460` a numeric vector
`X128_SRR037461` a numeric vector
`X128_SRR037462` a numeric vector
`X128_SRR037463` a numeric vector
`X128_SRR037464` a numeric vector
`X128_SRR037465` a numeric vector

Index

∗Topic **datasets**
length.gene.hg19, 4
length.isoform.hg19, 4
x, 6
∗Topic **package**
RoMA-package, 1

calcLibSizes, 2
calcNormRPKMs, 3

DEGList, 3

length.gene.hg19, 4
length.isoform.hg19, 4

plotMA, 5

RoMA *(RoMA-package)*, 1
roma, 5
RoMA-package, 1

x, 6