Package ‘erccdashboard’

March 28, 2017

Type Package

Title Assess Differential Gene Expression Experiments with ERCC Controls

Version 1.8.0

Author Sarah Munro, Steve Lund

Maintainer Sarah Munro <sarah.munro@nist.gov>

Description Technical performance metrics for differential gene expression experiments using External RNA Controls Consortium (ERCC) spike-in ratio mixtures.

URL http://www.nist.gov/mml/bbd/erccdashboard.cfm,
https://github.com/usnistgov/erccdashboard,
http://tinyurl.com/erccsrm

BugReports https://github.com/usnistgov/erccdashboard/issues

Depends R (>= 3.2), ggplot2 (>= 2.1.0), gridExtra (>= 2.0.0)

Imports edgeR, gplots, grid, gtools, limma, locfit, MASS, plyr, QuasiSeq, qvalue, reshape2, ROCR, scales, stringr

License GPL (>=2)


biocViews GeneExpression, Transcription, AlternativeSplicing, DifferentialExpression, DifferentialSplicing, Genetics, Microarray, mRNAMicroarray, RNASeq, BatchEffect, MultipleComparison, QualityControl

LazyData yes

Roxygen list(wrap = FALSE)

RoxygenNote 5.0.1

NeedsCompilation no
R topics documented:

annotLODR ........................................ 2
dynRangePlot ..................................... 3
ERCC ................................................. 4
ERCCDef ............................................ 4
ERCCMix1and2 ...................................... 5
erccROC ............................................. 5
estLODR ............................................. 6
est_r_m ............................................. 7
geneExprTest ....................................... 7
initDat ............................................. 8
maSignal ........................................... 9
MET.CTL.countDat ................................ 10
MET.CTL.totalReads ................................ 11
runDashboard ....................................... 11
saveERCCPlots .................................... 12
SEQC.Example ..................................... 13
UHRR.HBRR.arrayDat ............................... 14
UHRR.HBRR.countDat ............................... 14
UHRR.HBRR.totalReads ............................. 15

Index 16

annotLODR  Annotate signal-abundance and ratio-abundance plots with LODR

Description
Annotate signal-abundance and ratio-abundance plots with LODR

Usage
annotLODR(exDat)

Arguments
exDat  list, contains input data and stores analysis results

Examples
data(SEQC.Example)

exDat <- initDat(datType="array", isNorm=FALSE,
exTable=UHRR.HBRR.arrayDat,
filenameRoot="testRun", sample1Name="UHRR",
sample2Name="HBRR", erccmix="RatioPair",
erccdilution = 1, spikeVol = 50,
totalRNAmass = 2.5*10^(3), choseFDR=0.01)

exDat <- est_r_m(exDat)

exDat <- dynRangePlot(exDat)
dynRangePlot

produce signal-abundance plot to evaluate dynamic range

dynRangePlot(exDat, allPoints, labelReps)

Description
Produce signal-abundance plot to evaluate dynamic range

Usage

dynRangePlot(exDat, allPoints, labelReps)

Arguments

exDat list, contains input data and stores analysis results
allPoints boolean, default is false, means of replicates will be plotted. If true then all replicates will be plotted as individual points.
labelReps boolean, default is false. If true then replicates will be labeled.

Examples

data(SEQC.Example)
exDat <- initDat(datType="count", isNorm=FALSE, exTable=MET.CTL.countDat, filenameRoot="testRun", sample1Name="MET", sample2Name="CTL", erccmix="RatioPair", erccdilution=1/100, spikeVol=1, totalRNAmass=0.500, choseFDR=0.1)
exDat <- est_r_m(exDat)
exDat <- dynRangePlot(exDat, allPoints="FALSE", labelReps ="FALSE")
exDat$Figures$dynRangePlot
ERCCDef

**Description**

Contains 2 data frames: ERCCDef and ERCCMix1and2

**Usage**

```r
data(ERCC)
```

**Examples**

```r
data(ERCC)
```

---

**ERCCDef**

**ERCCDef dataframe**

**Description**

ERCC transcript lengths and GC content

**Format**

A data frame with 96 observations on the following 3 variables.

- **Feature** a factor vector
- **Length** a numeric vector
- **GC** a numeric vector

**Details**

Length and GC content of all 96 ERCC controls in NIST SRM 2374

**Source**

http://tinyurl.com/erccsrm
**ERCCMix1and2 dataframe**

**Description**

Ambion RatioPair ERCC Mixtures

**Format**

A data frame with 96 observations on the following 4 variables.

- **ERCC.AMB.Expected**: a factor vector of all 96 ERCC control IDs
- **Subpool**: a factor vector of the ERCC Ratios in each Subpool with levels 4:1 1:1 1:1.5 1:2
- **Mix1Conc.AttoMoles_ul**: a numeric vector of the ERCC concentrations in Mix 1
- **Mix2Conc.AttoMoles_ul**: a numeric vector of the ERCC concentrations in Mix 2

**Source**

http://www.lifetechnologies.com/order/catalog/product/4456739

---

**erccROC**

*Produce Receiver Operator Characteristic (ROC) Curves and AUC statistics*

**Description**

Produce Receiver Operator Characteristic (ROC) Curves and AUC statistics

**Usage**

`erccROC(exDat)`

**Arguments**

- **exDat**: list, contains input data and stores analysis results

**Examples**

```r
data(SEQC.Example)
exDat <- initDat(datType="array", isNorm=FALSE,
exTable=UHRR.HBRR.arrayDat,
filenameRoot="testRun", sample1Name="UHRR",
sample2Name="HBRR", erccmix="RatioPair",
erccdilution = 1, spikeVol = 50,
totalRNAmass = 2.5*10^3, choseFDR=0.01)
exDat <- est_r_m(exDat)
```
estLODR <- dynRangePlot(exDat)

exDat <- geneExprTest(exDat)

exDat <- erccROC(exDat)

exDat$Figures$rocPlot

---

estLODR

Estimate Limit of Detection of Ratios (LODR)

Description

Estimate Limit of Detection of Ratios (LODR)

Usage

estLODR(exDat, kind = "ERCC", prob = 0.9)

Arguments

exDat: list, contains input data and stores analysis results
kind: "ERCC" or "Sim"
prob: probability, ranging from 0 - 1, default is 0.9

Details

This is the function to estimate a limit of detection of ratios (LODR) for a chosen probability and threshold p-value for the fold changes in the ERCC control ratio mixtures.

Examples

data(SEQC.Example)

exDat <- initDat(datType="array", isNorm=FALSE,
  exTable=UHRR.HBRR.arrayDat,
  filenameRoot="testRun", sample1Name="UHRR",
  sample2Name="HBRR", erccmix="RatioPair",
  erccdilution = 1, spikeVol = 50,
  totalRNAmass = 2.5*10^3, choseFDR=0.01)

exDat <- est_r_m(exDat)

exDat <- dynRangePlot(exDat)

exDat <- geneExprTest(exDat)

exDat <- estLODR(exDat, kind = "ERCC", prob = 0.9)

exDat$Figures$lodrERCCPlot
Estimate the mRNA fraction differences for the pair of samples using replicate data

Description

Estimate the mRNA fraction differences for the pair of samples using replicate data

Usage

```r
est_r_m(exDat)
```

Arguments

- `exDat` list, contains input data and stores analysis results

Details

This is the first function to run after an exDat structure is initialized using initDat, because it is needed for all additional analysis. An \( r_m \) of 1 indicates that the two sample types under comparison have similar mRNA fractions of total RNA. The \( r_m \) estimate is used to adjusted the expected ERCC mixture ratios in this analysis and may indicate a need for a different sample normalization approach.

Examples

```r
data(SEQC.Example)

exDat <- initDat(datType="count", isNorm = FALSE, exTable=MET.CTL.countDat,
filenameRoot = "testRun", sample1Name = "MET",
sample2Name = "CTL", erccmix = "RatioPair",
erccdilution = 1/100, spikeVol = 1, totalRNAmass = 0.500,
choseFDR = 0.1)

exDat <- est_r_m(exDat)
```

geneExprTest Prepare differential expression testing results for spike-in analysis

Description

Prepare differential expression testing results for spike-in analysis

Usage

```r
geneExprTest(exDat)
```

Arguments

- `exDat` list, contains input data and stores analysis results
Details

This function wraps the QuasiSeq differential expression testing package for `dataType = "count"` or uses the limma package for differential expression testing if `dataType = "array"`. Alternatively, for count data only, if correctly formatted DE test results are provided, then `geneExprTest` will bypass DE testing (with reduced runtime).

Examples

data(SEQC.Example)

exDat <- initDat(datType="array", isNorm=FALSE, 
exTable=UHRR.HBRR.arrayDat, 
filenameRoot="testRun", sample1Name="UHRR", 
sample2Name="HBRR", erccmix="RatioPair", 
erccdilution = 1, spikeVol = 50, 
totalRNAmass = 2.5*10^3, choseFDR=0.01)

exDat <- est_r_m(exDat)

exDat <- dynRangePlot(exDat)

exDat <- geneExprTest(exDat)

initDat

**Initialize the exDat list**

Description

Initialize the exDat list

Usage

```r
initDat(datType = NULL, isNorm = FALSE, exTable = NULL, 
repNormFactor = NULL, filenameRoot = NULL, sample1Name = NULL, 
sample2Name = NULL, erccmix = "RatioPair", erccdilution = 1, 
isNorm = FALSE, if FALSE then the unnormalized input data will be normalized 
exTable = NULL, the first column contains names of genes or transcripts (Feature) 
and the remaining columns are counts for sample replicates spiked with ERCC controls
```

Arguments

datType
- type is "count" or "array", unnormalized data is expected (normalized data may be accepted in future version of the package). Default is "count" (integer count data), "array" is unnormalized fluorescent intensities from microarray fluorescent intensities (not log transformed or normalized)

isNorm
- default is FALSE, if FALSE then the unnormalized input data will be normalized in erccdashboard analysis. If TRUE then it is expected that the data is already normalized

exTable
- data frame, the first column contains names of genes or transcripts (Feature) and the remaining columns are counts for sample replicates spiked with ERCC controls
maSignal

repNormFactor  optional vector of normalization factors for each replicate, default value is NULL and 75th percentile normalization will be applied to replicates
filenameRoot  string root name for output files
sample1Name  string name for sample 1 in the gene expression experiment
sample2Name  string name for sample 2 in the gene expression experiment
erccmix  Name of ERCC mixture design, "RatioPair" is default, the other option is "Single"
erccdilution  unitless dilution factor used in dilution of the Ambion ERCC spike-in mixture solutions
spikeVol  volume in microliters of diluted ERCC mix spiked into the total RNA samples
totalRNAmass  mass in micrograms of total RNA spiked with diluted ERCC mixtures
choseFDR  False Discovery Rate for differential expression testing, default is 0.05
ratioLim  Limits for ratio axis on MA plot, default is c(-4,4)
signalLim  Limits for signal axis on dynamic range plot, default is c(-14,14)
userMixFile  optional filename input, default is NULL, if ERCC control ratio mixtures other than the Ambion product were used then a userMixFile can be used for the analysis

Examples

data(SEQC.Example)
exDat <- initDat(datType="count", isNorm = FALSE, exTable=MET.CTL.countDat,
filenameRoot = "testRun", sample1Name = "MET",
sample2Name = "CTL", erccmix = "RatioPair",
erccdilution = 1/100, spikeVol = 1, totalRNAmass = 0.500,
choseFDR = 0.1)
summary(exDat)

maSignal  Generate MA plots with or without annotation using LODR estimates

Description

Generate MA plots with or without annotation using LODR estimates

Usage

maSignal(exDat, alphaPoint = 0.8, r_mAdjust = TRUE, replicate = TRUE)

Arguments

exDat  list, contains input data and stores analysis results
alphaPoint  numeric value, for alpha (transparency) for plotted points, range is 0 - 1
r_mAdjust  default is TRUE, if FALSE then the r_m estimate will not used to offset dashed lines for empirical ratios on figure
replicate  default is TRUE, if FALSE then error bars will not be produced
Examples

data(SEQC.Example)

exDat <- initDat(datType="array", isNorm=FALSE,
               exTable=UHRR.HBRR.arrayDat,
               filenameRoot="testRun", sample1Name="UHRR",
               sample2Name="HBRR", erccmix="RatioPair",
               erccdilution = 1, spikeVol = 50,
               totalRNAmass = 2.5*10^3, choseFDR=0.01)

exDat <- est_r_m(exDat)

exDat <- dynRangePlot(exDat)

exDat <- geneExprTest(exDat)
# generate MA plot without LODR annotation
exDat <- maSignal(exDat)

exDat$Figures$maPlot

exDat <- estLODR(exDat, kind = "ERCC", prob = 0.9)
# Include LODR annotation
exDat <- annotLODR(exDat)

exDat$Figures$maPlot

MET.CTL.countDat

Rat toxicogenomics count data

Description

RNA-Seq count data from Methimazole and Control rat biological replicates

Format

A data frame with 16590 observations of the following 7 variables.

Feature a factor vector of all Endogenous and ERCC transcripts in the experiment
MET_1 a numeric vector of counts from Methimazole treatment biological replicate 1
MET_2 a numeric vector of counts from Methimazole treatment biological replicate 2
MET_3 a numeric vector of counts from Methimazole treatment biological replicate 3
CTL_1 a numeric vector of counts from Control biological replicate 1
CTL_2 a numeric vector of counts from Control biological replicate 2
CTL_3 a numeric vector of counts from Control biological replicate 3
MET.CTL.totalReads  Rat toxicogenomics total read data

Description
Total reads per biological replicate from FASTQ files

Format
The format is: int [1:6] 41423502 46016148 44320280 38400362 47511484 33910098

runDashboard  Run default erccdashboard analysis of ERCC control ratio mixtures

Description
Run default erccdashboard analysis of ERCC control ratio mixtures

Usage
runDashboard(datType = NULL, isNorm = FALSE, exTable = NULL, 
repNormFactor = NULL, filenameRoot = NULL, sample1Name = NULL, 
sample2Name = NULL, erccmix = "RatioPair", erccdilution = 1, 
spikeVol = 1, totalRNAmass = 1, choseFDR = 0.05, ratioLim = c(-4, 4), 
signalLim = c(-14, 14), userMixFile = NULL)

Arguments

datType  type is "count" (RNA-Seq) or "array" (microarray), "count" is unnormalized integer count data (normalized RNA-Seq data will be accepted in an updated version of the package), "array" can be normalized or unnormalized fluorescent intensities from a microarray experiment.

isNorm  default is FALSE, if FALSE then the unnormalized input data will be normalized in erccdashboard analysis. If TRUE then it is expected that the data is already normalized

exTable  data frame, the first column contains names of genes or transcripts (Feature) and the remaining columns are expression measures for sample replicates spiked with ERCC controls

repNormFactor  optional vector of normalization factors for each replicate, default value is NULL and 75th percentile normalization will be applied to replicates

filenameRoot  string root name for output files

sample1Name  string name for sample 1 in the gene expression experiment

sample2Name  string name for sample 2 in the gene expression experiment

erccmix  Name of ERCC mixture design, "RatioPair" is default, the other option is "Single"

erccdilution  unitless dilution factor used in dilution of the Ambion ERCC spike-in mixture solutions
saveERCCPlots

spikeVol volume in microliters of diluted ERCC mix spiked into the total RNA samples
totalRNAmass mass in micrograms of total RNA spiked with diluted ERCC mixtures
choseFDR False Discovery Rate for differential expression testing
ratioLim Limits for ratio axis on MA plot, default is c(-4,4)
signalLim Limits for ratio axis on MA plot, default is c(-14,14)
userMixFile optional filename input, default is NULL, if ERCC control ratio mixtures other than the Ambion product were used then a userMixFile can be used for the analysis

Examples

data(SEQC.Example)
exDat <- runDashboard(datType = "count", isNorm = FALSE,
exTable = MET.CTL.countDat,
filenameRoot = "COH.ILM",
sample1Name = "MET", sample2Name = "CTL",
erncmix = "RatioPair", erncdilution = 1/100,
spikeVol = 1, totalRNAmass = 0.500, choseFDR = 0.1)

summary(exDat)

Description

The function savePlots will save selected figures to a pdf file. The default is the 4 manuscript figures to a single page (plotsPerPg = "manuscript"). If plotsPerPg = "single" then each plot is placed on an individual page. If plotlist is not defined (plotlist = NULL) or if plotlist = exDat$Figures then all plots in exDat$Figures are printed to a PDF file.

Usage

saveERCCPlots(exDat, plotsPerPg = "main", savas = "pdf", outName, plotlist, res)

Arguments

exDat list, contains input data and stores analysis results
plotsPerPg string, if "main" then the 4 main plots are printed to one page, if "single" then a single plot is printed per page from the plotlist argument
savas Choose file format from "pdf", "jpeg" or "png"
outName Choose output file name, default will be fileName from exDat
plotlist list, contains plots to print
res Choose the file resolution
Examples

```r
data(SEQC.Example)
exDat <- initDat(datType="count", isNorm=FALSE, exTable=MET.CTL.countDat,
filenameRoot="testRun", sample1Name="MET",
sample2Name="CTL", erccmix="RatioPair",
erccdilution=1/100, spikeVol=1, totalRNAmass=0.500,
choreFDR=0.1)
exDat <- est_r_m(exDat)
exDat <- dynRangePlot(exDat)
exDat <- geneExprTest(exDat)
exDat <- erccROC(exDat)
exDat <- estLODR(exDat, kind="ERCC", prob=0.9)
exDat <- annotLODR(exDat)

#to print 4 main plots to a single page pdf file
saveERCCPlots(exDat, plotsPerPg = "manuscript", saveas = "pdf")

#to print 4 plots to a jpeg file
saveERCCPlots(exDat, plotsPerPg = "manuscript", saveas = "jpeg")

# or to create a multiple page pdf of all plots produced
saveERCCPlots(exDat, plotsPerPg = "single", plotlist = exDat$Figures)

# or to create a multiple page pdf of just 2 plots
saveERCCPlots(exDat, plotsPerPg = "single",
plotlist = list(exDat$Figures$lodrPlot, exDat$Figures$maPlot))
```

Description

Contains the following 5 items:
- MET.CTL.countDat - Rat toxicogenomics count data
- MET.CTL.totalReads - Rat toxicogenomics total read data
- UHRR.HBRR.arrayDat - UHRR and HBRR Illumina BeadArray data
- UHRR.HBRR.countDat - UHRR and HBRR RNA-Seq Illumina count data
- UHRR.HBRR.totalReads - UHRR and HBRR sample total read data

Usage

```r
data(SEQC.Example)
```
Examples

data(SEQC.Example)

**UHRR.HBRR.arrayDat**  
*UHRR and HBRR Illumina BeadArray data*

**Description**

Unnormalized microarray data from Lab 13 of reference sample interlaboratory study

**Format**

A data frame with 17627 observations of the following 7 variables.

- **Feature**  
  a factor vector of all Endogenous and ERCC transcripts in the experiment
- **UHRR_3**  
  a numeric vector of fluorescence intensities from UHRR microarray technical replicate 1
- **UHRR_2**  
  a numeric vector of fluorescence intensities from UHRR microarray technical replicate 2
- **UHRR_1**  
  a numeric vector of fluorescence intensities from UHRR microarray technical replicate 3
- **HBRR_3**  
  a numeric vector of fluorescence intensities from HBRR microarray technical replicate 1
- **HBRR_2**  
  a numeric vector of fluorescence intensities from HBRR microarray technical replicate 2
- **HBRR_1**  
  a numeric vector of fluorescence intensities from HBRR microarray technical replicate 3

**UHRR.HBRR.countDat**  
*UHRR and HBRR RNA-Seq Illumina count data*

**Description**

RNA-Seq count data from UHRR and HBRR interlaboratory study library replicates

**Format**

A data frame with 43919 observations of the following 9 variables.

- **Feature**  
  a character vector of all Endogenous and ERCC transcripts in the experiment
- **UHRR_1**  
  a numeric vector of counts from UHRR library preparation replicate 1
- **UHRR_2**  
  a numeric vector of counts from UHRR library preparation replicate 2
- **UHRR_3**  
  a numeric vector of counts from UHRR library preparation replicate 3
- **UHRR_4**  
  a numeric vector of counts from UHRR library preparation replicate 4
- **HBRR_1**  
  a numeric vector of counts from HBRR library preparation replicate 1
- **HBRR_2**  
  a numeric vector of counts from HBRR library preparation replicate 2
- **HBRR_3**  
  a numeric vector of counts from HBRR library preparation replicate 3
- **HBRR_4**  
  a numeric vector of counts from HBRR library preparation replicate 4
UHRR HBRR total Reads

UHRR and HBRR sample total read data

Description
Total reads per library replicate from FASTQ files

Format
The format is: int [1:8] 138786892 256006510 199468322 431933806 247985592 219383270 251265814 257508210
Index

*Topic datasets
  ERCC, 4
  ERCCDef, 4
  ERCCMix1and2, 5
  SEQC.Example, 13
  UHRR.HBRR.arrayDat, 14
  UHRR.HBRR.countDat, 14

annotLODR, 2

dynRangePlot, 3

ERCC, 4
ERCCDef, 4
ERCCMix1and2, 5
erccROC, 5
est_r_m, 7
estLODR, 6

geneExprTest, 7

initDat, 8

maSignal, 9
MET.CTL.countDat, 10
MET.CTL.totalReads, 11

runDashboard, 11

saveERCCPlots, 12
SEQC.Example, 13

UHRR.HBRR.arrayDat, 14
UHRR.HBRR.countDat, 14
UHRR.HBRR.totalReads, 15