Package ‘Anaquin’

May 29, 2024

Type Package
Title Statistical analysis of sequins
Version 2.28.0
Date 2017-08-08
Author Ted Wong
Maintainer Ted Wong <t.wong@garvan.org.au>
Description The project is intended to support the use of sequins (synthetic sequencing spike-in controls) owned and made available by the Garvan Institute of Medical Research. The goal is to provide a standard open source library for quantitative analysis, modelling and visualization of spike-in controls.
License BSD_3_clause + file LICENSE
VignetteBuilder knitr
URL www.sequin.xyz
Depends R (>= 3.3), ggplot2 (>= 2.2.0)
Imports ggplot2, ROCR, knitr, qvalue, locfit, methods, stats, utils, plyr, DESeq2
Suggests RUnit, rmarkdown
BugReports https://github.com/student-t/RAnaquin/issues
LazyData true
biocViews ImmunoOncology, DifferentialExpression, Preprocessing, RNASeq, GeneExpression, Software
git_url https://git.bioconductor.org/packages/A Anaquin
git_branch RELEASE_3_19
git_last_commit b547d11
git_last_commit_date 2024-04-30
Repository Bioconductor 3.19
Date/Publication 2024-05-29
plotConjoint

Contents

plotConjoint .......................................................... 2
plotLinear ............................................................. 3
plotLOD ................................................................. 4
plotLogistic ............................................................ 6
plotROC ................................................................. 7
RnaQuinGeneMixture .................................................... 9
RnaQuinIsoformMixture ................................................ 9
UserGuideData_5.4.5.1 ................................................ 10
UserGuideData_5.4.6.3 ................................................ 11
UserGuideData_5.6.3 ................................................... 11

Index 13

plotConjoint Create conjoint plots

Description

Create scatter plot for conjoint sequins.

Usage

plotConjoint(seqs, units, x, y, title=NULL, xlab=NULL, ylab=NULL)

Arguments

seqs Sequin names
units Copy units
x Expected copy number on the x-axis
y Measured abundance on the y-axis
title Label of the plot. Default to NULL.
xlab Label for the x-axis. Default to NULL.
ylab Label for the y-axis. Default to NULL.

Details

This is an experimental function for the conjoint sequins, and thus might not be fully utilized.

Value

This function does not return anything.

Author(s)

Ted Wong <t.wong@garvan.org.au>
plotLinear

Plot linear model for sequins

Description

Create linear model for sequins, between input concentration on the x-axis and measurement on the y-axis.

Usage

plotLinear(seqs, x, y, std, title, xlab, ylab, showSD, showLOQ, showStats, xBreaks, yBreaks, errors, showLinear, showAxis)

Arguments

- **seqs**: Sequin names
- **x**: Input concentration on the x-axis
- **y**: Measurement on the y-axis
- **std**: Standard deviation. (Default to NULL).
- **title**: Label of the plot. (Default to NULL).
- **xlab**: Label for the x-axis. (Default to NULL).
- **ylab**: Label for the y-axis. (Default to NULL).
- **xBreaks**: Breaks for the x-axis. (Default to NULL).
- **yBreaks**: Breaks for the y-axis. (Default to NULL).
- **showSD**: Display vertical standard deviation bars. (Default to FALSE).
- **showLOQ**: Display limit-of-quantification? Default to TRUE.
- **showStats**: Display regression statistics? Default to TRUE.
- **errors**: How errors bar should be calculated. SD or Range.
- **showLinear**: Display regression line. (Default to TRUE).
- **showAxis**: Display x-axis and y-axis. (Default to TRUE).

Details

The `plotLinear` function plots a scatter plot with input concentration on the x-axis, and measurement on the y-axis. The input concentration is typically the concentration level in ladder mixture, although other measures (such as expected copy number) are also possible. The function builds a linear regression between the two variables, and reports associated statistics (R2, correlation and regression parameters) on the plot.

The function also estimates limit-of-quantification (LOQ) breakpoint, and reports it on the plot if found. LOQ is defined as the lowest empirical detection limit, a threshold value beyond which stochastic behavior occur. LOQ is estimated by fitting segmented linear regression with two segments on the entire data set, while minimizing the total sum of squares of the differences between the variables.
Value

The function prints a scatter plot and return its LOQ statistics.

Author(s)

Ted Wong <t.wong@garvan.org.au>

Examples

library(Anaquin)

# Data set generated by Cufflinks and Anaquin. described in Section 5.4.6.3 of
# the user guide.
# data(UserGuideData_5.4.6.3)

title <- 'Gene Expression'

title <- 'Gene Expression'
xlab <- 'Input Concentration (log2)'
ylab <- 'FPKM (log2)'

# Sequin names
seqs <- row.names(UserGuideData_5.4.6.3)

# Input concentration
x <- log2(UserGuideData_5.4.6.3$Input)

# Measured FPKM
y <- log2(UserGuideData_5.4.6.3[,2:4])

plotLinear(seqs, x, y, title=title, xlab=xlab, ylab=ylab, showLOQ=TRUE)

plotLOD Create Limit-of-Detection Ratio (LOD) plot

Description

Create Limit-of-Detection Ratio (LOD) plot between measured abundance (x-axis) and p-value probability (y-axis).

Usage

plotLOD(measured, pval, ratio, qval, FDR, title, xlab, ylab, legTitle, showConf)
plotLOD

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>measured</td>
<td>Measured abundance</td>
</tr>
<tr>
<td>pval</td>
<td>P-value probability</td>
</tr>
<tr>
<td>ratio</td>
<td>How to group ROC points</td>
</tr>
<tr>
<td>qval</td>
<td>Q-value probability. (Default to NULL).</td>
</tr>
<tr>
<td>FDR</td>
<td>Chosen false-discovery-rate. Default to NULL.</td>
</tr>
<tr>
<td>title</td>
<td>Title of the plot. (Default to NULL).</td>
</tr>
<tr>
<td>xlab</td>
<td>Label for the x-axis. (Default to NULL).</td>
</tr>
<tr>
<td>ylab</td>
<td>Label for the y-axis. (Default to NULL).</td>
</tr>
<tr>
<td>legTitle</td>
<td>Title for the legend. (Default to 'Ratio').</td>
</tr>
<tr>
<td>showConf</td>
<td>Display confidence interval. (Default to FALSE).</td>
</tr>
</tbody>
</table>

Details

Create a Limit-of-Detection Ratio (LOD) plot between measured abundance (x-axis) and p-value probability (y-axis).

The LOD plot indicates the confidence in measurement relative to the magnitude of the measurement. For example, p-value should converge to zero as the sequencing depth increases.

The function also fits non-parametric curves for each sequin ratio group. The curves are modelled with local regression analysis, and are colored by the sequin group.

plotLODR is a simplification from the ERCC dashboard R-package. Further details on the statistical algorithm is available in the ERCC documentation at https://bioconductor.org/packages/release/bioc/html/erccdashboard.html.

Value

The function prints a LODR plot and return associated statistics.

Author(s)

Ted Wong <t.wong@garvan.org.au>

Examples

```r
library(Anaquin)

# Data set generated by DESeq2 and Anaquin. described in Section 5.6.3.3 of
# the user guide.
#
data(UserGuideData_5.6.3)

xlab <- 'Average Counts'
ylab <- 'P-value'
title <- 'LOD Curves'

# Sequin names
```r
seqs <- row.names(UserGuideData_5.6.3)

# Expected log-fold
group <- UserGuideData_5.6.3$ExpLFC

# Measured average abundance
measured <- UserGuideData_5.6.3$Mean

# P-value
pval <- UserGuideData_5.6.3$Pval

# Q-value
qval <- UserGuideData_5.6.3$Qval

plotLOD(measured, pval, group, qval, xlab=xlab, ylab=ylab, title=title, FDR=0.1)
```

### plotLogistic

**Plot logistic model for sequins**

**Description**

Create a scatter plot with input concentration on the x-axis, and measured proportion on the y-axis.

**Usage**

```r
plotLogistic(seqs, x, y, title, xlab, ylab, showLOA, threshold)
```

**Arguments**

- `seqs` Seuin names
- `x` Expected input concentration on the x-axis
- `y` Measured proportion on the y-axis
- `title` Title of the plot. (Default to NULL).
- `xlab` Label for the x-axis. (Default to NULL).
- `ylab` Label for the y-axis. (Default to NULL).
- `showLOA` Display limit-of-assembly. (Default to TRUE).
- `threshold` Threshold required for limit-of-assembly (LOA). (Default to 0.7).

**Details**

The `plotLogistic` function creates a scatter plot with input concentration on the x-axis, and measured proportion on the y-axis. Common measured statistics include p-value, percentage and sensitivity. The plot builds a logistic regression model between the two variables.

The function also estimates limit-of-assembly (LOA) breakpoint, and reports it on the plot if found. The LOA breakpoint is an empirical detection limit, and also the abundance whereby the fitted logistic curve exceeds a user-defined threshold.
plotROC

Value

The function returns the limit of quantification.

Author(s)

Ted Wong <t.wong@garvan.org.au>

Examples

library(Anaquin)

# Data set generated by Cufflinks and Anaquin. described in Section 5.4.5.1 of # the user guide.
# data(UserGuideData_5.4.5.1)

title <- 'Assembly Plot'

title <- title
xlab <- 'Input Concentration (log2)'

title <- title
ylab <- 'Sensitivity'

# Sequin names
seqs <- row.names(UserGuideData_5.4.5.1)

# Input concentration
x <- log2(UserGuideData_5.4.5.1$Input)

# Measured sensitivity
y <- UserGuideData_5.4.5.1$Sn

plotLogistic(seqs, x, y, title=title, xlab=xlab, ylab=ylab, showLOA=TRUE)

plotROC

Create ROC plot

Description

Create receiver operating characteristic (ROC) plot at various threshold settings.

Usage

plotROC(seqs, score, group, label, refGroup, title, legTitle)

Arguments

seqs  Sequin names
score How to rank ROC points

group How to group ROC points
plotROC

label          True-positive (TP) or false positive (FP)
refGroup       Reference ratio groups
title          Label of the plot. Default to NULL.
legTitle       Title of the legend. Default to Ratio.

Details
Create a receiver operating characteristic (ROC) plot at various threshold settings. The true positive rate (TPR) is plotted on the x-axis and false positive rate (FPR) is plotted on the y-axis.

The function requires a scoring threshold function, and illustrates the performance of the data as the threshold is varied. Common scoring threshold include p-value, sequencing depth and allele frequency, etc.

ROC plot is a useful diagnostic performance tool; it provides tools to select possibly optimal models and to discard suboptimal ones. In particularly, the AUC statistics indicate the performance of the model relatively to a random experiment (AUC 0.5).

Value
The function prints ROC plot and return it's AUC statistics.

Author(s)
Ted Wong <t.wong@garvan.org.au>

Examples

library(Anaquin)

# Data set generated by DESeq2 and Anaquin. described in Section 5.6.3.3 of
# the user guide.
data(UserGuideData_5.6.3)

# Sequin names
seqs <- row.names(UserGuideData_5.6.3)

# Expected log-fold
group <- abs(UserGuideData_5.6.3$ExpLFC)

# How the ROC curves are ranked
score <- 1-UserGuideData_5.6.3$Pval

# Classified labels (TP/FP)
label <- UserGuideData_5.6.3$Label

plotROC(seqs, score, group, label, title='ROC Plot', refGroup=0)
**RnaQuinGeneMixture**  
*RnaQuin mixture (gene level)*

**Description**

Individual sequins are combined across a range of precise concentrations to formulate mixtures. By modulating the concentration at which each sequin is present in the mixture, we can emulate quantitative features of genome biology.

This is the mixture A and B in RnaQuin. File name is A.R.6.csv on http://www.sequins.xyz.

**Usage**

data(RnaQuinGeneMixture)

**Format**

Data frame:

- Name: Sequin name
- Length: Gene length
- MixA: Input concentration for mixture A
- MixB: Input concentration for mixture B

**Value**

Data frame with columns defined in Format.

---

**RnaQuinIsoformMixture**  
*RnaQuin mixture (isoform level)*

**Description**

Individual sequins are combined across a range of precise concentrations to formulate mixtures. By modulating the concentration at which each sequin is present in the mixture, we can emulate quantitative features of genome biology.

This is the mixture A and B in RnaQuin. File name is A.R.5.csv on http://www.sequins.xyz.

**Usage**

data(RnaQuinIsoformMixture)
Format

Data frame:

- Name: Sequin name
- Length: Sequin length
- MixA: Input concentration for mixture A
- MixB: Input concentration for mixture B

Value

Data frame with columns defined in Format.

UserGuideData_5.4.5.1  Section 5.4.5.1 Assembly Dataset

Description

Assembly sensitivity estimated by Cuffcompare. Section 5.4.5.1 of the Anaquin user guide has details on the data set.

Usage

data(UserGuideData_5.4.5.1)

Format

Data frame:

- InputConcent: Input concentration in attomol/ul
- Sn: Measured sensitivity

Value

Data frame with columns defined in Format.

Source

UserGuideData_5.4.6.3  Gene expression (RnaQuin)

Description
Gene expression estimated by Cufflinks. Section 5.4.6.3 of the Anaquin user guide has details on the data set.

Usage
data(UserGuideData_5.4.6.3)

Format
Data frame:

• InputConcent: Input concentration in attomol/ul
• Observed1: Measured FPKM for the first replicate
• Observed2: Measured FPKM for the second replicate
• Observed3: Measured FPKM for the third replicate

Value
Data frame with columns defined in Format.

Source

UserGuideData_5.6.3  Differential expression (RnaQuin)

Description
Differential gene expression estimated by DESeq2. Section 5.6.3 has details on the data set.

Usage
data(UserGuideData_5.6.3)
Format

Data frame:

- ExpLFC: Expected log-fold change
- ObsLFC: Observed log-fold change
- SD: Standard deviation of the measurement
- Pval: P-value probability
- Qval: Q-value probability
- Mean: Average counts across the samples
- Label: Average counts across the samples

Value

Data frame with columns defined in Format.

Source

Index

plotConjoint, 2
plotLinear, 3
plotLOD, 4
plotLogistic, 6
plotROC, 7

RnaQuinGeneMixture, 9
RnaQuinIsoformMixture, 9

UserGuideData_5.4.5.1, 10
UserGuideData_5.4.6.3, 11
UserGuideData_5.6.3, 11