Package 'ChIPexoQual'

May 9, 2024

Type Package

Title ChIPexoQual

Version 1.28.0

Author Rene Welch, Dongjun Chung, Sunduz Keles

Maintainer Rene Welch <welch@stat.wisc.edu>

Description Package with a quality control pipeline for ChIP-exo/nexus data.

URL https:github.com/keleslab/ChIPexoQual

BugReports https://github.com/welch16/ChIPexoQual/issues

License GPL (>=2)

Depends R (>= 3.4.0), GenomicAlignments (>= 1.0.1)

Imports methods, utils, GenomeInfoDb, stats, BiocParallel, GenomicRanges (>= 1.14.4), ggplot2 (>= 1.0), data.table (>= 1.9.6), Rsamtools (>= 1.16.1), IRanges (>= 1.6), S4Vectors (>= 0.8), biovizBase (>= 1.18), broom (>= 0.4), RColorBrewer (>= 1.1), dplyr (>= 0.5), scales (>= 0.4.0), viridis (>= 0.3), hexbin (>= 1.27), rmarkdown

Suggests ChIPexoQualExample (>= 0.99.1), knitr (>= 1.10), BiocStyle, gridExtra (>= 2.2), testthat

VignetteBuilder knitr

biocViews ChIPSeq, Sequencing, Transcription, Visualization, QualityControl, Coverage, Alignment

RoxygenNote 6.0.1

git_url https://git.bioconductor.org/packages/ChIPexoQual

git_branch RELEASE_3_19

git_last_commit bd77a4d

git_last_commit_date 2024-04-30

Repository Bioconductor 3.19

Date/Publication 2024-05-09

Contents

`ARCvURCplot` *ARCvURCplot*

Description

`ARCvURCplot` returns a `ggplot` object with the ARC vs URC plot to analyze enrichment and library complexity in ChIP-exo data.

Usage

```
ARCvURCplot(..., names.input = NULL, both.strand = FALSE)
```

Arguments

`...`	a `list` of `ExoData` objects, or several `ExoData` objects by themselves.
`names.input`	a character vector with the names to use in the plot. If it is empty `ARCvURCplot` is going to create the names as the names of the list when they are available or is going to name them as Sample: 1 ,... , Sample: k.
`both.strand`	A logical value indicating if the `DataFrame` contains only regions with reads aligned to both strand or all. The default value is `FALSE`.

Value

A `ggplot2` object with the ARC vs URC plot.

Examples

```
data(exoExample)
ARCvURCplot(exoExample)
```

---

beta1 *beta1 methods*

Description

beta1 returns a vector with all the estimated values of the $d_i = \beta_1 u_i + \beta_2 w_i + \epsilon_i$ models fitted by ChIPexoQual

Usage

```
beta1(object)

## S4 method for signature 'ExoData'
beta1(object)
```

Arguments

object	a ExoData object.

Value

A numeric vector with estimated values for β_1.

Examples

```
data(exoExample)
beta1(exoExample)
```

---

beta2 *beta2 methods*

Description

beta2 returns a vector with all the estimated values of the $d_i = \beta 1 u_i + \beta 2 w_i + \epsilon_i$ models fitted by ChIPexoQual

Usage

```
beta2(object)

## S4 method for signature 'ExoData'
beta2(object)
```

Arguments

object	a ExoData object.

Value

A numeric vector with estimated values for β_2.

Examples

```
data(exoExample)
beta2(exoExample)
```

---

blacklists — list *of* GRanges *objects with the blacklists generated by the ENCODE and modENCODE projects.*

Description

list of GRanges objects with the blacklists generated by the ENCODE and modENCODE projects.

Usage

```
data(blacklists)
```

Format

list of GRanges objects.

Value

A list with the blacklists listed in https://sites.google.com/site/anshulkundaje/projects/blacklists.

---

calculateParamDist — *calculateParamDist* calculateParamDist *calculates the quality parameters of one iteration. This function samples* nregions *rows from the stat matrix and fits the linear model* lm(d ~ 0 + u + w)

Description

calculateParamDist

calculateParamDist calculates the quality parameters of one iteration. This function samples nregions rows from the stat matrix and fits the linear model lm(d ~ 0 + u + w)

Usage

```
calculateParamDist(i, stats, nregions)
```

Arguments

i	a numeric value indicating the current iteration.
stats	a data.table object with the response and covariates for the model
nregions	a numeric value indicating the number of regions sampled.

Value

a data.table with both parameters and some extra info

Examples

```
data("exoExample")
DT <- formatRegions(exoExample)
calculateParamDist(1,DT,100)
```

ExoData-class *ExoData object and constructors*

Description

ExoData is a subclass of GenomicRanges, used to asses the quality of ChIP-exo/nexus sample.

Usage

```
ExoData(file = NULL, reads = NULL, height = 1,
  mc.cores = getOption("mc.cores", 2L), save.reads = FALSE,
  nregions = 1000, ntimes = 100, verbose = TRUE)
```

Arguments

file	a character value with location of the bam file with the aligned reads.
reads	a GAlignments object with the aligned reads of a ChIP-exo sample. It is meant to be used instead of file.
height	a numeric value indicating the value used to slice the coverage of the experiment into a set of regions.
mc.cores	a numeric value with the number of cores to use, i.e. at most how many child processes will be run simultaneously.
save.reads	a logical value to indicate if the reads are stored in the ExoData object. The default value is FALSE.
nregions	a numeric value indicating the number of regions sampled to estimate the quality parameter distributions. The default value is 1e3.
ntimes	a numeric value indicating the number of times that regions are sampled to estimate the quality parameter distributions. The default value is 1e2.
verbose	a logical value indicating if the user want to receive progress details. The default value is FALSE.

Value

It returns an ExoData object with the regions obtained after partitioning the genome and the summary statistics for each region. If the save.reads parameter is TRUE then it contains a GRanges object with the reads of the ChIP-exo experiment.

Examples

```
files <- list.files(system.file("extdata",package = "ChIPexoQualExample"),
    full.names = TRUE)
ExoData(files[5],mc.cores = 2L)
```

ExoDataBlacklist *ExoDataBlacklist*

Description

ExoDataBlacklist separates the regions in an ExoData object by overlapping them with a set of blacklisted regions and calculates the quality parameters in both collections of islands.

Usage

```
ExoDataBlacklist(exo, blacklist, which.param = "beta1", nregions = NULL,
  ntimes = NULL)
```

Arguments

exo	a ExoData object.
blacklist	a GRanges object with the blacklisted regions or a character indicating which of the blacklist included in ChIPexoQual to use.
which.param	a character value with either "beta1" or "beta2" that determines which parameters in the model depth_i ~ uniquePos_i + width_i to plot. The default value is "beta1".
nregions	a numeric value indicating the number of regions sampled to estimate the quality parameter distributions. The default value is extracted from exo.
ntimes	a numeric value indicating the number of times that regions are sampled to estimate the quality parameter distributions. The default value is extracted from object.

Value

A ggplot object with a boxplot that compares the quality scores distribution when the regions overlap a pre-defined collection of blacklists.

Examples

```
data(exoExample)
data(blacklists)
ExoDataBlacklist(exoExample,blacklists[["mm9"]],ntimes = 10,nregions = 500)
```

ExoDataSubsampling *ExoDataSubsampling*

Description

ExoDataSubsampling samples sample.reads from the ChIP-exo experiment and creates a list of ExoData objects

Usage

```
ExoDataSubsampling(file = NULL, reads = NULL, sample.depth = NULL,
  height = 1, nregions = 1000, ntimes = 1000, verbose = TRUE,
  save.reads = FALSE, mc.cores = getOption("mc.cores", 2L))
```

Arguments

file	a character value with location of the bam file with the aligned reads.
reads	a GAlignments object with the aligned reads of a ChIP-exo sample. It is meant to be used instead of file.
sample.depth	a numeric vector with the number of reads to be sampled.
height	a numeric value indicating the value used to slice the coverage of the experiment into a set of regions.
nregions	a numeric value indicating the number of regions sampled to estimate the quality parameter distributions. The default value is 1e3.
ntimes	a numeric value indicating the number of times that regions are sampled to estimate the quality parameter distributions. The default value is 1e2.
verbose	a logical value indicating if the user want to receive progress details. The default value is FALSE.
save.reads	a logical value to indicate if the reads are stored in the ExoData object. The default value is FALSE.
mc.cores	a numeric value with the number of cores to use, i.e. at most how many child processes will be run simultaneously.

Value

It returns an ExoData object with the regions obtained after partitioning the genome and the summary statistics for each region. If the save.reads parameter is TRUE then it contains a GRanges object with the reads of the ChIP-exo experiment.

Examples

```
files <- list.files(system.file("extdata",package = "ChIPexoQualExample"),
    full.names = TRUE)
sample.depth <- seq(1e5,2e5,5e4)
ExoDataSubsampling(file = files[5],sample.depth = sample.depth)
```

`exoExample` `ExoData` *results for FoxA1 ChIP-exo experiment*

Description

`ExoData` object, generated with `ChIPexoQual` and the file:

Usage

```
data(exoExample)
```

Format

ExoData object, which are `GRanges` with additional columns.

Details

- ChIPexo_carroll_FoxA1_mouse_rep3_chr1.bam

Value

An `ExoData` object with the 3rd replicate of the FoxA1 experiment from `ChIPExoQualExample`.

`formatRegions` *formatRegions* `formatRegions` *separates the width, depth and uniquePos summary statistics from the* `ExoData` *object to calculate the quality parameters/*

Description

formatRegions

`formatRegions` separates the width, depth and uniquePos summary statistics from the `ExoData` object to calculate the quality parameters/

Usage

```
formatRegions(exo)
```

Arguments

exo	a ExoData object

Value

a data.table with the width, depth and uniquePos of the regions in exo.

Examples

```
data("exoExample")
formatRegions(exoExample)
```

FSRDistplot *FSRDistplot*

Description

FSRDistplot returns a ggplot object with the Forward Strand Ratio distribution plot to analyze strand imbalance in ChIP-exo data.

Usage

```
FSRDistplot(..., names.input = NULL, quantiles = c(0, 0.25, 0.5, 0.75, 1),
  depth.values = seq_len(30), both.strand = FALSE)
```

Arguments

...	a list of ExoData objects, or several ExoData objects by themselves.
names.input	a character vector with the names to use in the plot. If it is empty FSRDistplot is going to create the names as the names of the list when they are available or is going to name them as Sample: 1 ,... , Sample: k.
quantiles	a numeric vector with the quantiles used to estimate the FSR distribution at a given depth. The default value is c(0,.25,.5,.75,1))
depth.values	a numeric vector indicating the regions with depth less or equal to, that are going to be filtered out. The defaulta values are seq_len(50).
both.strand	a logical value indicating if the DataFrame contains only regions with reads aligned to both strand or all. The default value is FALSE.

Value

A ggplot2 object with the FSR distribution plot.

Examples

```
data(exoExample)
FSRDistplot(exoExample)
```

MAplot *MAplot*

Description

MAplot returns a ggplot object with the MA plot to analyze the strand imbalance in ChIP-exo data.

Usage

```
MAplot(..., names.input = NULL)
```

Arguments

...	a list of ExoData objects, or several ExoData objects by themselves.
names.input	a character vector with the names to use in the plot. If it is empty MAplot is going to create the names as the names of the list when they are available or is going to name them as Sample: 1 ,... , Sample: k.

Value

A ggplot2 object with the MA plot.

Examples

```
data(exoExample)
MAplot(exoExample)
```

nreads *nreads methods*

Description

nreads returns the number of reads in the object.

Usage

```
nreads(object)

## S4 method for signature 'ExoData'
nreads(object)
```

Arguments

object	A ExoData object.

Value

The number of reads in the ExoData object.

Examples

```
data(exoExample)
nreads(exoExample)
```

---

paramDist *paramDist methods*

Description

paramDist returns a DataFrame with all the estimated coefficients in the $d_i = \beta_1 u_i + \beta_2 w_i + \epsilon_i$ models fitted by ChIPexoQual

Usage

```
paramDist(object)

## S4 method for signature 'ExoData'
paramDist(object = "ExoData")
```

Arguments

object	a ExoData object.

Value

A DataFrame with the fitted values of β_1 and β_2.

Examples

```
data(exoExample)
paramDist(exoExample)
```

paramDistBoxplot *paramDistBoxplot*

Description

paramDistBoxplot returns a ggplot object with a boxplot comparing the ntimes estimations of the chosen parameter.

Usage

```
paramDistBoxplot(..., names.input = NULL, which.param = "beta1",
  sort.as.numeric = FALSE)
```

Arguments

...	a list of ExoData objects, or several ExoData objects by themselves.
names.input	a character vector with the names to use in the plot. If it is empty paramDistBoxplot is going to create the names as the names of the list when they are available or is going to name them as Sample: 1 ,... , Sample: k.
which.param	a character value with either "beta1" or "beta2" that determines which paramters in the model depth_i ~ uniquePos_i + width_i to plot. The default value is "beta1".
sort.as.numeric	a logical value indicating if the values of names.input are meant to be interpreted as numeric and sorted accordingly.

Value

A ggplot2 object with the boxplot of the chosen parameter

Examples

```
data(exoExample)
paramDistBoxplot(exoExample)
```

regionCompplot *regionCompplot*

Description

regionCompplot returns a ggplot object with the Region Composition plot to analyze strand imbalance in ChIP-exo data.

Usage

```
regionCompplot(..., names.input = NULL, depth.values = seq_len(15))
```

Arguments

...	a list of ExoData objects, or several ExoData objects by themselves.
names.input	a character vector with the names to use in the plot. If it is empty regionCompplot is going to create the names as the names of the list when they are available or is going to name them as Sample: 1 ,... , Sample: k.
depth.values	a numeric vector indicating the regions with depth less or equal to, that are going to be filtered out. The defaulta values are seq_len(50).

Value

A ggplot2 object with the Region Composition plot.

Examples

```
data(exoExample)
regionCompplot(exoExample)
```

Index