Package 'intePareto'

June 18, 2021

Type Package

Title Integrative Analysis of RNA-Seq and ChIP-Seq Data

Version 0.1.2

Date 2021-06-17

Description Integrative analysis of gene expression (RNA-Seq data), and histone modification data for user-defined sets of histone marks (ChIP-Seq data) to discover consistent changes in genes between biological conditions. Additionally, Pareto optimization is used to prioritize genes based on the level of consistent changes in both RNA-Seq and ChIP-Seq data. Method is described in Cao, Y. et al. (2020) <doi:10.1186/s12864-020-07205-6>.

License GPL (>= 2)

Depends R (>= 3.6.0)

Imports GenomicRanges, GenomeInfoDb, IRanges, GenomicAlignments, biomaRt, Rsamtools, rPref, DESeq2

Suggests knitr, rmarkdown

LazyLoad yes

NeedsCompilation no

ByteCompile true

Encoding UTF-8

RoxygenNote 7.1.1

VignetteBuilder knitr

Author Yingying Cao [aut, cre]

Maintainer Yingying Cao <yingying.cao@uni-due.de>

Repository CRAN

Date/Publication 2021-06-18 16:10:07 UTC

R topics documented:

`bam2counts` *Compute the number of reads fall into specific genomic region*

Description

`bam2counts` computes the number of reads fall into specific genomic region such as promoter, enhancer, genebody

Usage

```
bam2counts(bamFile, region, fragLength = 180)
```

Arguments

`bamFile`	Aligned bam file as input.
`region`	The GRanges object defined by user to calculate the number of reads fall into this specific region. For ChIP-Seq of histone modifications they are usually promoter, enhancer and genebody regions.
`fragLength`	Extend reads toward the 3'-end to the average DNA fragment size obtained after DNA size selection

Value

a vector of numbers

Examples

```
data("promoter")
file.bam <- system.file("extdata", "SRR925640.bam", package = "intePareto")
bam2counts(bamFile = file.bam, region = promoter, fragLength = 180)
```

bam2rpm *Compute the normalized number of reads (rpm) fall into specific genomic region*

Description

bam2rpm computes the normalized number of reads (rpm) fall into specific genomic region such as promoter, enhancer, genebody

Usage

```
bam2rpm(bamFile, region, fragLength = 180)
```

Arguments

bamFile	Aligned bam file as input.
region	The GRanges object defined by user to calculate the number of reads fall into this specific region. For ChIP-Seq of histone modifications they are usually promoter, enhancer and genebody regions.
fragLength	Extend reads toward the 3'-end to the average DNA fragment size obtained after DNA size selection

Value

a vector of numbers

Examples

```
data("promoter")
file.bam <- system.file("extdata", "SRR925640.bam", package = "intePareto")
bam2rpm(bamFile = file.bam, region = promoter, fragLength = 180)
```

counts2lFC *Calculate log2FoldChange*

Description

counts2lFC calculates log2FoldChange from counts.

Usage

```
counts2lFC(
  countData,
  colData,
  condition,
  ref,
  type = "apeglm",
  apeAdapt = FALSE,
  ...
)
```

Arguments

countData	a matrix of counts.
colData	a data.frame with at least a single column. Rows of colData correspond to columns of countData.
condition	the formula expresses how the counts for each gene depend on the variables in colData. The comparisons will be based on the alphabetical order of the levels by default. You can also specify the reference level by ref parameter
ref	specifying the reference level
type	shrinkage estimator, default is "apeglm", the adaptive t prior shrinkage estimator from the 'apeglm' package.
apeAdapt	logical, should apeglm use the MLE estimates of LFC to adapt the prior, or use default.
...	refer to DESeq2::lfcShrink() for more detailed parameters.

Value

resLFC a dataframe contains log2FoldChange.

Please note this is a downsampling of the original data.

---

df_final *an example of the result of doIntegration*

Description

a dataframe contatins log2FoldChange of RNA-Seq and ChIP-Seq and the Z score for each mark

Usage

```
data(df_final)
```

Format

An object of data.frame.

Examples

```
data(df_final)
```

`doIntegration` *Do integrative analysis*

Description

`doIntegration` calculate log2FoldChange of RNA-Seq and ChIP-Seq and then calculate Z scores for each marker.

Usage

```
doIntegration(res, ref, type = "apeglm", apeAdapt = FALSE)
```

Arguments

`res`	a list result from `doMatch` function.
`ref`	specifying the reference level
`type`	shrinkage estimator, default is "apeglm", the adaptive t prior shrinkage estimator from the 'apeglm' package.
`apeAdapt`	logical, should apeglm use the MLE estimates of LFC to adapt the prior, or use default.

Value

df_final a dataframe contains log2FoldChange of RNA-Seq and ChIP-Seq and Z scores for each marker.

Examples

```
data(res)
doIntegration(res = res,ref="wild.type")
```

doMatch *Match the RNA-Seq and ChIP-Seq data on the gene level*

Description

doMatch computes the number of reads (counts) fall into specific genomic region such as promoter or genebody for ChIP-Seq, and calculate the gene expression in counts, and then match the RNA-Seq and ChIP-Seq data on the gene level with the method of "weighetd mean" or "highest".

Usage

```
doMatch(
  rnaMeta,
  chipMeta,
  region,
  method,
  ensemblDataset,
  host,
  fragLength = 180,
  promoter.length = 5000
)
```

Arguments

rnaMeta	metadata for RNA-Seq include column named "condition" indicates the experiment condition or cell type, and column named "files" indicates the paths of cprresponing abundance.tsv file that is returned from Kallisto.
chipMeta	metadata for ChIP-Seq include column of "mark" column indicates the markers of histone modifications, column of "condition" indicates the experiment condition or cell type, and "files" column indicates the paths and the file names of the aligned bam files.
region	region has to be specified as "promoter" or "genebody".
method	method has to be specified as "weighted.mean" or"highest" if region is set as "promoter".
ensemblDataset	Ensembl Dataset you want to use. To see the different datasets available within a biomaRt you can e.g. do: mart = useMart('ensembl'), followed by listDatasets(mart).
host	specify the archived versions of Ensembl. To see the available archived versions do: biomaRt::listEnsemblArchives()
fragLength	extend reads toward the 3'-end to the average DNA fragment size obtained after DNA size selection.
promoter.length	the length of the promoter region.

Value

A list with the following three items.

res.rna a data frame contains RNA-Seq counts

res.chip a data frame contains ChIP-Seq counts

matched.data a dataframe contains matched RNA-Seq counts and ChIP-Seq counts.

Examples

```
data(test_rna_meta)
data(test_chip_meta)

for(i in test_rna_meta$SRR){
test_rna_meta$files <- system.file("extdata",paste0(i,".tsv"),
package = "inntePareto")
}
for(i in test_chip_meta$SRR){
test_chip_meta$files <- system.file("extdata", paste0(i,".bam"),
package = "inntePareto")
}
doMatch(rnaMeta = test_rna_meta,
chipMeta = test_chip_meta,
region = "promoter",
method = "weighted.mean",
host = "http://aug2017.archive.ensembl.org",
ensemblDataset = "mmusculus_gene_ensembl")
```

---

doPareto *Prioritization of genes based on Z scores*

Description

doPareto takes the Z scores of several different histone modifications as input, the prioritization of genes based on Z scores can be formulated as multiobjective optimization problem and solved with Pareto optimization.

Usage

```
doPareto(df_final, objective, nr.fronts)
```

Arguments

df_final	a data frame which is the output of doIntegration.
objective	a data frame which include column of "mark" column indicates the z scores of markers of histone modifications (e.g. "z.H3K4me3"), and a column named "obj" indicates the direction of the operation on the z scores, one of "max" and "min".
nr.fronts	the number of the pareto fronts you want to get.

Value

a data.frame ranked by the level of pareto fronts.

Examples

```
data("df_final")
objective <- data.frame(mark = c("z.H3K27ac","z.H3K4me3"),
                        obj=c("max","max"),stringsAsFactors=FALSE)
nr.fronts <- 3
doPareto(df_final = df_final,
         objective = objective,
         nr.fronts = nr.fronts)
```

---

promoter *an example of promoter region*

Description

an example of promoter region

Usage

```
data(promoter)
```

Format

An object of GRanges.

Examples

```
data(promoter)
```

---

res *an example of the result of doMatch*

Description

a list with the following three items. 1. res.rna, a data frame contains RNA-Seq counts 2. res.chip, a data frame contains ChIP-Seq counts 3. matched.data, a dataframe contains matched RNA-Seq counts and ChIP-Seq counts

Usage

```
data(res)
```

Format

An object of list.

Examples

```
data(res)
```

---

test_chip_meta *meta data of preprocessed ChIP-Seq data*

Description

The ChIP-Seq meta data.frame at least three columns: 1. mark: the mark of histone modifications (e.g. H3K4me3 or H3K27ac). 2. condition: identifier of the condition to which each sample belongs. 3. files: the exact address of the aligned bam files.

Usage

```
data(test_chip_meta)
```

Format

An object of data.frame.

Examples

```
data(test_chip_meta)
```

---

test_rna_meta *meta data of preprocessed RNA-Seq data*

Description

The RNA-Seq meta data.frame at least two columns: 1. condition: identifier of the condition to which each sample belongs. 2. files: the exact address of the files contains the tsv file which is the output of RNA-Seq preprocessed with Kallisto.

Usage

```
data(test_rna_meta)
```

Format

An object of data.frame.

Examples

```
data(test_rna_meta)
```

Index