Package 'enrichTF'

April 15, 2020

Type Package

Title Transcription Factors Enrichment Analysis

Version 1.2.3

Author Zheng Wei, Zhana Duren, Shining Ma

Maintainer Zheng Wei <wzwz@stanford.edu>

Description As transcription factors (TFs)
play a crucial role in regulating the
transcription process through binding on the
genome alone or in a combinatorial manner,
TF enrichment analysis is an efficient and important procedure to
locate the candidate functional TFs from a set of experimentally
defined regulatory regions.
While it is commonly accepted that structurally
related TFs may have similar binding preference to sequences (i.e. motifs)
and one TF may have multiple motifs, TF enrichment analysis is much more
challenging than motif enrichment analysis. Here we present a R package
for TF enrichment analysis
which combine motif enrichment with the PECA model.

License GPL-3

Encoding UTF-8

LazyData FALSE

Depends pipeFrame

Imports BSgenome, rtracklayer, motifmatchr, TFBSTools, R.utils,
methods, JASPAR2018, GenomeInfoDb, GenomicRanges, IRanges,
BiocGenerics, S4Vectors, utils, parallel, stats, ggpubr,
heatmap3, ggplot2, clusterProfiler, rmarkdown, grDevices,
magrittr

Suggests knitr, testthat, webshot

Collate EnrichStep.R ConnectTargetGene.R TFsEnrichInRegions.R
FindMotifsInRegions.R UnzipAndMergeBed.R
TissueOpennessSpecificity.R TissueOpennessConserve.R
SingleSampleReport.R GenBackground.R onLoad.R untils.R Method.R
GeneOntology.R

RoxygenNote 6.1.1

VignetteBuilder knitr

biocViews Software, GeneTarget, MotifAnnotation, GraphAndNetwork, Transcription

URL https://github.com/wzthu/enrichTF

BugReports https://github.com/wzthu/enrichTF/issues

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

git_branch RELEASE_3_10

git_last_commit f78ac0a

git_last_commit_date 2019-11-18

Date/Publication 2020-04-14

R topics documented:

`EnrichStep-class` *Base class of this package*

Description

This class is inherit from `Step` in pipeFrame package, no more method is extended or override. Please see `Step` class for detail.

`GenBackground` *Generate background regions and reset the size of foreground regions*

Description

Use uniform distribution to generate background sequence regions from genome. The size of foreground regions will be unified into the length specified in argument.

Usage

```
enrichGenBackground(prevStep, inputForegroundBed = NULL, genome = NULL,
  outputForegroundBed = NULL, outputBackgroundBed = NULL,
  outputRegionBed = NULL, regionLen = 1000, sampleNumb = 10000, ...)

## S4 method for signature 'Step'
enrichGenBackground(prevStep, inputForegroundBed = NULL,
  genome = NULL, outputForegroundBed = NULL,
  outputBackgroundBed = NULL, outputRegionBed = NULL,
  regionLen = 1000, sampleNumb = NULL, ...)

genBackground(inputForegroundBed, genome = NULL,
  outputForegroundBed = NULL, outputBackgroundBed = NULL,
  outputRegionBed = NULL, regionLen = 1000, sampleNumb = NULL, ...)
```

Arguments

prevStep	Step-class object scalar. It needs to be the return value of upstream process from other packages, such as esATAC.
inputForegroundBed	Character scalar. The directory of foreground BED file.
genome	Character scalar. Bioconductor supported genome such as "hg19", "mm10", etc. Default: NULL (e.g. after library (enrichTF), you can call function setGenome("hg19"))
outputForegroundBed	Character scalar. The BED file directory of reshaped foreground regions. Default: NULL (generated base on inputForegroundBed)
outputBackgroundBed	Character scalar. The BED file directory of reshaped background regions. Default: NULL (generated base on inputForegroundBed)
outputRegionBed	Character scalar. Foreground and background merged BED files. Default: NULL (generated base on inputForegroundBed)
regionLen	Character scalar. It sets the length of forground sequence regions. Default: 1000
sampleNumb	numeric scalar. It sets the number of background regions that will be sampled. Default: 10000
...	Additional arguments, currently unused.

Details

Use uniform distribution to generate background sequence regions from genome. The size of foreground regions will be unified into the length specified in argument.

Value

An invisible EnrichStep-class object (Step-class based) scalar for downstream analysis.

Author(s)

Zheng Wei

See Also

regionConnectTargetGene findMotifsInRegions tfsEnrichInRegions

Examples

```
setGenome("testgenome") #Use "hg19","hg38",etc. for your application
foregroundBedPath <- system.file(package = "enrichTF", "extdata","testregion.bed")
gen <- genBackground(inputForegroundBed = foregroundBedPath)
```

GeneOntology *Gene ontology enrichment analysis for provided gene list*

Description

User provide target gene list of forground region. This function will call function for gene ontology enrichment analysis

Usage

```
enrichGeneOntology(prevStep, inputTxt = NULL, outputTxt = NULL,
  outputPdf = NULL, orgDb = NULL, keyType = "SYMBOL", ...)

## S4 method for signature 'Step'
enrichGeneOntology(prevStep, inputTxt = NULL,
  outputTxt = NULL, outputPdf = NULL, orgDb = NULL,
  keyType = "SYMBOL", ...)

geneOntology(inputTxt, outputTxt = NULL, outputPdf = NULL,
  orgDb = NULL, keyType = "SYMBOL", ...)
```

Arguments

prevStep	Step-class object scalar. This parameter is available when the upstream step function (printMap() to see the previous functions) have been sucessfully called. Accepted value can be the object return by any step function or be feed by %>% from last step function.
inputTxt	Character scalar. Gene list text file. All gene names are in one column.
outputTxt	Character scalar. Gene ontology enrichment analysis result table. Each row contain one gene ontology information.
outputPdf	Character scalar. Gene ontology enrichment analysis result figure. It contains gene ontology network.
orgDb	Character scalar. Bioconductor OrgDb object name for gene ontology enrichment analysis.
keyType	Character scalar. Gene name type include "SYMBOL" and "ENSEMBLE"
...	Additional arguments, currently unused.

Details

Currently, this function call enrichGO from package clusterProfiler to implement this funtion.

Value

An invisible EnrichStep-class object (Step-class based) scalar for downstream analysis.

Author(s)

Zheng Wei

See Also

regionConnectTargetGene enrichRegionConnectTargetGene

Examples

```
genelist.txt <- system.file(package = "enrichTF", "extdata","genelist.txt")
geneOntology(inputTxt = genelist.txt, orgDb = "org.Hs.eg.db")
```

MotifsInRegions *Find motifs in all input sequence regions*

Description

Scan for motif occurrences using the prepared PWMs and obtain the promising candidate motifs in these regions.

Usage

```
enrichFindMotifsInRegions(prevStep, inputRegionBed = NULL,
  outputRegionMotifBed = NULL, motifRc = c("integrate", "jaspar",
  "pwmfile"), inputPwmFile = getRefFiles("motifpwm"),
  genome = getGenome(), threads = getThreads(), ...)

## S4 method for signature 'Step'
enrichFindMotifsInRegions(prevStep,
  inputRegionBed = NULL, outputRegionMotifBed = NULL,
  motifRc = c("integrate", "jaspar", "pwmfile"),
  inputPwmFile = getRefFiles("motifpwm"), genome = getGenome(),
  threads = getThreads(), ...)

findMotifsInRegions(inputRegionBed, outputRegionMotifBed = NULL,
  motifRc = c("integrate", "jaspar", "pwmfile"),
  inputPwmFile = getRefFiles("motifpwm"), genome = getGenome(),
  threads = getThreads(), ...)
```

Arguments

prevStep	Step-class object scalar. This parameter is available when the upstream step function (printMap() to see the previous functions) have been sucessfully called. Accepted value can be the object return by any step function or be feed by %>% from last step function.

`inputRegionBed` `Character` scalar. BED file for regions including foreground and background sequences.

`outputRegionMotifBed`
`Character` scalar. BED file for regions with motif candidates. Default: NULL (generated base on inputForegroundBed)

`motifRc` `Character` scalar. Motif Resources can be one of "integrate" (integrated by us and can be download from internet automatically if call the function `setGenome("hg19")`), "jaspar" package JASPAR2018, or "pwmfile" (User defined PWM file. inputPwmFile is required).

`inputPwmFile` `Character` scalar. when "pwmfile" is set for motifRc, use this argument to provide PWM file directory.

`genome` `Character` scalar. Bioconductor supported genome, such as "hg19", "mm10", etc. Default: NULL (e.g. after `library(enrichTF)`, you can call function `setGenome("hg19")`)

`threads` `Integer` scalar. The maximum threads that will be used in this step. Default: getThreads()

`...` Additional arguments, currently unused.

Details

Scan for motif occurrences using the prepared PWMs and obtain the promising candidate motifs in these regions.

Value

An invisible `EnrichStep-class` object (`Step-class` based) scalar for downstream analysis.

Author(s)

Zheng Wei

See Also

`genBackground` `findMotifsInRegions` `tfsEnrichInRegions`

Examples

```
setGenome("testgenome") #Use "hg19","hg38",etc. for your application
foregroundBedPath <- system.file(package = "enrichTF",
    "extdata","testregion.bed")
gen <- genBackground(inputForegroundBed = foregroundBedPath)
findMotif <- enrichFindMotifsInRegions(gen,motifRc="integrate")
```

Pipelines *ready-to-use pipelines*

Description

There are two ready-to-use pipelines in this package. One is the pipeline mainly for TF enrichment with PECA model. The ohter is the pipeline is for gene regulation analysis including the first pipeline and other related analysis like openness and gene ontology analysis.

Usage

```
PECA_TF_enrich(inputForegroundBed, genome, threads = 2,
  pipeName = "pipe", ...)

GeneReguPipe(inputForegroundBed, genome, threads = 2,
  pipeName = "pipe", ...)
```

Arguments

inputForegroundBed	Character scalar. Foreground BED file directory.
genome	Character scalar. Bioconductor supported genome like "hg19", "mm10", etc.
threads	Numeric scalar. The max number of threads that can be used by each step of the pipeline
pipeName	Character scalar or vector. Pipeline name.
...	Additional arguments to set arguments for each Steps. See below for details.

Details

This is a function for the pipeline. There are four steps in this pipeline: GenBackground, RegionConnectTarget, FindMotifsInRegions and TFsEnrichInRegions. Parameter setting is available for all these functions. For example, if you want to change the number of background regions (sampleNumb) into 1000, you can add the argument GenBackground.sampleNumb = 1000 into the function like this: PECA_TF_enrich(inputForegroundBed = "your_file.bed",genome="hg19",GenBackground.sa
= 1000). The number of arguments is not limited so you can add other arguments with the format (StepName.argumentName) in the same way.

Value

An invisible list scalar. A list containing all objects that belongs to the pipeline.

Author(s)

Zheng Wei

References

Zhana Duren, et al., Modeling gene regulation from paired expression and chromatin accessibility data. Proc Natl Acad Sci U S A. 2017 1;111(44):15675-80

Examples

```
foregroundBedPath <- system.file(package = "enrichTF", "extdata","testregion.bed")
# This is the whole pipeline example.
PECA_TF_enrich(inputForegroundBed = foregroundBedPath, genome = "testgenome")
```

RegionConnectTargetGene

Connect regions with their target genes

Description

Connect foreground and background regions to their target genes, which is predicted from PECA model.

Usage

```
enrichRegionConnectTargetGene(prevStep, inputForegroundBed = NULL,
  inputBackgroundBed = NULL, outputForegroundBed = NULL,
  outputBackgroundBed = NULL, regularGeneCorrBed = NULL,
  enhancerRegularGeneCorrBed = NULL, ouputForgroundGeneTxt = NULL, ...)

## S4 method for signature 'Step'
enrichRegionConnectTargetGene(prevStep,
  inputForegroundBed = NULL, inputBackgroundBed = NULL,
  outputForegroundBed = NULL, outputBackgroundBed = NULL,
  regularGeneCorrBed = NULL, enhancerRegularGeneCorrBed = NULL,
  ouputForgroundGeneTxt = NULL, ...)

regionConnectTargetGene(inputForegroundBed, inputBackgroundBed,
  outputForegroundBed = NULL, outputBackgroundBed = NULL,
  regularGeneCorrBed = NULL, enhancerRegularGeneCorrBed = NULL,
  ouputForgroundGeneTxt = NULL, ...)
```

Arguments

prevStep	Step-class object scalar. This parameter is available when the upstream step function (printMap() to see the previous functions) have been sucessfully called. Accepted value can be the object return by any step function or be feed by %>% from last step function.
inputForegroundBed	Character scalar. The BED file directory of foreground regions.
inputBackgroundBed	Character scalar. The BED file directory of background regions.
outputForegroundBed	Character scalar. The BED file directory of target genes connecting with foreground regions, which are derived from PECA model. Default: NULL (generated base on inputForegroundBed)

outputBackgroundBed
: Character scalar. The BED file directory of target genes connecting with background regions, which are derived from PECA model. Default: NULL (generated base on inputBackgroundBed)

regularGeneCorrBed
: Character scalar. The BED file directory of target genes which are predicted from PECA. Default: NULL (e.g. after library(enrichTF), you can call function setGenome("hg19"))

enhancerRegularGeneCorrBed
: Character scalar. The BED file directory of enhancer-targets predicted from PECA. Default: NULL (e.g. after library(enrichTF), you can call function setGenome("hg19"))

ouputForgroundGeneTxt
: Character scalar. The TXT file directory of target genes list connecting with foreground regions, which are derived from PECA model. Default: NULL (generated base on inputForegroundBed)

...
: Additional arguments, currently unused.

Details

Connect foreground and background regions to target genes, which are predicted from PECA.

Value

An invisible EnrichStep-class object (Step-class based) scalar for downstream analysis.

Author(s)

Zheng Wei

See Also

genBackground findMotifsInRegions tfsEnrichInRegions

Examples

```
setGenome("testgenome") #Use "hg19","hg38",etc. for your application
foregroundBedPath <- system.file(package = "enrichTF", "extdata","testregion.bed")
gen <- genBackground(inputForegroundBed = foregroundBedPath)
conTG <- enrichRegionConnectTargetGene(gen)
```

SingleSampleReport *Final report for single group of regions*

Description

When user call all steps in the pipeline, the final report can be generated.

Usage

```
enrichSingleSampleReport(prevStep, htmlOutput = NULL, ...)

## S4 method for signature 'Step'
enrichSingleSampleReport(prevStep, htmlOutput = NULL,
  ...)
```

Arguments

prevStep	Step-class object scalar. Any steps object in this package is acceptable when the pipeline is ready.
htmlOutput	Character scalar. HTML report file directory Default: NULL (generated base on bedInput)
...	Additional arguments, currently unused.

Details

The report is HTML format. All link in HTML file is the relative directory in report step folder and other step folder If user want to move HTML file and keep all link access available, they should move the whole pipeline folder at the same time.

Value

An invisible EnrichStep-class object (Step-class based) scalar for downstream analysis.

Author(s)

Zheng Wei

See Also

unzipAndMergeBed

TFsEnrichInRegions *Test each TF is enriched in regions or not*

Description

Test each TF is enriched in regions or not

Usage

```
enrichTFsEnrichInRegions(prevStep, inputRegionBed = NULL,
  inputForegroundGeneBed = NULL, inputBackgroundGeneBed = NULL,
  inputRegionMotifBed = NULL, outputTFsEnrichTxt = NULL,
  inputMotifWeights = NULL, inputTFgeneRelMtx = NULL,
  inputMotifTFTable = NULL, ...)

## S4 method for signature 'Step'
enrichTFsEnrichInRegions(prevStep,
  inputRegionBed = NULL, inputForegroundGeneBed = NULL,
```

```
  inputBackgroundGeneBed = NULL, inputRegionMotifBed = NULL,
  outputTFsEnrichTxt = NULL, inputMotifWeights = NULL,
  inputTFgeneRelMtx = NULL, inputMotifTFTable = NULL, ...)

tfsEnrichInRegions(inputRegionBed, inputForegroundGeneBed,
  inputBackgroundGeneBed, inputRegionMotifBed, outputTFsEnrichTxt = NULL,
  inputMotifWeights = NULL, inputTFgeneRelMtx = NULL,
  inputMotifTFTable = NULL, ...)
```

Arguments

prevStep	Step-class object scalar. This parameter is available when the upstream step function (printMap() to see the previous functions) have been sucessfully called. Accepted value can be the object return by any step function or be feed by %>% from last step function.
inputRegionBed	Character scalar. Directory of Regions BED file including foreground and background
inputForegroundGeneBed	Character scalar. Directory of BED file including foreground regions connected to related genes. The forth column is region ID
inputBackgroundGeneBed	Character scalar. Directory BED file including foreground regions connected to related genes. The forth column is region ID
inputRegionMotifBed	Character scalar. Directory BED file including foreground regions matched motifs. The forth column is region ID. The fifth column is motif calling score. The sixth column is motif name.
outputTFsEnrichTxt	Character scalar. Directory of Text result file with five columns. The first columns is transcription factor ,The second column is xxxx
inputMotifWeights	Character scalar. Directory of Text file contain motif weight. The first column is motif name. The second column is the weight. Default: NULL (if setGenome is called.)
inputTFgeneRelMtx	Character scalar. Directory of Text file contain a Transcription Factior(TF) and Gene relation weight matrix. Default: NULL (if setGenome is called.)
inputMotifTFTable	Character scalar. Directory of Text file contain Transcription Factior(TF) (the first column) and motif name(the second column). Default: NULL (if setGenome is called.)
...	Additional arguments, currently unused.

Details

Connect foreground and background regions to targetGene. If you only use this function without previous steps and you do not familiar with the data format of the input, you can run the example to see the example input from previous steps.

Value

An invisible EnrichStep-class object (Step-class based) scalar for downstream analysis.

Author(s)

Zheng Wei

See Also

genBackground findMotifsInRegions tfsEnrichInRegions

Examples

```
library(magrittr)
setGenome("testgenome") #Use "hg19","hg38",etc. for your application
foregroundBedPath <- system.file(package = "enrichTF", "extdata","testregion.bed")
gen <- genBackground(inputForegroundBed = foregroundBedPath)
conTG <- enrichRegionConnectTargetGene(gen)
findMotif <- enrichFindMotifsInRegions(gen,motifRc="integrate")
result <- enrichTFsEnrichInRegions(gen)

genBackground(inputForegroundBed = foregroundBedPath) %>%
    enrichRegionConnectTargetGene %>%
    enrichFindMotifsInRegions(motifRc="integrate") %>%
    enrichTFsEnrichInRegions
```

TissueOpennessConserve

Tissue's open conservation of the given region

Description

User provide region through a BED file. This function will provide tissue's open conservation analysis for these region.

Usage

```
enrichTissueOpennessConserve(prevStep, bedInput = NULL,
  openConserveBedInput = NULL, bedOutput = NULL,
  distrPdfOutput = NULL, ...)

## S4 method for signature 'Step'
enrichTissueOpennessConserve(prevStep, bedInput = NULL,
  openConserveBedInput = NULL, bedOutput = NULL,
  distrPdfOutput = NULL, ...)

tissueOpennessConserve(bedInput, openConserveBedInput = NULL,
  bedOutput = NULL, distrPdfOutput = NULL, ...)
```

Arguments

prevStep	Step-class object scalar. This parameter is available when the upstream step function (printMap() to see the previous functions) have been sucessfully called. Accepted value can be the object return by any step function or be feed by %>% from last step function.

bedInput	Character scalar. The directory of region BED file for analysis.
openConserveBedInput	Character scalar. The open level BED file for analysis. The first three columns are chromosome, start and end, The remaining columns are region name and conservation score.
bedOutput	Character scalar. The BED output file directory of merged BED files. Default: NULL (generated base on bedInput)
distrPdfOutput	Character scalar. The open conservation distribution figure for each tissue will be provided in PDF file.
...	Additional arguments, currently unused.

Details

We collected 201 DNase-seq or ATAC-seq sample from ENCODE and calculate their open level value. They can be download and install automatically. So users do not need to configure themselves.

Value

An invisible EnrichStep-class object (Step-class based) scalar for downstream analysis.

Author(s)

Zheng Wei

See Also

unzipAndMergeBed

Examples

```
foregroundBedPath <- system.file(package = "enrichTF", "extdata","testregion.bed")
tissueOpennessConserve(bedInput = foregroundBedPath)
```

TissueOpennessSpecificity

Tissue's open specificity of the given region

Description

User provide region through a BED file. This function will provide tissue's open specificity analysis for this region. Open level median, distribution and clustering result (heatmap) based on tissue and region will be provided.

Usage

```
enrichTissueOpennessSpecificity(prevStep, bedInput = NULL,
  openBedInput = NULL, sampleTxtInput = NULL, bedOutput = NULL,
  distPdfOutput = NULL, heatmapPdfOutput = NULL,
  sampleTxtOutput = NULL, ...)

## S4 method for signature 'Step'
enrichTissueOpennessSpecificity(prevStep,
  bedInput = NULL, openBedInput = NULL, sampleTxtInput = NULL,
  bedOutput = NULL, distPdfOutput = NULL, heatmapPdfOutput = NULL,
  sampleTxtOutput = NULL, ...)

tissueOpennessSpecificity(bedInput, openBedInput = NULL,
  sampleTxtInput = NULL, bedOutput = NULL, distPdfOutput = NULL,
  heatmapPdfOutput = NULL, sampleTxtOutput = NULL, ...)
```

Arguments

prevStep	Step-class object scalar. This parameter is available when the upstream step function (printMap() to see the previous functions) have been sucessfully called. Accepted value can be the object return by any step function or be feed by %>% from last step function.
bedInput	Character scalar. The directory of region BED file for analysis.
openBedInput	Character scalar. The open level BED file for analysis. The first three columns are chromosome, start and end, The remaining columns are the open level for each tissue. The order of tissue should be consistent with the order in the file provided by sampleTxtInput.
sampleTxtInput	Character scalar. The tissue sample information of in the file provided by openBedInput. There are 4 columns seperated by tab. The first column is the order number. The second column is the tissue detail information. The third column is the tissue name. The forth column is the code from source project like ENCODE
bedOutput	Character scalar. The BED output file directory of merged BED files. Default: NULL (generated base on bedInput)
distPdfOutput	Character scalar. The open level distribution figure for each tissue will be provided in PDF file. The order is strong to weak.
heatmapPdfOutput	Character scalar. The open level hiachical clustering heatmap base on region and tissue will be provided in this PDF file. The corresponding heatmap data will store at the same directory with suffix .Rdata
sampleTxtOutput	Character scalar. In this file, there are five columns seperated with tab. Fist four columns are the same with sampleTxtInput: The first column is the order number. The second column is the tissue detail information. The third column is the tissue name. The forth column is the code from source project like ENCODE The last column is the open level median level for each tissue. The table is in decreasing order of last column
...	Additional arguments, currently unused.

Details

We collect 201 DNase-seq or ATAC-seq sample from ENCODE and calculate their open level value. They can be download and install automatically. So users do not need to configure themselves.

Value

An invisible EnrichStep-class object (Step-class based) scalar for downstream analysis.

Author(s)

Zheng Wei

See Also

unzipAndMergeBed

Examples

```
foregroundBedPath <- system.file(package = "enrichTF", "extdata","testregion.bed")
tissueOpennessSpecificity(bedInput = foregroundBedPath)
```

UnzipAndMergeBed *Unzip all zipped BED files and merge them into one BED file*

Description

This function process region BED files in three step: First, unzip the gzip and bzip2 BED input files. Second, select first 3 columns of the BED files. Third, merge the BED files into one BED.

Usage

```
enrichUnzipAndMergeBed(prevStep, bedInput = NULL, bedOutput = NULL,
  ...)

## S4 method for signature 'Step'
enrichUnzipAndMergeBed(prevStep, bedInput = NULL,
  bedOutput = NULL, ...)

unzipAndMergeBed(bedInput, bedOutput = NULL, ...)
```

Arguments

prevStep	Step-class object scalar. It needs to be the return value of upstream process from other packages, such as ATAC-seq peak calling result from esATAC.
bedInput	Character scalar or vector. The directory of region BED files for analysis. BED, BED.gz, BED.bz2 formats are supported.
bedOutput	Character scalar. The BED output file directory of merged BED files. Default: NULL (generated base on first BED file in bedInput)
...	Additional arguments, currently unused.

Details

All compressed files will be de-compressed. Only first 3 columns (chromasomes, start and end) will be collected. All BED files will be merged into one BED file.

Value

An invisible EnrichStep-class object (inherit from Step-class) scalar for downstream analysis.

Author(s)

Zheng Wei

See Also

genBackground

Examples

```
foregroundBedPath <- system.file(package = "enrichTF", "extdata","testregion.bed.gz")
gen <- unzipAndMergeBed(bedInput = foregroundBedPath)
```

Index