Package 'enrichTF'

October 16, 2019

Type Package

Title Transcription Factors Enrichment Analysis

Version 1.0.0

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

Suggests knitr, magrittr, testthat

Collate EnrichStep.R ConnectTargetGene.R TFsEnrichInRegions.R
FindMotifsInRegions.R GenBackground.R onLoad.R untils.R
Method.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_9

git_last_commit 76268d9

git_last_commit_date 2019-05-02

Date/Publication 2019-10-15

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)
```

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(), ...)

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

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

Arguments

prevStep	Step-class object scalar. It needs to be the return value of upstream process from genBackground or enrichGenBackground when it is not used in a pipeline. If it is used in a pipeline or %>% is applied on this function, any steps in this package is acceptable.
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"))
...	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")
```

PECA_TF_enrich *TF enrichment with PECA model*

Description

This is a pipeline for TF enrichment with PECA model.

Usage

```
PECA_TF_enrich(inputForegroundBed, genome, threads = 2, ...)
```

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
...	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 contains all four steps EnrichTF objects

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, ...)

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

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

Arguments

prevStep	Step-class object scalar. It needs to be the return value of upstream process from genBackground or enrichGenBackground when it is not used in a pipeline. If it is used in a pipeline or %>% is applied on this function, any steps in this package is acceptable.
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")`)

`...`
: 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)
```

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

Description

Test each TF is enriched in regions or not

Usage

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

## S4 method for signature 'Step'
enrichTFsEnrichInRegions(enrichStep,
  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

enrichStep	Step-class object scalar. It has to be the return value of upstream process from regionConnectTargetGene, regionConnectTargetGene, findMotifsInRegions or enrichFindMotifsInRegions, If it is used in a pipeline or %>% is applied on this function, any steps in this package is acceptable.
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
```

Index