Package ‘ChIPseeker’

March 25, 2015

Type Package

Title ChIPseeker for ChIP peak Annotation, Comparison, and Visualization

Version 1.3.11

Author Guangchuang Yu

Maintainer Guangchuang Yu <guangchuangyu@gmail.com>

Description This package implements functions to retrieve the nearest genes around the peak, annotate genomic region of the peak, statistical methods for estimate the significance of overlap among ChIP peak data sets, and incorporate GEO database for user to compare the own dataset with those deposited in database. The comparison can be used to infer cooperative regulation and thus can be used to generate hypotheses. Several visualization functions are implemented to summarize the coverage of the peak experiment, average profile and heatmap of peaks binding to TSS regions, genomic annotation, distance to TSS, and overlap of peaks or genes.

Depends R (>= 3.1.0)

Imports BiocGenerics, AnnotationDbi, IRanges, GenomeInfoDb, GenomicRanges, GenomicFeatures, ggplot2, gplots, grDevices, gtools, methods, plotrix, dplyr, plyr, magrittr, RColorBrewer, rtracklayer, S4Vectors, TxDb.Hsapiens.UCSC.hg19.knownGene

Suggests clusterProfiler, DOSE, ReactomePA, org.Hs.eg.db, knitr, BiocStyle

URL https://github.com/GuangchuangYu/ChIPseeker

BugReports https://github.com/GuangchuangYu/ChIPseeker/issues

VignetteBuilder knitr

License Artistic-2.0

biocViews Annotation, ChIPSeq, Software, Visualization, MultipleComparison

NeedsCompilation no
R topics documented:

ChIPseeker-package ......................................................... 3
addGeneAnno ................................................................. 3
annotatePeak ................................................................. 4
as.data.frame.csAnno ......................................................... 6
as.GRanges ................................................................. 6
covplot ................................................................. 7
csAnno-class ................................................................. 8
downloadGEObedFiles ....................................................... 8
downloadGSMbedFiles ..................................................... 9
enrichAnnoOverlap ......................................................... 9
enrichPeakOverlap ......................................................... 10
genomicAnnotation ......................................................... 10
getGEOgenomeVersion ...................................................... 11
getGEOInfo ................................................................. 12
getGEOspecies ............................................................ 12
getNearestFeatureIndicesAndDistances ................................ 13
getPromoters ............................................................... 13
getSampleFiles ............................................................ 14
getTagMatrix ............................................................... 14
info ................................................................. 15
overlap ................................................................. 15
peakHeatmap ............................................................... 15
plotAnnoBar ............................................................... 16
plotAnnoBar.data.frame .................................................. 17
plotAnnoPie ............................................................... 18
plotAnnoPie.csAnno ......................................................... 19
plotAvgProf ............................................................... 20
plotAvgProf2 ............................................................. 20
plotChrCov ............................................................... 21
plotDistToTSS ............................................................ 22
plotDistToTSS.data.frame ................................................ 23
readPeakFile .............................................................. 24
show ................................................................. 24
shuffle ................................................................. 25
tagHeatmap ............................................................... 25
vennpie ................................................................. 26
vennplot ................................................................. 27
vennplot.peakfile ........................................................ 27

Index .......................... 29
ChIPseeker-package  

*ChIP-SEQ Annotation, Visualization and Comparison*

**Description**

This package is designed for chip-seq data analysis

**Details**

- **Package:** ChIPseeker
- **Type:** Package
- **Version:** 1.4.0
- **Date:** 2-01-2014
- **biocViews:** ChIPSeq, Annotation, Software
- **Depends:**
- **Imports:** methods, ggplot2
- **Suggests:** clusterProfiler, GOSemSim
- **License:** Artistic-2.0

**Author(s)**

Guangchuang Yu

Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>

---

**addGeneAnno**

**Description**

add gene annotation, symbol, gene name etc.

**Usage**

`addGeneAnno(annoDb, geneID, type)`

**Arguments**

- **annoDb**
  - annotation package
- **geneID**
  - query geneID
- **type**
  - gene ID type
Value
data.frame

Author(s)
G Yu

Description
Annotate peaks

Usage
annotatePeak(peak, tssRegion = c(-3000, 3000), TxDB = NULL,
level = "transcript", assignGenomicAnnotation = TRUE,
genomicAnnotationPriority = c("Promoter", "5UTR", "3UTR", "Exon", "Intron",
"Downstream", "Intergenic"), annoDb = NULL, addFlankGeneInfo = FALSE,
flankDistance = 5000, verbose = TRUE)

Arguments
peak peak file or GRanges object
tssRegion Region Range of TSS
TxDB TxDB object
level one of transcript and gene
assignGenomicAnnotation logical, assign peak genomic annotation or not
genomicAnnotationPriority genomic annotation priority
annoDb annotation package
addFlankGeneInfo logical, add flanking gene information from the peaks
flankDistance distance of flanking sequence
verbose print message or not
Value

data.frame or GRanges object with columns of:

all columns provided by input.

annotation: genomic feature of the peak, for instance if the peak is located in 5’UTR, it will annotated by 5’UTR. Possible annotation is Promoter-TSS, Exon, 5’ UTR, 3’ UTR, Intron, and Intergenic.

geneChr: Chromosome of the nearest gene
geneStart: gene start
geneEnd: gene end
geneLength: gene length
geneStrand: gene strand
geneId: entrezgene ID
distanceToTSS: distance from peak to gene TSS

if annoDb is provided, extra column will be included:

ENSEMBL: ensembl ID of the nearest gene
SYMBOL: gene symbol
GENENAME: full gene name

Author(s)

G Yu

See Also

plotAnnoBar plotAnnoPie plotDistToTSS

Examples

## Not run:
require(TxDb.Hsapiens.UCSC.hg19.knownGene)
txdb <- TxDb.Hsapiens.UCSC.hg19.knownGene
peakfile <- system.file("extdata", "sample_peaks.txt", package="ChIPseeker")
peakAnno <- annotatePeak(peakfile, tssRegion=c(-3000, 3000), TxDb=txdb)
peakAnno

## End(Not run)
### Description
as.data.frame.csAnno

### Usage
```r
## S3 method for class 'csAnno'
as.data.frame(x, row.names = NULL, optional = FALSE, ...)
```

### Arguments
- `x`: csAnno object
- `row.names`: row names
- `optional`: should be omitted.
- `...`: additional parameters

### Value
data.frame

### Author(s)
Guangchuang Yu [http://ygc.name](http://ygc.name)

---

### Description
as.GRanges

### Usage
```r
as.GRanges(x)
```

### Arguments
- `x`: csAnno object

### Value
GRanges object
Description
plot peak coverage

Usage
covplot(peak, weightCol = NULL, xlab = "Chromsome Size (bp)", ylab = "", title = "ChIP Peaks over Chromosomes", chrs = NULL, xlim = NULL)

Arguments
peak peak file or GRanges object
weightCol weight column of peak
xlab xlab
ylab ylab
title title
chrs selected chromosomes to plot, all chromosomes by default
xlim ranges to plot, default is whole chromosome

Value
ggplot2 object

Author(s)
G Yu
Class csAnno

Description

Class "csAnno" This class represents the output of ChIPseeker Annotation

Slots

anno annotation
tssRegion TSS region
level transcript or gene
detailGenomicAnnotation Genomic Annotation in detail
annoStat annotation statistics
peakNum number of peaks

Author(s)

Guangchuang Yu http://ygc.name

See Also

annotatePeak

downloadGEObedFiles

download all BED files of a particular genome version

Usage

downloadGEObedFiles(genome, destDir = getwd())

Arguments

genome genome version
destDir destination folder

Author(s)

G Yu
**downloadGSMbedFiles**  

**Description**  

Download BED supplementary files of a list of GSM accession numbers.

**Usage**  

```r  
downloadGSMbedFiles(GSM, destDir = getwd())  
```

**Arguments**

- `GSM`: GSM accession numbers.
- `destDir`: Destination folder.

**Author(s)**

G Yu

---

**enrichAnnoOverlap**  

**Description**  

Calculate overlap significance of ChIP experiments based on their nearest gene annotation.

**Usage**  

```r  
enrichAnnoOverlap(queryPeak, targetPeak, TxDB = NULL, pAdjustMethod = "BH", chainFile = NULL)  
```

**Arguments**

- `queryPeak`: Query bed file.
- `targetPeak`: Target bed file(s) or folder containing bed files.
- `TxDB`: TxDB object.
- `pAdjustMethod`: P-value adjustment method.
- `chainFile`: Chain file for liftOver.

**Value**

Data frame.

**Author(s)**

G Yu
enrichPeakOverlap  

**Description**

calculate overlap significant of ChIP experiments based on the genome coordinations

**Usage**

```r
enrichPeakOverlap(queryPeak, targetPeak, TxDB = NULL, pAdjustMethod = "BH", nShuffle = 1000, chainFile = NULL)
```

**Arguments**

- **queryPeak**: query bed file
- **targetPeak**: target bed file(s) or folder that containing bed files
- **TxDB**: TxDb
- **pAdjustMethod**: pvalue adjustment method
- **nShuffle**: shuffle numbers
- **chainFile**: chain file for liftOver

**Value**

data.frame

**Author(s)**

G Yu

---

getGenomicAnnotation  

**Description**

get Genomic Annotation of peaks

**Usage**

```r
getGenomicAnnotation(peaks, distance, tssRegion = c(-3000, 3000), TxDB, level, genomicAnnotationPriority)
```
**getGEOgenomeVersion**

**Arguments**
- **peaks**  peaks in GRanges object
- **distance**  distance of peak to TSS
- **tssRegion**  tssRegion, default is -3kb to +3kb
- **TxDb**  TxDb object
- **level**  one of gene or transcript
- **genomicAnnotationPriority**  genomic Annotation Priority

**Value**
- character vector

**Author(s)**
- G Yu

---

**Description**
- get genome version statistics collecting from GEO ChIPseq data

**Usage**
- `getGEOgenomeVersion()`

**Value**
- data.frame

**Author(s)**
- G Yu
getGEOInfo

Description
get subset of GEO information by genome version keyword

Usage
getGEOInfo(genome, simplify = TRUE)

Arguments
- genome: genome version
- simplify: simplify result or not

Value
data.frame

Author(s)
G Yu

getGEOspecies

Description
accessing species statistics collecting from GEO database

Usage
getGEOspecies()

Value
data.frame

Author(s)
G Yu
getNearestFeatureIndicesAndDistances

Description
get index of features that closest to peak and calculate distance

Usage
getNearestFeatureIndicesAndDistances(peaks, features)

Arguments
peaks      peak in GRanges
features   features in GRanges

Value
list

Author(s)
G Yu

getPromoters

Description
prepare the promoter regions

Usage
getPromoters(TxDb = NULL, upstream = 1000, downstream = 1000,
by = “gene”)

Arguments
TxDb         TxDb
upstream     upstream from TSS site
downstream   downstream from TSS site
by           one of gene or transcript

Value
GRanges object
**getSampleFiles**

**Description**
get filenames of sample files

**Usage**
getSampleFiles()

**Value**
list of file names

**Author(s)**
G Yu

---

**getTagMatrix**

**Description**
calculate the tag matrix

**Usage**
getTagMatrix(peak, weightCol = NULL, windows)

**Arguments**
- peak: peak file or GRanges object
- weightCol: column name of weight, default is NULL
- windows: a collection of region with equal size, eg. promoter region.

**Value**
tagMatrix
**Information Datasets**

**Description**

ucsc genome version, precalculated data and gsm information

**Overlap**

**Description**

calculate the overlap matrix, which is useful for vennplot

**Usage**

`overlapHsetsI

Arguments**

Sets a list of objects

**Value**

data.frame

**Author(s)**

G Yu

**PeakHeatmap**

**Description**

plot the heatmap of peaks align to flank sequences of TSS

**Usage**

`peakHeatmap(peak, weightCol = NULL, TxDB = NULL, upstream = 1000, downstream = 1000, xlab = "", ylab = "", title = NULL, color = NULL, verbose = TRUE)"
Arguments

peak         peak file or GRanges object
weightCol    column name of weight
TxDb          TxDb object
upstream     upstream position
downstream   downstream position
xlab          xlab
ylab          ylab
title         title
color         color
verbose       print message or not

Value

figure

Author(s)

G Yu

plotAnnoBar method generics

Description

plotAnnoBar method generics
plotAnnoBar method for list of csAnno instances
plotAnnoBar method for csAnno instance

Usage

plotAnnoBar(x, xlab = "", ylab = "Percentage(%)",
            title = "Feature Distribution", ...)

## S4 method for signature 'list'
plotAnnoBar(x, xlab = "", ylab = "Percentage(%)",
            title = "Feature Distribution", ...)

plotAnnoBar(x, xlab="", ylab='Percentage(%)',title="Feature Distribution", ...)
Arguments

- x: csAnno instance
- xlab: xlab
- ylab: ylab
- title: title
- ...: additional parameter

Value

- plot

Author(s)

Guangchuang Yu [http://ygc.name](http://ygc.name)

Description

Plot feature distribution based on their chromosome region

Usage

```r
plotAnnoBar.data.frame(anno.df, xlab = "", ylab = "Percentage(%)", title = "Feature Distribution", categoryColumn)
```

Arguments

- anno.df: annotation stats
- xlab: xlab
- ylab: ylab
- title: plot title
- categoryColumn: category column

Details

Plot chromosome region features

Value

Bar plot that summarize genomic features of peaks
Author(s)

Guangchuang Yu http://ygc.name

See Also

annotatePeak plotAnnoPie

-----------------

plotAnnoPie    plotAnnoPie method generics

Description

plotAnnoPie method generics
plotAnnoPie method for csAnno instance

Usage

plotAnnoPie(x, ndigit = 2, cex = 0.9, col = NA,
   legend.position = "rightside", pie3D = FALSE, ...)

   plotAnnoPie(x, ndigit = 2, cex = 0.9, col = NA,
   legend.position = "rightside", pie3D = FALSE, ...)

Arguments

  x            csAnno instance
  ndigit       number of digit to round
  cex          label cex
  col          color
  legend.position  topright or other.
  pie3D        plot in 3D or not
  ...          extra parameter

Value

  plot

Author(s)

Guangchuang Yu http://ygc.name
plotAnnoPie.csAnno  

Description

Pie plot from peak genomic annotation

Usage

```r
plotAnnoPie.csAnno(x, ndigit = 2, cex = 0.9, col = NA,
                  legend.position = "rightside", pie3D = FALSE, ...)
```

Arguments

- `x`: csAnno object
- `ndigit`: number of digit to round
- `cex`: label cex
- `col`: color
- `legend.position`: top right or other.
- `pie3D`: plot in 3D or not
- `...`: extra parameter

Value

Pie plot of peak genomic feature annotation

Author(s)

G Yu

See Also

annotatePeak plotAnnoBar

Examples

```r
## Not run:
require(TxDb.Hsapiens.UCSC.hg19.knownGene)
txdb <- TxDb.Hsapiens.UCSC.hg19.knownGene
peakfile <- system.file("extdata", "sample_peaks.txt", package="chipseeker")
peakAnno <- annotatePeak(peakfile, TxDb=txdb)
plotAnnoPie(peakAnno)
```

## End(Not run)
plotAvgProf

Description
plot the profile of peaks

Usage
plotAvgProf(tagMatrix, xlim, xlab = "Genomic Region (5'->3')", ylab = "Read Count Frequency")

Arguments
- tagMatrix
- xlim
- xlab
- ylab

Value
ggplot object

Author(s)
G Yu

plotAvgProf2

Description
plot the profile of peaks that align to flank sequences of TSS

Usage
plotAvgProf2(peak, weightCol = NULL, TxDB = NULL, upstream = 1000, downstream = 1000, xlab = "Genomic Region (5'->3')", ylab = "Read Count Frequency", verbose = TRUE)
plotChrCov

Arguments
- peak: peak file or GRanges object
- weightCol: column name of weight
- TxDB: TxDB object
- upstream: upstream position
- downstream: downstream position
- xlab: xlab
- ylab: ylab
- verbose: print message or not

Value
- ggplot object

Author(s)
- G Yu

Description
plot the Peak Regions over Chromosomes

Usage
plotChrCov(peak, weightCol = NULL, xlab = "Chromosome Size (bp)", ylab = "", title = "ChIP Peaks over Chromosomes")

Arguments
- peak: peak file or GRanges object
- weightCol: weight column of peak
- xlab: xlab
- ylab: ylab
- title: title

Value
- ggplot2 object

Author(s)
- G Yu
plotDistToTSS method generics

### Description

plotDistToTSS method generics
plotDistToTSS method for list of csAnno instances
plotDistToTSS method for csAnno instance

### Usage

```r
plotDistToTSS(x, distanceColumn = "distanceToTSS", xlab = "", ylab = "Binding sites (%) (5'->3')", title = "Distribution of transcription factor-binding loci relative to TSS", ...)
```

```r
## S4 method for signature 'list'
plotDistToTSS(x, distanceColumn = "distanceToTSS", xlab = "", ylab = "Binding sites (%) (5'->3')", title = "Distribution of transcription factor-binding loci relative to TSS", ...)
```

```r
plotDistToTSS(x, distanceColumn="distanceToTSS", xlab="", ylab="Binding sites (%) (5'->3')", title="Distribution of transcription factor-binding loci relative to TSS",...)
```

### Arguments

- `x`: csAnno instance
- `distanceColumn`: distance column name
- `xlab`: xlab
- `ylab`: ylab
- `title`: title
- `...`: additional parameter

### Value

plot

### Author(s)

Guangchuang Yu [http://ygc.name](http://ygc.name)
plotDistToTSS.data.frame

Description

plot feature distribution based on the distances to the TSS

Usage

plotDistToTSS.data.frame(peakDist, distanceColumn = "distanceToTSS",
                         xlab = "", ylab = "Binding sites (%) (5'->3')",
                         title = "Distribution of transcription factor-binding loci relative to TSS",
                         categoryColumn)

Arguments

peakDist        peak annotation
distanceColumn  column name of the distance from peak to nearest gene
xlab            x label
ylab            y label
title           figure title
categoryColumn  category column

Value

bar plot that summarize distance from peak to TSS of the nearest gene.

Author(s)

Guangchuang Yu http://ygc.name

See Also

annotatePeak

Examples

## Not run:
require(TxDb.Hsapiens.UCSC.hg19.knownGene)
txdb <- TxDb.Hsapiens.UCSC.hg19.knownGene
peakfile <- system.file("extdata", "sample_peaks.txt", package="ChIPseeker")
peakAnno <- annotatePeak(peakfile, TxDb=txdb)
plotDistToTSS(peakAnno)

## End(Not run)
Description

read peak file and store in data.frame or GRanges object

Usage

readPeakFile(peakfile, as = "GRanges")

Arguments

peakfile  peak file
as         output format, one of GRanges or data.frame

Value

peak information, in GRanges or data.frame object

Author(s)

G Yu

Examples

peakfile <- system.file("extdata", "sample_peaks.txt", package="ChIPseeker")
peak.gr <- readPeakFile(peakfile, as = "GRanges")
peak.gr

Description

show method for csAnno instance

Usage

show(object)

Arguments

object  A csAnno instance
**shuffle**

**Value**

message

**Author(s)**

Guangchuang Yu [http://ygc.name](http://ygc.name)

---

**shuffle**

**Description**

shuffle the position of peak

**Usage**

shuffle(peak.gr, TxDb)

**Arguments**

<table>
<thead>
<tr>
<th>peak.gr</th>
<th>GRanges object</th>
</tr>
</thead>
<tbody>
<tr>
<td>TxDb</td>
<td>TxDb</td>
</tr>
</tbody>
</table>

**Value**

GRanges object

**Author(s)**

G Yu

---

**tagHeatmap**

**Description**

plot the heatmap of tagMatrix

**Usage**

tagHeatmap(tagMatrix, xlim = "", ylab = "", title = NULL, color = "red")
vennpie

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>tagMatrix</td>
<td>tagMatrix or a list of tagMatrix</td>
</tr>
<tr>
<td>xlim</td>
<td>xlim</td>
</tr>
<tr>
<td>xlab</td>
<td>xlab</td>
</tr>
<tr>
<td>ylab</td>
<td>ylab</td>
</tr>
<tr>
<td>title</td>
<td>title</td>
</tr>
<tr>
<td>color</td>
<td>color</td>
</tr>
</tbody>
</table>

Value

figure

Author(s)

G Yu

Description

vennpie method generics
vennpie method generics

Usage

vennpie(x, r = 0.2, ...)

vennpie(x, r=0.2, ...)

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>A csAnno instance</td>
</tr>
<tr>
<td>r</td>
<td>initial radius</td>
</tr>
<tr>
<td>...</td>
<td>additional parameter</td>
</tr>
</tbody>
</table>

Value

plot

Author(s)

Guangchuang Yu [http://ygc.name](http://ygc.name)
vennplot

---

**Description**

plot the overlap of a list of object

**Usage**

```r
vennplot(Sets, by = "gplots")
```

**Arguments**

- `Sets`: a list of object, can be vector or GRanges object
- `by`: one of gplots or Vennerable

**Value**

venn plot that summarize the overlap of peaks from different experiments or gene annotation from different peak files.

**Author(s)**

G Yu

**Examples**

```r
## example not run
## require(TxDb.Hsapiens.UCSC.hg19.knownGene)
## txdb <- TxDb.Hsapiens.UCSC.hg19.knownGene
## peakfiles <- getSampleFiles()
## peakAnnoList <- lapply(peakfiles, annotatePeak)
## names(peakAnnoList) <- names(peakfiles)
## genes= lapply(peakAnnoList, function(i) as.data.frame(i)$geneId)
## vennplot(genes)
```

---

vennplot.peakfile

---

**Description**

vennplot for peak files

**Usage**

```r
vennplot.peakfile(files, labels = NULL)
```
Arguments

files  peak files
labels  labels for peak files

Value

figure

Author(s)

G Yu
Index

*Topic **classes**
csAnno-class, 8

*Topic **datasets**
info, 15

*Topic **package**
ChIPseeker-package, 3

addGeneAnno, 3
annotatePeak, 4, 8, 18, 19, 23
as.data.frame.csAnno, 6
as.GRanges, 6

ChIPseeker (ChIPseeker-package), 3
ChIPseeker-package, 3
covplot, 7
csAnno-class, 8

download GEO bedFiles, 8
download GSM bedFiles, 9

enrichAnnoOverlap, 9
enrichPeakOverlap, 10

getGenomicAnnotation, 10
get GEO genomeVersion, 11
get GEO Info, 12
get GEO species, 12
get Nearest Feature Indices And Distances, 13
get Promoters, 13
get Sample Files, 14
get Tag Matrix, 14
gsminfo (info), 15

info, 15

overlap, 15

peakHeatmap, 15
plot Anno Bar, 5, 16, 19
plot Anno Bar, csAnno, ANY-method
(plot Anno Bar), 16
plot Anno Bar, csAnno-method
(csAnno-class), 8
plot Anno Bar, list-method (plot Anno Bar), 16
plot Anno Bar.data.frame, 17
plot Anno Pie, 5, 18, 18
plot Anno Pie, csAnno, ANY-method
(plot Anno Pie), 18
plot Anno Pie, csAnno-method
(csAnno-class), 8
plot Anno Pie.csAnno, 19
plot AvgProf, 20
plot AvgProf2, 20
plot Chr Cov, 21
plot Dist To TSS, 5, 22
plot Dist To TSS, csAnno, ANY-method
(plot Dist To TSS), 22
plot Dist To TSS, csAnno-method
(csAnno-class), 8
plot Dist To TSS, list-method
(plot Dist To TSS), 22
plot Dist To TSS.data.frame, 23

read Peak File, 24

show, 24
show, csAnno, ANY-method (show), 24
show, csAnno-method (csAnno-class), 8
shuffle, 25

tag Heatmap, 25
tag Matrix List (info), 15

tag Pie, 26
tag Pie, csAnno-method (csAnno-class), 8
tag Plot, 27
tag Plot Peak File, 27

ucsc release (info), 15

venn pie, 26
venn pie, csAnno-method (csAnno-class), 8
venn plot, 27
venn plot peak file, 27