Package ‘genomation’

May 17, 2024

Type Package
Title Summary, annotation and visualization of genomic data
Version 1.36.0
Author Altuna Akalin [aut, cre], Vedran Franke [aut, cre], Katarzyna Wreczycka
[aut], Alexander Gosdschan [ctb], Liz Ing-Simmons [ctb], Bozena Mika-Gospodorz [ctb]
Maintainer Altuna Akalin <aakalin@gmail.com>, Vedran Franke
<vedran.franke@gmail.com>, Katarzyna Wreczycka <katwre@gmail.com>
Description A package for summary and annotation of genomic intervals. Users
can visualize and quantify genomic intervals over pre-defined functional
regions, such as promoters, exons, introns, etc. The genomic intervals
represent regions with a defined chromosome position, which may be associated
with a score, such as aligned reads from HT-seq experiments, TF binding sites,
methylation scores, etc. The package can use any tabular genomic feature data
as long as it has minimal information on the locations of genomic intervals. In
addition, It can use BAM or BigWig files as input.
License Artistic-2.0
LazyLoad yes
VignetteBuilder knitr
biocViews Annotation, Sequencing, Visualization, CpGIsland
Encoding latin1
URL http://bioinformatics.mdc-berlin.de/genomation/
BugReports https://github.com/BIMSBbioinfo/genomation/issues
Depends R (>= 3.0.0), grid
Imports Biostrings (>= 2.47.6), BSgenome (>= 1.47.3), data.table,
GenomeInfoDb, GenomicRanges (>= 1.31.8), GenomicAlignments (>=
1.15.6), S4Vectors (>= 0.17.25), ggplot2, gridBase, impute,
IRanges (>= 2.13.12), matrixStats, methods, parallel, plotrix,
plyr, readr, reshape2, Rsamtools (>= 1.31.2), seqPattern,
rtracklayer (>= 1.39.7), Rcpp (>= 0.12.14)
Suggests BiocGenerics, genomationData, knitr, RColorBrewer, rmarkdown,
RUnit
Collate 'combineScoreMatrixList.R' 'documentData.R'
  'genomation-classes.R' 'getRandomEnrichment.R'
  'findFeatureComb.R' 'patternMatrix.R' 'plotMatrix.R'
  'randomizeFeature.R' 'readAnnotate.R' 'readData.R'
  'scoreMatrix.R' 'scoreMatrixBin.R' 'scoreMatrixList.R' 'Ops.R'
  'test_genomation_package.R' 'deprecated_defunct.R'
  'enrichmentMatrix.R' 'RcppExports.R'

LinkingTo Rcpp

RoxygenNote 6.0.1.9000

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

git_branch RELEASE_3_19

git_last_commit 8a85f98

git_last_commit_date 2024-04-30

Repository Bioconductor 3.19

Date/Publication 2024-05-17

Contents

annotateWithFeature .................................................. 4
annotateWithFeatureFlank ................................................ 5
annotateWithFeatures .................................................... 6
annotateWithGeneParts .................................................. 7
annotatGrWithGeneParts ................................................ 8
AnnotationByFeature-class .............................................. 9
AnnotationByGeneParts-class ........................................... 9
bed12ToExons ............................................................ 10
bed12ToIntrons ........................................................... 10
binMatrix ................................................................. 10
binMax ...................................................................... 11
binMean .................................................................... 12
binMedian ................................................................. 12
binMin ...................................................................... 12
binner ...................................................................... 13
binSum ...................................................................... 13
c.ScoreMatrix ............................................................ 14
c.ScoreMatrixList ......................................................... 14
cage .......................................................................... 15
calculateOverlapSignificance .......................................... 15
checkBedValidity .......................................................... 16
checkClass ................................................................. 16
compressedAndUrl2temp .................................................. 17
constrainRanges ........................................................... 17
convertBed2Exons .......................................................... 18
convertBed2Introns ......................................................... 18
convertBedDf ............................................................... 19
**annotateWithFeature**

*Function to annotate given GRanges object with a given genomic feature*

**Description**

Function to annotate given GRanges object with a given genomic feature

**Usage**

```r
annotateWithFeature(target, feature, strand = FALSE, extend = 0,
                     feature.name = NULL, intersect.chr = FALSE)
```

```r
# S4 method for signature 'GRanges,GRanges'
annotateWithFeature(target, feature,
                     strand = FALSE, extend = 0, feature.name = NULL,
                     intersect.chr = FALSE)
```

**Arguments**

- `target` (a GRanges object storing chromosome locations to be annotated)
- `feature` (a GRanges object storing chromosome locations of a feature (can be CpG islands, ChIP-seq peaks, etc))
- `strand` (If set to TRUE, annotation features and target features will be overlapped based on strand (def:FAULT))
- `extend` (specifying a positive value will extend the feature on both sides as much as..."
### annotateWithFeatureFlank

Function to annotate a given GRanges object with promoter, exon, intron & intergenic values

#### Description

Function to annotate a given GRanges object with promoter, exon, intron & intergenic values

#### Usage

```r
annotateWithFeatureFlank(target, feature, flank, feature.name = NULL,
strand = FALSE, intersect.chr = FALSE)
```

#### Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>target</td>
<td>a granges object storing chromosome locations to be annotated</td>
</tr>
<tr>
<td>feature</td>
<td>a granges object storing chromosome locations of a feature (can be CpG islands, ChIP-seq peaks, etc)</td>
</tr>
<tr>
<td>flank</td>
<td>a granges object storing chromosome locations of the flanks of the feature</td>
</tr>
<tr>
<td>feature.name</td>
<td>string for the name of the feature</td>
</tr>
<tr>
<td>flank.name</td>
<td>string for the name of the flanks</td>
</tr>
<tr>
<td>strand</td>
<td>If set to TRUE, annotation features and target features will be overlapped based on strand (def:FALSE)</td>
</tr>
<tr>
<td>intersect.chr</td>
<td>boolean, whether to select only chromosomes that are common to feature and target. FALSE by default</td>
</tr>
</tbody>
</table>
Value

returns an AnnotationByFeature object

Examples

data(cpgi)
data(cage)
cpgi.flanks = getFlanks(cpgi)
flank.annot = annotateWithFeatureFlank(cage, cpgi, cpgi.flanks)

annotateWithFeatures
Annotate given ranges with genomic features

Description

The function annotates a target GRangesList or GRanges object as overlapping or not with the elements of a named GRangesList. This is useful to annotate your regions of interest with genomic features with arbitrary categories such as repeat classes or families, or output from genome segmentation algorithms such as chromHMM.

Usage

annotateWithFeatures(target, features, strand.aware = FALSE, intersect.chr = FALSE)

## S4 method for signature 'GRanges,GRangesList'
annotateWithFeatures(target, features, strand.aware = FALSE, intersect.chr = FALSE)

## S4 method for signature 'GRangesList,GRangesList'
annotateWithFeatures(target, features, strand.aware = FALSE, intersect.chr = FALSE)

Arguments

target GRanges or GRangesList object storing chromosome locations to be annotated (e.g. chipseq peaks)

features a named GRangesList object containing GRanges objects different set of features. The function calculates percent overlaps with and without precedence at the same time. The order of objects in GRangesList defines their precedence. If a range in target overlaps with a more precedent range in an element of features, the other overlaps from other less precedent elements will be discarded. This is useful for getting piecharts where percentages should add up to 100.

strand.aware if set to TRUE, annotation features and target features will be overlapped based on strand (def:FALSE)
annotateWithGeneParts

interact.chr logical value, whether to select only chromosomes that are common to feature and target. FALSE by default

Value

returns an AnnotationByFeature object or if target is a GRangesList, a list of AnnotationByFeature objects.

See Also

see getMembers, heatTargetAnnotation, plotTargetAnnotation

Examples

library(GenomicRanges)
data(cage)
data(cpgi)
cage$tpm=NULL
gl = GRangesList(cage=cage, cpgi=cpgi)

bed.file = system.file("extdata/chr21.refseq.hg19.bed", package = "genomation")
gene.parts = readTranscriptFeatures(bed.file)
annot = annotateWithFeatures(gl, gene.parts, intersect.chr=TRUE)
annotatGrWithGeneParts

Description

annotatGrWithGeneParts function

Usage

annotatGrWithGeneParts(gr, prom, exon, intron, strand = FALSE)

Arguments

gr  target object
prom  promotors
exon  exons
intron  introns
strand  logical
**AnnotationByFeature-class**

*An S4 class that information on overlap of target features with annotation features*

---

**Description**

This object is designed to hold statistics and information about genomic feature overlaps.

**Slots**

- `members` a matrix showing overlap of target features with annotation genomic features
- `annotation` a named vector of percentages
- `precedence` a named vector of percentages
- `num.annotation` vector
- `num.precedence` vector
- `no.of.OlapFeat` vector
- `perc.of.OlapFeat` vector

---

**AnnotationByGeneParts-class**

*An S4 class that information on overlap of target features with annotation features*

---

**Description**

This object is designed to hold statistics and information about genomic feature overlaps.

**Slots**

- `members` a matrix showing overlap of target features with annotation genomic features
- `annotation` a named vector of percentages
- `precedence` a named vector of percentages
- `num.annotation` vector
- `num.precedence` vector
- `no.of.OlapFeat` vector
- `perc.of.OlapFeat` vector
- `dist.to.TSS` a data frame showing distances to TSS and gene/TSS names and strand
binMatrix

Description
Bins the columns of a matrix using a user provided function

Usage
binMatrix(x, bin.num = NULL, fun = "mean")

## S4 method for signature 'ScoreMatrix'
binMatrix(x, bin.num = NULL, fun = "mean")

## S4 method for signature 'ScoreMatrixList'
binMatrix(x, bin.num = NULL, fun = "mean")

bed12ToExonsfunction

Description
extracts exons from a bed12 file and puts them into GRanges object

Usage
bed12ToExons(ref)

Arguments
ref data.frame object

bed12ToIntronsfunction

Description
extracts introns from a bed12 file and puts them into GRanges object

Usage
bed12ToIntrons(ref)

Arguments
ref data.frame object
binMax

Arguments

- **x**: ScoreMatrix or a ScoreMatrixList object
- **bin.num**: integer number of bins in the final matrix
- **fun**: character vector or an anonymous function that will be used for binning

Value

ScoreMatrix or ScoreMatrixList object

Examples

```r
# binning the columns in a ScoreMatrix object
library(GenomicRanges)
target = GRanges(rep(c(1,2),each=7), IRange(rep(c(1,1,2,3,7,8,9), times=2), width=5),
weight = rep(c(1,2),each=7),
strand=c('-','-','-', '+','-', '+', '-','-', '-','-', ' '+))
windows = GRanges(rep(c(1,2),each=2), IRange(rep(c(1,2), times=2), width=5),
strand=c('-','+','-','+'))
sm = ScoreMatrix(target, windows)
bin = binMatrix(sm, bin.num=2)
```

---

**Description**

Function that computes a maximum value for each bin

**Usage**

`binMax(x, n)`

**Arguments**

- **x**: NumericVector
- **n**: integer - number of bins
binMean

Function that computes a mean value for each bin

Description
Function that computes a mean value for each bin

Usage
binMean(x, n)

Arguments

x NumericVector
n integer - number of bins

binMedian

Function that computes a median value for each bin

Description
Function that computes a median value for each bin

Usage
binMedian(x, n)

Arguments

x NumericVector
n integer - number of bins

binMin

Function that computes a minimum value for each bin

Description
Function that computes a minimum value for each bin

Usage
binMin(x, n)

Arguments

x NumericVector
n integer - number of bins
**binner**

*binner function*

---

**Description**

given a vector and length smooths the vector to a given size

**Usage**

binner(start, end, nbins)

**Arguments**

<table>
<thead>
<tr>
<th>start</th>
<th>start position</th>
</tr>
</thead>
<tbody>
<tr>
<td>end</td>
<td>end position</td>
</tr>
<tr>
<td>nbins</td>
<td>number of bins</td>
</tr>
</tbody>
</table>

---

**binSum**

*Function that computes a sum of values in a bin*

---

**Description**

Function that computes a sum of values in a bin

**Usage**

binSum(x, n)

**Arguments**

<table>
<thead>
<tr>
<th>x</th>
<th>NumericVector - vector of values of a bin</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>integer - number of bins</td>
</tr>
</tbody>
</table>
**Description**

Combine a scoreMatrix into a scoreMatrixList object - when a ScoreMatrix is a first argument

**Usage**

```r
## S3 method for class 'ScoreMatrix'
c(..., recursive = FALSE, use.names = TRUE)
```

**Arguments**

- `...`: contains scoreMatrix and scoreMatrixList objects
- `recursive`: logical
- `use.names`: logical

**Value**

returns a scoreMatrixList object
Example CAGE data set.

**Description**

Location and tag per million values for CAGE TSS clusters on chr21 and chr22 of human genome (hg19 assembly). The clusters are downloaded from ENCODE project downloads for NHEK cells.

**Format**

`GRanges` object

---

**calculateOverlapSignificance**

*function that calculates the significance of overlaps of two sets of features using randomization*

---

**Description**

This function calculates the significance of overlaps of two sets of features using randomization. It returns a distribution of overlaps of a target set with a given randomized feature set. The randomization can be constrained by supplied arguments. The function is still in Beta mode - the regions can overlap excluded regions, and the randomized regions are not disjoint. Please take care that the excluded and included regions are not too strict when compared to the total width of the ranges.

**Usage**

```r
calculateOverlapSignificance(target, feature, chrom.sizes = NULL, stranded = TRUE, keep.strand.prop = TRUE, keep.chrom = TRUE, exclude = NULL, include = NULL, seed = NULL, nrand = 1)
```

```r
## S4 method for signature 'GRanges,GRanges'
calculateOverlapSignificance(target, feature, chrom.sizes = NULL, stranded = TRUE, keep.strand.prop = TRUE, keep.chrom = TRUE, exclude = NULL, include = NULL, seed = NULL, nrand = 1)
```

**Arguments**

- `target`: a GRanges object for which the overlap needs to be calculated
- `feature`: a GRanges object to be randomized
- `chrom.sizes`: sizes of chromosomes as a named vector (names are chromosomes names and elements of the vectors are lengths). If not given sizes in GRanges object will be used if no sizes there the end of each chr will be the end last feature on each chr
stranded  if FALSE, all of the returned features will be strandless (will have "*" in the strand slot)
keep.strand.prop  If TRUE strands will have the same proportion as the features
keep.chrom  If TRUE, number of features and randomized features for a chromosome will match. Currently setting this to FALSE is not supported.
exclude  A GRanges object where no randomized feature should overlap, can be gaps or unmappable regions in the genome as an example.
include  A GRanges object which defines the boundaries of randomized features
seed  random number generator seed
nrand  number of randomizations (default: 1)

Value
returns a GRanges object which is randomized version of the feature

cHECKBedValidity  checkBedValidity function

description
checks the validity of the bed data.frame if it is a legitimate bed columns

Usage
checkBedValidity(bed.df, type = "none")

Arguments
bed.df  data.frame object
type  type

cHECKClass  checkClass function

description
check whether the x object corresponds to the given class

Usage
checkClass(x, class.name, var.name = deparse(substitute(x)))
CompressedAndUrl2Temp

**Arguments**

- x: object
- class.name: class name
- var.name: uses x object

---

ConstrainRanges

**Description**

removes ranges that fell of the rle object does not check for the correspondence of the chromosome names - always check before using this function

**Usage**

constrainRanges(target, windows)

**Arguments**

- target: target file
- windows: windows
convertBed2Exons  

**Description**

convert a data frame read-in from a bed file to a GRanges object for exons

**Usage**

```r
convertBed2Exons(bed.df)
```

**Arguments**

- `bed.df`: a data.frame where column order and content resembles a bed file with 12 columns

**Value**

GRanges object

**Note**

one bed track per file is only accepted, the bed files with multiple tracks will cause an error

**Examples**

```r
file = system.file('extdata/chr21.refseq.hg19.bed', package='genomation')
bcd12 = read.table(file)
exons = convertBed2Exons(bed12)
head(exons)
```

convertBed2Introns  

**Description**

convert a data frame read-in from a bed file to a GRanges object for introns

**Usage**

```r
convertBed2Introns(bed.df)
```

**Arguments**

- `bed.df`: a data.frame where column order and content resembles a bed file with 12 columns

**Value**

GRanges object
## convertBedDf

### Usage

```r
convertBed2Introns(bed.df)
```

```r
## S4 method for signature 'data.frame'
convertBed2Introns(bed.df)
```

### Arguments

- `bed.df` a data.frame where column order and content resembles a bed file with 12 columns

### Value

`GRanges` object

### Note

one bed track per file is only accepted, the bed files with multiple tracks will cause en error

### Examples

```r
file = system.file('extdata/chr21.refseq.hg19.bed', package='genomation')
bed12 = read.table(file)
introns = convertBed2Introns(bed12)
head(introns)
```

---

## convertBedDf

### Description

convert a data frame read-in from a bed file to a GRanges object

### Usage

```r
convertBedDf(bed)
```

```r
## S4 method for signature 'data.frame'
convertBedDf(bed)
```

### Arguments

- `bed` a data.frame where column order and content resembles a bed file with 12 columns

### Value

`GRanges` object
Note
one bed track per file is only accepted, the bed files with multiple tracks will cause an error
bed files are expected not to have header lines

---

cpgi  
*Example CpG island data set.*

---

Description
CpG islands of hg19 assembly of human genome on chr21 and chr22. Downloaded from UCSC genome browser.

Format

*GRanges* object

---

detectUCSCheader  
*detectUCSCheader function*

---

Description
detects UCSC header (and first track)

Usage
detectUCSCheader(filename)

Arguments

| filename | file name |

---

distance2NearestFeature  
*distance2NearestFeature function*

---

Description
distance2NearestFeature function

Usage
distance2NearestFeature(g.idh, tss)

Arguments

| g.idh | target object |
| tss | TSSes |
enrichmentMatrix

Description

This is an enrichmentMatrix function for ScoreMatrix objects, that enables to normalize ChIP-seq signals with respect to IgG or input DNA control.

Usage

\S4method{enrichmentMatrix}{ScoreMatrix,ScoreMatrix}(IP, control)

Arguments

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP</td>
<td>ScoreMatrix object storing an IP sample</td>
</tr>
<tr>
<td>control</td>
<td>ScoreMatrix object storing a control sample</td>
</tr>
</tbody>
</table>

Value

ScoreMatrix object

Note

The function computes an enrichment of IP over control as follow: Suppose both IP and control are ScoreMatrix objects that have same dimensions. Then, the enrichment is calculated using a formula: \( \log_2 \left( \frac{IP + 1}{control + 1} \right) \).

See Also

ScoreMatrix

Examples

#load IP and control BAM files and create ScoreMatrix objects
class('library')('genomationData')
bam.file_IP <- system.file("extdata", "wgEncodeBroadHistoneH1hescSuz12851317AlnRep1.chr21.bam", package = "genomationData")
bam.file_c <- system.file("extdata", "wgEncodeBroadHistoneH1hescCtcfStdAlnRep1.chr21.bam", package = "genomationData")
data(promoters)
IP <- ScoreMatrix(target = bam.file_IP, windows = promoters, type = 'bam')
control <- ScoreMatrix(target = bam.file_c, windows = promoters, type = 'bam')

# compute an enrichment of IP over control
enrichmentMatrix(IP, control)
Compute an enrichment of IP (stored in ScoreMatrixList object) over control (stored in ScoreMatrix object)

Description

This is an enrichmentMatrix function for IP ScoreMatrixList object and control ScoreMatrix object, that enables to normalize ChIP-seq signals with respect to IgG or input DNA control.

Usage

\S4method{enrichmentMatrix}{ScoreMatrixList,ScoreMatrix}(IP, control)

Arguments

- **IP**: ScoreMatrixList object storing IP samples
- **control**: ScoreMatrix storing control sample

Value

ScoreMatrixList object

Note

The function computes an enrichment of IP over control as follow: Suppose both IP and control are ScoreMatrix objects that have same dimensions. Then, the enrichment is calculated using a formula: \( \log_2((IP + 1)/(control + 1)) \).

See Also

ScoreMatrixList, ScoreMatrix

Examples

```r
# load IP and control BAM files and create ScoreMatrix objects
library('genomationData')
data(promoters)
bam.file_IP_1 <- system.file("extdata",
"wgEncodeSydhFbshH1hescZnf143IggrabAlnRep1.chr21.bam", package = "genomationData")
IP_1 <- ScoreMatrix(target = bam.file_IP_1, windows = promoters, type = 'bam')

bam.file_IP_2 <- system.file("extdata",
"wgEncodeBroadHistoneH1hescSuz12051317AlnRep1.chr21.bam", package = "genomationData")
IP_2 <- ScoreMatrix(target=bam.file_IP_2, windows = promoters, type = 'bam')

bam.file_c <- system.file("extdata",
"wgEncodeBroadHistoneH1hescCtcfStdAlnRep1.chr21.bam", package = "genomationData")
control <- ScoreMatrix(target = bam.file_c, windows = promoters, type = 'bam')
```
# create a ScoreMatrixList object storing IP ScoreMatrix objects
sml_IP <- ScoreMatrixList(list(IP1 = IP_1, IP2 = IP_2))

# compute an enrichment of IP over control
enrichmentMatrix(sml_IP, control)

---

### Description

This is an `enrichmentMatrix` function for `ScoreMatrixList` objects, that enables to normalize ChIP-seq signals with respect to IgG or input DNA control.

### Usage

\[
\text{S4method}(\text{enrichmentMatrix})(\text{ScoreMatrixList}, \text{ScoreMatrixList})(\text{IP}, \text{control})
\]

### Arguments

- **IP**  
  `ScoreMatrixList` object storing IP samples
- **control**  
  `ScoreMatrixList` storing control samples

### Value

`ScoreMatrixList` object

### Note

The function computes an enrichment of IP over control as follow: Suppose both IP and control are `ScoreMatrix` objects that have same dimensions. Then, the enrichment is calculated using a formula: \( \log_2((\text{IP} + 1) / (\text{control} + 1)) \).

### See Also

`ScoreMatrixList`

### Examples

```r
# load IP and control BAM files and create ScoreMatrix objects
library('genomationData')
data(promoters)

bam.file_IP_1 <- system.file("extdata",  
"wgEncodeSydhTfbsH1hescZnf143IggrabAlnRep1.chr21.bam", package = "genomationData")

IP_1 <- ScoreMatrix(target = bam.file_IP_1, windows = promoters, type = 'bam')
```
bam.file.IP.2 <- system.file("extdata", "wgEncodeBroadHistoneH1hesc5z12851317A1nRep1.chr21.bam", package = "genomationData")
IP.2 <- ScoreMatrix(target = bam.file.IP.2, windows = promoters, type = 'bam')

bam.file.c <- system.file("extdata", "wgEncodeBroadHistoneH1hescCtcfStdA1nRep1.chr21.bam", package = "genomationData")
control <- ScoreMatrix(target = bam.file.c, windows = promoters, type = 'bam')

# create a ScoreMatrixList object storing IP ScoreMatrix objects
sml.IP <- ScoreMatrixList(list(IP1 = IP.1, IP2 = IP.2))

# create a ScoreMatrixList object storing control ScoreMatrix objects
sml.control <- ScoreMatrixList(list(c1 = control, c2 = control))

# compute an enrichment of IP over control
enrichmentMatrix(sml.IP, sml.control)

---

**file.ext**

**file.ext function**

**Description**

Extract file extension from file path

**Usage**

file.ext(x)

**Arguments**

x file

**findFeatureComb**

Find combinations of genomic features

**Description**

Provided a GRangesList, finds the combinations of sets of ranges. It is mostly used to look at the combinatorics of transcription factor binding. The function works by, firstly, constructing a union of all ranges in the list, which are then designated by the combinatorics of overlap with the original sets. A caveat of this approach is that the number of possible combinations increases exponentially, so we would advise you to use it with up to 6 data sets. If you wish to take a look at a greater number of factors, methods like self organizing maps or ChromHMM might be more appropriate.
Usage

findFeatureComb(gl, width=0, use.names=FALSE, collapse.char=': ')

## S4 method for signature 'GRangesList'
findFeatureComb(gl, width = 0, use.names = FALSE,
   collapse.char = "":")

Arguments

- **gl**: a GRangesList object, containing ranges for which represent regions enriched for transcription factor binding.
- **width**: integer is the requested width of each enriched region. If 0 the ranges are not resized, if a positive integer, the width of all ranges is set to that number. Ranges are resized relative to the center of original ranges.
- **use.names**: a boolean which tells the function whether to return the resulting ranges with a numeric vector which designates each class (the default), or to construct the names of each class using the names from the GRangesList.
- **collapse.char**: a character which will be used to separate the class names if use.names=TRUE. The default is ":".

Value

a GRanges object

Examples

library(GenomicRanges)
g = GRanges(paste('chr',rep(1:2, each=3), sep=''), IRanges(rep(c(1,5,9), times=2), width=3))
gl = GRangesList(g1=g, g2=g[2:5], g3=g[3:4])
findFeatureComb(gl)
findFeatureComb(gl, use.names=TRUE)

galpTo2Ranges
galpTo2Ranges function

description

galpTo2Ranges function

Usage
galpTo2Ranges(x)

Arguments

- **x**: object
genes  

Example RefSeq genes data set.

Description

RefSeq genes of hg19 assembly of human genome on chr21 and chr22. Downloaded from UCSC genome browser.

Format

GRanges object

getAssociationWithTSS  
Get distance to nearest TSS and gene id from AnnotationByGeneParts

Description

This accessor function gets the nearest TSS, its distance to target feature, strand and name of TSS/gene from AnnotationByGeneParts object

Usage

getAssociationWithTSS(x)

## S4 method for signature 'AnnotationByGeneParts'
gegetAssociationWithTSS(x)

Arguments

x  
a AnnotationByGeneParts object

Value

RETURNS a data.frame containing row number of the target features, distance of target to nearest TSS, TSS/Gene name, TSS strand

Examples

data(cage)
bed.file = system.file("extdata/chr21.refseq.hg19.bed", package = "genomation")
gene.parts = readTranscriptFeatures(bed.file)
cage.annot = annotateWithGeneParts(cage, gene.parts, intersect.chr=TRUE)
head(getAssociationWithTSS(cage.annot))
getColors

**getColors**

**getColors**

**Description**

gets colors for a factor variable

**Usage**

getColors(n)

**Arguments**

n

getFeatsWithTargetsStats

Get the percentage/count of annotation features overlapping with target features from AnnotationByFeature

**Description**

This function retrieves percentage/number of annotation features overlapping with targets. For example, if AnnotationByFeature object is containing statistics of differentially methylated regions overlapping with gene annotation. This function will return number/percentage of introns, exons and promoters overlapping with differentially methylated regions.

**Usage**

getFeatsWithTargetsStats(x, percentage=TRUE)

## S4 method for signature 'AnnotationByFeature'
getFeatsWithTargetsStats(x, percentage = TRUE)

**Arguments**

x a AnnotationByFeature object

percentage TRUE|FALSE. If TRUE percentage of annotation features will be returned. If FALSE, number of annotation features will be returned

**Value**

RETURNS a vector of percentages or counts showing quantity of annotation features overlapping with target features
Examples

```r
data(cage)
bed.file=system.file("extdata/chr21.refseq.hg19.bed", package = "genomation")
gene.parts = readTranscriptFeatures(bed.file)
cage.annot = annotateWithGeneParts(cage, gene.parts, intersect.chr=TRUE)
getFeatsWithTargetsStats(cage.annot)
```

---

**getFlanks**

*Function to get upstream and downstream adjacent regions to a genomic feature such as CpG islands*

Description

Function to get upstream and downstream adjacent regions to a genomic feature such as CpG islands

Usage

```r
getFlanks(grange, flank=2000, clean=TRUE)
```

## S4 method for signature 'GRanges'
getFlanks(grange, flank = 2000, clean = TRUE)

Arguments

- `grange` GRanges object for the feature
- `flank` number of basepairs for the flanking regions
- `clean` If set to TRUE, flanks overlapping with other main features will be trimmed, and overlapping flanks will be removed. This will remove multiple counts when other features overlap with flanks

Value

GRanges object for flanking regions

Examples

```r
data(cpgi)
cpgi.flanks = getFlanks(cpgi)
head(cpgi.flanks)
```
getMembers

*Get the membership slot of AnnotationByFeature*

**Description**

Membership slot defines the overlap of target features with annotation features. For example, if a target feature overlaps with an exon.

**Usage**

```r
getMembers(x)
```

## S4 method for signature 'AnnotationByFeature'
getMembers(x)

**Arguments**

- `x`: A `AnnotationByFeature` object

**Value**

A matrix showing the overlap of target features with annotation features. 1 for overlap, 0 for non-overlap.

getRandomEnrichment

*Get enrichment based on randomized feature overlap*

**Description**

This function measures the association between two genomic features by randomizing one feature and counting the overlaps in randomized sets. That is to say, query feature will be randomly distributed over the genome (constrained by provided options), and the overlap of target with these randomized features will be measured.

**Usage**

```r
getRandomEnrichment(target, query, randomizations = 1000, rand.set = NULL, ...)
```

## S4 method for signature 'GRanges,GRanges'
getRandomEnrichment(target, query,
randomizations = 1000, rand.set = NULL, ...)
getTargetAnnotationStats

Get the percentage of target features overlapping with annotation from AnnotationByFeature

Description

This function retrieves percentage/number of target features overlapping with annotation

Usage

getTargetAnnotationStats(x, percentage=TRUE, precedence=TRUE)

```r
# S4 method for signature 'AnnotationByFeature'
getTargetAnnotationStats(x, percentage = TRUE, precedence = TRUE)
```
gffToGRanges

Arguments

- `x`: An AnnotationByFeature object
- `percentage`: TRUE|FALSE. If TRUE percentage of target features will be returned. If FALSE, number of target features will be returned.
- `precedence`: TRUE|FALSE. If TRUE there will be a hierarchy of annotation features when calculating numbers (with promoter>exon>intron precedence).

That means if a feature overlaps with a promoter it will be counted as promoter overlapping only, or if it is overlapping with an exon but not a promoter, it will be counted as exon overlapping only whether or not it overlaps with an intron.

Value

A vector of percentages or counts showing quantity of target features overlapping with annotation.

Examples

```r
data(cage)
bed.file = system.file("extdata/chr21.refseq.hg19.bed", package = "genomation")
gene.parts = readTranscriptFeatures(bed.file)
cage.annot = annotateWithGeneParts(cage, gene.parts, intersect.chr = TRUE)
getTargetAnnotationStats(cage.annot)
```

gffToGRanges

Converts a gff formatted data.frame into a GenomicRanges object. The GenomicRanges object needs to be properly formatted for the function to work.

Description

Converts a gff formatted data.frame into a GenomicRanges object. The GenomicRanges object needs to be properly formatted for the function to work.

Usage

`gffToGRanges(gff.file, filter = NULL, zero.based = FALSE, ensembl = FALSE)`

Arguments

- `gff.file`: path to a gff formatted file. The file can end in .gz, .bz2,.xz, or .zip and/or start with http:// or ftp://. If the file is not compressed it can also start with https:// or ftps://.
- `filter`: a character designating which elements to retain from the gff file (e.g. exon, CDS, ...)
- `zero.based`: boolean whether the coordinates are 0 or 1 based. 0 is the default
- `ensembl`: boolean if TRUE, add the chr prefix to seqlevels. FALSE by default
heatMatrix

**Value**

returns a GenomicRanges object

**Examples**

```r
gff.file = system.file('extdata/chr21.refseq.hg19.gtf', package='genomation')
gff = gffToGRanges(gff.file)
```

---

heatMatrix  

**Draw a heatmap of a given ScoreMatrix object**

**Description**

The function makes a heatmap out of given ScoreMatrix object. If desired it can use clustering using given clustering function (e.g. k-means) and plot cluster color codes as a sidebar. In addition, user can define groups of rows using 'group' argument.

**Usage**

```r
heatMatrix(mat, grid = FALSE, col = NULL, xcoords = NULL, group = NULL,
          group.col = NULL, order = FALSE, user.order = FALSE, winsorize = c(0,
          100), clustfun = NULL, main = '', legend.name = NULL, cex.legend = 1,
          xlab = NULL, cex.main = 1, cex.lab = 1, cex.axis = 1,
          newpage = TRUE)
```

**Arguments**

- `mat`: a ScoreMatrix object
- `grid`: if TRUE, grid graphics will be used. if FALSE, base graphics will be used on the top level, so users can use par(mfrow) or par(mfcol) prior to calling the function. Default:FALSE
- `col`: a vector of colors, such as the ones created by heat.colors(10). If NULL (which is default), jet color scheme (common in matlab plots) will be used.
- `xcoords`: a vector of numbers showing relative positions of the bases or windows. It must match the number of columns in the ScoreMatrix. Alternatively, it could be a numeric vector of two elements. Such as c(0,100) showing the relative start and end coordinates of the first and last column of the ScoreMatrix object.
- `group`: a list of vectors of row numbers or a factor. This grouping is used for rowside colors of the heatmap. If it is a list, each element of the list must be a vector of row numbers. Names of the elements of the list will be used as names of groups. If `group` is a factor, it's length must match the number of rows of the matrix, and factor levels will be used as the names of the groups in the plot.
- `group.col`: a vector of color names to be used at the rowside colors if `group` argument is given or `clustfun` function is given.
heatMatrix

**order**  Logical indicating if the rows should be ordered or not (Default:FALSE). If order=TRUE the matrix will be ordered with rowSums(mat) values in descending order. If group argument is provided, first the groups will be ordered in descending order of sums of rows then, everything within the clusters will be ordered by sums of rows. If clustfun is given then rows within clusters will be order in descending order of sums of rows.

**user.order**  a numerical vector indicating the order of groups/clusters (it works only when group or clustfun argument is given).

**winsorize**  Numeric vector of two, defaults to c(0,100). This vector determines the upper and lower percentile values to limit the extreme values. For example, c(0,99) will limit the values to only 99th percentile, everything above the 99 percentile will be equalized to the value of 99th percentile. This is useful for visualization of matrices that have outliers.

**clustfun**  a function for clustering rows of mat that returns a vector of integers indicating the cluster to which each point is allocated (a vector of cluster membership), e.g. k-means algorithm with 3 centers: function(x) kmeans(x, centers=3)$cluster. By default FALSE.

**main**  a character string for the plot title

**legend.name**  a character label plotted next to the legend

**cex.legend**  A numerical value giving the amount by which legend axis marks should be magnified relative to the default

**xlab**  label a character string for x-axis of the heatmap

**cex.main**  A numerical value giving the amount by which plot title should be magnified

**cex.lab**  A numerical value giving the amount by which axis labels (including 'legend.name') should be magnified relative to the default.

**cex.axis**  A numerical value giving the amount by which axis marks should be magnified relative to the default.

**newpage**  logical indicating if grid.newpage() function should be invoked if grid=TRUE.

**Value**
returns clustering result invisibly, if clustfun is defined

**Examples**

```r
data(cage)
data(promoters)
scores1=ScoreMatrix(target=cage,windows=promoters,strand.aware=TRUE,
weight.col="tpm")

set.seed(1000)

heatMatrix(mat=scores1,legend.name="tpm",winsorize=c(0,99),xlab="region around TSS",
xcoords=-1000:1000,
cex.legend=0.8,main="CAGE clusters on promoters",cex.lab=1,
cex.axis=0.9,grid=FALSE)
```

```
heatMeta

Heatmap for meta-region profiles

Description

Function calculates meta-profile(s) from a ScoreMatrix or a ScoreMatrixList, then produces a heatmap or a set of stacked heatmaps for meta-region profiles.

Usage

heatMeta(mat, centralTend = "mean", profile.names = NULL, xcoords = NULL, col = NULL, meta.rescale = FALSE, winsorize = c(0, 100), legend.name = NULL, cex.legend = 1, xlab = NULL, main = "", cex.lab = 1, cex.axis = 1)

Arguments

- **mat**: ScoreMatrix or ScoreMatrixList to be plotted.
- **centralTend**: a character that determines central tendency of meta-profile(s). It takes "mean" (default) or "median".
- **profile.names**: a character vector for names of profiles. If NULL, the names will be taken from names(mat) if mat is a ScoreMatrixList object.
- **xcoords**: a vector of numbers showing relative positions of the bases or windows. It must match the number of columns in the ScoreMatrix. For example: if there are 2001 elements in the matrices which are base-pair resolution and they are centered around an anchor point like TSS, the xcoords argument should be -1000:1000. This argument is used to plot accurate x-axis labels for the plots. If NULL it will be equal to 1:ncol(mat).
**heatTargetAnnotation**

Plots the percentage of overlapping intervals with genomic features in a heatmap

**Description**

This function plots a heatmap for percentage of overlapping ranges with provided genomic features. The input object is a list of `AnnotationByFeature` objects, which contains necessary information about overlap statistics to make the plot.

**Arguments**

- `col` a vector of color pallete. color scheme to be used. If NULL, a version of jet colors will be used.
- `meta.rescale` if TRUE meta-region profiles are scaled to 0 to 1 range by subtracting the min from profiles and dividing them by max-min.
- `winsorize` Numeric vector of two, defaults to c(0,100). This vector determines the upper and lower percentile values to limit the extreme values. For example, c(0,99) will limit the values to only 99th percentile, everything above the 99 percentile will be equalized to the value of 99th percentile. This is useful for visualization of matrices that have outliers.
- `legend.name` a character label plotted next to the legend
- `cex.legend` A numerical value giving the amount by which legend axis marks should be magnified relative to the default
- `xlab` label a character string for x-axis
- `main` a character string for the plot title
- `cex.lab` A numerical value giving the amount by which axis labels (including 'legend.name') should be magnified relative to the default.
- `cex.axis` A numerical value giving the amount by which axis marks should be magnified relative to the default

**Value**

returns meta-profile matrix invisibly.

**Examples**

```r
data(cage)
data(promoters)
scores1=ScoreMatrix(target=cage, windows=promoters, strand.aware=TRUE)
data(cpgi)
scores2=ScoreMatrix(target=cpgi, windows=promoters, strand.aware=TRUE)
x=new("ScoreMatrixList", list(scores1, scores2))
heatMeta(mat=x, legend.name="fg", cex.legend=0.8, main="fdf", cex.lab=6, cex.axis=0.9)
```
heatTargetAnnotation

Usage

heatTargetAnnotation(l, cluster = FALSE, col = c("white", "blue"),
                       precedence = FALSE, plot = TRUE)

Arguments

l          a list of AnnotationByFeature objects. This could be returned by annotateWithFeatures function.
cluster    TRUE/FALSE. If TRUE the heatmap is going to be clustered using a default hierarchical clustering scheme.
col        a vector of two colors that will be used for interpolation. The first color is the lowest one, the second is the highest one
precedence TRUE/FALSE. If TRUE the precedence of annotation features will be used when plotting. The precedence will be taken from the GRangesList used as annotation. The first GRanges will be treated as most important, and the second as the second most important and so on. Such that, if an interval overlaps with features on that is part of the first GRanges object in the annotation GRangesList, the rest of its overlaps with other elements in the annotation GRangesList will be ignored. This feature is important to have if the user desires that percentage of overlaps adds up to 100. This is only possible when the annotation features are non-overlapping with each other or there is a hierarchy/precedence among them such as (with promoter>exon>intron precedence). In this case, anything that overlaps with a promoter annotation will only be counted as promoter even if it overlaps with exons or introns.
plot       If FALSE, does not plot the heatmap just returns the matrix used to make the heatmap

Value

returns the matrix used to make the heatmap when plot FALSE, otherwise returns ggplot2 object which can be modified further.

See Also

see getMembers, annotateWithFeatures

Examples

library(GenomicRanges)
data(cage)
data(cpgi)
cage$tpm=NULL
gl = GRangesList(cage=cage, cpgi=cpgi)

bed.file = system.file("extdata/chr21.refseq.hg19.bed", package = "genomation")
gene.parts = readTranscriptFeatures(bed.file)
annot = annotateWithFeatures(gl, gene.parts, intersect.chr=TRUE)
intersectScoreMatrixList

Description

Returns a intersection of rows for each matrix in a ScoreMatrixList object. This is done using the rownames of each element in the list.

Usage

intersectScoreMatrixList(sml, reorder = FALSE)

## S4 method for signature 'ScoreMatrixList'
intersectScoreMatrixList(sml, reorder = FALSE)

Arguments

sml 
 a ScoreMatrixList object

reorder 
 if TRUE ScoreMatrix objects in the list are sorted based on their common row ids.

Value

ScoreMatrixList object

Examples

library(GenomicRanges)
target = GRanges(rep(c(1,2),each=7),
    IRanges(rep(c(1,1,2,3,7,8,9), times=2), width=5),
    weight = rep(c(1,2),each=7))

windows1 = GRanges(rep(c(1,2),each=2),
    IRanges(rep(c(1,2), times=2), width=5),
    strand=c('-','-','+','+'))

windows2 = windows1[c(1,3)]
sml = as(list(ScoreMatrix(target, windows1),
    ScoreMatrix(target, windows2)), 'ScoreMatrixList')
sml
intersectScoreMatrixList(sml)
**Description**

listSliceMax() function calls the binMax() function

**Usage**

listSliceMax(xlist, n)

**Arguments**

- **xlist**: List of vectors storing values of a bin
- **n**: integer - number of bins

---

**Description**

listSliceMean() function calls the binMean() function

**Usage**

listSliceMean(xlist, n)

**Arguments**

- **xlist**: List of vectors storing values of a bin
- **n**: integer - number of bins
**listSliceMedian**

Function creates a matrix storing data with desirable number of bins for each window

---

**Description**

listSliceMean() function calls the binMedian() function

**Usage**

listSliceMedian(xlist, n)

**Arguments**

- xlist: List of vectors storing values of a bin
- n: Integer - number of bins

---

**listSliceMin**

Function creates a matrix storing data with desirable number of bins for each window

---

**Description**

listSliceMin() function calls the binMin() function

**Usage**

listSliceMin(xlist, n)

**Arguments**

- xlist: List of vectors storing values of a bin
- n: Integer - number of bins
listSliceSum

*Function creates a matrix storing data with desirable number of bins for each window*

**Description**

listSliceSum() function calls the binSum() function

**Usage**

listSliceSum(xlist, n)

**Arguments**

- **xlist**: List of vectors storing values of a bin
- **n**: integer - number of bins

Max_c

*Function that computes a max value*

**Description**

Function that computes a max value

**Usage**

Max_c(x)

**Arguments**

- **x**: NumericVector

Mean_c

*Function that computes a mean value*

**Description**

Function that computes a mean value

**Usage**

Mean_c(x)

**Arguments**

- **x**: NumericVector
**Median.c**

*Function that computes a median value*

**Description**

Function that computes a median value

**Usage**

Median.c(x)

**Arguments**

x NumericVector

---

**Min.c**

*Function that computes a min value*

**Description**

Function that computes a min value

**Usage**

Min.c(x)

**Arguments**

x NumericVector

---

**multiHeatMatrix**

*Draw multiple heatmaps from a ScoreMatrixList object*

**Description**

The function plots multiple heatmaps for a ScoreMatrixList object side by side. Each matrix can have different color schemes but it is essential that each matrix is obtained from same regions or neighbouring regions.
Usage

multiHeatMatrix(sml, grid = TRUE, col = NULL, xcoords = NULL, 
  group = NULL, group.col = NULL, order = FALSE, user.order = FALSE, 
  winsorize = c(0, 100), clustfun = FALSE, clust.matrix = NULL, 
  column.scale = TRUE, matrix.main = NULL, common.scale = FALSE, 
  legend = TRUE, legend.name = NULL, cex.legend = 0.8, xlab = NULL, 
  cex.lab = 1, cex.main = 1, cex.axis = 0.8, newpage = TRUE)

Arguments

sml  a ScoreMatrixList object

grid if TRUE, grid graphics will be used. if FALSE, base graphics will be used on the 
  top level, so users can use par(mfrow) or par(mfcol) prior to calling the function. 
  Default:FALSE

col a color palette or list of color palettes, such as list(heat.colors(10),topo.colors(10)). 
  If it is a list, it is length must match the number of matrices to be plotted. If it is 
  a single palette every heatmap will have the same colors.

xcoords a vector of numbers showing relative positions of the bases or windows or a list 
  of vectors. The elements of the list must match the number of columns in the 
  corresponding ScoreMatrix. Alternatively, the elements could be a numeric 
  vector of two elements. Such as c(0,100) showing the relative start and end co- 
  ordinates of the first and last column of the ScoreMatrix object. The remaining 
  coordinates will be automatically matched in this case. If the argument is not a 
  list but a single vector, then all heatmaps will have the same coordinate on their 
  x-axis.

group a list of vectors of row numbers or a factor. The rows will be reordered to match 
  their grouping. The grouping is used for rowside colors of the heatmap. If it is 
  a list, each element of the list must be a vector of row numbers. Names of the 
  elements of the list will be used as names of groups. If group is a factor, it’s 
  length must match the number of rows of the matrix, and factor levels will be 
  used as the names of the groups in the plot.

group.col a vector of color names to be used at the rowside colors if group and clustfun 
  arguments are given

order Logical indicating if the rows should be ordered or not (Default:FALSE). If 
  order=TRUE the matrix will be ordered with rowSums(mat) values in descend- 
  ing order. If group argument is provided, first the groups will be ordered in 
  descending order of sums of rows then, everything within the clusters will be 
  ordered by sums of rows. If clustfun is given then rows within clusters will be 
  order in descending order by sums of rows.

user.order a numerical vector indicating the order of groups/clusters (it works only when 
  group or clustfun argument is given).

winsorize Numeric vector of two, defaults to c(0,100). This vector determines the upper 
  and lower percentile values to limit the extreme values. For example, c(0,99) 
  will limit the values to only 99th percentile for a matrix, everything above the 
  99 percentile will be equalized to the value of 99th percentile. This is useful for 
  visualization of matrices that have outliers.
clustfun a function for clustering rows of mat that returns a vector of integers indicating the cluster to which each point is allocated (a vector of cluster membership), e.g. k-means algorithm with 3 centers: function(x) kmeans(x, centers=3)$cluster. By default FALSE.

clust.matrix a numerical vector of indexes or a character vector of names of the ScoreMatrix objects in 'sml' to be used in clustering (if clustfun argument is provided). By default all matrices are clustered. Matrices that are not indicated in clust.matrix are ordered according to result of clustering algorithm.

column.scale Logical indicating if matrices should be scaled or not, prior to clustering or ordering. Setting this to TRUE scales the columns of the matrices using, scale() function. scaled columns are only used for clustering or ordering. Original scores are displayed for heatmaps.

matrix.main a vector of strings for the titles of the heatmaps. If NULL titles will be obtained from names of the ScoreMatrix objects in the ScoreMatrixList objects.

common.scale if TRUE (Default:FALSE) all the heatmap colors will be coming from the same score scale, although each heatmap color scale can be different. The color intensities will be coming from the same scale. The scale will be determined by minimum of all matrices and maximum of all matrices. This is useful when all matrices are on the same score scale. If FALSE, the color scale will be determined by minimum and maximum of each matrix individually.

legend if TRUE and color legend for the heatmap is drawn.

legend.name a vector of legend labels to be plotted with legends of each heatmap. If it is a length 1 vector, all heatmaps will have the same legend label.

cex.legend A numerical value giving the amount by which legend axis marks should be magnified relative to the default.

xlab a vector of character strings for x-axis labels of the heatmaps. if it is length 1, all heatmaps will have the same label.

cex.lab A numerical value giving the amount by which axis labels (including 'legend.name') should be magnified relative to the default.

cex.main A numerical value giving the amount by which plot title should be magnified

cex.axis A numerical value giving the amount by which axis marks should be magnified relative to the default.

newpage logical indicating if grid.newpage() function should be invoked if grid=TRUE.

Value

invisibly returns the order of rows, if clustfun is provided and/or order=TRUE

Examples

data(cage)
data(promoters)
scores1=ScoreMatrix(target=cage, windows=promoters, strand.aware=TRUE)

data(cpgi)
scores2=ScoreMatrix(target=cpgi, windows=promoters, strand.aware=TRUE)
Ops,numeric,ScoreMatrixList-method

Ops method for a ScoreMatrixList object. It enables to use arithmetic, indicator and logic operations on ScoreMatrixList objects.

Description

Arithmetic method for ScoreMatrixList

Usage

## S4 method for signature 'numeric,ScoreMatrixList'
Ops(e1, e2)

Arguments

e1 the numeric value
e2 the ScoreMatrixList object
Description

Arithmetic method for ScoreMatrix and ScoreMatrixList classes

Usage

## S4 method for signature 'ScoreMatrix,ScoreMatrix'
Ops(e1, e2)

Arguments

e1 the ScoreMatrix object or numeric value
e2 the ScoreMatrix object or numeric value

Value

ScoreMatrix

Description

Arithmetic method for ScoreMatrixList

Usage

## S4 method for signature 'ScoreMatrixList,numeric'
Ops(e1, e2)

Arguments

e1 the ScoreMatrixList object
e2 the numeric value

Value

ScoreMatrixList
Ops, ScoreMatrixList, ScoreMatrixList-method

*Ops method for a ScoreMatrixList object. It enables to use arithmetic, indicator and logic operations on ScoreMatrixList objects.*

**Description**

Arithmetic methods for ScoreMatrixList

**Usage**

```r
## S4 method for signature 'ScoreMatrixList,ScoreMatrixList'
Ops(e1, e2)
```

**Arguments**

- `e1` the `ScoreMatrixList` object
- `e2` the `ScoreMatrixList` object

**Value**

ScoreMatrixList

---

**orderBy**

*Reorder all elements of a ScoreMatrixList to a given ordering vector*

**Description**

Reorder all elements of a ScoreMatrixList to a given ordering vector

**Usage**

```r
orderBy(sml, ord.vec)
```

```r
## S4 method for signature 'ScoreMatrixList'
orderBy(sml, ord.vec)
```

**Arguments**

- `sml` ScoreMatrixList object
- `ord.vec` an integer vector

**Value**

ScoreMatrixList object
Examples

```r
library(GenomicRanges)
data(cage)
data(cpgi)
data(promoters)
cage$tpm = NULL
targets = GRangesList(cage=cage, cpgi=cpgi)
sml = ScoreMatrixList(targets, promoters, bin.num=10)
kmeans.clust = kmeans(sml$cage, 3)

sml.ordered = orderBy(sml, kmeans.clust$cluster)
multiHeatMatrix(sml.ordered)
```

patternMatrix

Get scores that correspond to k-mer or PWM matrix occurrence for bases in each window

Description

The function produces a base-pair resolution matrix or matrices of scores that correspond to k-mer or PWM matrix occurrence over predefined windows that have equal width. It finds either positions of pattern hits above a specified threshold and creates score matrix filled with 1 (presence of pattern) and 0 (its absence) or matrix with scores themselves. If pattern is a character of length 1 or PWM matrix then the function returns a ScoreMatrix object, if character of length more than 1 or list of PWMs then ScoreMatrixList.

Usage

```r
patternMatrix(pattern, windows, genome = NULL, min.score = 0.8,
               asPercentage = FALSE, cores = 1)
```

\[\text{S4method\{patternMatrix\}\{character,DNAStringSet\}\{pattern, windows,
               asPercentage, cores\}}\]

\[\text{S4method\{patternMatrix\}\{character,GRanges,BSgenome\}\{pattern, windows, genome,
               cores\}}\]

\[\text{S4method\{patternMatrix\}\{matrix,DNAStringSet\}\{pattern, windows,
               min.score, asPercentage, cores\}}\]

\[\text{S4method\{patternMatrix\}\{matrix,GRanges,BSgenome\}\{pattern, windows, genome,
               min.score, asPercentage, cores\}}\]
patternMatrix

\S4method{patternMatrix}{list,DNAStringSet}(pattern, windows, 
  min.score, asPercentage, 
  cores)

\S4method{patternMatrix}{list,GRanges,BSgenome}(pattern, windows, genome, 
  min.score, asPercentage, 
  cores)

Arguments

pattern  matrix (a PWM matrix), list of matrices or a character vector of length 1 or more. A matrix is a PWM matrix that needs to have one row for each nucleotide ("A", "C", "G" and "T" respectively). IUPAC ambiguity codes can be used and it will match any letter in the subject that is associated with the code.

windows  GRanges object or DNAStringSet object that have equal width of ranges or sequences.

geno  BSgenome object

min.score  numeric or character indicating minimum score to count a match. It can be given as a character string containing a percentage of the highest possible score or a single number (by default "80%" or 0.8). If min.score is set to NULL then patternMatrix returns scores themselves (default).

asPercentage  boolean telling whether scores represent percentage of the maximal motif PWM score (default: TRUE) or raw scores (FALSE).

cores  the number of cores to use (default: 1). It is supported only on Unix-like platforms.

Details

patternMatrix is based on functions from the seqPattern package: getPatternOccurrenceList function to find position of pattern that is a character vector in a list of sequences (a DNAStringSet object) and adapted function motifScanHits to find pattern that is a PWM matrix in sequences (a DNAStringSet object).

If cores > 1 is provided then for every window occurrence of pattern is counted in parallel.

Value

returns a ScoreMatrix object or a ScoreMatrixList object

See Also

ScoreMatrix, ScoreMatrixList

Examples

library(Biostrings)

# consensus sequence of the ctcf motif
motif = "CCGCNGGNGGCAG"
# Creates 10 random DNA sequences
seqs = sapply(1:10,
    function(x) paste(sample(c("A","T","G","C"), 180, replace=TRUE), collapse=""))
windows = DNAStringSet(seqs)
p = patternMatrix(pattern=motif, windows=windows, min.score=0.8)
p

## plotMeta

__Line plot(s) for meta-region profiles__

### Description
Function calculates meta-profile(s) from a ScoreMatrix or a ScoreMatrixList, then produces a line plot or a set of line plots for meta-region profiles

### Usage

```
plotMeta(mat, centralTend = "mean", overlay = TRUE, winsorize = c(0, 100),
    profile.names = NULL, xcoords = NULL, meta.rescale = FALSE,
    smoothfun = NULL, line.col = NULL, dispersion = NULL,
    dispersion.col = NULL, ylim = NULL, ylab = "average score",
    xlab = "bases", ...)
```

### Arguments

- **mat** ScoreMatrix or ScoreMatrixList object. If it is a ScoreMatrixList object, all matrices in the ScoreMatrixList should have the same number of columns.
- **centralTend** a character that determines central tendency of meta-profile(s). It takes "mean" (default) or "median".
- **overlay** If TRUE multiple profiles will be overlayed in the same plot (Default:TRUE). If FALSE, and mat is a ScoreMatrixList, consider using par(mfrow=c(1,length(mat))) to see the plots from all matrices at once.
- **winsorize** Numeric vector of two, defaults to c(0, 100). This vector determines the upper and lower percentile values to limit the extreme values. For example, c(0.99) will limit the values to only 99th percentile, everything above the 99 percentile will be equalized to the value of 99th percentile. This is useful for visualization of matrices that have outliers.
- **profile.names** a character vector for names of the profiles. The order should be same as the as the order of ScoreMatrixList.
- **xcoords** a numeric vector which designates relative base positions of the meta-region profiles. For example, for a 2001 column ScoreMatrix, xcoord=-1000:1000 indicates relative positions of each column in the score matrix. If NULL (Default), xcoords equals to 1:ncol(mat)
- **meta.rescale** if TRUE meta-region profiles are scaled to 0 to 1 range by subtracting the min from profiles and dividing them by max-min. If dispersion is not NULL, then dispersion will be scaled as well.
smoothfun

a function to smooth central tendency and dispersion bands (Default: NULL), e.g. stats::lowess.

line.col
color of lines for centralTend of meta-region profiles. Defaults to colors from rainbow() function.

dispersion
shows dispersion interval bands around centralTend (default:NULL). It takes one of the character:

- "se" shows standard error of the mean and 95 percent confidence interval for the mean
- "sd" shows standard deviation and 2*(standard deviation)
- "IQR" shows 1st and 3rd quartile and confidence interval around the median based on the median +/- 1.57 * IQR/sqrt(n) (notches)

dispersion.col
color of bands of dispersion. Defaults to colors from rainbow() and transparency is set to 0.5 (rainbow(length(mat), alpha = 0.5)).

ylim
same as ylim at plot function. If NULL ylim is estimated from all meta-region profiles.

ylab
same as ylab at plot function. Default: "average score"

xlab
same as xlab at plot function. Default: "bases"

... other options to plot

Value
returns the meta-region profiles invisibly as a matrix.

Note
Score matrices are plotted according to ScoreMatrixList order. If ScoreMatrixList contains more than one matrix then they will overlap each other on a plot, i.e. the first one is plotted first and every next one overlays previous one(s) and the last one is the topmost.

Missing values in data slow down plotting dispersion around central tendency. The reason is that dispersion is plotted only for non-missing values, for each segment that contains numerical values graphics::polygon function is used to plot dispersion bands. There might be a situation, when in a column of ScoreMatrix is only one numeric number and the rest are NAs, then at corresponding position only central tendency will be plotted.

Notches show the 95 percent confidence interval for the median according to an approximation based on the normal distribution. They are used to compare groups - if notches corresponding to adjacent base pairs on the plot do not overlap, this is strong evidence that medians differ. Small sample sizes (5-10) can cause notches to extend beyond the interquartile range (IQR) (Martin Krzywinska et al. Nature Methods 11, 119-120 (2014))

Examples

data(cage)
data(promoters)
scores1=ScoreMatrix(target=cage, windows=promoters, strand.aware=TRUE)

data(cpgi)
scores2=ScoreMatrix(target=cpgi, windows=promoters, strand.aware=TRUE)

# create a new ScoreMatrixList
x=new("ScoreMatrixList",list(scores1,scores2))

plotMeta(mat=x,overlay=TRUE,main="my plotowski")

# plot dispersion nd smooth central tendency and variation interval bands
plotMeta(mat=x, centralTend="mean", dispersion="se", winsorize=c(0,99),
     main="Dispersion as interquartile band", lwd=4,
     smoothfun=function(x) stats::lowess(x, f = 1/5))

---

**plotTargetAnnotation**  
*Plot annotation categories from AnnotationByGeneParts or AnnotationByFeature*

**Description**

This function plots a pie or bar chart for showing percentages of targets annotated by genic parts or other query features.

**Usage**

```r
plotTargetAnnotation(x, precedence = TRUE, col = getColors(length(x@annotation)), cex.legend = 1, ...)
```

## S4 method for signature 'AnnotationByFeature'

```r
plotTargetAnnotation(x, precedence = TRUE,
     col = getColors(length(x@annotation)), cex.legend = 1, ...)
```

**Arguments**

- **x**  
a AnnotationByFeature or AnnotationByGeneParts object
- **precedence**  
TRUE|FALSE. If TRUE there will be a hierarchy of annotation features when calculating numbers (with promoter>exon>intron precedence). This option is only valid when x is a AnnotationByGeneParts object
- **col**  
a vector of colors for piechart or the bar plot
- **cex.legend**  
a numeric value of length 1 to specify the size of the legend. By default 1.
- **...**  
graphical parameters to be passed to pie or barplot functions

**Value**

plots a piechart or a barplot for percentage of the target features overlapping with annotation
Examples

data(cage)

bed.file = system.file("extdata/chr21.refseq.hg19.bed", package = "genomation")
gene.parts = readTranscriptFeatures(bed.file)
annot = annotateWithGeneParts(cage, gene.parts, intersect.chr=TRUE)

plotTargetAnnotation(annot)

promoters

Example promoter data set.

Description

promoters of hg19 assembly of human genome on chr21 and chr22. Promoter set is derived from refseq TSS.

Format

GRanges object

RandomEnrichment-class

An S4 class for storing getRandomEnrichment function results

Description

The resulting object stores the results of getRandomEnrichment function

Slots

orig.cnt: number of features overlapping with query at getRandomEnrichment
rand.olap.dist: set of number of features overlapping with randomized queries at getRandomEnrichment
log2fc: log2 fold change calculated by dividing orig.cnt by mean(rand.olap.dist) and taking log2 of that result
p.value: P-value assuming rand.olap.dist has a normal distribution and comparing orig.cnt with that distribution
rand.p.value: p-value from randomization by calculation the proportion of how many times a random number of overlap exceeds the original number of overlap

See Also

getRandomEnrichment
RandomizeFeature

**Description**

This function randomly distributes the coordinates of genomic features which is stored in a GRanges object. The randomization can be constrained by supplied arguments. The function is still in Beta mode - the regions can overlap excluded regions, and the randomized regions are not disjoint. Please take care that the excluded and included regions are not too strict when compared to the total width of the ranges.

**Usage**

```r
randomizeFeature(feature, chrom.sizes = NULL, stranded = TRUE, keep.strand.prop = TRUE, keep.chrom = TRUE, exclude = NULL, include = NULL, seed = NULL, nrand = 1)
```

## S4 method for signature 'GRanges'

```r
randomizeFeature(feature, chrom.sizes = NULL, stranded = TRUE, keep.strand.prop = TRUE, keep.chrom = TRUE, exclude = NULL, include = NULL, seed = NULL, nrand = 1)
```

**Arguments**

- `feature`: a GRanges object to be randomized
- `chrom.sizes`: sizes of chromosomes as a named vector (names are chromosomes names and elements of the vectors are lengths). If not given sizes in GRanges object will be used if no sizes there the end of each chr will be the end last feature on each chr
- `stranded`: if FALSE, all of the returned features will be strandless (will have "*" in the strand slot)
- `keep.strand.prop`: If TRUE, strands will have the same proportion as the features
- `keep.chrom`: If TRUE, number of features and randomized features for a chromosome will match. Currently setting this to FALSE is not supported.
- `exclude`: A GRanges object where no randomized feature should overlap, can be gaps or unmappable regions in the genome as an example.
- `include`: A GRanges object which defines the boundaries of randomized features. If not provided the whole genome is used, as defined using the chrom.sizes parameter.
- `seed`: random number generator seed
- `nrand`: number of randomizations (default:1)

**Value**

returns a GRanges object which is randomized version of the feature, along with a "set" column in the metadata which designates to which iteration of the randomization the range belong.
read.zip

Description
read.zip function

Usage
read.zip(file, ...)

Arguments
- file: zip file
- ...: additional objects and parameters

readBam

Description
given a big bam path reads the big wig file into a RleList to be used by ScoreMatrix:char,GRanges

Usage
readBam(target, windows, rpm = FALSE, unique = FALSE, extend = 0,
param = NULL, paired.end = FALSE, library.size = NULL, ...)

Arguments
- target: target object
- windows: windows
- rpm: logical
- unique: logical
- extend: numeric
- param: ScanBamParam object
- paired.end: logical
- library.size: numeric
- ...: additional parameters
**readBed**

*Read a BED file and convert it to GRanges.*

**Description**

The function reads a BED file that contains location and other information on genomic features and returns a `GRanges` object. The minimal information that the BED file has to have is chromosome, start and end columns. It can handle all BED formats up to 12 columns.

**Usage**

```r
readBed(file, track.line = FALSE, remove.unusual = FALSE, zero.based = TRUE)
```

**Arguments**

- `file` location of the file, a character string such as: "/home/user/my.bed" or the input itself as a string (containing at least one \n). The file can end in .gz, .bz2, .xz, or .zip and/or start with http:// or ftp://. If the file is not compressed it can also start with https:// or ftps://.
- `track.line` the number of track lines to skip, "auto" to detect them automatically or FALSE (default) if the bed file doesn’t have track lines
- `remove.unusual` if TRUE remove the chromosomes with unusual names, such as chrX_random (Default: FALSE)
- `zero.based` a boolean which tells whether the ranges in the bed file are 0 or 1 base encoded. (Default: TRUE)

**Value**

`GRanges` object

**Examples**

```r
my.file = system.file("extdata", "chr21.refseq.hg19.bed", package="genomation")
refseq = readBed(my.file, track.line=FALSE, remove.unusual=FALSE)
head(refseq)
```
**readBigWig**

*readBigWig function*

**Description**

given a big wig path reads the big wig file into a RleList to be used by ScoreMatrix:char,GRanges

**Usage**

```r
readBigWig(target, windows = NULL, ...)
```

**Arguments**

- `target`: target object
- `windows`: windows
- `...`: additional parameters

**readBroadPeak**

*A function to read the Encode formatted broad peak file into a GRanges object*

**Description**

A function to read the Encode formatted broad peak file into a GRanges object

**Usage**

```r
readBroadPeak(file, track.line=FALSE, zero.based=TRUE)
```

**Arguments**

- `file`: an absolute or relative path to a bed file formatted by the Encode broadPeak standard. The file can end in .gz, .bz2, .xz, or .zip and/or start with http:// or ftp://. If the file is not compressed it can also start with https:// or ftps://.
- `track.line`: the number of track lines to skip, "auto" to detect them automatically or FALSE(default) if the bed file doesn’t have track lines
- `zero.based`: a boolean which tells whether the ranges in the bed file are 0 or 1 base encoded. (Default: TRUE)

**Value**

a GRanges object
**readFeatureFlank**

**Examples**

```r
broad.peak.file = system.file('extdata','ex.broadPeak', package='genomation')

broad.peak = readBroadPeak(broad.peak.file)
head(broad.peak)
```

---

**readFeatureFlank**  
* A function to read-in genomic features and their upstream and downstream adjacent regions such as CpG islands and their shores

**Description**

A function to read-in genomic features and their upstream and downstream adjacent regions such as CpG islands and their shores

**Usage**

```r
readFeatureFlank(location, remove.unusual=TRUE, flank=2000, clean=TRUE, feature.flank.name=NULL)

## S4 method for signature 'character'
readFeatureFlank(location, remove.unusual = TRUE, flank = 2000, clean = TRUE, feature.flank.name = NULL)
```

**Arguments**

- `location` for the bed file of the feature.
- `remove.unusual` remove chromosomes with unusual names random, Un and antyhing with "_" character
- `flank` number of basepairs for the flanking regions
- `clean` If set to TRUE, flanks overlapping with other main features will be trimmed
- `feature.flank.name` the names for feature and flank ranges, it should be a character vector of length 2. example: c("CpGi","shores")

**Value**

a GRangesList object containing one GRanges object for flanks and one for GRanges object for the main feature. NOTE: This can not return a CompressedGRangesList at the moment because flanking regions do not have to have the same column name as the feature. CompressedGRangesList elements should resemble each other in the column content. We can not satisfy that criteria for the flanks
Examples

cgi.path = system.file('extdata/chr21.CpGi.hg19.bed', package='genomation')
cgi.shores = readFeatureFlank(cgi.path)
cgi.shores

readGeneric

Read a tabular file and convert it to GRanges.

Description

The function reads a tabular text file that contains location and other information on genomic features and returns a GRanges object. The minimal information that the file has to have is chromosome, start and end columns. Strand information is not compulsory.

Usage

readGeneric(file, chr = 1, start = 2, end = 3, strand = NULL,
meta.cols = NULL, keep.all.metadata = FALSE, zero.based = FALSE,
remove.unusual = FALSE, header = FALSE, skip = 0, sep = "\t")

Arguments

file location of the file, a character string such as: "/home/user/my.bed" or the input itself as a string (containing at least one \n).
chr number of the column that has chromosomes information in the table (Def:1)
start number of the column that has start coordinates in the table (Def:2)
end number of the column that has end coordinates in the table (Def:3)
strand number of the column that has strand information, only -/+ is accepted (Default:NULL)
meta.cols named list that maps column numbers to meta data columns. e.g. list(name=5, score=10), which means 5th column will be named "name", and 10th column will be named "score" and their contents will be a part of the returned GRanges object. If header = TRUE, meta.cols parameter will over-write the column names given by the header line of the data frame.
keep.all.metadata logical determining if the extra columns ( the ones that are not designated by chr,start,end,strand and meta.cols arguments ) should be kept or not. (Default:FALSE)
zero.based a boolean which tells whether the ranges in the bed file are 0 or 1 base encoded. (Default: FALSE)
remove.unusual if TRUE(default) remove the chromosomes with unusual names, such as chrX_random (Default:FALSE)
header whether the original file contains a header line which designates the column names. If TRUE header will be used to construct column names. These names can be over written by meta.cols argument.
**readNarrowPeak**

A function to read the Encode formatted narrowPeak file into a GRanges object

### Description

A function to read the Encode formatted narrowPeak file into a GRanges object

### Usage

`readNarrowPeak(file, track.line=FALSE, zero.based=TRUE)`

### Arguments

- **file**: an absolute or relative path to a bed file formatted by the Encode narrowPeak standard. The file can end in .gz, .bz2, .xz, or .zip and/or start with http:// or ftp://. If the file is not compressed it can also start with https:// or ftpps://.
- **track.line**: the number of track lines to skip, "auto" to detect them automatically or FALSE(default) if the bed file doesn’t have track lines
- **zero.based**: a boolean which tells whether the ranges in the bed file are 0 or 1 base encoded. (Default: TRUE)

### Value

a GRanges object

**skip**

number of lines to skip. If there is a header line(s) you do not wish to include you can use skip argument to skip that line.

**sep**

a single character which designates the separator in the file. The default value is tab.

### Examples

```r
my.file = system.file("extdata", "chr21.refseq.hg19.bed", package="genomation")
refseq = readGeneric(my.file, chr=1, start=2, end=3, strand=NULL, 
                      meta.cols=list(score=5, name=4),
                      keep.all.metadata=FALSE, zero.based=TRUE)
head(refseq)
```
Examples

narrow.peak.file = system.file('extdata','ex.narrowPeak', package='genomation')

narrow.peak = readBroadPeak(narrow.peak.file)
head(narrow.peak)

readTableFast

readTableFast function

Description

fast reading of big tables chr indicates index of column of chromosomes

Usage

readTableFast(filename, header = TRUE, skip = 0, sep = "\t", chr = 1)

Arguments

filename file name
header logical
skip numeric
sep character
chr numeric

readTranscriptFeatures

Function for reading exon intron and promoter structure from a given bed file

Description

Function for reading exon intron and promoter structure from a given bed file

Usage

readTranscriptFeatures(location,remove.unusual=TRUE,
                        up.flank=1000,down.flank=1000,unique.prom=TRUE)

## S4 method for signature 'character'
readTranscriptFeatures(location, remove.unusual = TRUE,
                        up.flank = 1000, down.flank = 1000, unique.prom = TRUE)
Arguments

location: location of the bed file with 12 or more columns. The file can end in .gz, .bz2, .xz, or .zip and/or start with http:// or ftp://. If the file is not compressed it can also start with https:// or ftps://.

remove.unusual: remove the chromosomes with unusual names, mainly random chromosomes etc.

up.flank: up-stream from TSS to detect promoter boundaries

down.flank: down-stream from TSS to detect promoter boundaries

unique.prom: get only the unique promoters, promoter boundaries will not have a gene name if you set this option to be TRUE

Value

a GRangesList containing locations of exon/intron/promoter/TSS

Note

one bed track per file is only accepted, the bed files with multiple tracks will cause an error

Examples

my.bed12.file = system.file("extdata/chr21.refseq.hg19.bed", package = "genomation")
my.bed12.file
feats = readTranscriptFeatures(my.bed12.file)
names(feats)
sapply(feats, head)

scaleScoreMatrix(mat, columns = FALSE, rows = TRUE, scalefun = NULL)

scaleScoreMatrix Scales the values in the matrix by rows and/or columns

Description

Scales the values in the matrix by rows and/or columns

Usage

scaleScoreMatrix(mat, columns = FALSE, rows = TRUE, scalefun = NULL)

## S4 method for signature 'ScoreMatrix'
scaleScoreMatrix(mat, columns = FALSE, rows = TRUE, scalefun = NULL)
Arguments
mat ScoreMatrix object
columns columns whether to scale the matrix by columns. Set by default to FALSE.
rows rows Whether to scale the matrix by rows. Set by default to TRUE
scalefun function object that takes as input a matrix and returns a matrix. By default the argument is set to \((x - \text{mean}(x))/(\max(x)-\text{min}(x)+1)\)

Value
ScoreMatrix object

Examples
# scale the rows of a scoreMatrix object
library(GenomicRanges)
target = GRanges(rep(c(1,2),each=7), IRanges(rep(c(1,1,2,3,7,8,9), times=2), width=5),
weight = rep(c(1,2),each=7),
strand=c('-','-','-','-','+','-','-', '', '', '', '-','+'))
windows = GRanges(rep(c(1,2),each=2), IRanges(rep(c(1,2), times=2), width=5),
strand=c('-','+','-', '+'))
sm = ScoreMatrix(target, windows)
ssm = scaleScoreMatrix(sm, rows=TRUE)

scaleScoreMatrixList Scale the ScoreMatrixList

Description
Scales each ScoreMatrix in the ScoreMatrixList object, by rows and/or columns

Usage
scaleScoreMatrixList(sml, columns, rows, scalefun)

## S4 method for signature 'ScoreMatrixList'
scaleScoreMatrixList(sml, columns = FALSE,
rows = TRUE, scalefun = NULL)

Arguments
sml a ScoreMatrixList object
columns a columns whether to scale the matrix by columns. Set by default to FALSE
rows a rows Whether to scale the matrix by rows. Set by default to TRUE
scalefun a function object that takes as input a matrix and returns a matrix. By default the argument is set to the R scale function with center=TRUE and scale=TRUE
ScoreMatrix

Value

ScoreMatrixList object

Examples

library(GenomicRanges)
data(cage)
data(cpgi)
data(promoters)
cage$tpm = NULL
targets = GRangesList(cage=cage, cpgi=cpgi)
sml = ScoreMatrixList(targets, promoters, bin.num=10, strand.aware=TRUE)
sml.scaled = scaleScoreMatrixList(sml, rows=TRUE)
sml.scaled

multiHeatMatrix(sml)

ScoreMatrix

Get base-pair score for bases in each window

Description

The function produces a base-pair resolution matrix of scores for given equal width windows of interest. The returned matrix can be used to draw meta profiles or heatmap of read coverage or wig track-like data. The windows argument can be a predefined region around transcription start sites or other regions of interest that have equal lengths. The function removes all window that fall off the Rle object - have the start coordinate < 1 or end coordinate > length(Rle). The function takes the intersection of names in the Rle and GRanges objects. On Windows OS the function will give an error if the target is a file in .bigWig format.

Usage

ScoreMatrix(target, windows, strand.aware = FALSE, weight.col = NULL, is.noCovNA = FALSE, type = "auto", rpm = FALSE, unique = FALSE, extend = 0, param = NULL, bam.paired.end = FALSE, library.size = NULL)
\S4method{ScoreMatrix}{RleList,GRanges}(target,windows,strand.aware)
\S4method{ScoreMatrix}{GRanges,GRanges}(target, windows, strand.aware, weight.col, is.noCovNA)
\S4method{ScoreMatrix}{character,GRanges}(target, windows, strand.aware, weight.col=NULL,is.noCovNA=FALSE, type='auto', rpm=FALSE, unique=FALSE, extend=0, param=NULL,
ScoreMatrix

\[
\text{bam.paired.end=FALSE,} \\
\text{library.size=NULL)}
\]

Arguments

- **target**: RleList, GRanges, a BAM file or a BigWig to be overlapped with ranges in windows.
- **windows**: GRanges object that contains the windows of interest. It could be promoters, CpG islands, exons, introns. However the sizes of windows have to be equal.
- **strand.aware**: If TRUE (default: FALSE), the strands of the windows will be taken into account in the resulting ScoreMatrix. If the strand of a window is - , the values of the bins for that window will be reversed.
- **weight.col**: If the object is GRanges object a numeric column in meta data part can be used as weights. This is particularly useful when genomic regions have scores other than their coverage values, such as percent methylation, conservation scores, GC content, etc.
- **is.noCovNA**: (Default:FALSE) if TRUE, and if 'target' is a GRanges object with 'weight.col' provided, the bases that are uncovered will be preserved as NA in the returned object. This useful for situations where you can not have coverage all over the genome, such as CpG methylation values.
- **type**: (Default:"auto") if target is a character vector of file paths, then type designates the type of the corresponding files (bam or bigWig).
- **rpm**: boolean telling whether to normalize the coverage to per milion reads. FALSE by default. See library.size.
- **unique**: boolean which tells the function to remove duplicated reads based on chr, start, end and strand.
- **extend**: numeric which tells the function to extend the reads to width=extend.
- **param**: ScanBamParam object.
- **bam.paired.end**: boolean indicating whether given BAM file contains paired-end reads (default:FALSE). Paired-reads will be treated as fragments.
- **library.size**: numeric indicating total number of mapped reads in a BAM file (rpm has to be set to TRUE). If is not given (default: NULL) then library size is calculated using the Rsamtools idxstatsBam function: sum(idxstatsBam(target)$mapped).

Value

returns a ScoreMatrix object.

Note

We assume that a paired-end BAM file contains reads with unique ids and we remove both mates of reads if they are repeated. Due to the fact that ScoreMatrix uses the GenomicAlignments:readGAlignmentPairs function to read paired-end BAM files a duplication of reads occurs when mates of one pair map into two different windows.

Strands of reads in a paired-end BAM are inferred depending on strand of first alignment from the pair. This is a default setting in the GenomicAlignments:readGAlignmentPairs function (see a
strandMode argument). This mode should be used when the paired-end data was generated using one of the following stranded protocols: Directional Illumina (Ligation), Standard SOLiD.

See Also

ScoreMatrixBin

Examples

# When target is GRanges
data(cage)
data(promoters)
scores1=ScoreMatrix(target=cage,windows=promoters,strand.aware=TRUE,
weight.col="tpm")

# When target is RleList
library(GenomicRanges)
covs = coverage(cage)
scores2 = ScoreMatrix(target=covs,windows=promoters,strand.aware=TRUE)
scores2

# When target is a bam file
bam.file = system.file('unitTests/test.bam', package='genomation')
windows = GRanges(rep(c(1,2),each=2), IRange(rep(c(1,2), times=2), width=5))
scores3 = ScoreMatrix(target=bam.file, windows=windows, type='bam')
scores3

ScoreMatrix-class

An S4 class for storing ScoreMatrix function results

Description

The resulting object is an extension of a matrix object, and stores values (typically genome-wide scores) for a predefined set of regions. Each row on the ScoreMatrix is a predefined region (Ex: CpG islands, promoters) and columns are values across those regions.

Constructors

see ScoreMatrix

Coercion

as(from, "matrix"): Creates a matrix from ScoreMatrix object. You can also use S3Part() function to extract the matrix from ScoreMatrix object.
Subsetting

In the code snippets below, x is a ScoreMatrix object. `x[i,j]`: Get or set elements from row i and column j and return a subset ScoreMatrix object.

See Also

ScoreMatrix

---

**Description**

The function first bins each window to equal number of bins, and calculates the a summary matrix for scores of each bin (currently, mean, max and min supported) A scoreMatrix object can be used to draw average profiles or heatmap of read coverage or wig track-like data. windows can be a predefined region such as CpG islands, gene bodies, transcripts or CDS (coding sequences) that are not necessarily equi-width. Each window will be chopped to equal number of bins based on bin.num option.

**Usage**

```r
ScoreMatrixBin(target, windows, bin.num = 10, bin.op = "mean", strand.aware = FALSE, weight.col = NULL, is.noCovNA = FALSE, type = "auto", rpm = FALSE, unique = FALSE, extend = 0, param = NULL, bam.paired.end = FALSE, library.size = NULL)
```

- `ScoreMatrixBin(RleList, GRanges)(target, windows, bin.num, bin.op, strand.aware)`
- `ScoreMatrixBin(GRanges, GRanges)(target, windows, bin.num, bin.op, strand.aware, weight.col, is.noCovNA)`
- `ScoreMatrixBin(character, GRanges)(target, windows, bin.num=10, bin.op='mean', strand.aware, is.noCovNA=FALSE, type='auto', rpm, unique, extend, param, bampaired.end=FALSE, library.size=NULL)`
- `ScoreMatrixBin(RleList, GRangesList)(target, windows, bin.num, bin.op, strand.aware)`
- `ScoreMatrixBin(GRanges, GRangesList)(target, windows, bin.num, bin.op, strand.aware)`
\S4method{ScoreMatrixBin}{character,GRangesList}(target, windows, bin.num=10, bin.op = "mean", strand.aware, weight.col=NULL, is.noCovNA=FALSE, type='auto', rpm, unique, extend, param, bam.paired.end=FALSE, library.size=NULL)

Arguments

target  RleList, GRanges, a BAM file or a bigWig file object to be overlapped with ranges in windows

windows  GRanges or GRangesList object that contains the windows of interest. It could be promoters, CpG islands, exons, introns as GRanges object or GrangesList object representing exons of each transcript. Exons must be ordered by ascending rank by their position in transcript. The sizes of windows does NOT have to be equal.

bin.num  single integer value denoting how many bins there should be for each window

bin.op  bin operation that is either one of the following strings: "max","min","mean","median","sum". The operation is applied on the values in the bin. Defaults to "mean"

strand.aware  If TRUE (default: FALSE), the strands of the windows will be taken into account in the resulting scoreMatrix. If the strand of a window is -, the values of the bins for that window will be reversed

weight.col  if the object is GRanges object a numeric column in meta data part can be used as weights. This is particularly useful when genomic regions have scores other than their coverage values, such as percent methylation, conservation scores, GC content, etc.

is.noCovNA  (Default:FALSE) if TRUE, and if ‘target’ is a GRanges object with ‘weight.col’ provided, the bases that are uncovered will be preserved as NA in the returned object. This useful for situations where you can not have coverage all over the genome, such as CpG methylation values.

type  (Default:"auto") if target is a character vector of file paths, then type designates the type of the corresponding files (bam or bigWig)

rpm  boolean telling whether to normalize the coverage to per milion reads. FALSE by default. See library.size.

unique  boolean which tells the function to remove duplicated reads based on chr, start, end and strand

extend  numeric which tells the function to extend the reads to width=extend

param  Scan BamParam object

bam.paired.end  boolean indicating whether given BAM file contains paired-end reads (default:FALSE). Paired-reads will be treated as fragments.
library.size numeric indicating total number of mapped reads in a BAM file (rpm has to be set to TRUE). If is not given (default: NULL) then library size is calculated using the Rsamtools idxstatsBam function: sum(idxstatsBam(target)$mapped).

Value
returns a scoreMatrix object

See Also
ScoreMatrix

Examples

data(cage)
data(cpgi)
data(promoters)
myMat=ScoreMatrixBin(target=cage,
windows=cpgi,bin.num=10,bin.op="mean",weight.col="tpm")

plot(colMeans(myMat,na.rm=TRUE),type="1")

myMat2=ScoreMatrixBin(target=cage,
windows=promoters,bin.num=10,bin.op="mean",
weight.col="tpm",strand.aware=TRUE)

plot(colMeans(myMat2,na.rm=TRUE),type="1")

# Compute transcript coverage of a set of exons.
library(GenomicRanges)
bed.file = system.file("extdata/chr21.refseq.hg19.bed",
package = "genomation")
gene.parts = readTranscriptFeatures(bed.file)
transcripts = split(gene.parts$exons, gene.parts$exons$name)
transcripts = transcripts[
myMat3 = ScoreMatrixBin(target=cage, windows=transcripts[1:250],
bin.num=10)
myMat3

ScoreMatrixList Make ScoreMatrixList from multiple targets

Description
The function constructs a list of ScoreMatrix objects in the form of ScoreMatrixList object. This object can be visualized using multiHeatMatrix, heatMeta or plotMeta
ScoreMatrixList

Usage

ScoreMatrixList(targets, windows = NULL, bin.num = NULL, bin.op = "mean", strand.aware = FALSE, weight.col = NULL, is.noCovNA = FALSE, type = "auto", rpm = FALSE, unique = FALSE, extend = 0, param = NULL, library.size = NULL, cores = 1)

Arguments

targets can be a list of scoreMatrix objects, that are coerced to the ScoreMatrixList, a list of RleList objects, or a character vector specifying the locations of multiple bam files or bigWig files that are used to construct the scoreMatrixList. If it is either a RleList object or a character vector of files, it is obligatory to give a windows argument.

windows GenomicRanges containing viewpoints for the scoreMatrix or ScoreMatrixList functions

bin.num an integer telling the number of bins to bin the score matrix

bin.op an name of the function that will be used for smoothing windows of ranges

strand.aware a boolean telling the function whether to reverse the coverage of ranges that come from - strand (e.g. when plotting enrichment around transcription start sites)

weight.col if the object is GRanges object a numeric column in meta data part can be used as weights. This is particularly useful when genomic regions have scores other than their coverage values, such as percent methylation, conservation scores, GC content, etc.

is.noCovNA (Default:FALSE) if TRUE, and if 'targets' is a GRanges object with 'weight.col' provided, the bases that are uncovered will be preserved as NA in the returned object. This useful for situations where you can not have coverage all over the genome, such as CpG methylation values.

type (Default:"auto") if targets is a character vector of file paths, then type designates the type of the corresponding files (bam or bigWig)

rpm boolean telling whether to normalize the coverage to per milion reads. FALSE by default. See library.size.

unique boolean which tells the function to remove duplicated reads based on chr, start, end and strand

extend numeric which tells the function to extend the features (i.e aligned reads) to total length of width+extend

param ScanBamParam object

library.size a numeric vector of the same length as targets indicating total number of mapped reads in BAM files (targets). If is not given (default: NULL) then library sizes for every target is calculated using the Rsamtools idxstatsBam function: sum(idxstatsBam(target)$mapped). rpm argument has to be set to TRUE.

cores the number of cores to use (default: 1)

Value

returns a ScoreMatrixList object
Examples

```r
# visualize the distribution of cage clusters and cpg islands around promoters
library(GenomicRanges)
data(cage)
data(cpgi)
data(promoters)

cage$tpm = NULL
targets = GRangesList(cage=cage, cpgi=cpgi)
sml = ScoreMatrixList(targets, promoters, bin.num=10, strand.aware=TRUE)
sml

multiHeatMatrix(sml)
```

ScoreMatrixList-class  An S4 class for storing a set of ScoreMatrix

Description

The resulting object is an extension of a list object, where each element corresponds to a score matrix object

Constructors

see ScoreMatrixList

Coercion

```r
as(from, "ScoreMatrixList"): Creates a ScoreMatrixList object from a list containing ScoreMatrix or ScoreMatrixBin objects.
```

Subsetting

In the code snippets below, x is a ScoreMatrixList object.

```r
x[[i]], x[[i]]: Get or set elements i, where i is a numeric or character vector of length 1.
```

```r
x$name, x$name: value: Get or set element name, where name is a name or character vector of length 1.
```

See Also

ScoreMatrixList
Description

show method for some of the genomatation classes

Usage

## S4 method for signature 'RandomEnrichment'
show(object)

## S4 method for signature 'AnnotationByGeneParts'
show(object)

## S4 method for signature 'AnnotationByFeature'
show(object)

## S4 method for signature 'ScoreMatrix'
show(object)

## S4 method for signature 'ScoreMatrixList'
show(object)

Arguments

object object of class RandomEnrichment

Value

Shows the dimension of the ScoreMatrix
Shows the number of matrices and their sizes

Sum_c

*Function that computes a sum value*

Description

Function that computes a sum value

Usage

Sum_c(x)
Arguments
  x       NumericVector

Description
Check if a target file is in bam and bigWig formats by looking at the file extension

Usage
target.type(target, type = "")

Arguments
  target     target file
  type       type name

[,] ScoreMatrix, ANY, ANY, ANY-method
Extract method for a ScoreMatrix object.

Description
Extract method for a ScoreMatrix object.

Usage
  ## S4 method for signature 'ScoreMatrix, ANY, ANY, ANY'
x[i, j]

Arguments
  x       the ScoreMatrix object
  i       numeric value
  j       numeric value
### 

Extract method for a `ScoreMatrixList` object.

**Usage**

```r
## S4 method for signature 'ScoreMatrixList,ANY,ANY,ANY'
x[i]
```

**Arguments**

- `x` the `ScoreMatrixList` object
- `i` numeric value
Index

* internal
  
annotatGrWithGeneParts, 8
bed12ToExons, 10
bed12ToIntrons, 10
binMax, 11
binMean, 12
binMedian, 12
binMin, 12
binner, 13
binSum, 13
checkBedValidity, 16
checkClass, 16
compressedAndUrl12temp, 17
constrainRanges, 17
detectUCSCheader, 20
distance2NearestFeature, 20
file.ext, 24
galpTo2Ranges, 25
getColors, 27
listSliceMax, 38
listSliceMean, 38
listSliceMedian, 39
listSliceMin, 39
listSliceSum, 40
Max_c, 40
Mean_c, 40
Min_c, 41
read.zip, 54
readBam, 54
readBigWig, 56
readTableFast, 60
Sum_c, 71
target.type, 72
[,] ,ScoreMatrix,ANY,ANY,ANY-method, 72
[,] ,ScoreMatrix-method
   ([,] ,ScoreMatrix,ANY,ANY,ANY-method), 72
[,] ,ScoreMatrixList,ANY,ANY,ANY-method,
BSgenome, 48
c.ScoreMatrix, 14
c.ScoreMatrixList, 14
cage, 15
calculateOverlapSignificance, 15
calculateOverlapSignificance, GRanges, GRanges-method
   (calculateOverlapSignificance), 15
checkBedValidity, 16
checkClass, 16
compressedAndUrl2temp, 17
constrainRanges, 17
convertBed2Exons, 18
convertBed2Exons, data.frame-method
   (convertBed2Exons), 18
convertBed2Introns, 18
convertBed2Introns, data.frame-method
   (convertBed2Introns), 18
convertBedDf, 19
convertBedDf, data.frame-method
   (convertBedDf), 19
cpgi, 20
detectUCSCheader, 20
distance2NearestFeature, 20
dnaStringSet, 48
enrichmentMatrix, 21
enrichmentMatrix, ScoreMatrix, ScoreMatrix-method
   (enrichmentMatrix), 21
enrichmentMatrix, ScoreMatrixList, ScoreMatrix-method
   22
enrichmentMatrix, ScoreMatrixList, ScoreMatrixList-method
   23
file.ext, 24
findFeatureComb, 24
findFeatureComb, GRangesList-method
   (findFeatureComb), 24
galpTo2Ranges, 25
genes, 26
getAssociationWithTSS, 26
getAssociationWithTSS, (getAssociationWithTSS), 26
getAssociationWithTSS, -methods
   (getAssociationWithTSS), 26
getAssociationWithTSS, AnnotationByGeneParts-method
   (getAssociationWithTSS), 26
getColors, 27
getFeatsWithTargetsStats, 27
getFeatsWithTargetsStats, AnnotationByFeature-method
   (getFeatsWithTargetsStats), 27
getFlanks, 28
getFlanks, GRanges-method (getFlanks), 28
getMembers, 7, 29, 36
getMembers, AnnotationByFeature-method
   (getMembers), 29
getRandomEnrichment, 29, 52
getRandomEnrichment, GRanges, GRanges-method
   (getRandomEnrichment), 29
getTargetAnnotationStats, 30
getTargetAnnotationStats, AnnotationByFeature-method
   (getTargetAnnotationStats), 30
gffToGRanges, 31
GRanges, 8, 15, 18–20, 26, 48, 52, 55, 58, 59
GRangesList, 8, 61
heatMatrix, 32
heatMeta, 34
heatTargetAnnotation, 7, 35
interactScoreMatrixList, 37
interactScoreMatrixList, ScoreMatrixList-method
   (interactScoreMatrixList), 37
listSliceMax, 38
listSliceMean, 38
listSliceMedian, 39
listSliceMin, 39
listSliceSum, 40
Max_c, 40
Mean_c, 40
Median_c, 41
Min_c, 41
multiHeatMatrix, 41
Ops, numeric, ScoreMatrixList-method, 44
Ops, ScoreMatrix, ScoreMatrix-method, 45
Ops, ScoreMatrixList, numeric-method, 45
Ops, ScoreMatrixList, ScoreMatrixList-method,
   46
orderBy, 46
orderBy, ScoreMatrixList-method
   (orderBy), 46
patternMatrix, 47
patternMatrix, character, DNAStringSet, ANY-method
(scaleScoreMatrixList), 62
(patternMatrix), 47

patternMatrix, character, DNAStringSet-method
(scaleScoreMatrixList, ScoreMatrixList-method
(patternMatrix), 47

ScoreMatrix, 21, 22, 45, 48, 63, 65, 66, 68,
70, 72

ScoreMatrix, character, GRanges-method
(scaleScoreMatrixList), 63

ScoreMatrix, GRanges, GRanges-method
(scaleScoreMatrixList, ScoreMatrixList-method
(patternMatrix), 47

ScoreMatrix, RleList, GRanges-method
(ScoreMatrix), 63

ScoreMatrix-class, 65

ScoreMatrixBin, 65, 66, 70

ScoreMatrixBin, character, GRanges-method
(ScoreMatrixBin), 66

ScoreMatrixBin, character, GRangesList-method
(ScoreMatrixBin), 66

ScoreMatrixBin, GRanges, GRanges-method
(ScoreMatrixBin), 66

ScoreMatrixBin, GRanges, GRangesList-method
(ScoreMatrixBin), 66

ScoreMatrixBin, RleList, GRanges-method
(ScoreMatrixBin), 66

ScoreMatrixBin, RleList, GRangesList-method
(ScoreMatrixBin), 66

ScoreMatrixList, 22, 23, 44–46, 48, 68, 70, 73

ScoreMatrixList-class, 70

show, AnnotationByFeature-method
(show, RandomEnrichment-method), 71

show, AnnotationByGeneParts-method
(show, RandomEnrichment-method), 71

show, RandomEnrichment-method, 71

show, ScoreMatrix-method
(show, RandomEnrichment-method), 71

show, ScoreMatrixList-method
(show, RandomEnrichment-method), 71

Sum_c, 71

target.type, 72

read.zip, 54

randomizeFeature, 30, 53

randomizeFeature, GRanges-method
(randomizeFeature), 53

read.Bam, 54

readBed, 55

readBigWig, 56

readBroadPeak, 56

readFeatureFlank, 57

readFeatureFlank, character-method
(readFeatureFlank), 57

readGenerics, 58

readNarrowPeak, 59

readTableFast, 60

readTranscriptFeatures, 60

readTranscriptFeatures, character-method
(readTranscriptFeatures), 60

show, AnnotationByFeature-method
(show, RandomEnrichment-method), 71

show, AnnotationByGeneParts-method
(show, RandomEnrichment-method), 71

show, RandomEnrichment-method, 71

show, ScoreMatrix-method
(show, RandomEnrichment-method), 71

show, ScoreMatrixList-method
(show, RandomEnrichment-method), 71

show, ScoreMatrixList-method
(show, RandomEnrichment-method), 71

Sum_c, 71

target.type, 72

scaleScoreMatrix, 61

scaleScoreMatrix, ScoreMatrix-method
(scaleScoreMatrix), 61