Package 'CSAR'

October 16, 2019

Type Package

Title Statistical tools for the analysis of ChIP-seq data

Version 1.36.0

Date 2009-11-09

Author Jose M Muino

Description Statistical tools for ChIP-seq data analysis. The package includes the statistical method described in Kaufmann et al. (2009) PLoS Biology: 7(4):e1000090. Briefly, Taking the average DNA fragment size subjected to sequencing into account, the software calculates genomic single-nucleotide read-enrichment values. After normalization, sample and control are compared using a test based on the Poisson distribution. Test statistic thresholds to control the false discovery rate are obtained through random permutation.

Depends R (>= 2.15.0), S4Vectors, IRanges, GenomeInfoDb, GenomicRanges

Maintainer Jose M Muino <jose.muino@live.com>

Suggests ShortRead, Biostrings

Imports stats, utils

License Artistic-2.0

LazyLoad yes

LazyData yes

biocViews ChIPSeq, Transcription, Genetics

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

git_branch RELEASE_3_9

git_last_commit f758569

git_last_commit_date 2019-05-02

Date/Publication 2019-10-15

R topics documented:

CSAR-package *Statistical tools for the analysis of ChIP-seq data*

Description

Statistical tools for ChIP-seq data analysis.
The package is oriented to plant organisms, and compatible with standard file formats in the plant research field.

Details

Package:	CSAR
Type:	Package
Version:	1.0
Date:	2009-11-09
License:	Artistic-2.0
LazyLoad:	yes

Author(s)

Jose M Muino

Maintainer: Jose M Muino <jose.muino@wur.nl>

References

Muino et al. (submitted). Plant ChIP-seq Analyzer: An R package for the statistcal detection of protein-bound genomic regions.
Kaufmann et al.(2009).Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biology; 7(4):e1000090.

Examples

```
##For this example we will use the a subset of the SEP3 ChIP-seq data (Kaufmann, 2009)
data("CSAR-dataset");
##We calculate the number of hits for each nucleotide posotion for the control and sample. We do that just for ch
nhitsS<-mappedReads2Nhits(sampleSEP3_test,file="sampleSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
nhitsC<-mappedReads2Nhits(controlSEP3_test,file="controlSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
```

```
##We calculate a score for each nucleotide position
test<-ChIPseqScore(control=nhitsC,sample=nhitsS)

##We calculate the candidate read-enriched regions
win<-sigWin(test)

##We generate a wig file of the results to visualize tehm in a genome browser
score2wig(test,file="test.wig")

##We calculate relative positions of read-enriched regions regarding gene position
d<-distance2Genes(win=win,gff=TAIR8_genes_test)

##We calculate table of genes with read-enriched regions, and their location
genes<-genesWithPeaks(d)

##We calculate two sets of read-enrichment scores through permutation
permutatedWinScores(nn=1,sample=sampleSEP3_test,control=controlSEP3_test,fileOutput="test",chr=c("CHR1v012
permutatedWinScores(nn=2,sample=sampleSEP3_test,control=controlSEP3_test,fileOutput="test",chr=c("CHR1v012

###Next function will get all permutated score values generated by permutatedWinScores function.
##This represent the score distribution under the null hypotesis and therefore it can be use to control the erro
nulldist<-getPermutatedWinScores(file="test",nn=1:2)

##From this distribution, several cut-off values can be calculated to control the error of our test.
##Several functions  in R can be used for this purpose.
##In this package we had implemented a simple method for the control of the error based on FDR"
getThreshold(winscores=values(win)$score,permutatedScores=nulldist,FDR=.01)
```

ChIPseqScore *Calculate read-enrichment scores for each nucleotide position*

Description

Calculate read-enrichment scores for each nucleotide position

Usage

```
ChIPseqScore(control, sample, backg = -1, file = NA, norm = 3 * 10^9,  test = "Ratio",times=1e6,digit
```

Arguments

control	data.frame structure obtained by mappedReads2Nhits
sample	data.frame structure obtained by mappedReads2Nhits
backg	Due low coverage in the control, there could be regions with no hits. Any region with a hit value lower than backg in the control will be set to the value of backg
file	Name of the file where you wan to save the results (if desired)
norm	Integer value. Number of hits will be reported by number of hits per norm nucleotides

test	Use a score based on the poisson distribution ("Poisson") or in the ratio ("Ratio")
times	To be memory efficient, CSAR will only upload to the RAM memory fragments of length times. A bigger value means more RAM memory needed but whole process will be faster
digits	Number of decimal digits used to report the score values

Details

Different sequencing efforts yield different number of sequenced reads, for this reason the "number of hits" at each nucleotide position is normalized by the total number of nucleotides sequenced. Subsequently, the number of hits for the sample is normalize to have the same mean and variance than the control, for each chromosome independently or for the whole set of chromosomes (depending of the value of normEachChrInd). Due low coverage, there could be regions with no hits. Any region with a hit value lower than backg in the control will be set to the value of backg For each nucleotide position, a read-enrichment score will be calculated with the Poisson test, or with the ratio.

Value

A list to be used for other functions of the CSAR package

chr	Chromosme names
chrL	Chromosme length (bp)
filenames	Name of the files where the score values are storaged
digits	Score values storaged on the files need to be divided by 10^digits

Author(s)

Jose M Muino, <jose.muino@wur.nl>

References

Muino et al. (submitted). Plant ChIP-seq Analyzer: An R package for the statistical detection of protein-bound genomic regions.
Kaufmann et al.(2009).Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biology; 7(4):e1000090.

See Also

CSAR-package

Examples

```
##For this example we will use the a subset of the SEP3 ChIP-seq data (Kaufmann, 2009)
data("CSAR-dataset");
##We calculate the number of hits for each nucleotide posotion for the control and sample. We do that just for ch
nhitsS<-mappedReads2Nhits(sampleSEP3_test,file="sampleSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
nhitsC<-mappedReads2Nhits(controlSEP3_test,file="controlSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))


##We calculate a score for each nucleotide position
test<-ChIPseqScore(control=nhitsC,sample=nhitsS)
```

distance2Genes *Calculate relative positions of read-enriched regions regarding gene position*

Description

Calculate relative positions of read-enrichment regions regarding gene position

Usage

```
distance2Genes(win, gff, t = 1, d1 = -3000, d2 = 1000)
```

Arguments

win	GRange structure obtained with the function sigWin
gff	Data.frame structure obtained after loading a desired gff file
t	Integer. Only distances of read-enriched regions with a score bigger than t will be considered
d1	Negative integer. Minimum relative position regarding the start of the gene to be considered
d2	Positive integer. Maximum relative position regarding the end of the gene to be considered

Value

data.frame structure where each row represents one relative position, and each column being:

peakName	read-enriched region name
p1	relative position regarding the start of the gene
p2	relative position regarding the end of the gene
gene	name of the gene
le	length (bp) of the gene

Author(s)

Jose M Muino, <jose.muino@wur.nl>

References

Muino et al. (submitted). Plant ChIP-seq Analyzer: An R package for the statistcal detection of protein-bound genomic regions.
Kaufmann et al.(2009).Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biology; 7(4):e1000090.

See Also

genesWithPeaks, CSAR-package

Examples

```
##For this example we will use the a subset of the SEP3 ChIP-seq data (Kaufmann, 2009)
data("CSAR-dataset");
##We calculate the number of hits for each nucleotide posotion for the control and sample. We do that just for ch
nhitsS<-mappedReads2Nhits(sampleSEP3_test,file="sampleSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
nhitsC<-mappedReads2Nhits(controlSEP3_test,file="controlSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))


##We calculate a score for each nucleotide position
test<-ChIPseqScore(control=nhitsC,sample=nhitsS)

##We calculate the candidate read-enriched regions
win<-sigWin(test)


##We calculate relative positions of read-enriched regions regarding gene position
d<-distance2Genes(win=win,gff=TAIR8_genes_test)
```

---

genesWithPeaks *Provide table of genes with read-enriched regions, and their location*

Description

Provide table of genes with read-enriched regions, and their location

Usage

```
genesWithPeaks(distances)
```

Arguments

distances	data.frame structure obtained by distances2Genes

Details

This function report for each gene, the maximum peak score in different regions near of the gene. The input of the function is the distances between genes and peaks calculated by distance2Genes

Value

data.frame structure with each coloumn being:

name	name of the gene
max3kb1kb	maximum score value for the region 3Kb upstream to 1Kb dowstream
u3000	maximum score value for the region 3Kb upstream to 2Kb upstream
u2000	maximum score value for the region 2Kb upstream to 1Kb upstream
u1000	maximum score value for the region 1Kb upstream to 0Kb upstream
d0	maximum score value for the region 0Kb upstream to 0Kb dowstream
d1000	maximum score value for the region 0Kb dowstream to 1Kb dowstream

Author(s)

Jose M Muino, <jose.muino@wur.nl>

References

Muino et al. (submitted). Plant ChIP-seq Analyzer: An R package for the statistcal detection of protein-bound genomic regions.
Kaufmann et al.(2009).Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biology; 7(4):e1000090.

See Also

distance2Genes,CSAR-package

Examples

```
##For this example we will use the a subset of the SEP3 ChIP-seq data (Kaufmann, 2009)
data("CSAR-dataset");
##We calculate the number of hits for each nucleotide posotion for the control and sample. We do that just for ch
nhitsS<-mappedReads2Nhits(sampleSEP3_test,file="sampleSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
nhitsC<-mappedReads2Nhits(controlSEP3_test,file="controlSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))


##We calculate a score for each nucleotide position
test<-ChIPseqScore(control=nhitsC,sample=nhitsS)

##We calculate the candidate read-enriched regions
win<-sigWin(test)

##We calculate relative positions of read-enriched regions regarding gene position
d<-distance2Genes(win=win,gff=TAIR8_genes_test)

##We calculate table of genes with read-enriched regions, and their location
genes<-genesWithPeaks(d)
```

---

getPermutatedWinScores

Obtain the read-enrichment score distribution under the null hypothesis

---

Description

Obtain the read-enrichment score distribution under the null hypothesis

Usage

```
getPermutatedWinScores(file, nn)
```

Arguments

file	Name of the file generated by permutatedWinScores
nn	ID for the multiple permutation process

Value

Numeric vector of score values under permutation

Author(s)

Jose M Muino, <jose.muino@wur.nl>

References

Muino et al. (submitted). Plant ChIP-seq Analyzer: An R package for the statistcal detection of protein-bound genomic regions.
Kaufmann et al.(2009).Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biology; 7(4):e1000090.

See Also

CSAR-package, permutatedWinScores

Examples

```
##For this example we will use the a subset of the SEP3 ChIP-seq data (Kaufmann, 2009)
data("CSAR-dataset");
##We calculate the number of hits for each nucleotide posotion for the control and sample. We do that just for ch
nhitsS<-mappedReads2Nhits(sampleSEP3_test,file="sampleSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
nhitsC<-mappedReads2Nhits(controlSEP3_test,file="controlSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))


##We calculate two sets of read-enrichment scores through permutation
permutatedWinScores(nn=1,sample=sampleSEP3_test,control=controlSEP3_test,fileOutput="test",chr=c("CHR1v012
permutatedWinScores(nn=2,sample=sampleSEP3_test,control=controlSEP3_test,fileOutput="test",chr=c("CHR1v012

###Next function will get all permutated score values generated by permutatedWinScores function.
##This represent the score distribution under the null hypotesis and therefore it can be use to control the erro
nulldist<-getPermutatedWinScores(file="test",nn=1:2)
```

---

getThreshold *Calculate the threshold value corresponding to control FDR at a desired level*

Description

Calculate the threshold value corresponding to control FDR at a desired level

Usage

```
getThreshold(winscores, permutatedScores, FDR)
```

Arguments

winscores	Numeric vector with score values obtained from the sigWin function
permutatedScores	Numeric vector with the permutated read-enrichment score values
FDR	Numeric value with the desired FDR control

Details

This is a very simple function to obtain the threshold value of our test statistic controlling FDR at a desired level. Other functions implemented in R (eg: multtest) could be more sophisticated. Basically, for each possible threshold value, the proportion of error type I is calculated assuming that the permutated score distribution is a optimal estimation of the score distribution under the null hypothesis. This is, the proportion of permutated scores exceding the considered threshold value is used as an estimation of the error type I of our statisitic. FDR is obtained as the ratio of the proportion of error type I by the proportion of significant tests.

Value

A table with the columns being:

threshold	The threshold value
p-value	The p-value obtained from the permutated score ditribution
FDR	The FDR control obtained using threshold

Author(s)

Jose M Muino, <jose.muino@wur.nl>

References

Muino et al. (submitted). Plant ChIP-seq Analyzer: An R package for the statistcal detection of protein-bound genomic regions.
Kaufmann et al.(2009).Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biology; 7(4):e1000090.

See Also

CSAR-package,getPermutatedWinScores, sigWin

Examples

```
##For this example we will use the a subset of the SEP3 ChIP-seq data (Kaufmann, 2009)
data("CSAR-dataset");
##We calculate the number of hits for each nucleotide posotion for the control and sample. We do that just for ch
nhitsS<-mappedReads2Nhits(sampleSEP3_test,file="sampleSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
nhitsC<-mappedReads2Nhits(controlSEP3_test,file="controlSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))


##We calculate a score for each nucleotide position
test<-ChIPseqScore(control=nhitsC,sample=nhitsS)

##We calculate the candidate read-enriched regions
win<-sigWin(test)
```

```
##We calculate two sets of read-enrichment scores through permutation
permutatedWinScores(nn=1,sample=sampleSEP3_test,control=controlSEP3_test,fileOutput="test",chr=c("CHR1v012
permutatedWinScores(nn=2,sample=sampleSEP3_test,control=controlSEP3_test,fileOutput="test",chr=c("CHR1v012

###Next function will get all permutated score values generated by permutatedWinScores function.
##This represent the score distribution under the null hypotesis and therefore it can be use to control the erro
nulldist<-getPermutatedWinScores(file="test",nn=1:2)

##From this distribution, several cut-off values can be calculated to control the error of our test.
##Several functions  in R can be used for this purpose.
##In this package we had implemented a simple method for the control of the error based on FDR"
getThreshold(winscores=values(win)$score,permutatedScores=nulldist,FDR=.01)
```

`loadMappedReads` *Load mapped reads*

Description

This function load the output file of a read mapping software (eg:SOAP)

Usage

```
loadMappedReads(file, format = "SOAP", header = FALSE)
```

Arguments

`file`	File name to load
`format`	Format of the file. "SOAP" for the output of the soap software and "MAQ" for the maq software. Other user formats can be provided as a character vector for the `file` column names. Columns named: "Nhits", "lengthRead", "strand", "chr", and "pos" are needed.
`header`	Logical value indicating if the first line of the file should be skipped (TRUE) or not (FALSE)

Value

data.frame structure that can be used by `mappedReads2Nhits`

Author(s)

Jose M Muino, <jose.muino@wur.nl>

References

Muino et al. (submitted). Plant ChIP-seq Analyzer: An R package for the statistcal detection of protein-bound genomic regions.
Kaufmann et al.(2009).Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biology; 7(4):e1000090.

See Also

CSAR-package

Examples

```
##We load the mapped reads:
#sample<-loadMappedReads(file=file,format="SOAP",w=300,header=F)
##where file is the name and path of the output file of the mapping process.
```

`mappedReads2Nhits` *Calculate number of overlapped extended reads per nucleotide position*

Description

Calculate number of overlapped extended reads per nucleotide position

Usage

```
mappedReads2Nhits(input, file , chr = c("chr1", "chr2", "chr3", "chr4", "chr5"), chrL = "TAIR9", w =
```

Arguments

`input`	data loaded with loadMappedReads or an AlignedRead object from the ShortRead package
`file`	Name of the file where the results will be saved. If NA the results will not be saved in a file.
`chr`	Character vector containing the chromosome names as identified on `input`.
`chrL`	Numeric vector containing the length (bp) of the chromosomes. It should be in the same order than `chr`
`w`	Integer corresponding to the desired length of the extended reads. An advised value will be the average fragment length of the DNA submitted to sequence (usually 300 bp).
`considerStrand`	Character value. "Minimum"=>Default value. Report the minimum number of hits at each nucleotide position for both strands. "Foward"=> Report the number of hits at each nucleotide position for the "foward" strands (the one denoted as "+" in `q`). "Reverse"=>Report the number of hits at each nucleotide position for the "reverse" strands (the one denoted as "-" in `q`). "Sum"=>Report the sum of number of hits at each nucleotide position for both strands.
`uniquelyMapped`	Logic value, If TRUE, only consider uniquely mapped reads.
`uniquePosition`	Logic value. If TRUE, only consider reads mapped in different positions.

Value

A list to be used for other functions of the CSAR package

`chr`	Chromosme names
`chrL`	Chromosme length (bp)

chrL_0	Number of nucleotide positions with at least one extended read
chrL_0	Number of nucleotide positions with at least one extended read
filenames	Name of the files where the Nhits values are storaged
c1	Sum of all the Nhits values for each chromosome
c2	Sum of all the Nhits square values for each chromosome

Author(s)

Jose M Muino, <jose.muino@wur.nl>

References

Muino et al. (submitted). Plant ChIP-seq Analyzer: An R package for the statistical detection of protein-bound genomic regions.
Kaufmann et al.(2009).Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biology; 7(4):e1000090.

See Also

CSAR-package

Examples

```
#For this example we will use the a subset of the SEP3 ChIP-seq data (Kaufmann, 2009)
data("CSAR-dataset");
#We calculate the number of hits for each nucleotide posotion for the sample. We do that just for chromosome chr1
nhitsS<-mappedReads2Nhits(sampleSEP3_test,file="sampleSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
```

---

permutatedWinScores *Calculate scores for permutated read-enriched regions*

Description

Calculate scores for permutated read-enriched regions

Usage

```
permutatedWinScores(nn = 1, control, sample, fileOutput, chr = c("chr1", "chr2", "chr3", "chr4", "ch
```

Arguments

nn	ID to identify each permutation
control	data.frame structure obtained by loading the mapped reads with the function LoadMappedReads()
sample	data.frame structure obtained by loading the mapped reads with the function LoadMappedReads()
fileOutput	Name of the file were the results will be written

chr	Character vector containing the chromosome names as identified on q.
chrL	Numeric vector containing the length (bp) of the chromosomes. It should be in the same order than chr
w	Integer corresponding to the desired length of the extended reads.
considerStrand	Character value. "Minimum"=>Default value. Report the minimum number of hits at each nucleotide position for both strands. "Foward"=> Report the number of hits at each nucleotide position for the "foward" strands (the one denoted as "+" in q). "Reverse"=>Report the number of hits at each nucleotide position for the "reverse" strands (the one denoted as "-" in q). "Sum"=>Report the sum of number of hits at each nucleotide position for both strands.
uniquelyMapped	Logic value, If TRUE, only consider unquely mapped reads.
uniquePosition	Logic value. If TRUE, only consider reads mapped in different positions.
norm	Integer value. Number of hits will be reported by number of hits per norm nucleotides
backg	Any region with a hit value lower than backg in the control will be set to the value of backg
t	Numeric value. Read-enriched regions are calculated as genomic regions with score values bigger than t
g	Integer value. The maximum gap allowed between regions. Regions that are less than g bps away will be merged.
times	To be memory efficient, CSAR will only upload to the RAM memory fragments of length times. A bigger value means more RAM memory needed but whole process will be faster
digits	Number of decimal digits used to report the score values
test	Use a score based on the poisson distribution ("Poisson") or in the ratio ("Ratio")

Details

The parameter values should be the same than the one used in sigWin, ChIPseqScore, and mappedReads2Nhits. The label "control" and "sample" is asigned to each read to identify from which group they came. Labels are randomly permutated, and read-enriched regions for this new permuated dataset are calculated.

Value

The file filePutput is created with its values being the permuated score values.

Author(s)

Jose M Muino, <jose.muino@wur.nl>

References

Muino et al. (submitted). Plant ChIP-seq Analyzer: An R package for the statistcal detection of protein-bound genomic regions.
Kaufmann et al.(2009).Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biology; 7(4):e1000090.

See Also

CSAR-package,getPermutatedWinScores

Examples

```
##For this example we will use the a subset of the SEP3 ChIP-seq data (Kaufmann, 2009)
data("CSAR-dataset");
##We calculate the number of hits for each nucleotide posotion for the control and sample. We do that just for ch
nhitsS<-mappedReads2Nhits(sampleSEP3_test,file="sampleSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
nhitsC<-mappedReads2Nhits(controlSEP3_test,file="controlSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))


##We calculate two sets of read-enrichment scores through permutation
permutatedWinScores(nn=1,sample=sampleSEP3_test,control=controlSEP3_test,fileOutput="test",chr=c("CHR1v0121
permutatedWinScores(nn=2,sample=sampleSEP3_test,control=controlSEP3_test,fileOutput="test",chr=c("CHR1v0121
```

---

sampleSEP3_test *Partial dataset of a ChIP-seq experiment*

Description

Partial dataset of a Solexa DNA library obtained from a ChIP-seq experiment in Arabidopsis

Source

Kaufmann et al. (2009) Target Genes of the MADS Transcription Factor SEPALLATA3: Integration of Developmental and Hormonal Pathways in the $Arabidopsis$ Flower. PLoS Biol 7:e1000090

Examples

```
data(CSAR-dataset)
```

---

score2wig *Save the read-enrichment scores at each nucleotide position in a .wig file format*

Description

Save the read-enrichment scores at each nucleotide position in a .wig file format that can be visualize by a genome browser (eg: Integrated Genome Browser)

Usage

```
score2wig(experiment, file, t = 2, times = 1e6,description="", name="")
```

Arguments

experiment	Output of the function ChIPseqScore
file	Name of the output .wig file
t	Only nucleotide positions with a read-enrichment score bigger than t will be reported
times	To be memory efficient, CSAR will only upload to the RAM memory fragments of length times. A bigger value means more RAM memory needed but whole process will be faster
description	Character. It adds a description to the wig file. The description will be shown by the genome browser used to visualize the wig file.
name	Character. It adds a wig to the wig file. The name will be shown by the genome browser used to visualize the wig file.

Value

None. Results are printed in a file

Author(s)

Jose M Muino, <jose.muino@wur.nl>

References

Muino et al. (submitted). Plant ChIP-seq Analyzer: An R package for the statistcal detection of protein-bound genomic regions.
Kaufmann et al.(2009).Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biology; 7(4):e1000090.

See Also

CSAR-package

Examples

```
##For this example we will use the a subset of the SEP3 ChIP-seq data (Kaufmann, 2009)
data("CSAR-dataset");
##We calculate the number of hits for each nucleotide position for the control and sample. We do that just for ch
nhitsS<-mappedReads2Nhits(sampleSEP3_test,file="sampleSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
nhitsC<-mappedReads2Nhits(controlSEP3_test,file="controlSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))

##Since we will not need the raw data anymore, we could delete it from the RAM memory
rm(sampleSEP3_test,controlSEP3_test);gc(verbose=FALSE)
##We calculate a score for each nucleotide position
test<-ChIPseqScore(control=nhitsC,sample=nhitsS)

##We generate a wig file of the results to visualize them in a genome browser
score2wig(test,file="test.wig")
```

sigWin *Calculate regions of read-enrichment*

Description

Calculate regions of read-enrichment

Usage

```
sigWin(experiment, t = 1, g = 100)
```

Arguments

experiment	Output of the function ChIPseqScore
t	Numeric value. Read-enriched regions are calculated as genomic regions with score values bigger than t
g	Integer value. The maximum gap allowed between regions. Regions that are less than g bps away will be merged.

Value

An object of type'GRange' with its values being:

seqnames	Chromosome name
ranges	An IRanges object indicating start and end of the read-enriched region
posPeak	Position of the maximum score value on the read-enriched region
score	Maximum score value on the read-enriched region

Author(s)

Jose M Muino, <jose.muino@wur.nl>

References

Muino et al. (submitted). Plant ChIP-seq Analyzer: An R package for the statistcal detection of protein-bound genomic regions.
Kaufmann et al.(2009).Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. PLoS Biology; 7(4):e1000090.

See Also

CSAR-package

Examples

```
##For this example we will use the a subset of the SEP3 ChIP-seq data (Kaufmann, 2009)
data("CSAR-dataset");
##We calculate the number of hits for each nucleotide posotion for the control and sample. We do that just for ch
nhitsS<-mappedReads2Nhits(sampleSEP3_test,file="sampleSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
nhitsC<-mappedReads2Nhits(controlSEP3_test,file="controlSEP3_test",chr=c("CHR1v01212004"),chrL=c(10000))
```

```
##We calculate a score for each nucleotide position
test<-ChIPseqScore(control=nhitsC,sample=nhitsS)

##We calculate the candidate read-enriched regions
win<-sigWin(test)
```

Index