

Package ‘TransView’

April 15, 2017

Type Package

Title Read density map construction and accession. Visualization of
ChIPSeq and RNASeq data sets

Version 1.18.0

Author Julius Muller

Maintainer Julius Muller <ju-mu@alumni.ethz.ch>

Description This package provides efficient tools to generate, access
and display read densities of sequencing based data sets such
as from RNA-Seq and ChIP-Seq.

URL <http://bioconductor.org/packages/release/bioc/html/TransView.html>

License GPL-3

LazyLoad yes

Depends methods, GenomicRanges

Imports BiocGenerics, S4Vectors (>= 0.9.25), IRanges, Rsamtools (>=
1.19.38), zlibbioc, gplots

Suggests RUnit, pasillaBamSubset

LinkingTo Rsamtools

biocViews DNAMethylation, GeneExpression, Transcription, Microarray,
Sequencing, Sequencing, ChIPSeq, RNASeq, MethylSeq, DataImport,
Visualization, Clustering, MultipleComparison

NeedsCompilation yes

R topics documented:

TransView-package	2
annotatePeaks	3
DensityContainer-class	4
gtf2gr	6
histogram-methods	7
macs2gr	8
meltPeak	9
parseReads	10
peak2tss	12
plotTV	13
plotTVData	15

rmTV	16
slice1	17
slice1T	19
TVResults-class	20
tvStats-methods	21

Index 23

TransView-package	<i>Read density map construction and accession. Visualization of ChIPSeq and RNASeq data sets.</i>
-------------------	--

Description

This package provides efficient tools to generate, access and display read densities of sequencing based data sets such as from RNA-Seq and ChIP-Seq.

Details

Package: TransView
 Type: Package
 Version: 1.7.4
 URL: <http://bioconductor.org/packages/release/bioc/html/TransView.html>
 License: GPL-3
 LazyLoad: yes
 Depends: methods, GenomicRanges
 Imports: Rsamtools, zlibbioc, gplots, IRanges
 Suggests: RUnit, pasillaBamSubset
 biocViews: Bioinformatics, DNAMethylation, GeneExpression, Transcription, Microarray, Sequencing, HighThroughputS
 LinkingTo: Rsamtools

Index:

DensityContainer-class

Class "DensityContainer"

TVResults-class

Class "TVResults"

TransView-package

The TransView package: Construction and visualisation of read density maps.

annotatePeaks

Associates peaks to TSS

gtf2gr

GTF file parsing

histogram-methods

Histogram of the read distribution

macs2gr

Convenience function for MACS output conversion

parseReads

User configurable efficient assembly of read density maps

peak2tss

Changes the peak center to the TSS

plotTV

Plot and cluster global read densities

plotTVData

Summarize plotTV results

rmTV

Free space occupied by DensityContainer

slice1

Slice read densities from a TransView dataset

slice1T

Slice read densities of whole transcripts from

tvStats-methods a TransView DensityContainer
 DensityContainer accessor function

Further information is available in the following vignettes:

TransView An introduction to TransView (source, pdf)

Author(s)

Julius Muller

Maintainer: Julius Muller <ju-mu@alumni.ethz.ch>

Examples

#see vignette

annotatePeaks *Associates peaks to TSS*

Description

A convenience function to associate the peak center to a TSS or gene body provided by a gtf file.

Usage

annotatePeaks(peaks, gtf, limit=c(-10e3,10e3), remove_unmatched=T, unifyBy=F, unify_fun="mean", m

Arguments

- peaks A [GRanges](#) object.
- gtf A [GRanges](#) object with a meta data column ‘transcript_id’ and ‘exon_id’ like e.g. from gtf2gr.
- limit Maximal distance range for a peak - TSS association in base pairs.
- remove_unmatched If TRUE, only TSS associated peaks will be returned.
- unifyBy If a transcript has multiple isoforms, the peak will be associated arbitrarily to the first ID found. In order associate a peak to an isoform with specific characteristics, a DensityContainer can be provided. The choice of the returned isoform will be made based on unify_fun.
- unify_fun A function which will choose the isoform in case of non unique peak - TSS associations. Defaults to the isoform with the highest mean score $\text{function}(x)\{\text{mean}(x)\}$.
- min_geneLength Genes with a total sum of all exons smaller than this value will not be associated to a peak.
- reference If set to ‘tss’, the transcript with the smallest distance from the TSS to the peak center will be returned. If set to ‘gene_body’ the transcript with the smallest distance from the gene body (TSS or TES) to the peak center will be returned and the distance will be zero if the peak center is located within the gene body.

Details

Convenience function to annotate a [GRanges](#) object having one row per peak from e.g. `macs2gr`. The resulting peak - TSS associations can be customized by the restricting the distance and resolving multiple matches using `unify_fun`.

Value

[GRanges](#) object with row names according to the peak names provided and an added or updated meta data column 'transcript_id' with the associated transcript IDs and distances.

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

Examples

```
exgtf<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="gtf.gz$")[2]
exls<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="xls$")

GTF<-gtf2gr(exgtf)
peaks<-macs2gr(exls,psize=500)
apeaks<-annotatePeaks(peaks=peaks,gtf=GTF)
apeaks.gb<-annotatePeaks(peaks=peaks,gtf=GTF,reference="gene_body")
```

DensityContainer-class

Class "DensityContainer"

Description

Container with the pointer of the actual density maps and a histogram. Inherits from internal classes storing informations about the origin and the details of the results.

Objects from the Class

Objects are created by the function `parseReads()` using an internal constructor.

Accessors

`dc` represents a "DensityContainer" instance in the following

`data_pointer(dc)`: A character string pointing to the read density map. It points to a variable in `.GlobalEnv` which is essentially a list resulting from a call to `parseReads`. The storage space can be freed with the `rmTV` function.

`ex_name(dc),ex_name(dc)<-value`: Get or set a string to define a name of this data set

`origin(dc)`: Filename of the original file

`histogram(dc)`: A histogram of read pile-ups generated across all read density maps after filtering excluding gaps.

`env(dc)`: The environment which holds the `data_pointer` target.

`spliced(dc),spliced(dc)<-bool`: This option will mark the object to be treated like a data set with spliced reads.

`readthrough_pairs(dc)`: If TRUE, paired reads will be connected from left to right and used as one long read.

`paired(dc)`: Does the source file contain reads with proper pairs?

`filtered(dc)`: Is there a range filter in place? If TRUE, slicing should be only conducted using the same filter!!

`strands(dc)`: Which strands were parsed at all. Can be "+", "-" or "both"

`filtered_reads(dc)`: FilteredReads class storing information about reads used for read density construction

`chromosomes(dc)`: Character string with the chromosomes used for map construction

`pos(dc)`: Reads used from the forward strand

`neg(dc)`: Reads used from the reverse strand

`lsize(dc)`: Total region covered by reads within the densities returned

`gsize(dc)`: Equals to the sum of the length of all ranges from 0 to the last read per chromosome within the chromosome.

`lcoverage(dc)`: Local coverage within the densities returned which is computed by local map-mass/lsize

`lmaxScore(dc)`: Maximum read pileup within the density maps after filtering

`fmapmass(dc)`: Total map mass after quality filtering present in the file. Equals to `filtered_reads*read length`

`nreads(dc)`: Total number of reads in the file.

`coverage(dc)`: Total coverage computed by total map mass/(chromosome end - chromosome start). Chromosome length derived from the SAM/BAM header

`maxScore(dc)`: Maximum read pileup found in file after quality filtering

`lowqual(dc)`: Amount of reads that did not pass the quality score set by `min_quality` or were not mapped

`paired_reads(dc)`: Amount of reads having multiple segments in sequencing

`proper_pairs(dc)`: Amount of pairs with each segment properly aligned according to the aligner

`collapsed(dc)`: If `maxDups` is in place, the reads at the same position and strand exceeding this value will be counted here.

`size(dc)`: Size in bytes occupied by the object.

Slice Methods

slice1 signature(`dc = "DensityContainer"`): Fetch a slice of read densities.

slice1T signature(`dc = "DensityContainer"`): Recover the structure of a gene from a provided pre-processed GTF and read densities.

sliceN signature(`dc = "DensityContainer"`, `ranges = "data.frame"`): Like `slice1` but optimized for repeated slicing.

sliceNT signature(`dc = "DensityContainer"`, `tnames = "character"`, `gtf = "data.frame"`): Like `slice1T` but optimized for repeated slicing.

Convenience Methods

tvStats signature(`dc = "DensityContainer"`): Returns a list of important metrics about the source file.

Extends

Class TransView, directly.

Note

Class TotalReads and FilteredReads are not exported but their slots can be fully accessed by several accessors and the tvStats() method.

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

See Also

[tvStats-methods](#), [slice1-methods](#), [sliceN-methods](#), [histogram-methods](#), [rmTV-methods](#)

Examples

```
showClass("DensityContainer")
```

 gtf2gr

GTF file parsing

Description

Conversion of a gtf file from UCSC or ENSEMBL to a [GRanges](#) object maintaining the exon structure per transcript.

Usage

```
gtf2gr(gtf_file, chromosomes=NA, refseq_nm=F, gtf_feature=c("exon"), transcript_id="transcript_id")
```

Arguments

gtf_file	Character string with the filename of the gtf file. Fileformats from USCS and ENSEMBL are supported and gzip compression is supported.
chromosomes	A character vector with the chromosomes. Restricts the output to the case insensitive matching chromosomes.
refseq_nm	An option for GTF files based on RefSeq annotation. If TRUE only identifiers beginning with NM_ will be used.
gtf_feature	Defines the GTF feature types to be returned.
transcript_id	Defines name of the attribute within the attribute list which should be used as transcript IDs.
gene_id	Defines name of the attribute within the attribute list which should be used as gene IDs.

Details

This function parses GTF files generated by the UCSC table browser or downloaded from the ENSEMBL ftp server. It uses only rows with a 'exon' tag in the feature column (3rd column). The transcript name will be generated from the 'transcript' entry in the attribute column (9th column). The exons of each transcript are numbered using the `make.unique` function on the transcript name and used as row names.

Value

GenomicRanges object with one row per exon. `rownames` are transcript IDs and an `exon_id` is provided.

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

Examples

```
exgtf<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="gtf.gz$")
GTF.mm9<-gtf2gr(exgtf[2])
head(GTF.mm9)
```

histogram-methods

Histogram of the read distribution

Description

Retrieves the histogram computed by the `parseReads` function

Usage

```
## S4 method for signature 'DensityContainer'
histogram(dc)
```

Arguments

`dc` An object of class [DensityContainer](#).

Details

The histogram is computed by taking the running average within a window of window size as specified by the argument `hwindow` to the function `parseReads()`. The histogram is only counting local reads within the read density maps and outside of gaps or outside of possible range filters that might be in place.

Value

Returns a numeric vector with the histogram in 1Bp resolution starting from 0.

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

macs2gr

Convenience function for MACS output conversion

Description

Parses the output of MACS Peak finding algorithm and returns a [GRanges](#) object compatible to the down stream functions of TransView

Usage

```
macs2gr(macs_peaks_xls, psize, amount="all", min_pileup=0, log10qval=0, log10pval=0, fenrichment=
```

Arguments

macs_peaks_xls	Full path to the file ending with ‘_peaks.xls’ located in the output folder of a MACS run.
psize	An integer setting the total length of the peaks. Setting psize to ‘preserve’ will keep the original peak lengths from the output file and override peak_mid. Note that this is not compatible with plotTV
amount	Amount of peaks returned. If an integer is provided, the returned peaks will be limited to this amount after sorting by pile up score.
min_pileup	Minimum pile up.
log10qval	Minimal log10 q-value
log10pval	Minimal log10 p-value
fenrichment	Minimal enrichment.
peak_mid	If set to ‘summit’, the peaks with length psize will centered on the peak summit. If set to ‘center’, the mid point of start and end will be used.

Details

Convenience function parsing the output of a MACS file. Tested with MACS v1.4 and v.2.09

Value

GRanges object with one row per peak and meta data score, enrichment and log10 pvalue.

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

Examples

```
exls<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="xls$")
peaks<-macs2gr(exls,psize=500)
head(peaks)
```

meltPeak	<i>Convenience function which returns a data frame with normalized peak densities suitable for plotting with ggplot2</i>
----------	--

Description

Returns a data frame with labels and normalized densities of the provided [DensityContainer](#)

Usage

```
meltPeak(..., region, control=FALSE, peak_windows = 0, bin_method="mean", rpm=TRUE, smooth=0)
```

Arguments

...	DensityContainer objects
region	Can be one entry of the annotated output of <code>annotatePeaks</code> or a GRanges object with one entry and with a <code>transcript_id</code> and <code>distance</code> metadata column.
control	An optional vector of DensityContainer objects, that have to match the order of experiments passed as a first argument. E.g. <code>plotTV(ex1.ChIP, ex2.ChIP, control=c(ex1.Input, ...))</code> . The content will be treated as background densities and subtracted from the matching experiment.
peak_windows	If set to an integer greater than 0, all binding profiles will be interpolated into this amount of windows by the method specified by <code>bin_method</code> .
bin_method	Specifies the function used to summarize the bins specified by <code>nbins</code> . Possible methods are 'max', 'mean', 'median' or 'approx' for linear interpolation.
rpm	If set to TRUE, all sample groups will be normalized to Reads Per Million mapped reads after quality filtering according to the <code>filtered_reads</code> slot of the DensityContainer . Should not be set in truncated density maps!
smooth	If greater than 0, <code>smooth</code> defines the smoother span as described in the function <code>lowess</code> . This function will be applied to reads or RPM values, depending on <code>rpm</code> and the results will be stored in the column 'Smooth'.

Details

Convenience function which returns a data frame with one row per BP or, if `peak_window` greater than zero, per `peak_window`. The label will be taken from the `ex_name` slot of the [DensityContainer](#). The slot should be set to meaningful names before using this function. All read densities will be normalized to the total map mass and if a control is provided also background subtracted.

Value

data.frame with 3 columns: 'NormalizedReads', 'Label' and 'Position'. Optionally a column 'Smooth' will be appended.

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

Examples

```

exbam<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="bam$")
exls<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="xls$")
exgtf<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="gtf.gz$")[2]
fn.macs<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="xls$")

exden.ctrl<-parseReads(exbam[1],verbose=0)
exden.chip<-parseReads(exbam[2],verbose=0)

peaks<-macs2gr(exls,psize=500)

GTF<-gtf2gr(exgtf)
peaks<-macs2gr(fn.macs,psize=500)
peaks.anno<-annotatePeaks(peaks=peaks,gtf=GTF)

peak1.df<-meltPeak(exden.chip,region=peaks.anno["Peak.1"],bin_method="mean",peak_windows=100,rpm=TRUE)
head(peak1.df)

```

 parseReads

User configurable efficient assembly of read density maps

Description

Generates density maps for further downstream processing. Constructs a DensityContainer.

Usage

```

parseReads( filename, spliced=F, read_stranded=0, paired_only=F, readthrough_pairs=F, set_filter=
description="NA", extendreads=0, unique_only=F,max_dups=0, hwindow=1, compression=1, verbose=1 )

```

Arguments

filename	Character string with the filename of the bam file. The bam file must be sorted according to genomic position.
spliced	This option will mark the object to be treated like a data set with spliced reads. Can be switched off also for spliced experiments for special purposes. If TRUE, switches off extendreads and readthrough_pairs.
read_stranded	0 will read tags from both strands. 1 will skip all tags from the '-' strand and -1 will only utilize tags from the '-' strand
paired_only	If TRUE, any reads which are not members of a proper pair according to the 0x0002 FLAG will be discarded. If FALSE all reads will be used individually.
set_filter	Optional GRanges object or data.frame with similar structure: data.frame(chromosomes,start,end). Providing this filter will limit density maps to these regions.
min_quality	Phred-scaled mapping quality threshold. If 0, all reads will pass this filter.
extendreads	If greater 0, this amount of base pairs will be added into the strand direction of each read during density map generation.
unique_only	If TRUE, only unique reads with no multiple alignments will be used. This filter relies on the aligner to use the corresponding flag (0x100).

max_dups	If greater 0, maximally this amount of reads are allowed per start position and read direction.
description	An optional character string describing the experiment for labeling purposes.
hwindow	A numeric defining the window size used to compute the histogram. This value cannot be bigger than compression
compression	Should be left at the default value. Defines the minimal threshold in base pairs which triggers indexing and collapsing of read free regions. A smaller value leads to faster slicing at the cost of a higher memory footprint.
readthrough_pairs	Currently <i>*experimental*</i> . If TRUE, parseReads will attempt to use the region from the left to the right read of the pair for density map assembly. Requires ISIZE to be set within the BAM/SAM file.
verbose	Verbosity level

Details

parseReads uses read information of one bam file and scans the entire file read wise. Every read contributes to the density track in a user configurable manner. The resulting track will be stored in indexed integer vectors within a list. Since each score is stored as a unsigned 16bit integer, the scores can only be accessed with one of the slice methods slice1 or sliceN and not directly. As a consequence of the storage format read pile ups greater than 2^{16} will be capped and a warning will be issued.

If memory space is limiting, a filter can be supplied which will limit the density track to these regions. Filtered DensityContainer should only be sliced with the **same** regions used for parsing, since all other positions are set to 0 and can produce artificially low read counts.

Value

S4 DensityContainer

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

Examples

```
exbam<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="bam$")

#store density maps of the whole sam/bam file in test_data
exden.chip<-parseReads(exbam[2],verbose=0)

#display basic information about the content of test.sam
exden.chip

#all data are easily accessible
test_stat<-tvStats(exden.chip)
test_stat$origin

# histogram of hwindow sized windows
## Not run: histogram(exden.chip)
```

peak2tss	<i>Changes the peak center to the next TSS according to previous annotation</i>
----------	---

Description

Sets the peak boundaries of an annotated [GRanges](#) object with peak locations to TSS centered ranges based on the transcript_id column.

Usage

```
peak2tss(peaks, gtf, peak_len=500)
```

Arguments

peaks	An annotated GRanges object with a meta data column 'transcript_id' and 'exon_id' like e.g. from gtf2gr.
gtf	A GRanges object with a meta data column 'transcript_id' like e.g. from annotatePeaks.
peak_len	The desired total size of the region with the TSS located in the middle.

Details

Convenience function to change the peak centers to TSS for e.g. plotting with plotTV.

Value

A [GRanges](#) object

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

Examples

```
exgtf<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="gtf.gz$")[2]
fn.macs<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="xls$")

GTF<-gtf2gr(exgtf)
peaks<-macs2gr(fn.macs,psize=500)

peaks.anno<-annotatePeaks(peaks=peaks,gtf=GTF)

peak2tss(peaks.anno, GTF, peak_len=500)
```

plotTV	<i>Plot and cluster global read densities</i>
--------	---

Description

Plotting facility for DensityContainer.

Usage

```
plotTV( ..., regions, gtf=NA, scale="global", cluster="none", control = F, peak_windows = 0, ex_wi
bin_method="mean", show_names=T, label_size=1, zero_alpha=0.5, colr=c("white", "blue", "red"),
colr_df="redgreen", colour_spread=c(0.05, 0.05), key_limit="auto", key_limit_rna="auto",
set_zero="center", rowv=NA, gclust="peaks", norm_readc=T, no_key=F, stranded_peak=T,
ck_size=c(2,1), remove_lowex=0, verbose=1, showPlot=T, name_width=2, pre_mRNA=F)
```

Arguments

...	Depending on the combination of arguments and limited by the layout up to 20 DensityContainer and maximally one matrix can be supplied. The elements will be plotted in the order they were passed with the expression profiles and the peak profiles on the right hand and the left hand side respectively. The spliced slot determines about the kind of plot. If a matrix is provided, it will be plotted as a heatmap.
regions	GRanges object with uniformly sized regions used for plotting or character vector with IDs matching column 'transcript_id' in the GTF.
gtf	A GRanges object with a meta data column 'transcript_id' and 'exon_id' like e.g. from gtf2gr.
scale	A character string that determines the row scaling of the colors. Defaults to 'global' which results in a global maximum and minimum read value to be plotted across experiments. Alternative is 'individual' for individual scaling.
cluster	Sets the clustering method of the read densities. Defaults to 'none'. If an integer is passed, kmeans clustering will be performed with cluster defining the amount of clusters. A colour coded bar will be plotted to the left. For hierarchical clustering the options 'hc_sp' and 'hc_pe' for spearman or pearson correlation coefficient based distances respectively, or 'hc_rm' for distances based on row means are accepted and the results will be displayed as a dendrogram.
control	A vector of DensityContainer objects, matching the order of experiments passed as a first argument. E.g. plotTV(ex1.ChIP, ex2.ChIP, ex3.RNA_KO, control=c(ex1.Input, ex2.I). The content will be treated as background densities and subtracted from the matching experiment.
show_names	If TRUE, peak labels and transcript IDs will be displayed on the left and the right of the plot respectively.
label_size	Font size of the row and axis labels.
zero_alpha	Determines the alpha level of the line indicating the zero point within the peaks.
colr	A vector containing the 3 colors used for the lowest, middle and highest values respectively.
colr_df	Determines the color in case a matrix is provided and uses greenred(100) from gplots by default. If changed, the arguments should be formatted analogous to colr.

colour_spread	sets the distance of the maximum and minimum value to the saturation levels of the plot. The first value for the left side (Peak profiles) and the right for the expression plots. Can be used to adjust the contrast.
key_limit	If left at the default, the upper and lower saturation levels the peak profile colour keys will be automatically determined based on colour_spread. Can be manually overridden by a numeric vector with upper and lower levels.
key_limit_rna	If left at the default, the upper and lower saturation levels the transcript profile colour keys will be automatically determined based on colour_spread. Can be manually overridden by a numeric vector with upper and lower levels.
set_zero	if set to an integer, it determines the zero point of the x axis below the plot. E.g. a value of 250 will scale the x-axis of a 500bp peak from -250 to +250.
rowv	If a numeric vector is provided, no clustering will be performed and all rows will be ordered based on the values of this vector. Alternatively a TVResults object can be provided to reproduce previous k-means clustering.
peak_windows	If set to an integer greater than 0, all binding profiles will be interpolated into this amount of windows by the method specified by bin_method.
ex_windows	An integer that determines the amount of points at which the read densities of an expression experiment will get interpolated by the method specified by bin_method.
bin_method	Specifies the function used to summarize the bins specified by nbins. Possible methods are 'max', 'mean', 'median' or 'approx' for linear interpolation.
gclust	If cluster is not set to 'none', this character string determines the cluster group. If set to 'expression' or 'peaks', only the expression profile or peak profile data sets will be used to perform the clustering respectively. All data sets passed will be reordered based on the results of the clustering. If set to 'both', all data sets will be treated as one matrix and clustered altogether.
norm_readc	If set to TRUE, all sample groups will be normalized based on the map mass which is defined here as all mapped reads after quality filtering multiplied by their individual read length.
no_key	If TRUE, no color keys will be displayed.
stranded_peak	If TRUE and strand informations are provided in regions, peak profiles will flipped if located on the negative strand.
ck_size	Determines the size of the colour key in the form c(height,width)
remove_lowex	Numeric that sets the threshold for the average read density per base pair for expression data sets. Transcripts not passing will be filtered out and a message will be displayed.
verbose	Verbosity level
showPlot	If FALSE, plotting will be suppressed and only the TVResults will be returned.
name_width	Determines the width of the space for the peak and gene names.
pre_mRNA	All expression data will be plotted from the start of the first exon to the end of the last exon including all introns.

Details

Plots a false color image using the image function similar to heatmap.2 of **gplots** but based on read densities. There are 2 different kind of plots, that can be combined or plotted individually: expression profiles and peak profiles.

- "Peak profile plots": Peak profiles are plotted if a `DensityContainer` instance is supplied with the spliced slot set to `FALSE`. The image consists of color coded, optionally total read normalized read pileups as a stacked false color image with one peak per row. The size of the peaks is solely relying on the genomic range passed with peaks. If strand information is available through peaks, all peaks on the reverse strand will be reversed.
- "Transcript profile plots": If the spliced slot of the respective `DensityContainer` is set to `TRUE`, an expression profile will be plotted. First, each expression profile will be normalized to the total amount of reads of the source BAM/SAM file and reduced to `ex_windows` as calculated by the `approx` function. The optional clustering will then be performed and subsequently all expression profiles will be scaled across rows so that each row has a mean of zero and standard deviation of one.
- "Heatmap": Instead of a `DensityContainer` with spliced set to `TRUE`, one matrix can be provided. The data will be scaled analogous to 'Expression profile plots' and plotted as a heatmap using the `image` command.
- "Mixed plots": If `DensityContainer` instances with spliced slot set to `TRUE` or a matrix are combined with `DensityContainer` with the spliced slot set to `FALSE`, the peak profiles will be plotted on the left and the expression plots will be plotted on the right. The `gclust` argument determines the clustered groups.

Value

Returns a `TVResults` class object with the results of the clustering.

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

Examples

```
exbam<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="bam$")
exls<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="xls$")

exden.ctrl<-parseReads(exbam[1],verbose=0)
exden.chip<-parseReads(exbam[2],verbose=0)

peaks<-macs2gr(exls,psize=500)

cluster_res<-plotTV(exden.chip,exden.ctrl,regions=peaks,cluster=5,norm_readc=FALSE,showPlot=FALSE)
summary(cluster_res)
```

plotTVData

Summarize plotTV results

Description

plotTVData returns the ordering and clustering results as internally calculated by plotTV.

Usage

```
## S4 method for signature 'TVResults'
plotTVData(tvr)
```

Arguments

tvr A [TVResults](#) object as returned by plotTV

Details

If k-means or manual clustering was performed, row means per cluster will be returned in a data.frame. Otherwise row means over the whole data will be returned.

Value

Returns a data.frame of the clustering results with five columns: Position, Cluster, Sample, Average_scores and Plot

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

See Also

- [TVResults-class](#).

Examples

```
exbam<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="bam$")
exls<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="xls$")

exden.ctrl<-parseReads(exbam[1],verbose=0)
exden.chip<-parseReads(exbam[2],verbose=0)

peaks<-macs2gr(exls,psize=500)

cluster_res<-plotTV(exden.chip,exden.ctrl,regions=peaks,cluster=5,norm_readc=FALSE,showPlot=FALSE)
summaryTV(cluster_res)
tvdata<-plotTVData(cluster_res)
```

rmTV

Free space occupied by DensityContainer

Description

Free space occupied by DensityContainer

Usage

```
## S4 method for signature 'DensityContainer'
rmTV(dc)
```

Arguments

dc An object of class [DensityContainer](#).

Value

None

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

Examples

```
exbam<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="bam$")

#store density maps of the whole sam/bam file in test_data
exden.chip<-parseReads(exbam[2])

rmTV(exden.chip)
```

 slice1

Slice read densities from a TransView dataset

Description

slice1 returns read densities of a genomic interval. sliceN takes a GRanges object or a data.frame with genomic coordinates and returns a list of read densities.

Usage

```
## S4 method for signature 'DensityContainer,character,numeric,numeric'
slice1(dc, chrom, start, end, control=FALSE, input_method="-",treads_norm=TRUE, nbins=0, bin_method)
## S4 method for signature 'DensityContainer'
sliceN(dc, ranges, toRle=FALSE, control=FALSE, input_method="-",treads_norm=TRUE, nbins=0, bin_method)
```

Arguments

dc	Source DensityContainer object
chrom	A case sensitive string of the chromosome
start,end	Genomic start and end of the slice
ranges	A GRanges object or a data.frame.
toRle	The return values will be converted to a RleList.
control	An optional DensityContainer which will used as control and by default subtracted from dc.
input_method	Defines the handling of the optional control DensityContainer . '-' will subtract the control from the actual data and '/' will return log2 fold change ratios with an added pseudo count of 1 read.
treads_norm	If TRUE, the input densities are normalized to the read counts of the data set. Should not be used if one of the DensityContainer objects does not contain the whole amount of reads by e.g. placing a filter in parseReads.
nbins	If all input regions have equal length and nbins greater than 0, all densities will be summarized using the method specified by bin_method into nbins windows of approximately equal size.
bin_method	Character string that specifies the function used to summarize or expand the bins specified by nbins. Valid methods are 'max', 'mean' or 'median'.

Details

slice1 is a fast method to slice a vector of read densities from a [DensityContainer](#) object. The vector can be optionally background subtracted. If the query region exceeds chromosome boundaries or if a non matching chromosome name will be passed, a warning will be issued and a NULL vector will be returned.

sliceN returns a list with N regions corresponding to N rows in the [GRanges](#) object or the data.frame. A list with the corresponding read densities will be returned and row names will be conserved. Optionally the return values can be converted to a [RleList](#) for seamless integration into the [IRanges](#) package.

Value

slice1 returns a numeric vector of read densities sliceN returns a list of read densities and optionally an [RleList](#)

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

See Also

- [slice1T](#).
- [DensityContainer-class](#).

Examples

```
exbam<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="bam$")
exls<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="xls$")

#store density maps of the whole sam/bam file in test_data
exden.ctrl<-parseReads(exbam[1],verbose=0)
exden.chip<-parseReads(exbam[2],verbose=0)

peaks<-macs2gr(exls,psize=500)

#returns vector of read counts per base pair
slice1(exden.chip,"chr2",30663080,30663580)[300:310]
slice1(exden.ctrl,"chr2",30663080,30663580)[300:310]
slice1(exden.chip,"chr2",30663080,30663580,control=exden.ctrl,treads_norm=FALSE)[300:310]

xout<-sliceN(exden.chip,ranges=peaks)
lapply(xout,function(x)sum(x)/length(x))
xout<-sliceN(exden.ctrl,ranges=peaks)
lapply(xout,function(x)sum(x)/length(x))
xout<-sliceN(exden.chip,ranges=peaks,control=exden.ctrl,treads_norm=FALSE)
lapply(xout,function(x)sum(x)/length(x))
```

slice1T	<i>Slice read densities of whole transcripts from a TransView Density-Container</i>
---------	---

Description

slice1T returns read densities of a transcript. sliceNT takes the output of with genomic coordinates and returns a list of read densities.

Usage

```
## S4 method for signature 'DensityContainer,character'
slice1T(dc, tname, gtf, control=FALSE, input_method="-", concatenate=T, stranded=T, treads_norm=
## S4 method for signature 'DensityContainer,character'
sliceNT(dc, tnames, gtf, toRle=FALSE, control=FALSE, input_method="-", concatenate=T, stranded=T
```

Arguments

dc	Source DensityContainer object
tname, tnames	A character string or a character vector with matching identifiers of the provided gtf
gtf	A GRanges object with a meta data column 'transcript_id' and 'exon_id' like e.g. from gtf2gr.
toRle	The return values will be converted to a RleList .
control	An optional DensityContainer which will used as control and by default subtracted from dc.
input_method	Defines the handling of the optional control DensityContainer . '-' will subtract the control from the actual data and '/' will return log2 fold change ratios with an added pseudo count of 1 read.
concatenate	Logical that determines whether exons will be concatenated to one numeric vector (default) or returned as a list of vectors per exon.
stranded	If TRUE, the resulting vector will be reversed for reads on the reverse strand.
treads_norm	If TRUE, the input densities are normalized to the read counts of the data set. Should not be used if one of the DensityContainer objects does not contain the whole amount of reads by e.g. placing a filter in <code>parseReads</code> .
nbins	If all input regions have equal length and nbins greater than 0, all densities will be summarized using the method specified by <code>bin_method</code> into nbins windows of approximately equal size.
bin_method	Character string that specifies the function used to summarize or expand the bins specified by nbins. Valid methods are 'max', 'mean' or 'median'.

Details

slice1T and sliceNT provide a convenient method to access the read densities from a [Density-Container](#) of spliced reads. The transcript structure will be constructed based on the provided gtf information.

slice1T is a fast alternative to sliceNT to slice one vector of read densities corresponding to the structure of one transcript and reads can be optionally background subtracted. If the query region

exceeds chromosome boundaries or if a non matching chromosome name will be passed, a warning will be issued and a NULL vector will be returned.

sliceN slices N regions corresponding to N rows in the range GRanges object. A list with the corresponding read densities will be returned and row names will be conserved. Optionally the return values can be converted to a RleList for seamless integration into the IRanges package.

Value

sliceIT returns a numeric vector of read densities sliceNT returns a list of read densities and optionally an RleList

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

Examples

```
library("pasillaBamSubset")

exgtf<-dir(system.file("extdata", package="TransView"),full=TRUE,patt="gtf.gz$")[1]
fn.pas_paired<-untreated1_chr4()

exden.exprs<-parseReads(fn.pas_paired,spliced=TRUE,verbose=0)

GTF.dm3<-gtf2gr(exgtf)

sliceIT(exden.exprs,tname="NM_001014688",gtf=GTF.dm3,concatenate=FALSE)

my_genes<-sliceNT(exden.exprs,unique(mcols(GTF.dm3)$transcript_id[101:150]),gtf=GTF.dm3)
lapply(my_genes,function(x)sum(x)/length(x))
```

TVResults-class	Class "TVResults"
-----------------	-------------------

Description

Container holding the results of a call to plotTV().

Objects from the Class

Objects are created by the function plotTV() using an internal constructor.

Accessors

tvr represents a "TVResults" instance in the following

parameters(tvr): Holds all parameters used to call plotTV

clusters(tvr): Returns numeric vector with the cluster of each cluster.

cluster_order(tvr): Ordering of the rows within the original regions passed to plotTV with regard to the clusters.

`scores_peaks(tvr)`: Scores of the peaks. Corresponds to the values within the plot after interpolation and normalization.

`scores_rna(tvr)`: Scores of the transcripts. Corresponds to the values within the plot after interpolation and normalization.

`summaryTV(tvr)`: Returns a data frame with the clustering results of the internal data.

Convenience Methods

plotTVData `signature(tvr = "TVResults")`: Returns a data frame with summarized clustering results.

Note

Not all slots are currently being exported.

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

See Also

[plotTVData-methods](#)

Examples

```
showClass("TVResults")
```

tvStats-methods

DensityContainer accessor function

Description

Retrieve important metrics from the outcome of `parseReads()` stored in class `DensityContainer` and its super classes.

Usage

```
## S4 method for signature 'DensityContainer'  
tvStats(dc)
```

Arguments

`dc` An object of class [DensityContainer](#).

Value

Returns a list with the slots of the DensityContainer and its super classes. In detail:

- "ex_name": A user provided string to define a name of this dataset
- "origin": Filename of the original file
- "spliced": Should the class be treated like an RNA-Seq experiment for e.g. plotTV?
- "paired": Does the source file contain reads with proper pairs?
- "readthrough_pairs": If TRUE, paired reads will be connected from left to right as one long read.
- "filtered": Is there a range filter in place? If yes, slicing should be **only** conducted using the same filter!!
- "strands": Which strands were parsed at all. Can be "+", "-" or "both"
- "nreads": Total number of reads
- "coverage": Total coverage computed by total map mass/(chromosome end - chromosome start). Chromosome length derived from the SAM/BAM header
- "maxScore": Maximum read pileup found in file
- "lowqual": Amount of reads that did not pass the quality score set by min_quality or were not mapped
- "paired_reads": Amount of reads having multiple segments in sequencing
- "proper_pairs": Amount of pairs with each segment properly aligned according to the aligner
- "collapsed": If maxDups is in place, the reads at the same position and strand exceeding this value will be counted here.
- "compression": Size of a gap triggering an index event
- "chromosomes": Character string with the chromosomes with reads used for map construction
- "filtered":_reads Amount of reads
- "pos": Reads used from the forward strand
- "neg": Reads used from the reverse strand
- "lcoverage": Local coverage which is computed by filtered map mass/covered region
- "lmaxScore": Maximum score of the density maps
- "size": Size in bytes occupied by the object

Author(s)

Julius Muller <ju-mu@alumni.ethz.ch>

Index

- *Topic **classes**
 - DensityContainer-class, 4
 - TVResults-class, 20
- *Topic **methods**
 - histogram-methods, 7
 - tvStats-methods, 21
- *Topic **package**
 - TransView-package, 2
- annotatePeaks, 3
- chromosomes (DensityContainer-class), 4
- chromosomes, DensityContainer-method (DensityContainer-class), 4
- class: DensityContainer (DensityContainer-class), 4
- class: TVResults (TVResults-class), 20
- cluster_order (TVResults-class), 20
- cluster_order, TVResults-method (TVResults-class), 20
- clusters (TVResults-class), 20
- clusters, TVResults-method (TVResults-class), 20
- collapsed (DensityContainer-class), 4
- collapsed, DensityContainer-method (DensityContainer-class), 4
- compression (DensityContainer-class), 4
- compression, DensityContainer-method (DensityContainer-class), 4
- data_pointer (DensityContainer-class), 4
- data_pointer, DensityContainer-method (DensityContainer-class), 4
- DensityContainer, 7, 9, 13, 15–19, 21
- DensityContainer-class, 4, 18
- env (DensityContainer-class), 4
- env, DensityContainer-method (DensityContainer-class), 4
- ex_name (DensityContainer-class), 4
- ex_name, DensityContainer-method (DensityContainer-class), 4
- ex_name<- (DensityContainer-class), 4
- ex_name<-, DensityContainer-method (DensityContainer-class), 4
- filtered (DensityContainer-class), 4
- filtered, DensityContainer-method (DensityContainer-class), 4
- filtered_reads (DensityContainer-class), 4
- filtered_reads, DensityContainer-method (DensityContainer-class), 4
- fmapmass (DensityContainer-class), 4
- fmapmass, DensityContainer-method (DensityContainer-class), 4
- gcoverage (DensityContainer-class), 4
- gcoverage, DensityContainer-method (DensityContainer-class), 4
- GRanges, 3, 4, 6, 8–10, 12, 13, 17–19
- gsize (DensityContainer-class), 4
- gsize, DensityContainer-method (DensityContainer-class), 4
- gtf2gr, 6
- histogram (histogram-methods), 7
- histogram, DensityContainer-method (histogram-methods), 7
- histogram-methods, 7
- lcoverage (DensityContainer-class), 4
- lcoverage, DensityContainer-method (DensityContainer-class), 4
- lmaxScore (DensityContainer-class), 4
- lmaxScore, DensityContainer-method (DensityContainer-class), 4
- lowqual (DensityContainer-class), 4
- lowqual, DensityContainer-method (DensityContainer-class), 4
- lsize (DensityContainer-class), 4
- lsize, DensityContainer-method (DensityContainer-class), 4
- macs2gr, 8
- maxScore (DensityContainer-class), 4
- maxScore, DensityContainer-method (DensityContainer-class), 4
- meltPeak, 9
- neg (DensityContainer-class), 4

- neg, DensityContainer-method
(DensityContainer-class), 4
- nreads (DensityContainer-class), 4
- nreads, DensityContainer-method
(DensityContainer-class), 4
- origin (DensityContainer-class), 4
- origin, DensityContainer-method
(DensityContainer-class), 4
- paired (DensityContainer-class), 4
- paired, DensityContainer-method
(DensityContainer-class), 4
- paired_reads (DensityContainer-class), 4
- paired_reads, DensityContainer-method
(DensityContainer-class), 4
- parameters (TVResults-class), 20
- parameters, TVResults-method
(TVResults-class), 20
- parseReads, 10
- peak2tss, 12
- plotTV, 13
- plotTVData, 15
- plotTVData, TVResults (plotTVData), 15
- plotTVData, TVResults-method
(plotTVData), 15
- plotTVData-methods (plotTVData), 15
- pos (DensityContainer-class), 4
- pos, DensityContainer-method
(DensityContainer-class), 4
- proper_pairs (DensityContainer-class), 4
- proper_pairs, DensityContainer-method
(DensityContainer-class), 4
- readthrough_pairs
(DensityContainer-class), 4
- readthrough_pairs, DensityContainer-method
(DensityContainer-class), 4
- rmTV, 16
- rmTV, DensityContainer-method (rmTV), 16
- rmTV-methods (rmTV), 16
- scores_peaks (TVResults-class), 20
- scores_peaks, TVResults-method
(TVResults-class), 20
- scores_rna (TVResults-class), 20
- scores_rna, TVResults-method
(TVResults-class), 20
- show, DensityContainer-method
(DensityContainer-class), 4
- show, TVResults-method
(TVResults-class), 20
- size (DensityContainer-class), 4
- size, DensityContainer-method
(DensityContainer-class), 4
- slice1, 17
- slice1, DensityContainer, character, numeric, numeric-method
(slice1), 17
- slice1-methods (slice1), 17
- slice1T, 18, 19
- slice1T, DensityContainer, character-method
(slice1T), 19
- slice1T-methods (slice1T), 19
- sliceN (slice1), 17
- sliceN, DensityContainer-method
(slice1), 17
- sliceN-methods (slice1), 17
- sliceNT (slice1T), 19
- sliceNT, DensityContainer, character-method
(slice1T), 19
- sliceNT-methods (slice1T), 19
- spliced (DensityContainer-class), 4
- spliced, DensityContainer-method
(DensityContainer-class), 4
- spliced<- (DensityContainer-class), 4
- spliced<- , DensityContainer-method
(DensityContainer-class), 4
- strands (DensityContainer-class), 4
- strands, DensityContainer-method
(DensityContainer-class), 4
- summaryTV (TVResults-class), 20
- summaryTV, TVResults-method
(TVResults-class), 20
- TransView (TransView-package), 2
- TransView-package, 2
- TVResults, 14–16
- TVResults-class, 16, 20
- tvStats (tvStats-methods), 21
- tvStats, DensityContainer-method
(tvStats-methods), 21
- tvStats-methods, 21