

Package ‘epistack’

July 3, 2022

Title Heatmaps of Stack Profiles from Epigenetic Signals

Version 1.3.0

Description The epistack package main objective is the visualizations of stacks of genomic tracks (such as, but not restricted to, ChIP-seq, ATAC-seq, DNA methylation or genomic conservation data) centered at genomic regions of interest.

License MIT + file LICENSE

Encoding UTF-8

LazyData false

Imports GenomicRanges, SummarizedExperiment, BiocGenerics, S4Vectors, IRanges, viridisLite, graphics, plotrix, grDevices, stats, methods

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.2

Depends R (>= 4.1)

Suggests testthat (>= 3.0.0), BiocStyle, knitr, rmarkdown, EnrichedHeatmap, biomaRt, rtracklayer, covr, vdiff, magick

Config/testthat/edition 3

VignetteBuilder knitr

biocViews RNASeq, Preprocessing, ChIPSeq, GeneExpression

git_url <https://git.bioconductor.org/packages/epistack>

git_branch master

git_last_commit e2c21b2

git_last_commit_date 2022-04-26

Date/Publication 2022-07-03

Author SACI Safia [aut],
DEVAILLY Guillaume [cre]

Maintainer DEVAILLY Guillaume <gdevailly@hotmail.com>

R topics documented:

addBins	2
addMetricAndArrangeGRanges	3
addMetricAndArrangeRSE	4
GRanges2RSE	5
meanColor	6
plotAverageProfile	7
plotBinning	8
plotBoxMetric	9
plotEpistack	10
plotMetric	12
plotStackProfile	13
plotStackProfileLegend	15
redimMatrix	15
stackepi	17
stackepi_gr	17
Index	19

addBins	<i>addBins()</i>
---------	------------------

Description

Add an optional bin metadata column to `gr`, to serve as annotations for the epistack plots.

Usage

```
addBins(rse, nbins = 5L, bin = NULL)
```

Arguments

<code>rse</code>	a SummarizedExperiment or a GRanges object.
<code>nbins</code>	an integer number, the number of bins.
<code>bin</code>	a vector containing pre-determined bins, in the same order as <code>gr</code> .

Details

`nbins` is taken into account only if `bin` is `NULL`. `rse` should be sorted first, usually with the `addMetricAndArrangeGRanges()` function. `addBin(rse, bin = vec)` is equivalent to `rse$bin <- vec`, while `addBin(rse, nbins = 5)` will create 5 bins of equal size based on `rse` order.

Value

the RangedSummarizedExperiment or GRanges object with a new bin metadata column

See Also

[addMetricAndArrangeGRanges plotBinning](#)

Examples

```
data("stackepi")
addBins(stackepi)

# 3 bins instead of 5
addBins(stackepi, nbins = 3)

# assign bins using a vector
addBins(stackepi, bin = rep(c("a", "b", "c"),
  length.out = length(stackepi)))
```

```
addMetricAndArrangeGRanges
      addMetricAndArrangeGRanges()
```

Description

Perform an inner join between a GRanges object and a data.frame. Sort the resulting GRanges based on a metric column.

Usage

```
addMetricAndArrangeGRanges(
  gr,
  order,
  gr_key = "name",
  order_key = "name",
  order_value = "exp",
  shuffle_tie = TRUE
)
```

Arguments

<code>gr</code>	a GRanges object.
<code>order</code>	a data.frame with at least two columns: keys and values.
<code>gr_key</code>	name of the gr metadata column containing unique names for each genomic region in gr. Usually gene names/id or peak id.
<code>order_key</code>	name of the order column that will be used as key for the inner join.
<code>order_value</code>	name of the order column that contain value used for sorting.
<code>shuffle_tie</code>	a boolean Value (TRUE / FALSE). When TRUE, shuffle the GRanges before sorting, mixing the ties.

Details

This utility function allow the addition of a metric column to genomic regions of interest. One of its common use case is to add gene expression values on a set of transcription start sites. The resulting GRanges object will only contain regions presents in both gr and order.

Value

a GRanges sorted in descending order.

Examples

```
data("stackepi_gr")
randomOrder <- data.frame(gene_id = stackepi_gr$gene_id,
  value = rnorm(length(stackepi_gr)))
addMetricAndArrangeGRanges(stackepi_gr,
  randomOrder, gr_key = "gene_id",
  order_key = "gene_id", order_value = "value")
```

addMetricAndArrangeRSE

addMetricAndArrangeRSE()

Description

Perform an inner join between a rangedSummarizedExperiment object and a data.frame. Sort the resulting rangedSummarizedExperiment based on a metric column.

Usage

```
addMetricAndArrangeRSE(
  rse,
  order,
  rse_key = "name",
  order_key = "name",
  order_value = "exp",
  shuffle_tie = TRUE
)
```

Arguments

rse	a rangedSummarizedExperiment object.
order	a data.frame with at least two columns: keys and values.
rse_key	name of the gr metadata column containing unique names for each genomic region in rowRanges(rse). Usually gene names/id or peak id.
order_key	name of the order column that will be used as key for the inner join.

order_value name of the order column that contain value used for sorting.
 shuffle_tie a boolean Value (TRUE / FALSE). When TRUE, shuffle the GRanges before sorting, mixing the ties.

Details

This utility function allow the addition of a metric column to genomic regions of interest. One of its common use case is to add gene expression values on a set of transcription start sites. The resulting GRanges object will only contain regions presents in both rse and order.

Value

a rangedSummarizedExperiment sorted in descending order.

Examples

```
data("stackepi")
randomOrder <- data.frame(
  gene_id = SummarizedExperiment::rowRanges(stackepi)$gene_id,
  value = rnorm(length(stackepi))
)
addMetricAndArrangeRSE(stackepi,
  randomOrder, rse_key = "gene_id",
  order_key = "gene_id", order_value = "value")
```

GRanges2RSE

GRanges2RSE()

Description

Convert objects from the old input format (GRanges object) to the new recommended input format RangedSummarizedExperiment.

Usage

```
GRanges2RSE(gr, patterns, names = patterns)
```

Arguments

gr a GRanges object with matrix embeded as metadata columns.
 patterns A character vector of column prefixes (can be regular expressions) that should match columns of gr.
 names specify the desired names of the assays (if different from patterns).

Details

Mostly used for backward compatibilities and unit testing.

Value

a RangedSummarizedExperiment.

Examples

```
data("stackepi_gr")
GRanges2RSE(stackepi_gr, patterns = c("window"))
GRanges2RSE(stackepi_gr, patterns = c("^window_"), names = c("DName"))
```

meanColor

meanColor

Description

Return the average color of a vector of colors, computed in the RGB space.

Usage

```
meanColor(colors)
```

Arguments

colors a vector of colors

Details

Input colors can be either in html or color name formats. The alpha channel is supported but optional.

Value

a single color value

See Also

[redimMatrix](#)

Examples

```
meanColor(c("#000000FF", "#FFFFFF00", "#FFFF00FF", "#FF0000FF"))

# works with color names
meanColor(c("blue", "red"))

# Mix color names and HTML codes
meanColor(c("blue", "red", "#FFFF00FF"))

# works without alpha channel in inputs (but outputs an alpha channel):
meanColor(c("#000000", "#FFFFFF", "#FFFF00", "#FF0000"))
```

```
plotAverageProfile    plotAverageProfile()
```

Description

Plot the average stack profiles +/- error (sd or sem). If a bin column is present in `rowRanges(rse)`, one average profile is drawn for each bin.

Usage

```
plotAverageProfile(
  rse,
  assay = NULL,
  x_labels = c("Before", "Anchor", "After"),
  palette = colorRampPalette(c("magenta", "black", "green")),
  alpha_for_se = 0.25,
  error_type = c("sd", "sem", "ci95"),
  reversed_z_order = FALSE,
  ylim = NULL,
  pattern = NULL
)
```

Arguments

<code>rse</code>	a <code>RangedSummarizedExperiment</code> input. Alternatively: can be a <code>GRanges</code> object (for backward compatibility, <code>pattern</code> will be required).
<code>assay</code>	specify the name of the assay to plot, that should match one of <code>assayNames(rse)</code> .
<code>x_labels</code>	x-axis labels.
<code>palette</code>	a color palette function, by default: <code>colorRampPalette(c("magenta", "black", "green"))</code>
<code>alpha_for_se</code>	the transparency (alpha) value for the error band.
<code>error_type</code> ,	can be either "sd" (standard deviation), "sem" (standard error of the mean), or "ci95" (95% confidence interval). Default: "sd".
<code>reversed_z_order</code>	should the z-order of the curves be reversed (i.e. first or last bin on top?)
<code>ylim</code>	a vector of two numbers corresponding to the y-limits of the plot
<code>pattern</code>	only if <code>rse</code> is of class <code>GRanges</code> . A single character that should match metadata of <code>rse</code> (can be a regular expression).

Value

Display a plot.

Examples

```
data("stackepi")
plotAverageProfile(stackepi)
```

<code>plotBinning</code>	<i>plotBinning()</i>
--------------------------	----------------------

Description

Plot a vertical color bar of the bin column.

Usage

```
plotBinning(
  rse,
  target_height = 650,
  palette = colorRampPalette(c("magenta", "black", "green"))
)
```

Arguments

<code>rse</code>	a <code>RangedSummarizedExperiment</code> input with a column bin in <code>rowRanges(rse)</code> . Alternatively (for backward compatibility), a <code>GRanges</code> object or any object such as <code>rse\$bin</code> exists.
<code>target_height</code>	an integer, the approximate height (in pixels) of the final plot. Used to avoid overplotting artefacts.
<code>palette</code>	a function taking a number as a first argument, and returning a vector of colors.

Value

Display a plot.

Examples

```
data("stackepi")
rse <- stackepi
rse <- addBins(rse, nbins = 3)
plotBinning(rse)

gr2 <- data.frame(bin = rep(c(1,2,3,4), each = 5))
plotBinning(gr2, palette = colorRampPalette(c("blue4", "forestgreen", "coral3", "goldenrod")))
```

plotBoxMetric	<i>plotBoxMetric()</i>
---------------	------------------------

Description

Plot distribution of a metric values as boxplots depending of bins. If the bin is absent from gr, a single boxplot is drawn.

Usage

```
plotBoxMetric(  
  rse,  
  metric = "expr",  
  title = "Metric",  
  trans_func = function(x) x,  
  ylim = NULL,  
  ylab = "metric",  
  palette = colorRampPalette(c("magenta", "black", "green"))  
)
```

Arguments

rse	a RangedSummarizedExperiment input. Aletrnatively: can be a GRanges object (for backward compatibility).
metric	name of the column in rse metadata containing scores.
title	title of the plot.
trans_func	A function to transform value of x before plotting. Useful to apply log10 transformation (i.e. with trans_func = function(x) log10(x+1)).
ylim	limit of the y axis; format: ylim = c(min, max)
ylab	y-axis title
palette	a function that returns a palette of n colors.

Value

Display a plot.

Examples

```
data("stackepi")  
plotBoxMetric(  
  stackepi,  
  trans_func = function(x) x,  
  metric = "exp",  
  title = "Metric"  
)
```

plotEpistack	<i>plotEpistack()</i>
--------------	-----------------------

Description

Given a list of genomic regions, epigenetic signals surrounding these regions, and a score for each regions, plot epigenetic stacks depending on the score. An optional bin column allow the grouping of several genomic regions to produce average profiles per bins.

Usage

```
plotEpistack(
  rse,
  assays = NULL,
  tints = "gray",
  titles = NULL,
  legends = "",
  main = NULL,
  x_labels = c("Before", "Anchor", "After"),
  zlim = c(0, 1),
  ylim = NULL,
  metric_col = "exp",
  metric_title = "Metric",
  metric_label = "metric",
  metric_transfunc = function(x) x,
  bin_palette = colorRampPalette(c("magenta", "black", "green")),
  npix_height = 650,
  n_core = 1,
  high_mar = c(2.5, 0.6, 4, 0.6),
  low_mar = c(2.5, 0.6, 0.3, 0.6),
  error_type = c("sd", "sem", "ci95"),
  patterns = NULL,
  ...
)
```

Arguments

rse	a RangedSummarizedExperiment input. Aletrnatively: can be a GRanges object (for backward compatibility, patterns will be required).
assays	specify the name(s) and order of assay(s) to plot. A vector of names that should match assayNames(rse).
tints	a vector of colors to tint the heatmaps.
titles	titles of each heatmap. Defaults to assays.
legends	legend names for the epistacks.
main	Main title for the figure.

x_labels	a character vector of length 3 used as x-axis labels.
zlim	the minimum and maximum z values the heatmap. Format: zlim = c (min, max)
ylim	limits of the y axis for bottom plots. Format: ylim = c (min, max)
metric_col	a character, name of a column in gr such as expression value, peak height, pvalue, fold change, etc.
metric_title	title to be display on the leftmost plots.
metric_label	label of the leftmost plots.
metric_transfunc	a function to transform value of metric_col before plotting. Useful to apply log10 transformation (i.e. with trans_func = function(x) log10(x+1)).
bin_palette	a palette of color, (i.e. a function of parameter n that should retrun n colors), used to color average profiles per bin in the bottom plots.
npix_height	The matrix height is reduced to this number of rows before plotting. Useful to limit overplotting artefacts. It should roughlyly be set to the pixel height in the final heatmaps
n_core	number of core used to speedup the martrix resizing.
high_mar	a vector of numerical values corresponding to the margins of the top figures. c(bottom, left, top, right)
low_mar	a vector of numerical values corresponding to the margins of the bottom figures. c(bottom, left, top, right)
error_type,	error_type, can be either "sd" (standard deviation), "sem" (standard error of the mean), or "ci95" (95% confidence interval). Default: "sd".
patterns	only if rse is of class GRanges. A character vector of column prefixes (can be regular expressions) that should match columns of rse.
...	Arguments to be passed to par such as cex

Details

This function produce a comprehensive figure including epigenetic heatmaps and average epigenetic profiles from a well formatted RangedSummarizedExperiment object with expected rowData metadata columns. It scales resonably well up to hundreds of thousands of genomic regions.

The visualisation is centered on an anchor, a set of genomic coordinated that can be transcription start sites or peak center for example. Anchor coordinates are those of the GRanges used as a rowData in the input RangedSummarizedExperiment object (hereafter rse).

Anchors are plotted from top to bottom in the same order as in rse. One should sort rse before plotting if needed.

rse's rowData should have a metric column that is used in the leftmost plots. The name of the metric column must be specified to metric_col. The metric can be transformed before plotting if needed using the metric_transfunc parameter.

The matrix or matrices used to display the heatmap(s) should be passed as assay(s) in rse. Such matrix can be obtained using `EnrichedHeatmap::normalizeToMatrix()` for example. The assay names are then specified through assays.

If an optional bin column is present in `rse`'s `rowData`, it will be used to group genomic regions to performed average profile per bins in the bottom plots.

Epistack are multipanel plots build using `layout()`. Margins for the panels can be specified using `high_mar` and `low_mar` parameters if needed, especially to avoid text overlaps. The default value should be appropriate in most situations. Individual component can be plotted using several epistack functions such as `plotStackProfile()` or `plotAverageProfile()`.

Plotting more than > 1000 regions can lead to overplotting issued as well as some plotting artefacts (such as horizontal white strips). Both issues can be resolved with fiddling with the `npix_height` parameter. `npix_height` should be smaller than the number of regions, and in the same order of magnitude of the final heatmap height in pixels. Last minutes call to the `redimMatrix()` function will happen before plotting using `npix_height` as target height. Parameter `n_core` is passed to `redimMatrix()` to speed up the down-scaling.

The input can also be a `GRanges` object for backward compatibility. See `GRanges2RSE`. patterns would then be required.

Value

Display a plot.

See Also

[plotStackProfile](#), [plotAverageProfile](#), [redimMatrix](#), [normalizeToMatrix](#), [addMetricAndArrangeGRanges](#), [addBins](#)

Examples

```
data("stackepi")
plotEpistack(stackepi,
  metric_col = "exp",
  ylim = c(0, 1),
  metric_transfunc = function(x) log10(x+1))
```

plotMetric

plotMetric()

Description

Plot a vertical line chart of the metric column, in the same order as the input.

Usage

```
plotMetric(
  x,
  trans_func = function(x) x,
  title = "Metric",
  ylim = NULL,
  xlab = NULL
)
```

Arguments

x	a numeric vector.
trans_func	a function to transform x values before plotting. Useful to apply log10 transformation (i.e. with trans_func = function(x) log10(x+1)).
title	Title of the plot.
ylim	limit of the y axis; format: ylim = c(min, max)
xlab	x-axis title

Value

Display a plot.

See Also

[plotEpistack](#), [plotBoxMetric](#)

Examples

```
data("stackepi")
plotMetric(SummarizedExperiment::rowRanges(stackepi)$exp)
```

plotStackProfile *plotStackProfile()*

Description

Display a heatmap of an epigenetic track centered at genomic anchors such as Transcription Start Sites or peak center.

Usage

```
plotStackProfile(
  rse,
  assay = NULL,
  x_labels = c("Before", "Anchor", "After"),
  title = "",
  zlim = NULL,
  palette = function(n) viridisLite::viridis(n, direction = -1),
  target_height = 650,
  summary_func = function(x) mean(x, na.rm = TRUE),
  n_core = 1,
  pattern = NULL
)
```

Arguments

rse	a RangedSummarizedExperiment input. Alternatively: can be a GRanges object (for backward compatibility, pattern will be required).
assay	specify the name of the assay to plot, that should match one of assayNames(rse).
x_labels	a character vectors of length 3 used to label the x-axis.
title	The title of the heatmap
zlim	The minimum and maximum z values to match color to values. Format: zlim = c (min, max)
palette	a palette of color, (i.e. a function of parameter n that should return n colors).
target_height	The matrix height is reduced to this number of rows before plotting. Useful to limit overplotting artefacts. It should roughly be set to the pixel height in the final heatmap.
summary_func	function passed to redimMatrix(). Usually mean, but can be set to median or max for sparse matrices.
n_core	multicore option, passed to redimMatrix().
pattern	only if rse is of class GRanges. A character vector of length 1 of a column prefix (can be regular expressions) that should match columns of rse.

Details

The visualisation is centered on an anchor, a set of genomic coordinates that can be transcription start sites or peak center for example. Anchor coordinates are those of the RangedSummarizedExperiment object used as an input (hereafter rse).

Anchors are plotted from top to bottom in the same order as in rse. One should sort rse before plotting if needed.

The matrix used to display the heatmap should be passed as assay of rse. Such matrix can be obtained using `EnrichedHeatmap::normalizeToMatrix()` for example.

This function scale reasonably wells up to hundred thousands of regions. Overplotting issues are solved by last-minute reduction of the matrix size using `redimMatrix()`.

Value

Display a plot.

See Also

[plotAverageProfile](#), [plotEpistack](#), [normalizeToMatrix](#), [plotStackProfileLegend](#)

Examples

```
data("stackepi")
plotStackProfile(stackepi,
  target_height = 650,
  zlim = c(0, 1),
  palette = colorRampPalette(c("white", "dodgerblue", "black")),
  title = "DNA methylation")
```

plotStackProfileLegend
plotStackProfileLegend()

Description

Utility function to plot the corresponding legend key of plotStackProfile()'s plots.

Usage

```
plotStackProfileLegend(  
  zlim,  
  palette = colorRampPalette(c("white", "grey", "black")),  
  title = NA  
)
```

Arguments

zlim	the limits of the values to be displayed. Format: c(min, max)
palette	a palette of color, (i.e. a function of parameter n that should return n colors).
title	an optional title to be displayed below the color legend.

Value

Display a plot.

See Also

[plotStackProfile](#)

Examples

```
plotStackProfileLegend(zlim = c(0, 2),  
  palette = colorRampPalette(c("white", "grey", "black")))
```

redimMatrix *redimMatrix()*

Description

Reduce the input matrix size by applying a summary function on cells to be fused.

Usage

```
redimMatrix(
  mat,
  target_height = 100,
  target_width = 100,
  summary_func = function(x) mean(x, na.rm = TRUE),
  output_type = 0,
  n_core = 1
)
```

Arguments

<code>mat</code>	the input matrix.
<code>target_height</code>	height of the output matrix (should be smaller than or equal to <code>nrow(mat)</code>).
<code>target_width</code>	width of the output matrix (should be smaller than or equal to <code>ncol(mat)</code>).
<code>summary_func</code>	how to summarize cells? A function such as <code>mean</code> , <code>median</code> , <code>max</code> , or <code>meanColors</code> .
<code>output_type</code>	Type of the output, to be passed to <code>vapply</code> 's <code>FUN.VALUE</code> .
<code>n_core</code>	number of core to use for parallel processing.

Details

This function is used to reduce matrix right before plotting them in order to avoid overplotting issues as well as other plotting artefacts.

Value

a resized matrix of size `target_width` x `target_height` where the `summary_fun` was apply to adjacent cells.

See Also

[meanColor](#)

Examples

```
data("stackepi")
mat <- SummarizedExperiment::assay(stackepi, "DName")
dim(mat)
smallMat <- redimMatrix(mat, target_height = 10, target_width = ncol(mat))
dim(smallMat)

# changing the summary function
mat <- matrix(sample(1:40,100,replace=TRUE),nrow=10,ncol=10)
dim(mat)
smallMat <- redimMatrix(mat, target_height = 5, target_width = ncol(mat),
  summary_func = function(x) max(x, na.rm = TRUE))
dim(smallMat)

# working with colors
```



```
colmat <- matrix(
  c("red", "red", "blue", "blue", "red", "blue", "blue", "green"),
  ncol = 2
)
redimMatrix(colmat, target_height = 2, target_width = 2,
  summary_func = meanColor, output_type = "color")
```

stackepi

epistack example and test dataset

Description

DNA methylation profiles (from MBD-seq data) around transcription start sites of the 693 chr18 genes annotated on the pig genome (Sscrofa11.1), as well as gene expression levels in Transcript Per Million (TPM) measured by RNA-seq in the same duodenum sample.

Usage

```
data("stackepi")
```

Format

A RangedSummarizedExperiment of the 693 rows, 2 rows metadata columns, and one assay containing the DNA methylation signal.

Source

This dataset was generated from ENSEMBL annotation data and data generated by our lab (publicly available soon).

stackepi_gr

epistack backward compatibility dataset

Description

DNA methylation profiles (from MBD-seq data) around transcription start sites of the 693 chr18 genes annotated on the pig genome (Sscrofa11.1), as well as gene expression levels in Transcript Per Million (TPM) measured by RNA-seq in the same duodenum sample.

Usage

```
data("stackepi_gr")
```

Format

A GRanges of the 693 rows and 54 metadata columns, kept for unit-testing backward-compatibility.

Source

This dataset was generated from ENSEMBL annotation data and data generated by our lab (publicly available soon).

See Also

[GRanges2RSE](#)

Index

* datasets

stackepi, [17](#)

stackepi_gr, [17](#)

addBins, [2](#), [12](#)

addMetricAndArrangeGRanges, [3](#), [3](#), [12](#)

addMetricAndArrangeRSE, [4](#)

GRanges2RSE, [5](#), [12](#), [18](#)

meanColor, [6](#), [16](#)

normalizeToMatrix, [12](#), [14](#)

par, [11](#)

plotAverageProfile, [7](#), [12](#), [14](#)

plotBinning, [3](#), [8](#)

plotBoxMetric, [9](#), [13](#)

plotEpistack, [10](#), [13](#), [14](#)

plotMetric, [12](#)

plotStackProfile, [12](#), [13](#), [15](#)

plotStackProfileLegend, [14](#), [15](#)

redimMatrix, [6](#), [12](#), [15](#)

stackepi, [17](#)

stackepi_gr, [17](#)