Package 'vasp'

March 16, 2020

Type Package

Title Quantification and Visulization of Variations of Splicing in Population

Author Huihui Yu, Qian Du, Chi Zhang

Maintainer Huihui Yu <yuhuihui2011@foxmail.com>

Description Discovery of genome-wide variable alternative splicing events from short-read RNA-seq data and visualizations of gene splicing information for publication-quality multi-panel figures.

URL https://github.com/yuhuihui2011/vasp

BugReports https://github.com/yuhuihui2011/vasp/issues

License GPL (>= 2.0)

Depends R (>= 4.0), ballgown

Imports IRanges, GenomicRanges, S4Vectors, Sushi, parallel, matrixStats, GenomeInfoDb, Rsamtools, cluster, stats, graphics, methods

Suggests knitr, rmarkdown

VignetteBuilder knitr

biocViews RNASeq, AlternativeSplicing, DifferentialSplicing, StatisticalMethod, Visualization, KEGG, Preprocessing, Clustering, ImmunoOncology, DifferentialExpression

Encoding UTF-8

LazyData false

RoxygenNote 7.0.2

R topics documented:

BMfinder *Discover bimodal distrubition features*

Description

Find bimodal distrubition features and divide the samples into 2 groups by k-means clustering.

Usage

```
BMfinder(x, p.value = 0.01, maf = 0.05, miss = 0.05, fold = 2, log = FALSE,
    cores = detectCores() - 1)
```

Arguments

x	a numeric matrix with feature rows and sample columns, e.g., splicing score matrix from spliceGenome or spliceGene function.
p.value	p.value threshold for bimodal distrubition test
maf	minor allele frequency threshold in k-means clustering
miss	missing grouping rate threshold in k-means clustering
fold	fold change threshold between the two groups
log	whether the scores are to be logarithmic. If TRUE, all the scores are log2 tranformed before k-means clustering: x = log2(x+1).
cores	threads to be used. This value is passed to **?mclapply** in **parallel** package

Details

The matrix contains 1, 2 and NA, and values of 'x' in group 2 are larger than group 1.

Value

a matrix with feature rows and sample columns.

Examples

```
data(rice.bg)
score<-spliceGene(rice.bg,'MSTRG.183',junc.type='score')
score<-round(score,2)
as<-BMfinder(score,cores=1) # 4 bimodal distrubition features found

##compare
as
score[rownames(score)%in%rownames(as),]
```

getDepth *Get Read Depth*

Description

Get read depth from a BAM file (in bedgraph format)

Usage

```
getDepth(x, chrom, start, end)
```

Arguments

x	path to a BAM file
chrom	chromosome of a region to be searched
start	start position
end	end position

Value

a data.frame in bedgraph file format which can be used as input for plotBedgraph in the **SuShi** package.

See Also

splicePlot

Examples

```
path <- system.file('extdata',package='vasp')
bam_files<-list.files(path,'bam$')
bam_files

depth<-getDepth(file.path(path, bam_files[1]), 'Chr1',
                start=1171800, end=1179400)
head(depth)

library(Sushi)
plotBedgraph(depth,'Chr1',chromstart=1171800, chromend=1179400,yaxt='s')
mtext('Depth',side=2,line=2.5,cex=1.2,font=2)
labelgenome('Chr1',1171800,1179400,side=1,scipen=20,n=5,scale='Kb')
```

`getGeneinfo` *Get Gene Informaton from a ballgown object*

Description

Get gene informaton from a ballgown object by genes or by genomic regions

Usage

```
getGeneinfo(genes = NA, bg, chrom, start, end, samples = sampleNames(bg),
    trans.select = NA)
```

Arguments

`genes`	a character vector specifying gene IDs in 'bg'. Any values other than NA override genomic region (chrom, start, stop)
`bg`	ballgown object
`chrom`	chromosome of a region
`start`	start postion
`end`	stop postion
`samples`	names of samples. The transcrpts in these samples are subjected to 'trans.select'
`trans.select`	logical expression-like string, indicating transcript rows to select from a matrix of transcript coverages: NA value keeps all transcripts.

Value

a data.frame in bed-like file format that can be used as input for `plotGenes` in the **SuShi** package

See Also

`splicePlot`; `plotGenes` in **Sushi** package

Examples

```
data(rice.bg)
unique(geneIDs(rice.bg))

gene_id <- c('MSTRG.181', 'MSTRG.182', 'MSTRG.183')
geneinfo <- getGeneinfo(genes=gene_id,rice.bg)
trans <- table(geneinfo$name) # show how many exons each transcript has
trans

library(Sushi)
chrom = geneinfo$chrom[1]
chromstart = min(geneinfo$start) - 1e3
chromend = max(geneinfo$stop) + 1e3
color = rep(SushiColors(2)(length(trans)), trans)

par(mar=c(3,1,1,1))
plotGenes(geneinfo, chrom, chromstart, chromend, col = color, bheight = 0.2,
    bentline = FALSE, plotgenetype = 'arrow', labeloffset = 0.5)
labelgenome(chrom, chromstart , chromend, side = 1, n = 5, scale = 'Kb')
```

rice.bg *Rice Ballgown Object*

Description

Small ballgown object created with a subset of rice RNAseq data, for demonstration purposes

Usage

```
rice.bg
```

Format

a ballgown object: 33 transcripts and 6 samples

Examples

```
data(rice.bg)
rice.bg
# ballgown instance with 33 transcripts and 6 samples
```

spliceGene *Calculate Splicing Scores for One Gene*

Description

Calculate splicing Scores from ballgown object for a given gene. This function can only calculate one gene. Please use function spliceGenome to obtain genome-wide splicing scores.

Usage

```
spliceGene(bg, gene, samples = sampleNames(bg),junc.type = c("score", "count"),
    trans.select = "rowMaxs(x)>=1", junc.select = "rowMaxs(x)>=5")
```

Arguments

bg	ballgown object
gene	a character string specifying gene id
samples	names of samples
junc.type	type of junction estimate ('score' for junction score; 'count' for junction read count)
trans.select	logical expression-like string, indicating transcript rows to select from a matrix of transcript coverages: NA value keeps all transcripts. e.g. use trans.select='rowMaxs(x)>=1' to filter the transcrpts with the maximium coverage among all the samples less than 1.
junc.select	logical expression-like string, indicating junction rows to select from a matrix of junction counts: NA value keeps all junctions. e.g. use junc.select='rowMaxs(x)>=5' to filter the junctions with the maximium read count among all the samples less than 5.

Details

score = junction count/gene-level per base read coverage. Row functions for matrices are useful to select transcripts and junctions. See matrixStats package.

Value

a matrix of junction scores with intron rows and sample columns.

See Also

spliceGenome, which calculates splicing scores in whole genome.

Examples

```
data(rice.bg)
rice.bg
head(geneIDs(rice.bg))

score<-spliceGene(rice.bg,'MSTRG.183',junc.type='score')
count<-spliceGene(rice.bg,'MSTRG.183',junc.type='count')

## compare
tail(score)
tail(count)

## get intron structrue
intron<-structure(rice.bg)$intron
intron[intron$id%in%rownames(score)]
```

---

spliceGenome *Calculate Genome-wide Splicing Scores*

Description

Calculate splicing scores from ballgown objects for all genes.

Usage

```
spliceGenome(bg, gene.select = "rowQuantiles(x,probs = 0.05)>=1",
    intron.select = "rowQuantiles(x,probs = 0.95)>=5")
```

Arguments

bg	ballgown object
gene.select	logical expression-like string, indicating genes to select from a matrix of gene-level coverages: NA value keeps all genes. e.g. gene.select = 'rowQuantiles(x,probs = 0.05)>=1' keeps the genes with the read coverage greater than or equal to 1 in at least 95 (0.05 quantile). Used to filter low expressed genes.
intron.select	logical expression-like string, indicating introns to select from a matrix of junction counts: NA value keeps all introns. e.g. intron.select = 'rowQuantiles(x,probs = 0.95)>=5' keeps the introns with the read count greater than or euqal to 5 in at least 5 (0.95 quantile). Used to filter introns with very few junction reads supporting.

Details

score = junction count/gene-level per base read coverage. Row functions for matrices in `matrixStats` package are useful to select genes and introns.

Value

a list of two elelments: 'score' is matrix of intron splicing scores with intron rows and sample columns and 'intron' is a `GRanges` object of intron structure. See `structure` in **ballgown** package

See Also

`spliceGene`, which calculates splicing scores in one gene.

Examples

```
data(rice.bg)
rice.bg

splice<-spliceGenome(rice.bg,gene.select=NA,intron.select=NA)
names(splice)

head(splice$score)
splice$intron
```

---

splicePlot *Gene Splicing Plot*

Description

Visualization of read coverage, splicing information and gene information in a gene region. This function is a wrapper of `getDepth`, `getGeneinfo`, `spliceGene`, `plotBedgraph` and `plotGenes`.

Usage

```
splicePlot(bg, gene, samples, bam.dir = NA, start = NA, end = NA,
    labels = samples, junc.type = c("score", "count"), junc.text = TRUE,
    trans.select = "rowMaxs(x)>=1", junc.select = "rowMaxs(x)>=5",
    col.depth = SushiColors(2)(length(samples) + 1)[-1], scale = "Kb",
    plotgenetype = "arrow", ...)
```

Arguments

bg	ballgown object. See `ballgown`.
gene	string indicating a gene ID (must be in the 'bg')
samples	names of the samples to be shown (must be in the 'bg' and have bam files in the 'bam.dir')
bam.dir	bam file directory of the samples. If NA, instead of read depth, conserved exons are drawn.
start	start position to be shown. If NA, start position of the gene will be used.
end	stop position to be shown. If NA, end position of the gene will be used.

labels	labels for samples (default: sample names). If it is NA, neigher sample names nor gene names will be labeled
junc.type	type of junction estimates to be shown ('score' for junction score; count' for junction read count)
junc.text	TRUE/FALSE indicating whether junction estimates should be labeled
trans.select	logical expression-like string, indicating transcript rows to select from a matrix of transcript coverages: NA value keeps all transcripts. See spliceGene
junc.select	logical expression-like string, indicating junction rows to select from a matrix of junction counts: NA value keeps all junctions. See spliceGene
col.depth	a vector of length(samples) specifying colors of read depth.
scale	scale of the labelgenome ('bp','Kb','Mb')
plotgenetype	string specifying whether the genes should resemble a 'box' or a 'arrow'. See plotGenes.
...	values to be passed to plotGenes.

Value

see plotGenes.

Examples

```
data(rice.bg)
rice.bg

samples <- paste('Sample', c('027','102','237'),sep='_')
bam.dir <- system.file('extdata',package = 'vasp')

## plot the whole gene region
splicePlot(rice.bg,samples,bam.dir,gene='MSTRG.183',bheight=0.2)

## plot the alternative splicing region
splicePlot(rice.bg,samples,bam.dir,gene='MSTRG.183',start=1179000)
```

Index