Package 'SCANVIS'

April 25, 2024

Type Package

Title SCANVIS - a tool for SCoring, ANnotating and VISualizing splice junctions

Version 1.17.0

Date 2019-07-11

Author Phaedra Agius <pagius@nygenome.org>

Maintainer Phaedra Agius <pagius@nygenome.org>

Depends R (>= 3.6)

Description SCANVIS is a set of annotation-dependent tools for analyzing splice junctions and their read support as predetermined by an alignment tool of choice (for example, STAR aligner). SCANVIS assesses each junction's relative read support (RRS) by relating to the context of local split reads aligning to annotated transcripts. SCANVIS also annotates each splice junction by indicating whether the junction is supported by annotation or not, and if not, what type of junction it is (e.g. exon skipping, alternative 5' or 3' events, Novel Exons). Unannotated junctions are also futher annotated by indicating whether it induces a frame shift or not. SCANVIS includes a visualization function to generate static sashimi-style plots depicting relative read support and number of split reads using arc thickness and arc heights, making it easy for users to spot well-supported junctions. These plots also clearly delineate unannotated junctions from annotated ones using designated color schemes, and users can also highlight splice junctions of choice. Variants and/or a read profile are also incoroporated into the plot if the user supplies variants in bed format and/or the BAM file. One further feature of the visualization function is that users can submit multiple samples of a certain disease or cohort to generate a single plot - this occurs via a ``merge" function wherein junction details over multiple samples are merged to generate a single sashimi plot, which is useful when contrasting cohorots (eg. disease vs control).

Imports IRanges,plotrix,RCurl,rtracklayer

License file LICENSE

Suggests knitr, rmarkdown

VignetteBuilder knitr

biocViews Software,ResearchField,Transcriptomics,WorkflowStep,Annotation,Visualization

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

git_branch devel

git_last_commit 8541bed

git_last_commit_date 2023-10-24

Repository Bioconductor 3.19

Date/Publication 2024-04-24

Contents

SCANVIS-package . . . 2
GBM . . . 3
gbm3 . . . 4
gbm3.vcf . . . 4
gen19 . . . 4
gene2roi . . . 5
IR2Mat . . . 5
ls_url . . . 6
LUAD . . . 7
LUSC . . . 7
SCANVISannotation . . . 8
SCANVISexamples . . . 8
SCANVISlinkvar . . . 9
SCANVISmerge . . . 10
SCANVISread_STAR . . . 11
SCANVISscan . . . 12
SCANVISvisual . . . 13

Index **16**

`SCANVIS-package` *SCANVIS - a tool for SCoring, ANnotating and VISualizing splice junctions*

Description

SCANVIS is a set of annotation-dependent tools for analyzing splice junctions and their read support as predetermined by an alignment tool of choice (for example, STAR aligner). SCANVIS assesses each junction's relative read support (RRS) by relating to the context of local split reads aligning to annotated transcripts. SCANVIS also annotates each splice junction by indicating whether the junction is supported by annotation or not, and if not, what type of junction it is (e.g. exon skipping, alternative 5' or 3' events, Novel Exons). Unannotated junctions are also futher annotated by indicating whether it induces a frame shift or not. SCANVIS includes a visualization function to generate static sashimi-style plots depicting relative read support and number of split reads using arc thickness and arc heights, making it easy for users to spot well-supported junctions. These plots also clearly delineate unannotated junctions from annotated ones using designated color schemes, and users can also highlight splice junctions of choice. Variants and/or a read profile are also incoroporated into the plot if the user supplies variants in bed format and/or the BAM file. One further feature of the visualization function is that users can submit multiple samples of a certain disease

or cohort to generate a single plot - this occurs via a "merge" function wherein junction details over multiple samples are merged to generate a single sashimi plot, which is useful when contrasting cohorots (eg. disease vs control).

Details

This package was not yet installed at build time.

Index: This package was not yet installed at build time.
SCANVIS is a set of tools for SCoring and ANnotating splice junctions using gencode annotation. It also has a VISualization component that allows users to quickly view one or more samples in sashimi style plots, showing splice junctions (SJs) and, optionally, a read coverage profile as well as mutations in one figure. These sashimi style plots are novel in that unannotated splice junctions are highlighted in various colours to delineate various junction types, with line styles indicating whether unannotated junctions are in frame or not.

Author(s)

Phaedra Agius <pagius@nygenome.org>

Maintainer: Phaedra Agius <pagius@nygenome.org>

`GBM` *list of 3 TCGA glioblastoma samples, parts thereof, outputs of SCANVISscan and SCANVISlinkvar functions with toy variants supplied for the variant-SJ mapping*

Description

list of parts of 3 matrices output by SCANVISscan

Usage

```
GBM
```

Format

list

gbm3 *part of a TCGA glioblastoma sample from STAR alignment SJ.tab file*

Description

matrix with chr,start,end,uniq.reads indicating genomic coordinates for each splice junction and corresonding number of supporting split reads

Usage

```
gbm3
```

Format

matrix

gbm3.vcf *a toy set of 6 variants that pair up with the gbm3 data example*

Description

matrix with chr,start,end,passedMUT indicating genomic coordinates for each toy variant that has supposedly passed some threshold

Usage

```
gbm3
```

Format

matrix

gen19 *parts of the annotation object created by the SCANVISannotation function when used with the url ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_19/ which references the gencode v19 GTF file for human hg19*

Description

list with various genomic components for use in SCANVISfunctions

Usage

```
gen19
```

Format

list

`gene2roi` *gene name/s to region of interest*

Description

Converts gene name/s to genomic coordinates using gene annotation file from SCANVISannotation

Usage

```
gene2roi(g,gen)
```

Arguments

`g`	vector of one or more gene names or gene ids in the same chromosome
`gen`	gene annotation object as output by SCANIVS.annotation

Details

This function is called upon by SCANVISlinkvar and SCANVISvisual

Value

chr, start and end of the union of genomic intervals that overlap the genes in g

Examples

```
data(SCANVISexamples)
g=c('TDRD6','PLA2G7')
roi=gene2roi(g,gen19)
```

`IR2Mat` *IRanges to Matrix*

Description

converts IRanges interval object to matrix

Usage

```
IR2Mat(I)
```

Arguments

`I`	IRanges interval object

Details

This function is called upon by SCANVISscan

Value

a matrix with start and end coordinates for the intervals in I

Examples

```
library(IRanges)
I=IRanges(1:10,21:30)
m=IR2Mat(I)
```

`ls_url` *list files available at annotation/gencode url*

Description

Function called upon by SCANVISannotation for GTF file pulldown from url supplied

Usage

```
ls_url(url)
```

Arguments

`url`	url to GTF files

Details

calls upon functions in Rcurl and rtracklayer

Value

a list of files for download at url

Examples

```
ftpfiles=ls_url('ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_19/')
```

LUAD — *list of 3 TCGA lung adenocarcinoma samples, parts thereof, outputs of SCANVISscan*

Description

list of parts of 3 matrices output by SCANVISscan

Usage

```
LUAD
```

Format

list

LUSC — *list of 3 TCGA lung squamous cell carcinoma samples, parts thereof, both outputs of SCANVISscan with the second sample being variant-mapped via SCANVISlinkvar*

Description

list of parts of 3 matrices as output by SCANVISscan

Usage

```
LUSC
```

Format

list

SCANVISannotation *assembles annotation from gtf file into SCANVISreadable format*

Description

This function ftps to the supplied gtf url, downloads gtf to current directory and assembles annotation details into a SCANVISreadable object

Usage

```
SCANVISannotation(ftp.url)
```

Arguments

```
ftp.url
```

Value

a gencode object compatible (and required) for use with most SCANVISfunctions

Note

Web access required. If variants are available and intended for use with SCANVISlinkvar, the gencode reference genome must be the same as that used for the variant calls.

Examples

```
## Not run: gen19=SCANVISannotation('ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_19/')
```

SCANVISexamples *Data for running SCANVISexamples*

Description

This data is a list of parts of SJ.tab files as output by the STAR alignment algorithm on a number of TCGA samples (GBM, LUSC and LUAD). It also includes part of the annotation file as derived by SCANVISannotation on hg19, gencode v19 that was used to generate the included examples.

Usage

```
SCANVISexamples
```

Format

Contains the following data pieces: GBM, LUAD, LUSC, gbm3, gbm3.vcf,gen19

SCANVISlinkvar *maps variants to SCANVISscored junctions*

Description

This function maps variants to SJs by overlapping the union of gene coordinates that harbor the SJs (optionally, with some gene interval expansion) with variant coordinates

Usage

```
SCANVISlinkvar(scn, bed, gen, p = 0)
```

Arguments

scn	matrix output by SCANVISscan
bed	matrix with variants in bed format with colnames chr, start, end and with and additional description column (eg. ssSNP for splice site mutations)
gen	gencode object as generated by the function SCANVISannotation
p	expands gene intervals up/downstream by p (default=0, no padding)

Value

Returns the input scn matrix with an additional column showing variants, if any, that occur in/near the listed genes. For instances where multiple variants map to a SJ, the variants are | separated (eg. chr7:145562;A>G|chr7:145592;C>G)

Note

The reference genome used to align RNA-seq reads that generated the initial set of SJs should be the same reference genome used for the variant calls.

See Also

SCANVISscan, SCANVISannotation, SCANVISvisual

Examples

```
data(SCANVISexamples)
gbm3.scn<-SCANVISscan(sj=gbm3,gen=gen19,Rcut=5)
### Variant format required (these are toy variants)
head(gbm3.vcf)
gbm3.scnv<-SCANVISlinkvar(gbm3.scn,gbm3.vcf,gen19)
table(gbm3.scnv[,'passedMUT'])
### Expand variant intervals by p
gbm3.scnvp<-SCANVISlinkvar(gbm3.scn,gbm3.vcf,gen19,p=100)
### Observe variant chr6:46820148;Z>AA which was not previously matched to any SJ
table(gbm3.scnvp[,'passedMUT'])
```

SCANVISmerge *merges multiple SCANVISsamples*

Description

With this function, the RRS scores and number of supporting reads across a number of samples are collected into matrices by collecting the union of all SJs. Furthermore, a representative sample is assembled by computing the mean (or median) of RRSs and supporting reads across all samples - this may be used to visualize a cohort in one figure (see SCANVISvisual).

Usage

```
SCANVISmerge(scn, method = "mean", roi = NULL, gen = NULL)
```

Arguments

scn	list of SCANVISmatrices OR character vector of urls pointing to SCANVISmatrix outputs
method	method for computing a RRS/uniq.reads representative, either "mean" or "median" (default="mean")
roi	NULL for all SJs OR chromosome name for a query chromosome (eg. chr1) OR 3 bit vector (chr, start, end) indicating region of interest OR a vector with one or more gene names (default=NULL in which case all SJs are merged)
gen	gencode object as generated by SCANVISannotation which must be supplied if roi is a list of one or more gene names, otherwise NULL (default=NULL)

Value

Returns a list object ready for use in SCANVISvisual with the following details:

RRS	a matrix with RRS scores for each sample (columns) and the union of SJs across all samples (rows)
NR	a matrix with number of SJ reads each sample (columns) and the union of SJs across all samples (rows)
MUTS	a binary matrix with 1 indicating presence of a mutation (row) in a sample (column), generated only if samples submitted were variant-mapped SJs
SJ	a representative sample with mean/median RRS and uniq.reads that can be used in SCANVISvisual to visualize sample cohort
roi	genomic coordinates for region of interest used to derive resulting data

Note

For 50 or more samples, roi cannot be NULL as resulting matrices may be too large. For cohort agglomeration, please consider agglomerating chromosome by chromosome.

See Also

SCANVISscan, SCANVISlinkvar, SCANVISvisual

Examples

```
data(SCANVISexamples)
### merge all SJs across in sample list GBM
GBM.merged<-SCANVISmerge(GBM)
### only merge SJs intersecting with gene PTGDS
GBM.merged<-SCANVISmerge(GBM,'mean','PTGDS',gen19)
```

SCANVISread_STAR *upload SJ.tab STAR file in SCANVISuse*

Description

This function is a little wrapper for reading in splice junction details from the SJ.tab file output by the STAR alignment tool.

Usage

```
SCANVISread_STAR(sj_file)
```

Arguments

sj_file	url to SJ file output by STAR aligner

Value

SJ data in matrix format as required for SCANVISfunctions

Examples

```
#set up toy example with chr,start,end,strand
tmp=cbind(rep('chr1',10),seq(100,1000,100),seq(100,1000,100)+99,rep(2,10))
#add in intron motif, annot, num read, num multimap reads, max overhang
#see STAR manual for details
tmp=cbind(tmp,rep(2,10),rep(0,10),c(rep(500,5),rep(8,5)),rep(0,10),rep(50,10))
write.table(tmp,'tmp',sep='\t',quote=FALSE,row.names=FALSE,col.names=FALSE)
sj=SCANVISread_STAR('tmp')
#sj is now suitable as input for SCANVISscan
```

SCANVISscan *SCore and ANnotate splice junctions*

Description

This function annotates and scores splice junctions (SJs) supplied in bed format (coordinates plus read support) and gene annotations (see SCANVISannotation). Each SJ will get annotated by gene name and junction type, with unannotated SJs (USJs) falling into one of the following groups: exon.skip, alt3p, alt5p, IsoSwitch, Unknown and NE (Novel Exons) - see below. USJs are also checked and marked for in or out of frame shifts. Each SJ is scored by a Relative Read Support (RRS) score defined as the ratio of the junction read supportto the median read support of annotated SJs within a RRS genomic region, that being defined as the minimum interval that contains at least one gene overlapping the SJ and at least one annotated SJ overlapping the gene/s wtihin the interval. Novel Exons (NEs) are defined by USJ pairs that coincide in annotated intronic regions and are scored by the mean RRS of the supporting USJs and by a Relative Read Coverage (RRC) score when the bam file is supplied.

Usage

```
SCANVISscan(sj, gen, Rcut = 5, bam = NULL, samtools = NULL)
```

Arguments

sj	SJ matrix with colnames chr,start,end,uniq.reads
gen	gencode object as generated by SCANVISannotation
Rcut	min read cutoff; only SJs with >=Rcut reads are retained (Default=5)
bam	url to bam file for NE RRC computation (default=NULL)
samtools	url to samtools function, MUST be specified if bam is supplied (default=NULL)

Value

An extension of the input SJ matrix for relevant SJs, with additional rows for NE junction pairs, as well as the following additional columns:

JuncType	describes junction type as annot for annotated SJs and one of the following for unannotated SJs: exon.skip, alt3p, alt5p, IsoSwitch, Unknown and NE (Novel Exons) where exon.skip refers to SJs that skip an exon present in all isoforms, alt3p refers to an alternative 3 prime acceptor site, alt5p refers to an alternative 5 prime donor sites, IsoSwitch refers to SJs aligning to mutually exclusive isoforms such that a novel unannotated isoform is incurred, Unknown SJs have coordinates that do not align to any exons and NE (Novel Exons) refers to SJ pairs with the start of one SJ and the end coordinate of the other SJ coinciding in an intronic region
gene_name	genes that intersect with the SJ (multiple genes are comma separated

RRS — Relative Read Support score defined as x/(x+y) where x is the query junction read support and y is the median read support of annotated SJs in the RRS genomic_interval

genomic_interval — interval used for the RRS computation

FrameStatus — frame shifts induced by unannotated SJs, where INframe indicates no frame-shift in any gene isoforms, OUTframe indicates frame-shifting in ALL gene isoforms and all other entries indicating frame shifts for specified isoforms. FrameStatus is marked NA for annotated SJs)

RRC — Relative Read Coverage score generated for NEs only, and computed only if the bam file is supplied

See Also

SCANVISannotation, SCANVISlinkvar, SCANVISvisual

Examples

```
data(SCANVISexamples)
head(gbm3) #required SJ format
gbm3.scn<-SCANVISscan(sj=gbm3,gen=gen19,Rcut=5)
head(gbm3.scn)
### to compute RRC scores for NEs, run as follows:
#gbm3.scn<-SCANVISscan(gbm3,gen19,5,bam=<BAM>,samtools=<SAMTOOLS>)
```

SCANVISvisual *a sashimi-style visualization tool*

Description

This function quickly generates color-coded sashimi plots for SCANVIS outputs showing SJs for a query gene or a specific genomic region. Annotated SJs are depicted with grey arcs, while different colors segregate unannotated SJs. Arc height and thickness correspond to the junction read support and RRS score respectively. If the supplied junction file is output from SCANVISlinkvar output, variants are also plotted. If the bam file is supplied, a normalized read coverage profile is shown as an inverted read profile for a single sample. A bam file can only be supplied with one sample, and when supplied users must operate this SCANVISvisual function from a writeable directory. Users may submit multiple samples in which case the SCANVISmerge function kicks in to merge the samples, so that the resulting sashimi plot shows the union of SJs over the submitted sample cohort, with RRS scores and read support averaged over the samples. This is useful for comparing disease cohorts.

Usage

```
SCANVISvisual(roi, gen, scn, SJ.special = NULL, TITLE = NULL, bam =
                NULL, samtools = NULL, full.annot = FALSE, USJ = "NR")
```

Arguments

roi	gene name OR region of interest (chr,start,end as 3-bit vector)
gen	gen annotation object as generated by the function SCANVISannotation.R
scn	matrix OR list of url/s to output from SCANVISscan/linkvar (which will be submitted to SCANVISmerge) OR output from SCANVISmerge for a set of samples already merged
SJ.special	3 col matrix indicating chr,start,end of any SJs of interest to be highlighted in cyan (default=NULL)
TITLE	figure name/title (default=NULL)
bam	url to one bam file corresponding to the input scn (not applicable for multiple/merged samples, default=NULL); the bam url is used to create a read profile in your plot, and during the processing of the bam file, temporary read pileup files are written to your current working directory where you must have write permission
samtools	url to samtools which MUST be specified if bam is supplied (default=NULL)
full.annot	TRUE for each isoform listed separately, FALSE for concise format (default=FALSE)
USJ	"NR" or "RRS" where NR induces the function to print the Number of supporting Reads above unannotated junction arcs, while RRS induces the function to print the RRS score as computed by SCANVISscan (default="NR")

Value

Returns a sashimi-style plot depicting the relevant SJs, as well as an object with the coordinates of the genomic region, the SJs and any variants in the figure

See Also

SCANVISscan, SCANVISlinkvar

Examples

```
data(SCANVISexamples)
### exon skip events in PPA2 in two LUSC samples
par(mfrow=c(2,1),mar=c(1,1,1,1))
vis.lusc1<-SCANVISvisual('PPA2',gen19,LUSC[[1]],TITLE=names(LUSC)[1],full.annot=TRUE)
vis.lusc2<-SCANVISvisual('PPA2',gen19,LUSC[[2]],TITLE=names(LUSC)[2],full.annot=TRUE,USJ='RRS')
### if bam file were available for LUSC1 ...
#vis.lusc1<-SCANVISvisual('PPA2',gen19,LUSC[[1]],TITLE=names(LUSC)[1],full.annot=TRUE,bam=<BAM4LUSC1>,samtools=

### sashimi plots with variants
gbm3.scn<-SCANVISscan(sj=gbm3,gen=gen19,Rcut=5)
gbm3.scnv<-SCANVISlinkvar(gbm3.scn,gbm3.vcf,gen19)
vis.gbm3<-SCANVISvisual('PTGDS',gen19,gbm3.scnv,TITLE='gbm3')
roi<-vis.gbm3$roi
d<-diff(as.numeric(roi[2:3]))
roi2<-c(roi[1],round(as.numeric(roi[2])+(d*0.1)),round(as.numeric(roi[3])-(d*0.5)))
### Supply exact coordinates instead of gene names ... Zooming in for gbm3
vis.gbm3.zoom<-SCANVISvisual(roi2,gen19,gbm3.scnv)
```

```
### plot multiple genes ... PTGDS and neighbors
vis.gbm3.multiple_genes<-SCANVISvisual(c('FBXW5','PTGDS','C9orf142'),gen19,gbm3.scnv,TITLE='gbm3')

par(mfrow=c(2,1),mar=c(1,1,1,1))
### see PTGDS in merge of 3 GBMs
GBM.PTGDS<-SCANVISvisual('PTGDS',gen19,GBM,TITLE='GBM, merged',full.annot=TRUE)
#### see PTGDS in merge of 3 LUADs ... no exon skips
LUAD.PTGDS<-SCANVISvisual('PTGDS',gen19,LUAD,TITLE='LUAD, merged',full.annot=TRUE)

### NEs in GPR116 in LUAD, but not in GBM
par(mfrow=c(2,1),mar=c(1,1,1,1))
GBM.GPR116<-SCANVISvisual('GPR116',gen19,GBM,TITLE='GBM, merged',full.annot=TRUE)
LUAD.GPR116<-SCANVISvisual('GPR116',gen19,LUAD,TITLE='LUAD, merged',full.annot=TRUE)
```

Index

∗ **GTF**
ls_url, 6
∗ **RRC**
SCANVISscan, 12
∗ **RRS**
SCANVISscan, 12
∗ **SCANVISannotation**
gen19, 4
gene2roi, 5
∗ **SCANVISexamples**
GBM, 3
gbm3, 4
gbm3.vcf, 4
gen19, 4
LUAD, 7
LUSC, 7
∗ **SCANVISlinkvar**
gene2roi, 5
∗ **SCANVISscan**
IR2Mat, 5
∗ **SCANVISvisual**
gene2roi, 5
∗ **STAR**
SCANVISread_STAR, 11
∗ **annotation**
gen19, 4
ls_url, 6
SCANVISannotation, 8
∗ **bam**
SCANVISscan, 12
∗ **cohort**
SCANVISmerge, 10
SCANVISvisual, 13
∗ **datasets**
GBM, 3
gbm3, 4
gbm3.vcf, 4
gen19, 4
LUAD, 7
LUSC, 7
SCANVISexamples, 8
∗ **frameshift**
SCANVISscan, 12
∗ **gencode**
gen19, 4
ls_url, 6
SCANVISannotation, 8
∗ **gene**
gene2roi, 5
∗ **interval**
gene2roi, 5
∗ **merge**
SCANVISmerge, 10
∗ **sashimi**
SCANVISvisual, 13

GBM, 3
gbm3, 4
gbm3.vcf, 4
gen19, 4
gene2roi, 5

IR2Mat, 5

ls_url, 6
LUAD, 7
LUSC, 7

SCANVIS *(*SCANVIS-package*)*, 2
SCANVIS-package, 2
SCANVISannotation, 8
SCANVISexamples, 8
SCANVISlinkvar, 9
SCANVISmerge, 10
SCANVISread_STAR, 11
SCANVISscan, 12
SCANVISvisual, 13