Package ‘EventPointer’

October 17, 2017

Type Package

Title An effective identification of alternative splicing events using junction arrays and RNA-Seq data

Version 1.0.0

Author Juan Pablo Romero, Ander Muniategui, Fernando Carazo, Ander Aramburu, Angel Rubio

Maintainer Juan Pablo Romero <jpromero@ceit.es>

Description EventPointer is an R package to identify alternative splicing events that involve either simple (case-control experiment) or complex experimental designs such as time course experiments and studies including paired-samples. The algorithm can be used to analyze data from either junction arrays (Affymetrix Arrays) or sequencing data (RNA-Seq).
The software returns a data.frame with the detected alternative splicing events: gene name, type of event (cassette, alternative 3',...,etc), genomic position, statistical significance and increment of the percent spliced in (Delta PSI) for all the events.
The algorithm can generate a series of files to visualize the detected alternative splicing events in IGV. This eases the interpretation of results and the design of primers for standard PCR validation.

Depends R (>= 3.4), SGSeq, Matrix, SummarizedExperiment

Imports GenomicFeatures, stringr, GenomeInfoDb, igraph, MASS, nnls, limma, matrixStats, RBGL, prodlim, graph, methods, utils, stats, doParallel, foreach, affxparser, GenomicRanges, S4Vectors

Suggests knitr, rmarkdown, BiocStyle, RUnit, BiocGenerics

License Artistic-2.0

LazyData true

RoxygenNote 6.0.1

biocViews AlternativeSplicing, DifferentialSplicing, mRNAMicroarray, RNASeq, Transcription, Sequencing

VignetteBuilder knitr

Url https://github.com/jpromeror/EventPointer

BugReports https://github.com/jpromeror/EventPointer/issues

NeedsCompilation no

R topics documented:

AllEvents_RNASeq 2
ArraysData 2
CDFfromGTF 3
EventDetection 4
EventPointer 4
EventPointer_IGV 5
EventPointer_RNASeq 7
EventPointer_RNASeq_IGV 7
PrepareBam_EP 8
SG_RNASeq 9

Index **10**

`AllEvents_RNASeq` *Alternative splicing events detected by EventPointer*

Description

Alternative splicing events detected by EventPointer

Usage

```
data(AllEvents_RNASeq)
```

Format

A `list` object `AllEvents_RNASeq[[i]][[j]]` displays the jth splicing event for the ith gene.

Value

AllEvents_RNASeq object contains all the detected alternative splicing events using EventPointer methodology. The splicing events where detected using the BAM files from the dataset published in Seshagiri et al. 2012 and used in the `SGSeq` R package vignette.

`ArraysData` *Preprocessed arrays data*

Description

Preprocessed arrays data

Usage

```
data(ArraysData)
```

Format

A `data.frame` with preprocessed arrays data. The preprocessing was done using `aroma.affymetrix`. See the package vignette for the preprocessing pipeline

Value

ArraysData object contains preprocessed junction arrays data. The preprocessing was done using `aroma.affymetrix` R package, refer to EventPointer vignette for the pipeline used for the preprocessing. The data corresponds to 4 samples from the SUM149 Cell line hybridized to the HTA 2.0 Affymetrix array. The first two samples are control and the second ones are treated.

`CDFfromGTF` *CDF file creation for EventPointer*

Description

Generates the CDF file to be used under the aroma.affymetrix framework

Usage

```
CDFfromGTF(input = "Ensembl", inputFile = NULL, PSR, Junc, PathCDF,
  microarray = NULL)
```

Arguments

`input`	Reference transcriptome used to build the CDF file. Must be one of Ensembl, UCSC or GTF.
`inputFile`	If input is GTF, inputFile should point to the GTF file to be used.
`PSR`	Path to the Exon probes txt file
`Junc`	Path to the Junction probes txt file
`PathCDF`	Directory where the output will be saved
`microarray`	Microarray used to create the CDF file. Must be one of: HTA-2_0, ClariomD, RTA or MTA

Value

The function displays a progress bar to show the user the progress of the function. However, there is no value returned in R as the function creates three files that are used later by other EventPointer functions. 1) EventsFound.txt : Tab separated file with all the information of all the alternative splcing events found. 2) .flat file : Used to build the corresponding CDF file. 3) .CDF file: Output required for the aroma.affymetrix preprocessing pipeline. Both the .flat and .CDF file take large ammounts of memory in the hard drive, it is recommended to have at least 1.5 GB of free space.

Examples

```
   PathFiles<-system.file('extdata',package='EventPointer')
   DONSON_GTF<-paste(PathFiles,'/DONSON.gtf',sep='')
   PSRProbes<-paste(PathFiles,'/PSR_Probes.txt',sep='')
   JunctionProbes<-paste(PathFiles,'/Junction_Probes.txt',sep='')
   Directory<-tempdir()
   microarray<-"HTA-2_0"

 # Run the function

 CDFfromGTF(input='GTF',inputFile=DONSON_GTF,PSR=PSRProbes,Junc=JunctionProbes,PathCDF=Directory,microarr
```

EventDetection — *Detect splicing events using EventPointer methodology*

Description

Identification of all the alternative splicing events in the splicing graphs

Usage

```
EventDetection(Input, cores, Path)
```

Arguments

Input	Output of the PrepareBam_EP function
cores	Number of cores used for parallel processing
Path	Directory where to write the EventsFound_RNASeq.txt file

Value

list with all the events found for all the genes present in the experiment. It also generates a file called EventsFound_RNASeq.txt with the information for every detected event.

Examples

```
# Run EventDetection function
 data(SG_RNASeq)
 TxtPath<-tempdir()
 AllEvents_RNASeq<-EventDetection(SG_RNASeq,cores=1,Path=TxtPath)
```

EventPointer — *EventPointer*

Description

Statistical analysis of alternative splcing events

Usage

```
EventPointer(Design, Contrast, ExFit, Eventstxt, Filter = TRUE, Qn = 0.25,
  Statistic = "LogFC", PSI = FALSE)
```

Arguments

Design	The design matrix for the experiment.
Contrast	The contrast matrix for the experiment.
ExFit	aroma.affymetrix pre-processed variable after using extractDataFrame(affy, addNames=TRUE)
Eventstxt	Path to the EventsFound.txt file generated by CDFfromGTF function.
Filter	Boolean variable to indicate if an expression filter is applied
Qn	Quantile used to filter the events (Bounded between 0-1, Q1 would be 0.25).
Statistic	Statistical test to identify differential splicing events, must be one of : LogFC, Dif_LogFC or DRS.
PSI	Boolean variable to indicate if Delta PSI should be calculated for every splicing event.

Value

Data.frame ordered by the splicing p.value . The object contains the different information for each splicing event such as Gene name, event type, genomic position, p.value, z.value and delta PSI.

Examples

```
 data(ArraysData)

 Dmatrix<-matrix(c(1,1,1,1,0,0,1,1),nrow=4,ncol=2,byrow=FALSE)
 Cmatrix<-t(t(c(0,1)))
EventsFound<-paste(system.file('extdata',package='EventPointer'),'/EventsFound.txt',sep='')

Events<-EventPointer(Design=Dmatrix,
                     Contrast=Cmatrix,
                     ExFit=ArraysData,
                     Eventstxt=EventsFound,
                     Filter=TRUE,
                     Qn=0.25,
                     Statistic='LogFC',
                     PSI=TRUE)
```

EventPointer_IGV *EventPointer IGV Visualization*

Description

Generates of files to be loaded in IGV for visualization and interpretation of events

Usage

```
EventPointer_IGV(Events, input, inputFile = NULL, PSR, Junc, PathGTF,
  EventsFile, microarray = NULL)
```

Arguments

`Events`	Data.frame generated by EventPointer with the events to be included in the GTF file.
`input`	Reference transcriprome. Available options are : Ensembl, UCSC and GTF.
`inputFile`	If input is GTF, inputFile should point to the GTF file to be used.
`PSR`	Path to the Exon probes txt file.
`Junc`	Path to the Junction probes txt file.
`PathGTF`	Directory where to write the GTF files.
`EventsFile`	Path to EventsFound.txt file generated with CDFfromGTF function.
`microarray`	Microarray used to create the CDF file. Must be one of: HTA-2_0, ClariomD, RTA or MTA

Value

The function displays a progress bar to show the user the progress of the function. Once the progress bar reaches 100 files are written to the specified directory in PathGTF. The created files are: 1) paths.gtf : GTF file representing the alternative splicing events and 2) probes.gtf : GTF file representing the probes that measure each event and each path.

Examples

```
 PathFiles<-system.file('extdata',package='EventPointer')
 DONSON_GTF<-paste(PathFiles,'/DONSON.gtf',sep='')
 PSRProbes<-paste(PathFiles,'/PSR_Probes.txt',sep='')
 JunctionProbes<-paste(PathFiles,'/Junction_Probes.txt',sep='')
 Directory<-tempdir()

 data(ArraysData)

 Dmatrix<-matrix(c(1,1,1,1,0,0,1,1),nrow=4,ncol=2,byrow=FALSE)
 Cmatrix<-t(t(c(0,1)))
EventsFound<-paste(system.file('extdata',package='EventPointer'),'/EventsFound.txt',sep='')

 Events<-EventPointer(Design=Dmatrix,
                      Contrast=Cmatrix,
                      ExFit=ArraysData,
                      Eventstxt=EventsFound,
                      Filter=TRUE,
                      Qn=0.25,
                      Statistic='LogFC',
                      PSI=TRUE)

EventPointer_IGV(Events=Events[1,,drop=FALSE],
                 input='GTF',
                 inputFile=DONSON_GTF,
                 PSR=PSRProbes,
                 Junc=JunctionProbes,
                 PathGTF=Directory,
                EventsFile= EventsFound,
                microarray="HTA-2_0")
```

EventPointer_RNASeq *Statistical analysis of alternative splcing events for RNASeq data*

Description

Statistical analysis of all the alternative splicing events found in the given bam files.

Usage

```
EventPointer_RNASeq(Events, Design, Contrast, Statistic = "LogFC",
  PSI = FALSE)
```

Arguments

Events	Output from EventDetection function
Design	The design matrix for the experiment.
Contrast	The contrast matrix for the experiment.
Statistic	Statistical test to identify differential splicing events, must be one of : LogFC, Dif_LogFC and DRS.
PSI	Boolean variable to indicate if PSI should be calculated for every splicing event.

Value

Data.frame ordered by the splicing p.value . The object contains the different information for each splicing event such as Gene name, event type, genomic position, p.value, z.value and delta PSI.

Examples

```
 data(AllEvents_RNASeq)
 Dmatrix<-matrix(c(1,1,1,1,1,1,1,1,0,0,0,0,1,1,1,1),ncol=2,byrow=FALSE)
 Cmatrix<-t(t(c(0,1)))
Events <- EventPointer_RNASeq(AllEvents_RNASeq,Dmatrix,Cmatrix,Statistic='LogFC',PSI=TRUE)
```

EventPointer_RNASeq_IGV
EventPointer RNASeq IGV Visualization

Description

Generates of files to be loaded in IGV for visualization and interpretation of events

Usage

```
EventPointer_RNASeq_IGV(Events, SG_RNASeq, EventsTxt, PathGTF)
```

Arguments

Events	Data.frame generated by EventPointer_RNASeq with the events to be included in the GTF file.
SG_RNASeq	Output from PrepareBam_EP function. Contains splicing graphs components.
EventsTxt	Path to EventsFound.txt file generated with EventDetection function
PathGTF	Directory where to write the GTF files.

Value

The function displays a progress bar to show the user the progress of the function. Once the progress bar reaches 100 file is written to the specified directory in PathGTF. The created file: 1) paths_RNASeq.gtf : GTF file representing the alternative splicing events.

Examples

```
data(AllEvents_RNASeq)
data(SG_RNASeq)

 # Run EventPointer

 Dmatrix<-matrix(c(1,1,1,1,1,1,1,1,0,0,0,0,1,1,1,1),ncol=2,byrow=FALSE)
 Cmatrix<-t(t(c(0,1)))
Events <- EventPointer_RNASeq(AllEvents_RNASeq,Dmatrix,Cmatrix,Statistic='LogFC',PSI=TRUE)

 # IGV Visualization

EventsTxt<-paste(system.file('extdata',package='EventPointer'),'/EventsFound_RNASeq.txt',sep='')
 PathGTF<-tempdir()
 EventPointer_RNASeq_IGV(Events,SG_RNASeq,EventsTxt,PathGTF)
```

PrepareBam_EP — *Bam files preparation for EventPointer*

Description

Prepares the information contained in .bam files to be analyzed by EventPointer

Usage

```
PrepareBam_EP(Samples, SamplePath, Ref_Transc = "Ensembl",
  fileTransc = NULL, cores = 1, Alpha = 2)
```

Arguments

Samples	Name of the .bam files to be analyzed (Sample1.bam,Sample2.bam,...,etc).
SamplePath	Path where the bam files are stored.
Ref_Transc	Reference transcriptome used to name the genes found in bam files. Options are: Ensembl, UCSC or GTF.
fileTransc	Path to the GTF reference transcriptome ff Ref_Transc is GTF.
cores	Number of cores used for parallel processing.
Alpha	Internal SGSeq parameter to include or exclude regions

Value

SGFeaturesCounts object. It contains a GRanges object with the corresponding elements to build the different splicing graphs found and the counts related to each of the elements.

Examples

```
## Not run:
 # Obtain the samples and directory for .bam files

   BamInfo<-si
   Samples<-BamInfo[,2]
   PathToSamples <- system.file('extdata/bams', package = 'SGSeq')
   PathToGTF<-paste(system.file('extdata',package='EventPointer'),'/FBXO31.gtf',sep='')

  # Run PrepareBam function
   SG_RNASeq<-PrepareBam_EP(Samples=Samples,
                            SamplePath=PathToSamples,
                            Ref_Transc='GTF',
                            fileTransc=PathToGTF,
                            cores=1)

## End(Not run)
```

SG_RNASeq — *Splicing graph elements predicted from BAM files*

Description

Splicing graph elements predicted from BAM files

Usage

```
data(SG_RNASeq)
```

Format

A `SGFeatureCounts` objects with predicted splicing graph features and counts

Value

SG_RNASeq object displays the predicted features found in the BAM files from the dataset published in Seshagiri et al. 2012 and used in the `SGSeq` R package vignette.

Index

AllEvents_RNASeq, 2
ArraysData, 2

CDFfromGTF, 3

EventDetection, 4
EventPointer, 4
EventPointer_IGV, 5
EventPointer_RNASeq, 7
EventPointer_RNASeq_IGV, 7

PrepareBam_EP, 8

SG_RNASeq, 9