Package 'proActiv'

March 30, 2021

Title Estimate Promoter Activity from RNA-Seq data

Version 1.0.0

Description Most human genes have multiple promoters that control the
expression of different isoforms. The use of these alternative promoters
enables the regulation of isoform expression pre-transcriptionally.
Alternative promoters have been found to be important in a wide number of cell
types and diseases. proActiv is an R package that enables the analysis of
promoters from RNA-seq data. proActiv uses aligned reads as input, and
generates counts and normalized promoter activity estimates for each annotated
promoter. In particular, proActiv accepts junction files from TopHat2 or STAR
or BAM files as inputs. These estimates can then be used to identify which
promoter is active, which promoter is inactive, and which promoters change
their activity across conditions.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Depends R (>= 4.0.0)

RoxygenNote 7.1.1

Imports dplyr, GenomicRanges, GenomicFeatures, GenomicAlignments,
GenomeInfoDb, IRanges, S4Vectors, methods, rlang,
SummarizedExperiment, AnnotationDbi, DESeq2, data.table,
tibble, Gviz, BiocParallel

Suggests testthat, mockery, BSgenome, BiocManager,
BSgenome.Hsapiens.UCSC.hg38, knitr, rmarkdown, DEXSeq, Rtsne,
ggplot2, tidyr, vdiffr

URL https://github.com/GoekeLab/proActiv

biocViews RNASeq, GeneExpression, Transcription, AlternativeSplicing,
GeneRegulation, DifferentialSplicing, FunctionalGenomics,
Epigenetics, Transcriptomics, Preprocessing

VignetteBuilder knitr

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

git_branch RELEASE_3_12

git_last_commit a9bc3ac

git_last_commit_date 2020-10-27

Author Deniz Demircioglu [aut] (<https://orcid.org/0000-0001-7857-0407>),
      Jonathan Göke [aut],
      Joseph Lee [cre]

Maintainer Joseph Lee <joseph.lee@u.nus.edu>

R topics documented:

---

calculateJunctionReadCounts

Calculate the total number of junction reads overlapping with the introns of each promoter for the input junction file

---

Description

Calculate the total number of junction reads overlapping with the introns of each promoter for the input junction file

Usage

```
calculateJunctionReadCounts(
  promoterCoordinates,
  intronRanges,
  file = "",
  fileType = "",
  genome = ""
)
```

Arguments

promoterCoordinates

A GRanges object containing promoter coordinates and reduced exon coordinates by gene

intronRanges A Granges object containing the annotated unique intron ranges. These ranges will be used for counting the reads

file character path for the input junction bed or bam file

fileType	character type of the junction bed file. Either 'tophat', 'star' or 'bam'
genome	character genome version

Value

The total number of junction reads overlapping with each promoter for the input annotated intron ranges

---

calculatePromoterReadCounts

Calculate the promoter read counts using junction read counts approach for all the input junction files

---

Description

Calculate the promoter read counts using junction read counts approach for all the input junction files

Usage

```
calculatePromoterReadCounts(
  promoterAnnotation,
  files = NULL,
  fileLabels = NULL,
  fileType = NULL,
  genome = NULL,
  numberOfCores = 1
)
```

Arguments

promoterAnnotation

A PromoterAnnotation object containing the reduced exon ranges, annotated intron ranges, promoter coordinates and the promoter id mapping

files	A character vector. The list of junction or BAM files for which the junction read counts will be calculated
fileLabels	A character vector. The labels of junction or BAM files for which the junction read counts will be calculated. These labels will be used as column names for the output data.frame object
fileType	A character. Type of the junction bed or bam file, either 'tophat', 'star' or 'bam
genome	A character. Genome version used. Must be specified if input is a BAM file. Defaults to NULL
numberOfCores	A numeric value. The number of cores to be used for counting junction reads. Defaults to 1 (no parallelization). This parameter will be used as an argument to BiocParallel::bplapply

Value

A data.frame object. The number of junction reads per promoter (rows) for each sample (cols)

getAbsolutePromoterActivity
Prepare the absolute promoter activity table including the promoter and gene ids

Description

Prepare the absolute promoter activity table including the promoter and gene ids

Usage

```
getAbsolutePromoterActivity(
  junctionReadCounts,
  promoterAnnotation,
  log2 = TRUE,
  pseudocount = 1
)
```

Arguments

junctionReadCounts
Matrix of junction read counts (rows: promoters, cols: samples)

promoterAnnotation
A PromoterAnnotation object containing the intron ranges, promoter coordinates and the promoter id mapping

log2
Logical indicating whether log2 read counts should be used (default: TRUE) or not

pseudocount
Number to be used for log2 as pseudocount if log2 is TRUE

Value

data.frame of absolute promoter activity with promoter and gene ids

getGeneExpression   *Prepare the gene expression table including the gene ids*

Description

Prepare the gene expression table including the gene ids

Usage

```
getGeneExpression(absolutePromoterActivity)
```

Arguments

absolutePromoterActivity
data.frame of absolute promoter activity with promoter and gene ids

Value

data.frame of gene expression with gene ids#'

---

getRelativePromoterActivity

Prepare the relative promoter activity table including the promoter and gene ids

---

Description

Prepare the relative promoter activity table including the promoter and gene ids

Usage

getRelativePromoterActivity(absolutePromoterActivity, geneExpression)

Arguments

absolutePromoterActivity

data.frame of absolute promoter activity with promoter and gene ids

geneExpression   data.frame of gene expression with gene ids

Value

data.frame of relative promoter activity with promoter and gene ids

---

normalizePromoterReadCounts

Normalize promoter read counts using DESeq2

Description

Normalize promoter read counts using DESeq2

Usage

normalizePromoterReadCounts(promoterReadCounts)

Arguments

promoterReadCounts

A data.frame object. The number of junction reads per promoter (rows) for each sample (cols)

Value

A data.frame object. The normalized number of junction reads per promoter (rows) for each sample (cols) using DESeq2 counts function. Requires 'DESeq2' package to be installed

---

plotPromoters   *Visualizes promoter activity and transcript model for a gene of interest*

Description

Visualizes promoter activity and transcript model for a gene of interest

Usage

```
plotPromoters(
  result,
  gene,
  txdb,
  ranges,
  cex.title = 0.9,
  cex.axis = 0.9,
  cex.main = 1,
  blk.width = 500,
  blk.fill = "grey",
  blk.border = "darkgrey",
  label.col = "black",
  label.size = 0.7,
  arrow.width = NULL,
  arrow.fill = "transparent",
  arrow.border = "grey"
)
```

Arguments

result	A SummarizedExperiment object with assays giving promoter counts and activity with gene expression stored as column data and promoter gene id mapping stored as row data
gene	A character vector of length 1. Single gene of interest to be plotted
txdb	A TxDb object. The txdb must correspond to the genome version used in running proActiv. Here, it is recommended to use the same txdb in generating promoter annotations
ranges	A list of GRanges. Each entry in the list should correspond to a transcript that will be visualized, with Genomic Ranges giving the exons corresponding to that transcript
cex.title	A numeric value. Size of axis labels. Defaults to 0.9
cex.axis	A numeric value. Size of axis and axis ticks. Defaults to 0.9
cex.main	A numeric value. Size of plot name. Defaults to 1
blk.width	A numeric value. The width of promoters blocks in the data track. Defaults to 500 (bases)
blk.fill	A character vector of length 1. The fill colour of the promoter blocks in the data track. Defaults to 'grey'
blk.border	A character vector of length 1. The border colour of the promoter blocks in the data track. Defaults to 'darkgrey'

label.col	A character vector of length 1. The font colour of the promoter ID label in the annotation track. Defaults to 'black'
label.size	A numeric value. The size of the promoter ID label in the annotation track. Defaults to 0.7
arrow.width	A numeric value. The width of promoter arrows in the annotation track. This value is internally calculated based on the gene of interest
arrow.fill	A character vector of length 1. The fill colour of the promoter arrows in the annotation track. Defaults to 'transparent'
arrow.border	A character vector of length 1. The border colour of the promoter arrows in the annotation track. Defaults to 'grey'

Value

Outputs a plot of the promoters of the gene of interest across conditions, along with a model of transcripts belonging to the gene

Examples

```
## First, run proActiv to generate a summarizedExperiment result
files <- list.files(system.file('extdata/vignette/junctions',
                        package = 'proActiv'),
                        full.names = TRUE)
promoterAnnotation <- promoterAnnotation.gencode.v34.subset
result <- proActiv(files = files,
                        promoterAnnotation  = promoterAnnotation,
                        condition = rep(c('A549','HepG2'), each=3),
                        ncores = 1)
## Read in pre-computed ranges
txdb <- AnnotationDbi::loadDb(system.file('extdata/vignette/annotations',
                                'gencode.v34.annotation.rap1gap.sqlite',
                                package = 'proActiv'))
## Declare a gene of interest
gene <- 'ENSG00000076864.19'
## Call plot
plotPromoters(result = result, gene = gene, txdb = txdb)
```

---

preparePromoterAnnotation

Prepares promoter annotation from a gtf or txdb

Description

Prepares promoter annotation from a gtf or txdb

Usage

```
preparePromoterAnnotation(txdb, file, species)
```

Arguments

txdb	A txdb object. The txdb of the annotation version for which promoters will be identified. Either 'txdb' or 'file' argument must be specified, but not both.
file	A character object. The file path to a gtf/gff or txdb of the annotation version for which promoters will be identified. Either 'txdb' or 'file' argument must be specified, but not both.
species	A character object. The genus and species of the organism to be used in keep-StandardChromosomes(). Supported species can be seen with names(genomeStyles()).

Value

A PromoterAnnotation object. The annotated intron ranges, promoter coordinates and the promoter id mapping are attributes of the promoter annotation data.

Examples

```
txdbPath <- system.file('extdata/vignette/annotations/',
                        'gencode.v34.annotation.subset.sqlite',
                        package = 'proActiv')
txdb <- AnnotationDbi::loadDb(txdbPath)
promoterAnnotation <- preparePromoterAnnotation(txdb = txdb,
                                                species = 'Homo_sapiens')
```

---

proActiv   *Estimates promoter counts and activity in a single command*

Description

Estimates promoter counts and activity in a single command

Usage

```
proActiv(
  files,
  promoterAnnotation,
  fileLabels = NULL,
  condition = NULL,
  genome = NULL,
  ncores = 1
)
```

Arguments

files	A character vector. The list of input files for which the junction read counts will be calculated
promoterAnnotation	
	A PromoterAnnotation object containing the intron ranges, promoter coordinates and the promoter id mapping

fileLabels	A character vector. The labels of input files for which the junction read counts will be calculated. These labels will be used as column names for each output data.frame object. If not provided, filenames will be used as labels. Defaults to NULL
condition	A character vector. The condition to which each sample belong to. Must correspond to the order of the files. If supplied, results are summarized by condition. Defaults to NULL
genome	A character. Genome version. Must be specified if input file type is a BAM file. Defaults to NULL
ncores	A numeric value. The number of cores to be used for counting junction reads. Defaults to 1 (no parallelization). This parameter will be used as an argument to BiocParallel::bplapply

Value

A SummarizedExperiment object with assays giving promoter counts and activity with gene expression stored as metadata. rowData contains promoter metadata and absolute promoter activity summarized across conditions (if condition is provided)

Examples

```
files <- list.files(system.file('extdata/vignette/junctions',
                      package = 'proActiv'),
                      full.names = TRUE, pattern = 'replicate5')
promoterAnnotation <- promoterAnnotation.gencode.v34.subset
result <- proActiv(files = files,
                      promoterAnnotation  = promoterAnnotation,
                      condition = rep(c('A549', 'HepG2'), each=1),
                      fileLabels = NULL,
                      ncores = 1)
```

PromoterAnnotation-class

S4 class for promoter annotation data for a specific annotation version

Description

S4 class for promoter annotation data for a specific annotation version

Usage

```
PromoterAnnotation(
  intronRanges = GRanges(),
  promoterIdMapping = data.frame(),
  promoterCoordinates = GRanges()
)

intronRanges(x)
```

```
## S4 method for signature 'PromoterAnnotation'
intronRanges(x)

promoterIdMapping(x)

## S4 method for signature 'PromoterAnnotation'
promoterIdMapping(x)

promoterCoordinates(x)

## S4 method for signature 'PromoterAnnotation'
promoterCoordinates(x)

intronRanges(x) <- value

## S4 replacement method for signature 'PromoterAnnotation'
intronRanges(x) <- value

promoterIdMapping(x) <- value

## S4 replacement method for signature 'PromoterAnnotation'
promoterIdMapping(x) <- value

promoterCoordinates(x) <- value

## S4 replacement method for signature 'PromoterAnnotation'
promoterCoordinates(x) <- value
```

Arguments

intronRanges   A GRanges object containing annotated intron ranges

promoterIdMapping

                 A data.frame containing mapping between transcript, TSS, promoter and gene ids

promoterCoordinates

                 A GRanges object containing promoter coordinates

x   A PromoterAnnotation object

value   intronRanges, promoterIdMapping or promoterCoordinates to be assigned

Value

A promoter annotation object with three slots: intronRanges, promoterIdMapping and promoter Coordinates

Functions

- intronRanges: Getter for intronRanges

- intronRanges,PromoterAnnotation-method: Getter for intronRanges

- promoterIdMapping: Getter for promoterIdMapping

- promoterIdMapping,PromoterAnnotation-method: Getter for promoterIdMapping

- promoterCoordinates: Getter for promoterCoordinates

- `promoterCoordinates,PromoterAnnotation-method`: Getter for promoterCoordinates
- `intronRanges<-`: Setter for intronRanges
- `intronRanges<-,PromoterAnnotation-method`: Setter for intronRanges
- `promoterIdMapping<-`: Setter for promoterIdMapping
- `promoterIdMapping<-,PromoterAnnotation-method`: Setter for promoterIdMapping
- `promoterCoordinates<-`: Setter for promoterCoordinates
- `promoterCoordinates<-,PromoterAnnotation-method`: Setter for promoterCoordinates

Slots

`intronRanges` A GRanges object. The intron ranges annotated with the promoter information.

`promoterIdMapping` A data.frame object. The id mapping between transcript ids, names, TSS ids, promoter ids and gene ids.

`promoterCoordinates` A GRanges object. Promoter coordinates (TSS) with gene id and internal promoter state

Examples

```
promoterAnnotation <- PromoterAnnotation()
intronRanges(promoterAnnotation) <- intronRanges(
                                promoterAnnotation.gencode.v34.subset)
promoterIdMapping(promoterAnnotation) <- promoterIdMapping(
                                promoterAnnotation.gencode.v34.subset)
promoterCoordinates(promoterAnnotation) <- promoterCoordinates(
                                promoterAnnotation.gencode.v34.subset)
```

---

promoterAnnotation.gencode.v34.subset
Promoter annotation for Gencode.v34 (subset)

Description

Promoter annotation for Gencode.v34 (chr1:10,000,000 - 30,000,000)

Usage

```
promoterAnnotation.gencode.v34.subset
```

Format

A PromoterAnnotation (S4 Class) object containing all promoter annotation objects for Gencode.v34 chr1:10,000,000-30,000,000. The object has 3 slots:

intronRanges A GRanges object of 4,523 ranges corresponding to introns, annotated with the associated transcript.

promoterIdMapping The id mapping between transcript names, promoter ids and gene ids for Gencode v34.

promoterCoordinates A GRanges object of 1,380 ranges showing the tss coordinate for each promoter of Gencode v34 chr1:10,000,000-30,000,000, annotated with the associated gene id, coordinate of the 3' end of the first reduced exon, and intron id.

Index