Package 'RIPAT'

April 12, 2022

Title Retroviral Integration Pattern Analysis Tool (RIPAT)

Version 1.4.0

Description RIPAT is developed as an R package for retroviral integration sites annotation and distribution analysis. RIPAT needs local alignment results from BLAST and BLAT. Specific input format is depicted in RIPAT manual. RIPAT provides RV integration pattern analysis result as forms of R objects, excel file with multiple sheets and plots.

biocViews Annotation

Depends R (>= 4.0)

Imports biomaRt (>= 2.38.0), GenomicRanges (>= 1.34.0), ggplot2 (>= 3.1.0), grDevices (>= 3.5.3), IRanges (>= 2.16.0), karyoploteR (>= 1.6.3), openxlsx (>= 4.1.4), plyr (>= 1.8.4), regioneR (>= 1.12.0), rtracklayer (>= 1.42.2), stats (>= 3.5.3), stringr (>= 1.3.1), utils (>= 3.5.3)

Suggests knitr (>= 1.28)

VignetteBuilder knitr

License Artistic-2.0

URL https://github.com/bioinfo16/RIPAT/

Encoding UTF-8

RoxygenNote 7.1.0

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

git_branch RELEASE_3_14

git_last_commit 8b46270

git_last_commit_date 2021-10-26

Date/Publication 2022-04-12

Author Min-Jeong Baek [aut, cre]

Maintainer Min-Jeong Baek <mjbaek16@korea.ac.kr>

R topics documented:

`annoByCpG` *Annotate integration sites by CpG sites.*

Description

Annotate vector integration sites by CpG site data.

Usage

```
annoByCpG(hits, mapTool = 'blast', organism = 'GRCh37', interval = 5000,
          range = c(-20000, 20000), doRandom = TRUE,
          randomSize = if(doRandom){10000}else{NULL},
          includeUndecided = FALSE, outPath = getwd(),
          outFileName = paste0('RIPAT', round(unclass(Sys.time()))))
```

Arguments

`hits`	a GR object. This object made by `makeInputObj` function.
`mapTool`	a single character. Function serves two types of object such as outputs from BLAST and BLAT. Default is 'blast'. If you want to use BLAT result, use 'blat'.
`organism`	a single character. This function can run by two versions of organisms such as GRCh37, GRCh38 (Human). Default is 'GRCh37'.
`interval`	an integer vector. This number means interval number for distribution analysis. Default is 5000.

range	an integer array. The range of highlight region for analysis. Default range is c(-20000, 20000).
doRandom	TRUE or FALSE. If user types TRUE, random set is generated and user can do random distribution analysis. Default is TRUE. If this value is FALSE, random distribution analysis is not executed.
randomSize	an integer vector. A random set size. Default is 10000.
includeUndecided	TRUE or FALSE. If user want to use undecided hits in analysis, enter TRUE. Default is FALSE.
outPath	an string vector. Plots are saved in this path. Default value is R home directory.
outFileName	a character vector. Attached ID to the result file name.

Value

Return a result list that is made up of insertion and distribution result tables and GenomicRange object of CpG data.

Examples

```
data(blast_obj); data(cpg_exam_db)
saveRDS(cpg_exam_db, paste0(system.file("extdata", package = 'RIPAT'), '/GRCh37_cpg.rds'))

blast_cpg = annoByCpG(hits = blast_obj, doRandom = FALSE, outFileName = 'blast_res')
```

---

annoByGene *Annotate integration sites by genes and TSSs.*

Description

Annotate vector integration sites by gene data.

Usage

```
annoByGene(hits, mapTool = 'blast', organism = 'GRCh37', interval = 5000,
           range = c(-20000, 20000), doRandom = TRUE,
           randomSize = if(doRandom){10000}else{NULL},
           includeUndecided = FALSE, outPath = getwd(),
           outFileName = paste0('RIPAT', round(unclass(Sys.time()))))
```

Arguments

hits	a GR object. This object made by makeInputObj function.
mapTool	a single character. Function serves two types of object such as outputs from BLAST and BLAT. Default is 'blast'. If you want to use BLAT result, use 'blat'.

organism	a single character. This function can run by two versions of organisms such as GRCh37, GRCh38 (Human). Default is 'GRCh37'.
interval	an integer vector. This number means interval number for distribution analysis. Default is 5000.
range	an integer array. The range of highlight region for analysis. Default range is c(-20000, 20000).
doRandom	TRUE or FALSE. If user types TRUE, random set is generated and user can do random distribution analysis. Default is TRUE. If this value is FALSE, random distribution analysis is not executed.
randomSize	an integer vector. A random set size. Default is 10000.
includeUndecided	TRUE or FALSE. If user want to use undecided hits in analysis, enter TRUE. Default is FALSE.
outPath	an string vector. Plots are saved in this path. Default value is R home directory.
outFileName	a character vector. Attached ID to the result file name.

Value

Return a result list that is made up of insertion and distribution result tables and GenomicRange object of Gene and TSS data.

Examples

```
data(blast_obj); data(gene_exam_db); data(tss_exam_db)
saveRDS(gene_exam_db, paste0(system.file("extdata", package = 'RIPAT'), '/GRCh37_gene.rds'))
saveRDS(tss_exam_db, paste0(system.file("extdata", package = 'RIPAT'), '/GRCh37_TSS.rds'))

blast_gene = annoByGene(hits = blast_obj, doRandom = FALSE, outFileName = 'blast_res')
```

---

annoByRepeat *Annotate integration sites by repeats and microsatellites.*

Description

Annotate vector integration sites by repeat and microsatellite data.

Usage

```
annoByRepeat(hits, mapTool = 'blast', organism = 'GRCh37', interval = 5000,
             range = c(-20000, 20000), doRandom = TRUE,
             randomSize = if(doRandom){10000}else{NULL},
             includeUndecided = FALSE, outPath = getwd(),
             outFileName = paste0('RIPAT', round(unclass(Sys.time()))))
```

Arguments

hits	a GR object. This object made by makeInputObj function.
mapTool	a single character. Function serves two types of object such as outputs from BLAST and BLAT. Default is 'blast'. If you want to use BLAT result, use 'blat'.
organism	a single character. This function can run by two versions of organisms such as GRCh37, GRCh38 (Human). Default is 'GRCh37'.
interval	an integer vector. This number means interval number for distribution analysis. Default is 5000.
range	an integer array. The range of highlight region for analysis. Default range is c(-20000, 20000).
doRandom	TRUE or FALSE. If user types TRUE, random set is generated and user can do random distribution analysis. Default is TRUE. If this value is FALSE, random distribution analysis is not executed.
randomSize	an integer vector. A random set size. Default is 10000.
includeUndecided	TRUE or FALSE. If user want to use undecided hits in analysis, enter TRUE. Default is FALSE.
outPath	an string vector. Plots are saved in this path. Default value is R home directory.
outFileName	a character vector. Attached ID to the result file name.

Value

Return a result list that is made up of insertion and distribution result tables and GenomicRange object of Rpeat and microsatellite data.

Examples

```
data(blast_obj); data(repeat_exam_db); data(micro_exam_db)
saveRDS(repeat_exam_db, paste0(system.file("extdata", package = 'RIPAT'), '/GRCh37_repeat.rds'))
saveRDS(micro_exam_db, paste0(system.file("extdata", package = 'RIPAT'), '/GRCh37_microsat.rds'))

blast_repeat = annoByRepeat(hits = blast_obj, doRandom = FALSE, outFileName = 'blast_res')
```

---

annoByVar *Annotate integration sites by clinical variants.*

Description

Annotate vector integration sites by clinical variant data.

Usage

```
annoByVar(hits, mapTool = 'blast', organism = 'GRCh37', interval = 5000,
          range = c(-20000, 20000), doRandom = TRUE,
          randomSize = if(doRandom){10000}else{NULL},
          includeUndecided = FALSE, outPath = getwd(),
          outFileName = paste0('RIPAT', round(unclass(Sys.time()))))
```

Arguments

hits	a GR object. This object made by makeInputObj function.
mapTool	a single character. Function serves two types of object such as outputs from BLAST and BLAT. Default is 'blast'. If you want to use BLAT result, use 'blat'.
organism	a single character. This function can run by two versions of organisms such as GRCh37, GRCh38 (Human). Default is 'GRCh37'.
interval	an integer vector. This number means interval number for distribution analysis. Default is 5000.
range	an integer array. The range of highlight region for analysis. Default range is c(-20000, 20000).
doRandom	TRUE or FALSE. If user types TRUE, random set is generated and user can do random distribution analysis. Default is TRUE. If this value is FALSE, random distribution analysis is not executed.
randomSize	an integer vector. A random set size. Default is 10000.
includeUndecided	TRUE or FALSE. If user want to use undecided hits in analysis, enter TRUE. Default is FALSE.
outPath	an string vector. Plots are saved in this path. Default value is R home directory.
outFileName	a character vector. Attached ID to the result file name.

Value

Return a result list that is made up of insertion and distribution result tables and GenomicRange object of clinical variant data.

Examples

```
data(blast_obj); data(var_exam_db)
saveRDS(var_exam_db, paste0(system.file("extdata", package = 'RIPAT'), '/GRCh37_clinvar.rds'))

blast_clivar = annoByVar(hits = blast_obj, doRandom = FALSE, outFileName = 'blast_res')
```

blast_gene *Integration site annotation by gene data*

Description

Data used in example code of drawingKaryo and makeDocument

Usage

```
data(blast_gene)
```

Format

List object

Examples

```
data(blast_gene)
```

blast_obj *Retroviral vector integration site object*

Description

Data used in example code of integration site annotation functions

Usage

```
data(blast_obj)
```

Format

GRange object

Examples

```
data(blast_obj)
```

cpg_exam_db *Data file for annoByCpG vignette*

Description

Data used in example code of annoByCpG

Usage

```
data(cpg_exam_db)
```

Format

Data table object

Examples

```
data(cpg_exam_db)
```

drawingKaryo *Draw the karyogram plot.*

Description

Draw a karyogram plot and show integration site.

Usage

```
drawingKaryo(hits, feature, organism = 'GRCh37',
             includeUndecided = FALSE, outPath = getwd(),
             outFileName = paste0('RIPAT', round(unclass(Sys.time()))))
```

Arguments

hits	a GR object. This object made from makeInputObj function.
feature	a GR object. This object made from annotation function.
organism	a character vector. This function serves 2 versions of organisms such as GRCh37, GRCh38 (Human). Default is 'GRCh37'.
includeUndecided	TRUE or FALSE. If user want to use undecided hits in analysis, enter TRUE. Default is FALSE.
outPath	a string vector. Type path to save a plot.
outFileName	a character vector. This value used when saving the idegoram image file.

Value

Return the ideogram plot and object.

Examples

```
data(blast_obj)
data(blast_gene)
drawingKaryo(hits = blast_obj, feature = blast_gene$Gene_data, outFileName = 'blast_res')
```

---

gene_exam_db *Data file for annoByGene vignette*

Description

Data used in example code of annoByGene

Usage

```
data(gene_exam_db)
```

Format

Data table object

Examples

```
data(gene_exam_db)
```

---

makeData *Make data files for RIPAT.*

Description

Download datafiles for running RIPAT.

Usage

```
makeData(organism = 'GRCh37', dataType = 'gene')
```

Arguments

organism	a single character. Two versions of organism such as GRCh37, GRCh38 (Human). Default is 'GRCh37'.
dataType	a single character. Data type what user needs (gene, cpg, repeat and variant). Default is 'gene'.

Value

Database files are saved in the extdata directory of RIPAT.

Examples

```
makeData(organism = 'GRCh37')
```

`makeDocument` *Make the result object and document.*

Description

Rearrange the result from annotation functions.

Usage

```
makeDocument(res, dataType, excelOut = TRUE,
             includeUndecided = FALSE, outPath = getwd(),
             outFileName = paste0('RIPAT', round(unclass(Sys.time()))))
```

Arguments

`res`	a GR object. This object is output of `annoByGene`, `annoByCpG`, `annoByRepeat`, `annoByVar` function.
`dataType`	a character vector. User enter the annotation type of input such as gene, cpg, repeat and variant.
`excelOut`	TRUE or FALSE. If user want to make excel file, enter TRUE. Default is TRUE.
`includeUndecided`	TRUE or FALSE. If user want to use undecided hits in analysis, enter TRUE. Default is FALSE.
`outPath`	an string vector. Plots are saved in this path. Default value is R home directory.
`outFileName`	a character vector. Attached ID to the result file name.

Value

Make output table and excel files about vector integration sites and proportion test result.

Examples

```
data(blast_gene)
makeDocument(res = blast_gene, dataType = 'gene', outFileName = 'blast_gene_res')
```

makeInputObj *Make the retroviral vector integration site object.*

Description

Make an input object for annotation functions.

Usage

```
makeInputObj(inFile, mapTool = 'blast',
             vectorPos = 'front', outPath = getwd(),
             outFileName = paste0('RIPAT', round(unclass(Sys.time()))))
```

Arguments

inFile	a string vector. The path of a local alignment result file. File do not include any header and comment.
mapTool	a character vector. Function serves two types of file such as outputs from BLAST and BLAT. Default is 'blast'. If you want to use BLAT result, use 'blat'.
vectorPos	a character vector. Sets the position of vector on sequences. Default value is 'front'. If the vector is located at the behind of sequence, you can change it to 'behind'.
outPath	a string vector. Directory path of tab-deliminated hit files generated by this function.
outFileName	a character vector. Attached character to the result file name.

Value

Return two types of outputs. Text file and R object. Available hit data from input is written in text file and generated as a list of GenomicRange(GR) format object.

Examples

```
blast_obj = makeInputObj(inFile = paste0(.libPaths()[1], '/RIPAT/scripts/A5_15856M_BLASTn.txt'))
```

makeInputObj2 *Make the retroviral vector integration site object.*

Description

Make an input object for annotation functions.

Usage

```
makeInputObj2(inDir, id, mapTool = 'blast',
              vectorPos = 'front', outPath = getwd(),
              outFileName = paste0('RIPAT', round(unclass(Sys.time()))))
```

Arguments

inDir	a string vector. Location of a directory that has a local alignment result files. All alignment result files do not include any header and comment.
id	a character vector. the specific words that can appoint alignment file names exclusively.
mapTool	a character vector. Function serves two types of file such as outputs from BLAST and BLAT. Default is 'blast'. If you want to use BLAT result, use 'blat'.
vectorPos	a character vector. Sets the position of vector on sequences. Default value is 'front'. If the vector is located at the behind of sequence, you can change it to 'behind'.
outPath	a string vector. Directory path of tab-deliminated hit files generated by this function.
outFileName	a character vector. Attached character to the result file name.

Value

Return two types of outputs. Text file and R object. Available hit data from input is written in text file and generated as a list of GenomicRange(GR) format object.

Examples

```
blast_obj = makeInputObj2(inDir = system.file("scripts", package = "RIPAT"), id = 'BLASTn')
```

micro_exam_db *Data file for annoByRepeat vignette*

Description

Data used in example code of annoByRepeat

Usage

```
data(micro_exam_db)
```

Format

Data table object

Examples

```
data(micro_exam_db)
```

repeat_exam_db *Data file for annoByRepeat vignette*

Description

Data used in example code of annoByRepeat

Usage

```
data(repeat_exam_db)
```

Format

Data table object

Examples

```
data(repeat_exam_db)
```

tss_exam_db *Data file for annoByGene vignette*

Description

Data used in example code of annoByGene

Usage

```
data(tss_exam_db)
```

Format

Data table object

Examples

```
data(tss_exam_db)
```

var_exam_db *Data file for annoByVar vignette*

Description

Data used in example code of annoByVar

Usage

```
data(var_exam_db)
```

Format

Data table object

Examples

```
data(var_exam_db)
```

Index