Package 'SNPannotator'

April 27, 2022

Type Package

Title Investigating the Functional Characteristics of Selected SNPs and Their Vicinity Genomic Region

Version 0.2.2.0

Date 2022-04-27

Author Alireza Ani [aut, cre],
Zoha Kamali [aut],
Ahmad Vaez [aut]

Maintainer Alireza Ani <a.ani@umcg.nl>

Depends R (>= 3.5)

Imports data.table, httr, jsonlite, xml2, openxlsx, progress, doParallel, foreach, ggplot2, kableExtra, methods, rmarkdown

Description To investigate the functional characteristics of selected SNPs and their vicinity genomic region. Linked SNPs in moderate to high linkage disequilibrium (e.g. r2>0.50) with the corresponding index SNPs will be selected for further analysis.

License GPL-3

Encoding UTF-8

RoxygenNote 7.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2022-04-27 19:40:02 UTC

R topics documented:

annotate *Run the annotation pipeline on a list of variants*

Description

This function receives a list of variants and checks their information on Ensembl website via the Ensembl REST API server.

Usage

```
annotate(
  rslist,
  server,
  db,
  outputPath,
  window_size = 500,
  r2 = 0.5,
  LDlist = TRUE,
  cadd = FALSE,
  geneNames.file = NULL,
  regulatoryType.file = NULL,
  cores = 0
)
```

Arguments

rslist	A vector of rs numbers.
server	Name of the server. "https://rest.ensembl.org" can be used for GRCh38 and "https://grch37.rest.ensembl.org" for GRCh37.
db	The population database for calculating LD scores. This can be found using `listDatabases` function.
outputPath	The report file will be saved in this path as an Excel file (.xlsx)
window_size	Number of base pairs around the variant for checking LD scores (max = 500kb)
r2	The LD threshold for selecting variants around the target SNP.
LDlist	If set to TRUE, variants in high LD will be found and added to the output.
cadd	If set to TRUE, the CADD scores will be added to variant information.
geneNames.file	path the gene information file (*.rds). Default value is NULL and ENSEMBL website will be checked if no file is provided.
regulatoryType.file	path the variants regulatory type information file (*.rds). Default value is NULL and this step will be skipped if no file is provided.
cores	set to a value above 0 for parallel processing.

Value

a data table with all variant information is returned.

Examples

```
## Not run:
# select the required server
server <- "https://grch37.rest.ensembl.org"

# select the database for population data
# this can be selected from listDatabases() function
db <- "1000GENOMES:phase_3:EUR"

# create a vector of required SNPs
rslist=c('rs236349')

output <- annotate(rslist,server,db,
  outputPath = paste(tempdir(),'sampleOutput.xlsx',sep="/"),
  window_size = 500,
  r2 = .9,
  cadd = FALSE)

## End(Not run)
```

`LDlist` *Find variants in high LD with the lead SNP.*

Description

This function returns a list of variables that are in high LD with the lead variant.

Usage

```
LDlist(rslist, server, db, window_size, r2)
```

Arguments

`rslist`	A vector of rs numbers.
`server`	Name of the server. "https://rest.ensembl.org" can be used for GRCh38 and "https://grch37.rest.ensembl.org" for GRCh37.
`db`	The population database for calculating LD scores. This can be found using `listDatabases` function.
`window_size`	Number of base pairs around the variant for checking LD scores (max = 500kb)
`r2`	The LD threshold for selecting variants around the target SNP.

Value

a data table with variant information.

Examples

```
# select the required server
server <- "https://grch37.rest.ensembl.org"

# select the database for population data
# this can be selected from listDatabases() function
db <- "1000GENOMES:phase_3:EUR"

# create a vector of required SNPs
rslist=c('rs236349')

# fetch the LD list
output <- LDlist(rslist,server,db,500,0.8)
```

`listDatabases` *List population from human database (1000 Genomes project)*

Description

This function list the name, description and size of the available populations in 1000 Genomes project database. This database will be used for returning variables in high LD with the target SNP.

Usage

```
listDatabases(server)
```

Arguments

`server`	name of the server. "https://rest.ensembl.org" can be used for GRCh38 and "https://grch37.rest.ensembl.org" for GRCh37.

Value

A data table is returned which includes the name, description and size of the available populations in 1000 Genomes project database.

Examples

```
# select the required Ensembl server
server = "https://rest.ensembl.org"

# check the available population data for the selected server
listDatabases("https://rest.ensembl.org")
```

pingServer *Checks if the service is alive*

Description

This function test whether the Ensembl server is accessible or not

Usage

```
pingServer(server)
```

Arguments

server	name of the server. "https://rest.ensembl.org" can be used for GRCh38 and "https://grch37.rest.ensembl.org" for GRCh37.

Value

a message is displayed to the user

Examples

```
# select the required Ensembl server
server = "https://rest.ensembl.org"

# check if server is accessible
pingServer(server)
```

releaseVersion *Shows the data releases available on this REST server.*

Description

Shows the data releases available on this REST server. May return more than one release (unfrequent non-standard Ensembl configuration).

Usage

```
releaseVersion(server)
```

Arguments

server	name of the server.

Value

a message is displayed to the user

Examples

```
# select the required Ensembl server
server = "https://rest.ensembl.org"

# check the data releases of the selected server
releaseVersion(server)
```

Index