Package 'LDlinkR'

August 8, 2022

Type Package

Title Calculating Linkage Disequilibrium (LD) in Human Population Groups of Interest

Version 1.2.2

Maintainer Timothy A. Myers <myersta@mail.nih.gov>

Description
Provides access to the 'LDlink' API (<https://ldlink.nci.nih.gov/?tab=apiaccess>) using the R console. This programmatic access facilitates researchers who are interested in perform-
ing batch queries in 1000 Genomes Project (2015) <doi:10.1038/nature15393> data using 'LDlink'. 'LDlink' is an interactive and powerful suite of web-based tools for querying germline variants in human population groups of interest. For more details, please see Machiela et al. (2015) <doi:10.1093/bioinformatics/btv402>.

License GPL (>= 2)

URL https://ldlink.nci.nih.gov

BugReports https://github.com/CBIIT/LDlinkR/issues

Encoding UTF-8

Imports httr (>= 1.4.0), utils (>= 3.4.2)

Suggests testthat, knitr, rmarkdown, spelling

VignetteBuilder knitr

RoxygenNote 7.2.0

Language en-US

NeedsCompilation no

Author Timothy A. Myers [aut, cre] (<https://orcid.org/0000-0001-8127-3446>),
Stephen J. Chanock [aut],
Mitchell J. Machiela [aut] (<https://orcid.org/0000-0001-6538-9705>)

Repository CRAN

Date/Publication 2022-08-08 16:10:02 UTC

R topics documented:

`LDexpress` *Determine if genomic variants are associated with gene expression.*

Description

Search if a list of genomic variants (or variants in LD with those variants) is associated with gene expression in tissues of interest. Quantitative trait loci data is downloaded from the GTEx Portal (https://gtexportal.org/home/).

Usage

```
LDexpress(
  snps,
  pop = "CEU",
  tissue = "ALL",
  r2d = "r2",
  r2d_threshold = 0.1,
  p_threshold = 0.1,
  win_size = 5e+05,
  genome_build = "grch37",
  token = NULL,
  file = FALSE
)
```

Arguments

`snps`	between 1 - 10 variants, using an rsID or chromosome coordinate (e.g. "chr7:24966446")
`pop`	a 1000 Genomes Project population, (e.g. YRI or CEU), multiple allowed, default = "CEU". Use the 'list_pop' function to see a list of available human reference populations.

tissue	select from 1 - 54 non-diseased tissue sites collected for the GTEx project, multiple allowed. Acceptable user input is taken either from "tissue_name_ldexpress" or "tissue_abbrev_ldexpress" (tissue abbreviation) code listed in available GTEx tissue sites using the `list_getex_tissues()` function (e.g. "ADI_SUB" for Adipose Subcutaneous). Input is case sensitive. Default = "ALL" for all available tissue types.
r2d	either "r2" for LD R2 or "d" for LD D', default = "r2".
r2d_threshold	R2 or D' (depends on 'r2d' user input parameter) threshold for LD filtering. Any variants within -/+ of the specified genomic window and R^2 or D' less than the threshold will be removed. Value needs to be in the range 0 to 1. Default value is 0.1.
p_threshold	define the eQTL significance threshold used for returning query results. Default value is 0.1 which returns all GTEx eQTL associations with P-value less than 0.1.
win_size	set genomic window size for LD calculation. Specify a value greater than or equal to zero and less than or equal to 1,000,000 basepairs (bp). Default value is -/+ 500,000bp.
genome_build	Choose between one of the three options...'grch37' for genome build GRCh37 (hg19), 'grch38' for GRCh38 (hg38), or 'grch38_high_coverage' for GRCh38 High Coverage (hg38) 1000 Genome Project data sets. Default is GRCh37 (hg19).
token	LDlink provided user token, default = NULL, register for token at https://ldlink.nci.nih.gov/?tab=apiaccess
file	Optional character string naming a path and file for saving results. If file = FALSE, no file will be generated, default = FALSE.

Value

A data frame of all query variant RS numbers, respective QTL which are in LD with query variant, and associated gene expression.

Examples

```
## Not run: LDexpress(snps = c("rs345", "rs456"),
                   pop = c("YRI", "CEU"),
                   tissue = c("ADI_SUB", "ADI_VIS_OME"),
                   r2d = "r2",
                   r2d_threshold = "0.1",
                   p_threshold = "0.1",
                   win_size = "500000",
                   genome_build = "grch37"
                   token = Sys.getenv("LDLINK_TOKEN")
                  )

## End(Not run)
```

LDhap *Calculates population specific haplotype frequencies of all haplotypes observed for a list of query variants.*

Description

Calculates population specific haplotype frequencies of all haplotypes observed for a list of query variants.

Usage

```
LDhap(
  snps,
  pop = "CEU",
  token = NULL,
  file = FALSE,
  table_type = "haplotype",
  genome_build = "grch37"
)
```

Arguments

snps	list of between 1 - 30 variants, using an rsID or chromosome coordinate (e.g. "chr7:24966446")
pop	a 1000 Genomes Project population, (e.g. YRI or CEU), multiple allowed, default = "CEU"
token	LDlink provided user token, default = NULL, register for token at https://ldlink.nci.nih.gov/?tab=apiaccess
file	Optional character string naming a path and file for saving results. If file = FALSE, no file will be generated, default = FALSE.
table_type	Choose from one of four options available to determine output format type...'haplotype', 'variant', 'both' and 'merged'. Default = "haplotype".
genome_build	Choose between one of the three options...'grch37' for genome build GRCh37 (hg19), 'grch38' for GRCh38 (hg38), or 'grch38_high_coverage' for GRCh38 High Coverage (hg38) 1000 Genome Project data sets. Default is GRCh37 (hg19).

Value

a data frame or list

Examples

```
## Not run: LDhap(c("rs3", "rs4", "rs148890987"), "CEU", token = Sys.getenv("LDLINK_TOKEN"))
## Not run: LDhap("rs148890987", c("YRI", "CEU"), token = Sys.getenv("LDLINK_TOKEN"))
```

LDmatrix *Generates a data frame of pairwise linkage disequilibrium statistics.*

Description

Generates a data frame of pairwise linkage disequilibrium statistics.

Usage

```
LDmatrix(
  snps,
  pop = "CEU",
  r2d = "r2",
  token = NULL,
  file = FALSE,
  genome_build = "grch37"
)
```

Arguments

snps	list of between 2 - 1,000 variants, using an rsID or chromosome coordinate (e.g. "chr7:24966446")
pop	a 1000 Genomes Project population, (e.g. YRI or CEU), multiple allowed, default = "CEU"
r2d	r2d, either "r2" for LD R2 or "d" for LD D', default = "r2"
token	LDlink provided user token, default = NULL, register for token at https://ldlink.nci.nih.gov/?tab=apiaccess
file	Optional character string naming a path and file for saving results. If file = FALSE, no file will be generated, default = FALSE.
genome_build	Choose between one of the three options...'grch37' for genome build GRCh37 (hg19), 'grch38' for GRCh38 (hg38), or 'grch38_high_coverage' for GRCh38 High Coverage (hg38) 1000 Genome Project data sets. Default is GRCh37 (hg19).

Value

a data frame

Examples

```
## Not run: LDmatrix(c("rs3", "rs4", "rs148890987"),
                  "YRI", "r2",
                  token = Sys.getenv("LDLINK_TOKEN"))

## End(Not run)
```

LDpair *Investigates potentially correlated alleles for a pair of variants.*

Description

Investigates potentially correlated alleles for a pair of variants.

Usage

```
LDpair(
  var1,
  var2,
  pop = "CEU",
  token = NULL,
  output = "table",
  file = FALSE,
  genome_build = "grch37"
)
```

Arguments

var1	the first RS number or genomic coordinate (e.g. "chr7:24966446")
var2	the second RS number or genomic coordinate (e.g. "ch7:24966446")
pop	a 1000 Genomes Project population(s), (e.g. YRI or CEU), multiple allowed, default = "CEU"
token	LDlink provided user token, default = NULL, register for token at https://ldlink.nci.nih.gov/?tab=apiaccess
output	two output options available, "text", which displays a two-by-two matrix displaying haplotype counts and allele frequencies along with other statistics, or "table", which displays the same data in rows and columns, default = "table"
file	Optional character string naming a path and file for saving results. If file = FALSE, no file will be generated, default = FALSE.
genome_build	Choose between one of the three options...'grch37' for genome build GRCh37 (hg19), 'grch38' for GRCh38 (hg38), or 'grch38_high_coverage' for GRCh38 High Coverage (hg38) 1000 Genome Project data sets. Default is GRCh37 (hg19).

Value

text or data frame, depending on the output option

Examples

```
## Not run: LDpair(var1 = "rs3", var2 = "rs4", pop = "YRI", token = Sys.getenv("LDLINK_TOKEN"))
## Not run: LDpair("rs3", "rs4", "YRI", token = Sys.getenv("LDLINK_TOKEN"), "text")
```

LDpop *Investigates allele frequencies and linkage disequilibrium patterns across 1000 Genomes Project populations.*

Description

Investigates allele frequencies and linkage disequilibrium patterns across 1000 Genomes Project populations.

Usage

```
LDpop(
  var1,
  var2,
  pop = "CEU",
  r2d = "r2",
  token = NULL,
  file = FALSE,
  genome_build = "grch37"
)
```

Arguments

var1	the first RS number or genomic coordinate (e.g. "chr7:24966446")
var2	the second RS number or genomic coordinate (e.g. "ch7:24966446")
pop	a 1000 Genomes Project population(s), (e.g. YRI or CEU), multiple allowed, default = "CEU"
r2d	either "r2" for LD R2 or "d" for LD D', default = "r2"
token	LDlink provided user token, default = NULL, register for token at https://ldlink.nci.nih.gov/?tab=apiaccess
file	Optional character string naming a path and file for saving results. If file = FALSE, no file will be generated, default = FALSE.
genome_build	Choose between one of the three options...'grch37' for genome build GRCh37 (hg19), 'grch38' for GRCh38 (hg38), or 'grch38_high_coverage' for GRCh38 High Coverage (hg38) 1000 Genome Project data sets. Default is GRCh37 (hg19).

Value

a data frame

Examples

```
## Not run: LDpop(var1 = "rs3", var2 = "rs4",
               pop = "YRI", r2d = "r2",
               token = Sys.getenv("LDLINK_TOKEN"))

## End(Not run)
```

LDproxy *Explore proxy and putative functional variants for a single query variant.*

Description

Explore proxy and putative functional variants for a single query variant.

Usage

```
LDproxy(
  snp,
  pop = "CEU",
  r2d = "r2",
  token = NULL,
  file = FALSE,
  genome_build = "grch37"
)
```

Arguments

snp	an rsID or chromosome coordinate (e.g. "chr7:24966446"), one per query
pop	a 1000 Genomes Project population, (e.g. YRI or CEU), multiple allowed, default = "CEU"
r2d	either "r2" for LD R2 or "d" for LD D', default = "r2"
token	LDlink provided user token, default = NULL, register for token at https://ldlink.nci.nih.gov/?tab=apiaccess
file	Optional character string naming a path and file for saving results. If file = FALSE, no file will be generated, default = FALSE.
genome_build	Choose between one of the three options...'grch37' for genome build GRCh37 (hg19), 'grch38' for GRCh38 (hg38), or 'grch38_high_coverage' for GRCh38 High Coverage (hg38) 1000 Genome Project data sets. Default is GRCh37 (hg19).

Value

a data frame

Examples

```
## Not run: LDproxy("rs456", "YRI", "r2", token = Sys.getenv("LDLINK_TOKEN"))
```

LDproxy_batch *Query LDproxy using a list of query variants, one per line.*

Description

Query LDproxy using a list of query variants, one per line.

Usage

```
LDproxy_batch(
  snp,
  pop = "CEU",
  r2d = "r2",
  token = NULL,
  append = FALSE,
  genome_build = "grch37"
)
```

Arguments

snp	a character string or data frame listing rsID's or chromosome coordinates (e.g. "chr7:24966446"), one per line
pop	a 1000 Genomes Project population, (e.g. YRI or CEU), multiple allowed, default = "CEU"
r2d	either "r2" for LD R2 or "d" for LD D', default = "r2"
token	LDlink provided user token, default = NULL, register for token at https://ldlink.nci.nih.gov/?tab=apiaccess
append	A logical. If TRUE, output for each query variant is appended to a text file. If FALSE, output of each query variant is saved in its own text file. Default is FALSE.
genome_build	Choose between one of the three options...'grch37' for genome build GRCh37 (hg19), 'grch38' for GRCh38 (hg38), or 'grch38_high_coverage' for GRCh38 High Coverage (hg38) 1000 Genome Project data sets. Default is GRCh37 (hg19).

Value

text file(s) are saved to the current working directory.

Examples

```
## Not run: snps_to_upload <- c("rs3", "rs4")
## Not run: LDproxy_batch(snp = snps_to_upload, token = Sys.getenv("LDLINK_TOKEN"), append = FALSE)
```

`LDtrait` *Determine if genomic variants are associated with a trait or disease.*

Description

Search if a list of variants (or variants in LD with those variants) have been previously associated with a trait or disease. Trait and disease data is updated nightly from the GWAS Catalog (https://www.ebi.ac.uk/gwas/docs/file-downloads.

Usage

```
LDtrait(
  snps,
  pop = "CEU",
  r2d = "r2",
  r2d_threshold = 0.1,
  win_size = 5e+05,
  token = NULL,
  file = FALSE,
  genome_build = "grch37"
)
```

Arguments

`snps`	between 1 - 50 variants, using an rsID or chromosome coordinate (e.g. "chr7:24966446"). All input variants must match a bi-allelic variant.
`pop`	a 1000 Genomes Project population, (e.g. YRI or CEU), multiple allowed, default = "CEU". Use the 'list_pop' function to see a list of available human reference populations.
`r2d`	use "r2" to filter desired output from a threshold based on estimated LD R2 (R squared) or "d" for LD D' (D-prime), default = "r2".
`r2d_threshold`	R2 or D' (depends on 'r2d' user input parameter) threshold for LD filtering. Any variants within -/+ of the specified genomic window and R^2 or D' less than the threshold will be removed. Value needs to be in the range 0 to 1. Default value is 0.1.
`win_size`	set genomic window size for LD calculation. Specify a value greater than or equal to zero and less than or equal to 1,000,000bp. Default value is -/+ 500,000 bp.
`token`	LDlink provided user token, default = NULL, register for token at https://ldlink.nci.nih.gov/?tab=apiaccess
`file`	Optional character string naming a path and file for saving results. If file = FALSE, no file will be generated, default = FALSE.
`genome_build`	Choose between one of the three options...'grch37' for genome build GRCh37 (hg19), 'grch38' for GRCh38 (hg38), or 'grch38_high_coverage' for GRCh38 High Coverage (hg38) 1000 Genome Project data sets. Default is GRCh37 (hg19).

Value

A data frame of all query variant RS numbers with a list of queried variants in LD with a variant reported in the GWAS Catalog (https://www.ebi.ac.uk/gwas/docs/file-downloads.

Examples

```
## Not run: LDtrait(snps = "rs456",
                    pop = c("YRI", "CEU"),
                    r2d = "r2",
                    r2d_threshold = "0.1",
                    win_size = "500000",
                    token = Sys.getenv("LDLINK_TOKEN")
                   )

## End(Not run)
```

`list_chips` *Provides a data frame listing the names and abbreviation codes for available commercial SNP Chip Arrays from Illumina and Affymetrix.*

Description

Provides a data frame listing the names and abbreviation codes for available commercial SNP Chip Arrays from Illumina and Affymetrix.

Usage

```
list_chips()
```

Value

a data frame listing the names and abbreviation codes for available SNP Chip Arrays from Illumina and Affymetrix

Examples

```
list_chips()
```

`list_gtex_tissues` *Provides a data frame listing the GTEx full names, 'LDexpress' full names (without spaces) and acceptable abbreviation codes of the 54 non-diseased tissue sites collected for the GTEx Portal and used as input for the 'LDexpress' function.*

Description

Provides a data frame listing the GTEx full names, 'LDexpress' full names (without spaces) and acceptable abbreviation codes of the 54 non-diseased tissue sites collected for the GTEx Portal and used as input for the 'LDexpress' function.

Usage

```
list_gtex_tissues()
```

Value

a data frame listing the GTEx tissues, their names and abbreviation codes used as input for LDexpress.

Examples

```
list_gtex_tissues()
```

`list_pop` *Provides a data frame listing the available reference populations from the 1000 Genomes Project.*

Description

Provides a data frame listing the available reference populations from the 1000 Genomes Project.

Usage

```
list_pop()
```

Value

a data frame listing the available reference populations, continental (ex: European, African, and Admixed American) and sub-populations (ex: Finnish, Gambian, and Peruvian)

Examples

```
list_pop()
```

SNPchip *Find commercial genotyping chip arrays for variants of interest.*

Description

Find commercial genotyping chip arrays for variants of interest.

Usage

```
SNPchip(
  snps,
  chip = "ALL",
  token = NULL,
  file = FALSE,
  genome_build = "grch37"
)
```

Arguments

snps	between 1 - 5,000 variants, using an rsID or chromosome coordinate (e.g. "chr7:24966446")
chip	chip or arrays, platform code(s) for a SNP chip array, ALL_Illumina, ALL_Affy or ALL, default=ALL
token	LDlink provided user token, default = NULL, register for token at https://ldlink.nci.nih.gov/?tab=apiaccess
file	Optional character string naming a path and file for saving results. If file = FALSE, no file will be generated, default = FALSE.
genome_build	Choose between one of the three options...'grch37' for genome build GRCh37 (hg19), 'grch38' for GRCh38 (hg38), or 'grch38_high_coverage' for GRCh38 High Coverage (hg38) 1000 Genome Project data sets. Default is GRCh37 (hg19).

Value

a data frame

Examples

```
## Not run: SNPchip(c("rs3", "rs4", "rs148890987"), "ALL",
                 token = Sys.getenv("LDLINK_TOKEN"))

## End(Not run)
## Not run: SNPchip(c("rs3", "rs4", "rs148890987"),
                 c("A_CHB2", "A_SNP5.0"),
                 token = Sys.getenv("LDLINK_TOKEN"))

## End(Not run)
## Not run: SNPchip("rs148890987", "ALL_Affy", token = Sys.getenv("LDLINK_TOKEN"))
```

SNPclip *Prune a list of variants by linkage disequilibrium.*

Description

Prune a list of variants by linkage disequilibrium.

Usage

```
SNPclip(
  snps,
  pop = "CEU",
  r2_threshold = "0.1",
  maf_threshold = "0.01",
  token = NULL,
  file = FALSE,
  genome_build = "grch37"
)
```

Arguments

snps	a list of between 1 - 5,000 variants, using an rsID or chromosome coordinate (e.g. "chr7:24966446")
pop	a 1000 Genomes Project population, (e.g. YRI or CEU), multiple allowed, default = "CEU"
r2_threshold	LD R2 threshold between 0-1, default = 0.1
maf_threshold	minor allele frequency threshold between 0-1, default = 0.01
token	LDlink provided user token, default = NULL, register for token at https://ldlink.nci.nih.gov/?tab=apiaccess
file	Optional character string naming a path and file for saving results. If file = FALSE, no file will be generated, default = FALSE.
genome_build	Choose between one of the three options...'grch37' for genome build GRCh37 (hg19), 'grch38' for GRCh38 (hg38), or 'grch38_high_coverage' for GRCh38 High Coverage (hg38) 1000 Genome Project data sets. Default is GRCh37 (hg19).

Value

a data frame

Examples

```
## Not run: SNPclip(c("rs3", "rs4", "rs148890987"), "YRI", "0.1", "0.01",
                    token = Sys.getenv("LDLINK_TOKEN"))

## End(Not run)
```

Index