Package: gwasvcf (via r-universe)

September 24, 2024

Title Tools for Dealing with GWAS Summary Data in VCF Format

Version 0.1.2

Description Tools for dealing with GWAS summary data in VCF format. Includes reading, querying, writing, as well as helper functions such as LD proxy searches.

License MIT + file LICENSE

URL https://github.com/mrcieu/gwasvcf

BugReports https://github.com/mrcieu/gwasvcf/issues

Depends R (>= 4.0.0)

Imports BiocGenerics, Biostrings, data.table, dplyr, genetics.binaRies, GenomeInfoDb, GenomicRanges, gwasglue2, IRanges, magrittr, RCurl, rlang, Rsamtools, RSQLite, S4Vectors, stringr, SummarizedExperiment, utils, VariantAnnotation

Suggests knitr, rmarkdown, testthat

VignetteBuilder knitr

Remotes github::mrcieu/genetics.binaRies, github::mrcieu/gwasglue2

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

SystemRequirements GNU unzip

Repository https://mrcieu.r-universe.dev

RemoteUrl https://github.com/MRCIEU/gwasvcf

RemoteRef HEAD

RemoteSha 477b365da8522e9a47f3bce51993d5f36df49ceb

Contents

`check_bcftools` *Check if the tools_bcftools option is set*

Description

See set_bcftools() for more information

Usage

```
check_bcftools()
```

Value

TRUE or FALSE

check_plink *Check if the tools_plink option is set*

Description

See set_plink() for more information

Usage

```
check_plink()
```

Value

TRUE or FALSE

create_ldref_sqlite *Create LD reference sqlite database for tags*

Description

This is used for looking up proxies

Usage

```
create_ldref_sqlite(bfile, dbname, tag_r2 = 0.6)
```

Arguments

bfile	path to plink file
dbname	dbname to produce (overwrites existing if exists)
tag_r2	minimum tag r2

`create_pval_index_from_vcf` *Create pval index from GWAS-VCF file*

Description

Create a separate file called `<id>.pvali` which is used to speed up p-value queries.

Usage

```
create_pval_index_from_vcf(vcffile, maximum_pval, indexname)
```

Arguments

`vcffile`	VCF filename
`maximum_pval`	Maximum p-value to include. Default = 0.05
`indexname`	index file name to create. Deletes existing file if exists.

`create_rsidx_index_from_vcf` *Create RSID index from VCF*

Description

Create RSID index from VCF

Usage

```
create_rsidx_index_from_vcf(vcf, indexname)
```

Arguments

`vcf`	VCF filename
`indexname`	index file name to create. Deletes existing file if exists.

create_rsidx_sub_index *Create new index from existing index using a subset of rsids*

Description

Note this requires a modified version of plink that allows ld-window-r2 flag for –r option. Available here: https://github.com/explodecomputer/plink-ng

Usage

```
create_rsidx_sub_index(rsid, rsidx, newindex)
```

Arguments

rsid	Vector of rsids
rsidx	Existing index
newindex	New index (Note: will delete existing file if exists)

Value

NULL, creates new index file

create_vcf *Create GWAS vcf*

Description

Create GWAS vcf

Usage

```
create_vcf(
  chrom,
  pos,
  nea,
  ea,
  snp = NULL,
  ea_af = NULL,
  effect = NULL,
  se = NULL,
  pval = NULL,
  n = NULL,
  ncase = NULL,
  name = NULL
)
```

Arguments

chrom	chrom vector
pos	pos vector
nea	nea vector
ea	ea vector
snp	Optional vector
ea_af	Optional vector
effect	Optional vector
se	Optional vector
pval	Optional vector
n	Optional vector
ncase	Optional vector
name	Optional vector

Value

vcf object

---

get_ld_proxies *Find LD proxies for a set of SNPs*

Description

Find LD proxies for a set of SNPs

Usage

```
get_ld_proxies(
  rsid,
  bfile,
  searchspace = NULL,
  tag_kb = 5000,
  tag_nsnp = 5000,
  tag_r2 = 0.6,
  threads = 1,
  out = tempfile()
)
```

Arguments

rsid	list of rs IDs
bfile	ld reference panel
searchspace	Optional list of rs IDs to use as potential proxies
tag_kb	=5000 Proxy parameter
tag_nsnp	=5000 Proxy parameter
tag_r2	=0.6 Proxy parameter
threads	Number of threads to use (=1)
out	temporary output file

Value

data frame

---

gwasvcf_to_summaryset *Create a SummarySet*

Description

Returns a gwasglue2 SummarySet object

Usage

```
gwasvcf_to_summaryset(vcf)
```

Arguments

vcf	Path or URL to GWAS-VCF file or VCF object e.g. output from `VariantAnnotation::readVcf()`, `create_vcf()` or `query_gwas()`

---

merge_vcf *Merge two GWAS VCF objects*

Description

Returns merged intersection of two VCF objects

Usage

```
merge_vcf(a, b)
```

Arguments

a	VCF object
b	VCF object

Value

SimpleList of VCF objects

---

parse_chrompos *Parse chromosome:position*

Description

Takes data frame or vector of chromosome position ranges and parses to granges object

Usage

```
parse_chrompos(chrompos, radius = NULL)
```

Arguments

chrompos	Either vector of chromosome and position ranges e.g. "1:1000" or "1:1000-2000", or data frame with columns chrom, start, end.
radius	Add radius to the specified positions. Default = NULL

Value

GRanges object

---

proxy_match *Extract SNPs from vcf file*

Description

Finds proxies if necessary

Usage

```
proxy_match(
  vcf,
  rsid,
  bfile = NULL,
  proxies = "yes",
  tag_kb = 5000,
  tag_nsnp = 5000,
  tag_r2 = 0.6,
  threads = 1,
  rsidx = NULL,
  dbfile = NULL
)
```

Arguments

vcf	vcf file name
rsid	list of rs IDs
bfile	ld reference panel (plink)
proxies	="yes" If SNPs are absent then look for proxies (yes) or not (no). Can also mask all target SNPs and only return proxies (only), for testing purposes
tag_kb	=5000 Proxy parameter
tag_nsnp	=5000 Proxy parameter
tag_r2	=0.6 Proxy parameter
threads	Number of threads to use (=1)
rsidx	Path to rsidx index
dbfile	ld tag database (sqlite)

Value

data frame

---

query_chrompos_bcftools

Query chromosome and position using bcftools

Description

Query chromosome and position using bcftools

Usage

```
query_chrompos_bcftools(chrompos, vcffile, id = NULL)
```

Arguments

chrompos	Either vector of chromosome and position ranges e.g. "1:1000" or "1:1000-2000", or data frame with columns chrom, start, end.
vcffile	Path to .vcf.gz GWAS summary data file
id	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter

Value

vcf object

---

query_chrompos_file *Query vcf file, extracting by chromosome and position*

Description

Query vcf file, extracting by chromosome and position

Usage

```
query_chrompos_file(chrompos, vcffile, id = NULL, build = "GRCh37")
```

Arguments

chrompos	Either vector of chromosome and position ranges e.g. "1:1000" or "1:1000-2000", or data frame with columns chrom, start, end.
vcffile	Path to .vcf.gz GWAS summary data file
id	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter
build	Default="GRCh37" Build of vcffile

Value

VCF object

query_chrompos_vcf *Query chrompos from vcf object*

Description

Query chrompos from vcf object

Usage

```
query_chrompos_vcf(chrompos, vcf, id = NULL)
```

Arguments

chrompos	Either vector of chromosome and position ranges e.g. "1:1000" or "1:1000-2000", or data frame with columns chrom, start, end.
vcf	VCF object (e.g. from readVcf)
id	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter

Value

VCF object

query_gwas *Query data from vcf file*

Description

Read in GWAS summary data with filters on datasets (if multiple datasets per file) and/or chromosome/position, rsids or pvalues. Chooses most optimal choice for the detected operating system. Typically chrompos searches are the fastest. On Windows, rsid or pvalue filters from a file will be slow.

Usage

```
query_gwas(
  vcf,
  chrompos = NULL,
  rsid = NULL,
  pval = NULL,
  id = NULL,
  rsidx = NULL,
  pvali = NULL,
  build = "GRCh37",
  os = Sys.info()["sysname"],
```

```
  proxies = "no",
  bfile = NULL,
  dbfile = NULL,
  tag_kb = 5000,
  tag_nsnp = 5000,
  tag_r2 = 0.6,
  threads = 1
)
```

Arguments

vcf	Path or URL to GWAS-VCF file or VCF object e.g. output from `VariantAnnotation::readVcf()` or `create_vcf()`
chrompos	Either vector of chromosome and position ranges e.g. "1:1000" or "1:1000-2000", or data frame with columns `chrom`, `start`, `end`.
rsid	Vector of rsids
pval	P-value threshold (NOT -log10)
id	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter
rsidx	Path to rsidx index file
pvali	Path to pval index file
build	="GRCh37" Build of vcffile
os	The operating system. Default is as detected. Determines the method used to perform query
proxies	="no" If SNPs are absent then look for proxies (yes) or not (no). Can also mask all target SNPs and only return proxies (only), for testing purposes. Currently only possible if querying rsid.
bfile	=path to plink bed/bim/fam ld reference panel
dbfile	=path to sqlite tag snp database
tag_kb	=5000 Proxy parameter
tag_nsnp	=5000 Proxy parameter
tag_r2	=0.6 Proxy parameter
threads	=1 NUmber of threads

Value

vcf object

`query_pvali` *Query pvali*

Description

Query pvali

Query pvali

Usage

```
query_pvali(pval, pvali)

query_pvali(pval, pvali)
```

Arguments

`pval`	pval threshold
`pvali`	Path to pval index file

Value

data frame

data frame

`query_pval_bcftools` *Query p-value using bcftools*

Description

Query p-value using bcftools

Usage

```
query_pval_bcftools(pval, vcffile, id = NULL)
```

Arguments

`pval`	P-value threshold (NOT -log10)
`vcffile`	Path to .vcf.gz GWAS summary data file
`id`	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter

Value

vcf object

query_pval_file *Query pval from vcf file*

Description

Query pval from vcf file

Usage

```
query_pval_file(pval, vcffile, id = NULL, build = "GRCh37")
```

Arguments

pval	P-value threshold (NOT -log10)
vcffile	Path to tabix indexed vcf file
id	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter
build	Default="GRCh37"

Value

VCF object

query_pval_sqlite3 *Query pval from file using pvali index*

Description

See create_pvali_index

See create_pvali_index

Usage

```
query_pval_sqlite3(pval, vcffile, id = NULL, pvali)

query_pval_sqlite3(pval, vcffile, id = NULL, pvali)
```

Arguments

pval	pval threshold
vcffile	Path to .vcf.gz GWAS summary data file
id	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter
pvali	Path to pval index file

Value

vcf object

vcf object

`query_pval_vcf` *Query based on p-value threshold from vcf*

Description

Query based on p-value threshold from vcf

Usage

```
query_pval_vcf(pval, vcf, id = NULL)
```

Arguments

`pval`	P-value threshold (NOT -log10)
`vcf`	VCF object (e.g. from readVcf)
`id`	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter

Value

VCF object

`query_rsidx` *Query rsidx*

Description

Query rsidx

Usage

```
query_rsidx(rsid, rsidx)
```

Arguments

`rsid`	Vector of rsids
`rsidx`	Path to rsidx index file

Value

data frame

query_rsid_bcftools *Query*

Description

Query

Usage

```
query_rsid_bcftools(rsid, vcffile, id = NULL)
```

Arguments

rsid	Vector of rsids
vcffile	Path to .vcf.gz GWAS summary data file
id	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter

Value

VCF object

query_rsid_file *Query vcf file, extracting by rsid*

Description

Query vcf file, extracting by rsid

Usage

```
query_rsid_file(rsid, vcffile, id = NULL, build = "GRCh37")
```

Arguments

rsid	Vector of rsids. Use DBSNP build (???)
vcffile	Path to .vcf.gz GWAS summary data file
id	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter
build	Default="GRCh37" Build of vcffile

Value

VCF object

query_rsid_rsidx *Query rsid from file using rsidx index*

Description

See create_rsidx_index

Usage

```
query_rsid_rsidx(rsid, vcffile, id = NULL, rsidx)
```

Arguments

rsid	Vector of rsids
vcffile	Path to .vcf.gz GWAS summary data file
id	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter
rsidx	Path to rsidx index file

Value

vcf object

query_rsid_vcf *Query rsid from vcf object*

Description

Query rsid from vcf object

Usage

```
query_rsid_vcf(rsid, vcf, id = NULL)
```

Arguments

rsid	Vector of rsids
vcf	VCF object (e.g. from readVcf)
id	If multiple GWAS datasets in the vcf file, the name (sample ID) from which to perform the filter

Value

VCF object

set_bcftools *Set bcftools binary location*

Description

Set bcftools binary location

Usage

```
set_bcftools(path = "")
```

Arguments

path	If "" (default), then will use the MRCIEU/genetics.binaRies to get binaries that are appropriate for the detected operating system. Otherwise, provide the path to the bcftools binary. If NULL then will set the option to NULL.

Value

NULL, sets option 'tools_bcftools'

set_plink *Set plink binary location*

Description

Set plink binary location

Usage

```
set_plink(path = "")
```

Arguments

path	If "" (default), then will use the MRCIEU/genetics.binaRies to get binaries that are appropriate for the detected operating system. Otherwise, provide the path to the plink binary. If NULL then will set the option to NULL.

Value

NULL, sets option 'tools_plink'

sqlite_ld_proxies *Lookup LD proxies from sqlite database*

Description

Lookup LD proxies from sqlite database

Usage

```
sqlite_ld_proxies(rsids, dbfile, tag_r2)
```

Arguments

rsids	List of rsids
dbfile	path to dbfile
tag_r2	minimum r2 value

Value

data frame

VariantAnnotation *VariantAnnotation*

Description

VariantAnnotation

vcflist_overlaps *Reduce list of VCFs to intersecting regions*

Description

Reduce list of VCFs to intersecting regions

Usage

```
vcflist_overlaps(vcflist, chrompos)
```

Arguments

vcflist	List of VCF objects, or list of VCF filenames, or mix of VCF objects and filenames
chrompos	Either vector of chromosome and position ranges e.g. "1:1000" or "1:1000-2000", or data frame with columns chrom, start, end.

Value

List of VCFs

---

vcf_to_granges *Convert vcf format to granges format*

Description

Convert vcf format to granges format

Usage

```
vcf_to_granges(vcf, id = NULL)
```

Arguments

vcf	Output from readVcf
id	Only accepts one ID, so specify here if there are multiple GWAS datasets in the vcf

Value

GRanges object

---

vcf_to_tibble *Convert vcf format to tibble (data frame)*

Description

Convert vcf format to tibble (data frame)

Usage

```
vcf_to_tibble(vcf, id = NULL)
```

Arguments

vcf	Output from readVcf
id	Only accepts one ID, so specify here if there are multiple GWAS datasets in the vcf

Value

GRanges object

Index