Package: gpmapr (via r-universe)

June 3, 2026

Title Query the OpenGWAS genotype-phenotype map

Version 0.0.1.0

Description This package is a simple wrapper around the OpenGWAS genotype-phenotype map API.

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Imports httr, jsonlite, dplyr, readr, tidyr

Suggests ggplot2, ggrepel, knitr, RColorBrewer, reshape2, rmarkdown, stringr, testthat (>= 3.0.0)

Config/testthat/edition 3

VignetteBuilder knitr

Config/pak/sysreqs libicu-dev libssl-dev libx11-dev

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

Date/Publication 2026-06-03 10:55:47 UTC

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

RemoteRef HEAD

RemoteSha b084bedcc48b8ce50bbf6f2146903a0314b4cc63

Contents

`all_genes` *All genes*

Description

Get all genes from the API

Usage

```
all_genes()
```

Value

A dataframe containing all genes with the following columns:

- id: the id of the gene
- gene: the name of the gene
- description: the description of the gene
- gene_biotype: the gene biotype
- chr: the chromosome of the gene
- start: the start position of the gene
- stop: the end position of the gene
- strand: the strand of the gene
- source: the source of the gene
- distinct_trait_categories: the number of trait categories that the gene is associated with via coloc groups
- distinct_protein_coding_genes: the number of genes that the gene is associated with via coloc groups
- num_study_extractions: the number of study extractions for this gene
- num_coloc_groups: the number of coloc groups for this gene
- num_coloc_studies: the number of studies that have coloc results for this gene
- num_rare_groups: the number of rare groups for this gene

`all_traits` *All traits*

Description

Get all traits from the API

Usage

```
all_traits()
```

Value

A dataframe containing all traits with the following columns:

- id: the id of the trait
- data_type: the data type of the trait
- trait: the internal string id of the trait
- trait_name: the name of the trait
- trait_category: the trait category of the trait
- variant_type: the type of variant
- sample_size: the sample size of the trait
- category: the category of the trait (continuous, categorical)
- ancestry: the ancestry of the trait
- heritability: the LDSC heritability score of the trait
- heritability_se: the standard error of the LDSC heritability score of the trait
- num_study_extractions: the number of study extractions for this trait
- num_coloc_groups: the number of coloc groups for this trait
- num_coloc_studies: the number of studies that have coloc results for this trait
- num_rare_results: the number of rare results for this trait

associations *Get Associations by SNP ID and Study ID*

Description

Get associations from the API by SNP id and study id

Usage

```
associations(variant_ids, study_ids)
```

Arguments

variant_ids	A vector of numeric values specifying the SNP IDs
study_ids	A vector of numeric values specifying the Study IDs

Value

A dataframe containing the associations

associations_dataframe

The associations dataframe contains information about which studies have association results. It has the following columns:

- variant_id: the id of the SNP associated with this association
- study_id: the id of the study associated with this association
- beta: the beta value of the association
- se: the standard error of the association
- p: the p-value of the association
- eaf: the estimated allele frequency of the association
- imputed: whether the association is imputed

gene *Gene*

Description

A collection of studies that are associated with a particular gene.

Usage

```
gene(
  gene_id,
  include_associations = FALSE,
  include_coloc_pairs = FALSE,
  include_trans = TRUE,
  h4_threshold = 0.8
)
```

Arguments

gene_id	A numeric value specifying the gene id
include_associations	A logical value specifying whether to include associations (BETA, SE, P), defaults to FALSE
include_coloc_pairs	A logical value specifying whether to include coloc pairs, defaults to FALSE
include_trans	A logical value specifying whether to include trans genetic effects, defaults to TRUE
h4_threshold	A numeric value specifying the h4 threshold for coloc pairs, defaults to 0.8

Details

The dataframes returned by this function are as follows:

Value

A list which contains the following elements:

- gene: A list containing metadata about the gene, including region, and neighboring genes.
- coloc_groups: a dataframe containing information about which studies have coloc results for this gene. See below for details.
- study_extractions: a list of dataframes containing the study extractions for this trait. See below for details.
- rare_results: (optional) a list of dataframes containing the rare results for this trait
- coloc_pairs: (optional) a dataframe containing all pairwise coloc results for this trait.
- variants: a dataframe containing the variants for each associated coloc group or rare group.

See below for details.

coloc_groups_dataframe

The coloc_groups dataframe contains information about which studies have coloc results. It has the following columns:

- coloc_group_id: the unique id for this group of colocalised results
- study_id: the id of the study

- study_extraction_id: the id of the study extraction
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

study_extractions_dataframe

The study_extractions dataframe contains information about which studies have coloc results. It has the following columns:

- id: the unique id for this study extraction
- study_id: the id of the study associated with this study extraction
- variant_id: the id of the SNP
- snp: the SNP name
- ld_block_id: the id of the LD block
- unique_study_id: the unique id for this study
- study: the study name
- file: the file name
- svg_file: the SVG file name
- file_with_lbfs: the file name with lbfs
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- gene: the gene associated with the SNP
- gene_id: the id of the gene

- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

rare_results_dataframe

The rare_results dataframe contains information about which studies have coloc results. It has the following columns:

- rare_result_group_id: the unique id for this rare result group
- study_id: the id of the study associated with this rare result
- study_extraction_id: the id of the study extraction associated with this rare result
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait
- ld_block: the LD block of the SNP

coloc_pairs_dataframe

The coloc_pairs dataframe contains information about which studies have coloc pairs. It has the following columns:

- study_extraction_a_id: the id of the study extraction associated with this coloc pair
- study_extraction_b_id: the id of the study extraction associated with this coloc pair
- ld_block_id: the id of the LD block
- h3: the h3 value for this coloc pair
- h4: the h4 value for this coloc pair
- spurious: whether this coloc pair is spurious

variants_dataframe

The variants dataframe contains variant information that is pulled from the Variant Effect Predictor (VEP) database. It has the following columns, along side many more columns from VEP:

- id: the id of the SNP
- gene_id: the id of the gene as predicted by VEP
- gene: the gene name as predicted by VEP

`genes` *Genes*

Description

Get specific genes from the API. The API returns collapsed/combined data for all requested genes.

Usage

```
genes(
  gene_ids,
  include_associations = FALSE,
  include_coloc_pairs = FALSE,
  include_trans = TRUE,
  h4_threshold = 0.8
)
```

Arguments

`gene_ids`	A vector of gene ids (1 or more)
`include_associations`	A logical value specifying whether to include associations (BETA, SE, P), defaults to FALSE
`include_coloc_pairs`	A logical value specifying whether to include coloc pairs, defaults to FALSE
`include_trans`	A logical value specifying whether to include trans genetic effects, defaults to TRUE
`h4_threshold`	A numeric value specifying the h4 threshold for coloc pairs, defaults to 0.8

Details

The dataframes returned by this function are as follows:

Value

A list which contains the following elements:

- genes: gene metadata for the requested genes
- coloc_groups: a dataframe containing information about which studies have coloc results for all genes
- study_extractions: a dataframe containing the study extractions for all genes
- rare_results: a dataframe containing the rare results for all genes

coloc_groups_dataframe

The coloc_groups dataframe contains information about which studies have coloc results. It has the following columns:

- coloc_group_id: the unique id for this group of colocalised results
- study_id: the id of the study
- study_extraction_id: the id of the study extraction
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

study_extractions_dataframe

The study_extractions dataframe contains information about which studies have coloc results. It has the following columns:

- id: the unique id for this study extraction
- study_id: the id of the study associated with this study extraction
- variant_id: the id of the SNP
- snp: the SNP name

- ld_block_id: the id of the LD block
- unique_study_id: the unique id for this study
- study: the study name
- file: the file name
- svg_file: the SVG file name
- file_with_lbfs: the file name with lbfs
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

rare_results_dataframe

The rare_results dataframe contains information about which studies have coloc results. It has the following columns:

- rare_result_group_id: the unique id for this rare result group
- study_id: the id of the study associated with this rare result
- study_extraction_id: the id of the study extraction associated with this rare result
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait
- ld_block: the LD block of the SNP

coloc_pairs_dataframe

The coloc_pairs dataframe contains information about which studies have coloc pairs. It has the following columns:

- study_extraction_a_id: the id of the study extraction associated with this coloc pair
- study_extraction_b_id: the id of the study extraction associated with this coloc pair
- ld_block_id: the id of the LD block
- h3: the h3 value for this coloc pair
- h4: the h4 value for this coloc pair
- spurious: whether this coloc pair is spurious

`get_all_gene_pleiotropies` *Get All Gene Pleiotropies*

Description

Get gene pleiotropy from the API by gene id

Usage

```
get_all_gene_pleiotropies()
```

Value

A list containing the gene pleiotropy

- gene_id: the id of the gene
- gene: the name of the gene
- distinct_trait_categories: the number of trait categories that the gene is associated with via coloc groups
- distinct_protein_coding_genes: the number of genes that the gene is associated with via coloc groups

get_all_variant_pleiotropies *Get All SNP Pleiotropies*

Description

Get all SNP pleiotropies from the API

Usage

```
get_all_variant_pleiotropies()
```

Value

A list containing the SNP pleiotropies

- variant_id: the id of the SNP
- display_snp: the name of the SNP
- distinct_trait_categories: the number of trait categories that the SNP is associated with via coloc groups
- distinct_protein_coding_genes: the number of genes that the SNP is associated with via coloc groups

get_gwas *Get a GWAS from the API*

Description

Get a GWAS from the API

Usage

```
get_gwas(gwas_id, include_associations = FALSE, include_summary_stats = FALSE)
```

Arguments

gwas_id	The ID of the GWAS
include_associations	Whether to include associations
include_summary_stats	Whether to include summary statistics

Value

A list containing the GWAS information

health_api *Get API Health*

Description

Get the health status of the API

Usage

```
health_api()
```

Value

A list containing the health status

ld_matrix *LD Matrix*

Description

Get LD matrix from the API by Variant ID

Usage

```
ld_matrix(variant_ids = c())
```

Arguments

variant_ids	A character string specifying the Variant ID. Variant IDs can be SNP IDs or variant IDs.

Value

A list containing the LD matrix

ld_dataframe

The ld dataframe contains information about the LD matrix. It has the following columns:

- lead_variant_id: the id of the lead SNP
- proxy_variant_id: the id of the variant SNP
- ld_block_id: the id of the LD block
- r: the r value between the lead and variant SNPs

ld_proxies *LD Proxies*

Description

Get LD proxies from the API by Variant ID

Usage

```
ld_proxies(variant_ids = c())
```

Arguments

variant_ids	A character string specifying the Variant ID. Variant IDs can be SNP IDs or variant IDs.

Value

A list containing the LD proxies

ld_dataframe

The ld dataframe contains information about the LD matrix. It has the following columns:

- lead_variant_id: the id of the lead SNP
- proxy_variant_id: the id of the variant SNP
- ld_block_id: the id of the LD block
- r: the r value between the lead and variant SNPs

region *Region*

Description

A collection of studies that are associated with a particular region.

Usage

```
region(
  region_id,
  include_associations = FALSE,
  include_coloc_pairs = FALSE,
  h4_threshold = 0.8
)
```

Arguments

`region_id`	A numeric value specifying the region id
`include_associations`	A logical value specifying whether to include associations (BETA, SE, P), defaults to FALSE
`include_coloc_pairs`	A logical value specifying whether to include coloc pairs, defaults to FALSE
`h4_threshold`	A numeric value specifying the h4 threshold for coloc pairs, defaults to 0.8

Details

The dataframes returned by this function are as follows:

Value

A list which contains the following elements:

- gene: A list containing metadata about the gene, including region, and neighboring genes.
- coloc_groups: a dataframe containing information about which studies have coloc results for this gene. See below for details.
- study_extractions: a list of dataframes containing the study extractions for this trait. See below for details.
- rare_results: (optional) a list of dataframes containing the rare results for this trait
- coloc_pairs: (optional) a dataframe containing all pairwise coloc results for this trait.
- variants: a dataframe containing the variants for each associated coloc group or rare group.

See below for details.

coloc_groups_dataframe

The coloc_groups dataframe contains information about which studies have coloc results. It has the following columns:

- coloc_group_id: the unique id for this group of colocalised results
- study_id: the id of the study
- study_extraction_id: the id of the study extraction
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- display_snp: the display SNP name

- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

genes_in_region_dataframe

The genes_in_region dataframe contains information about which genes are in a region. It has the following columns:

- id: the id of the gene
- ensembl_id: the ensembl id of the gene
- gene: the name of the gene
- description: the description of the gene
- gene_biotype: the gene biotype
- chr: the chromosome of the gene
- start: the start position of the gene
- stop: the stop position of the gene
- strand: the strand of the gene
- source: the source of the gene

study_extractions_dataframe

The study_extractions dataframe contains information about which studies have coloc results. It has the following columns:

- id: the unique id for this study extraction
- study_id: the id of the study associated with this study extraction
- variant_id: the id of the SNP
- snp: the SNP name
- ld_block_id: the id of the LD block
- unique_study_id: the unique id for this study
- study: the study name
- file: the file name
- svg_file: the SVG file name
- file_with_lbfs: the file name with lbfs
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP

- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

rare_results_dataframe

The rare_results dataframe contains information about which studies have coloc results. It has the following columns:

- rare_result_group_id: the unique id for this rare result group
- study_id: the id of the study associated with this rare result
- study_extraction_id: the id of the study extraction associated with this rare result
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait
- ld_block: the LD block of the SNP

coloc_pairs_dataframe

The coloc_pairs dataframe contains information about which studies have coloc pairs. It has the following columns:

- study_extraction_a_id: the id of the study extraction associated with this coloc pair
- study_extraction_b_id: the id of the study extraction associated with this coloc pair
- ld_block_id: the id of the LD block
- h3: the h3 value for this coloc pair
- h4: the h4 value for this coloc pair
- spurious: whether this coloc pair is spurious

variants_dataframe

The variants dataframe contains variant information that is pulled from the Variant Effect Predictor (VEP) database. It has the following columns, along side many more columns from VEP:

- id: the id of the SNP
- gene_id: the id of the gene as predicted by VEP
- gene: the gene name as predicted by VEP

`search_gpmap` *Search the Genotype-Phenotype Map*

Description

Search the GP Map for Traits, Genes or Variants

Usage

```
search_gpmap(search_text, rsquared_threshold = 0.8)
```

Arguments

`search_text`	A character string specifying the search text
`rsquared_threshold`	A numeric value specifying the rsquared threshold for proxy variants, defaults to 0.8

Details

After calling search, you can use call the subsequent data as described in the `call` column of the search results.

Value

A dataframe containing the search results with the following columns:

- type: the type of the search result: "original_variant", "proxy_variant", "trait", "gene"
- name: the name of the search result
- type_id: the type_id of the search result. This is the internal id in which the data can be accessed.
- call: the call to get the search result: "variant(type_id)", "trait(type_id)", "gene(type_id)"
- info: a string containing informaiton about the search result, which may include:
  - Extractions: the number of extractions
  - Colocalisation Groups: the number of colocalisation groups
  - Colocalisation Studies: the number of colocalisation studies
  - Rare Results: the number of rare results
  - Rsquared: the rsquared of the proxy variant compared to the original variant

`trait` *Trait*

Description

A collection of studies that are associated with a particular phenotype. A trait will include a common study and occasionally a rare study. When trait_id is a GUID (from GWAS upload), fetches the upload result instead.

Usage

```
trait(
  trait_id,
  include_associations = FALSE,
  include_coloc_pairs = FALSE,
  h4_threshold = 0.8
)
```

Arguments

`trait_id`	A numeric value or GUID (from GWAS upload) specifying the trait id
`include_associations`	A logical value specifying whether to include associations (BETA, SE, P), defaults to FALSE
`include_coloc_pairs`	A logical value specifying whether to include coloc pairs, defaults to FALSE
`h4_threshold`	A numeric value specifying the h4 threshold for coloc pairs, defaults to 0.8

Details

The dataframes returned by this function are as follows:

Value

A list which contains the following elements:

- trait: A list containing metadata about the trait, including common and rare studies associated with the trait
- coloc_groups: a dataframe containing information about which studies have coloc results for this trait. See below for details.
- study_extractions: a list of dataframes containing the study extractions for this trait. See below for details.
- rare_results: (optional) a list of dataframes containing the rare results for this trait
- coloc_pairs: (optional) a dataframe containing all pairwise coloc results for this trait.

See below for details.

coloc_groups_dataframe

The coloc_groups dataframe contains information about which studies have coloc results. It has the following columns:

- coloc_group_id: the unique id for this group of colocalised results
- study_id: the id of the study
- study_extraction_id: the id of the study extraction
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

study_extractions_dataframe

The study_extractions dataframe contains information about which studies have coloc results. It has the following columns:

- id: the unique id for this study extraction
- study_id: the id of the study associated with this study extraction
- variant_id: the id of the SNP
- snp: the SNP name
- ld_block_id: the id of the LD block
- unique_study_id: the unique id for this study
- study: the study name
- file: the file name
- svg_file: the SVG file name
- file_with_lbfs: the file name with lbfs
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP

- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

rare_results_dataframe

The rare_results dataframe contains information about which studies have coloc results. It has the following columns:

- rare_result_group_id: the unique id for this rare result group
- study_id: the id of the study associated with this rare result
- study_extraction_id: the id of the study extraction associated with this rare result
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait
- ld_block: the LD block of the SNP

coloc_pairs_dataframe

The coloc_pairs dataframe contains information about which studies have coloc pairs. It has the following columns:

- study_extraction_a_id: the id of the study extraction associated with this coloc pair
- study_extraction_b_id: the id of the study extraction associated with this coloc pair
- ld_block_id: the id of the LD block
- h3: the h3 value for this coloc pair
- h4: the h4 value for this coloc pair
- spurious: whether this coloc pair is spurious

variants_dataframe

The variants dataframe contains variant information that is pulled from the Variant Effect Predictor (VEP) database. It has the following columns, along side many more columns from VEP:

- id: the id of the SNP
- gene_id: the id of the gene as predicted by VEP
- gene: the gene name as predicted by VEP

`traits` *Traits*

Description

Get specific traits from the API. The API returns collapsed/combined data for all requested traits. When a trait ID is a GUID (from GWAS upload), fetches the upload result instead.

Usage

```
traits(
  trait_ids,
  include_associations = FALSE,
  include_coloc_pairs = FALSE,
  h4_threshold = 0.8
)
```

Arguments

`trait_ids`	A vector of trait ids (numeric) or GUIDs (from GWAS upload)
`include_associations`	A logical value specifying whether to include associations (BETA, SE, P), defaults to FALSE
`include_coloc_pairs`	A logical value specifying whether to include coloc pairs, defaults to FALSE. Coloc pairs are fetched from a separate endpoint per trait.
`h4_threshold`	A numeric value specifying the h4 threshold for coloc pairs, defaults to 0.8

Details

The dataframes returned by this function are as follows:

Value

A list which contains the following elements:

- traits: trait metadata for the requested traits
- coloc_groups: a dataframe containing information about which studies have coloc results for all traits. See below for details.
- study_extractions: a dataframe containing the study extractions for all traits. See below for details.
- rare_results: a dataframe containing the rare results for all traits
- coloc_pairs: (optional) a dataframe containing all pairwise coloc results for all traits.

coloc_groups_dataframe

The coloc_groups dataframe contains information about which studies have coloc results. It has the following columns:

- coloc_group_id: the unique id for this group of colocalised results
- study_id: the id of the study
- study_extraction_id: the id of the study extraction
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

study_extractions_dataframe

The study_extractions dataframe contains information about which studies have coloc results. It has the following columns:

- id: the unique id for this study extraction
- study_id: the id of the study associated with this study extraction

- variant_id: the id of the SNP
- snp: the SNP name
- ld_block_id: the id of the LD block
- unique_study_id: the unique id for this study
- study: the study name
- file: the file name
- svg_file: the SVG file name
- file_with_lbfs: the file name with lbfs
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

rare_results_dataframe

The rare_results dataframe contains information about which studies have coloc results. It has the following columns:

- rare_result_group_id: the unique id for this rare result group
- study_id: the id of the study associated with this rare result
- study_extraction_id: the id of the study extraction associated with this rare result
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait
- ld_block: the LD block of the SNP

coloc_pairs_dataframe

The coloc_pairs dataframe contains information about which studies have coloc pairs. It has the following columns:

- study_extraction_a_id: the id of the study extraction associated with this coloc pair
- study_extraction_b_id: the id of the study extraction associated with this coloc pair
- ld_block_id: the id of the LD block
- h3: the h3 value for this coloc pair
- h4: the h4 value for this coloc pair
- spurious: whether this coloc pair is spurious

`upload_gwas` *Upload a GWAS to the API*

Description

Upload a GWAS to the API

Usage

```
upload_gwas(
  file,
  name,
  p_value_threshold = 5e-08,
  column_names = list(),
  email = NA,
  category = "continuous",
  is_published = FALSE,
  doi = NA,
  should_be_added = FALSE,
  ancestry = "EUR",
  sample_size = NA,
  reference_build = "GRCh38",
  compare_with_upload_guids = NA
)
```

Arguments

`file`	The path to the GWAS file, maximum size is 1GB
`name`	The name of the GWAS
`p_value_threshold`	The p-value threshold for the GWAS
`column_names`	A list of column names in the format of: list(CHR = "chr", BP = "pos"...) • CHR: chromosome

- BP: base pair position
- P: p-value
- EA: allele 1
- OA: allele 2
- EAF: allele frequency And either BETA and SE, or OR, LB, and UB
- BETA: beta
- SE: standard error
- OR: odds ratio
- LB: lower bound of the confidence interval
- UB: upper bound of the confidence interval

`email` The email of the user

`category` The category of the GWAS. Only "continuous" and "categorical" are accepted.

`is_published` Whether the GWAS is published

`doi` The DOI of the GWAS

`should_be_added`
Whether the GWAS should be added to the API

`ancestry` The ancestry of the GWAS. Currently only "EUR" is accepted.

`sample_size` The sample size of the GWAS

`reference_build`
The reference build of the GWAS. Only "GRCh37" and "GRCh38" are accepted.

`compare_with_upload_guids`
A vector of GUIDs of uploads to compare with

Value

A list containing the GWAS information

`variant` *Variant*

Description

A collection of studies that are associated with a particular variant.

Usage

```
variant(
  variant_id,
  include_coloc_pairs = FALSE,
  h4_threshold = 0.8,
  include_summary_stats = FALSE
)
```

Arguments

`variant_id`	A character string specifying the SNP ID
`include_coloc_pairs`	A logical value specifying whether to include coloc pairs
`h4_threshold`	A numeric value specifying the cutoff for included coloc pairs, defaults to 0.8. Only used if include_coloc_pairs is TRUE.
`include_summary_stats`	A logical value specifying whether to include summary stats

Details

The dataframes returned by this function are as follows:

Value

A list which contains the following elements:

- variant: named list containing the variant information
- coloc_groups: a dataframe containing information about which studies have coloc results for this variant
- rare_results: a list of dataframes containing the rare variants
- study_extractions: a list of dataframes containing the study extractions
- summary_stats (optional): a list of dataframes containing the summary stats for each study, where the name of each element is the study_id. Column names are uppercase (e.g. SNP, BP, BETA, SE, LBF_1).
- coloc_pairs (optional): a dataframe containing information about which studies have coloc pairs for this variant where the study_extraction_a_id and study_extraction_b_id are the study_extraction_ids of the two studies. h4_threshold is the cutoff for included coloc pairs, defaults to 0.8

coloc_groups_dataframe

The coloc_groups dataframe contains information about which studies have coloc results. It has the following columns:

- coloc_group_id: the unique id for this group of colocalised results
- study_id: the id of the study
- study_extraction_id: the id of the study extraction
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP

- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

rare_results_dataframe

The rare_results dataframe contains information about which studies have coloc results. It has the following columns:

- rare_result_group_id: the unique id for this rare result group
- study_id: the id of the study associated with this rare result
- study_extraction_id: the id of the study extraction associated with this rare result
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait
- ld_block: the LD block of the SNP

study_extractions_dataframe

The study_extractions dataframe contains information about which studies have coloc results. It has the following columns:

- id: the unique id for this study extraction
- study_id: the id of the study associated with this study extraction
- variant_id: the id of the SNP
- snp: the SNP name

- ld_block_id: the id of the LD block
- unique_study_id: the unique id for this study
- study: the study name
- file: the file name
- svg_file: the SVG file name
- file_with_lbfs: the file name with lbfs
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

summary_statistics_dataframe

The summary_statistics dataframe contains information about which studies have summary statistics. From the API, column names are typically uppercase (SNP, CHR, BP, EA, OA, EAF, Z, BETA, SE, P, LBF_1, etc.). It has the following columns (names may be upper or lower case depending on source):

- SNP / variant_id: the id of the SNP
- CHR / chr: the chromosome of the SNP
- BP / bp: the base pair position of the SNP
- EA / ea: the effect allele
- OA / oa: the other allele
- EAF / eaf: the estimated allele frequency
- Z / z: the z-score
- BETA / beta: the beta value
- SE / se: the standard error
- P / p: the p-value
- imputed: whether the summary statistics are imputed
- LBF_* / lbf_*: all different finemapped log-bayes factors for each credible set. Each credible set is numbered from 1 to 10. If finemapped failed or only returned 1 credible set, the LBF_1 column is just converted directly from the z-score.

coloc_pairs_dataframe

The coloc_pairs dataframe contains information about which studies have coloc pairs. It has the following columns:

- study_extraction_a_id: the id of the study extraction associated with this coloc pair
- study_extraction_b_id: the id of the study extraction associated with this coloc pair
- ld_block_id: the id of the LD block
- h3: the h3 value for this coloc pair
- h4: the h4 value for this coloc pair
- spurious: whether this coloc pair is spurious

---

variants *Variants*

Description

Get specific variants from the API. The API accepts variant identifiers (variant_ids, rsids, or strings) and returns collapsed/combined data. The API distinguishes between identifier types automatically. Max 10 variants when expand=TRUE.

Usage

```
variants(
  variants,
  expand = FALSE,
  include_associations = FALSE,
  include_coloc_pairs = FALSE,
  h4_threshold = 0.8
)
```

Arguments

variants	A vector of variant identifiers (variant_ids, rsids, or strings)
expand	Logical. FALSE (default) returns minimal data. TRUE returns full VariantResponse (max 10)
include_associations	Logical. Whether to include associations (BETA, SE, P). Only when expand=TRUE
include_coloc_pairs	Logical. Whether to include coloc pairs. Only when expand=TRUE
h4_threshold	Numeric. H4 threshold for coloc pairs, defaults to 0.8

Details

The dataframes returned by this function are as follows:

Value

A list which contains the following elements:

- variants: a dataframe containing the variants for all requested variants
- coloc_groups: (if expanded) a dataframe containing the coloc groups for all variants
- study_extractions: (if expanded) a dataframe containing the study extractions for all variants
- rare_results: (if expanded) a dataframe containing the rare results for all variants

coloc_groups_dataframe

The coloc_groups dataframe contains information about which studies have coloc results. It has the following columns:

- coloc_group_id: the unique id for this group of colocalised results
- study_id: the id of the study
- study_extraction_id: the id of the study extraction
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

study_extractions_dataframe

The study_extractions dataframe contains information about which studies have coloc results. It has the following columns:

- id: the unique id for this study extraction
- study_id: the id of the study associated with this study extraction
- variant_id: the id of the SNP
- snp: the SNP name
- ld_block_id: the id of the LD block

- unique_study_id: the unique id for this study
- study: the study name
- file: the file name
- svg_file: the SVG file name
- file_with_lbfs: the file name with lbfs
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- cis_trans: the cis/trans status of the SNP
- ld_block: the LD block of the SNP
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait

rare_results_dataframe

The rare_results dataframe contains information about which studies have coloc results. It has the following columns:

- rare_result_group_id: the unique id for this rare result group
- study_id: the id of the study associated with this rare result
- study_extraction_id: the id of the study extraction associated with this rare result
- variant_id: the id of the SNP
- ld_block_id: the id of the LD block
- chr: the chromosome of the SNP
- bp: the base pair position of the SNP
- min_p: the minimum p-value related to the study_extraction_id
- display_snp: the display SNP name
- gene: the gene associated with the SNP
- gene_id: the id of the gene
- trait_id: the id of the trait
- trait_name: the name of the trait
- trait_category: the category of the trait
- data_type: the data type of the trait
- tissue: the tissue of the trait
- ld_block: the LD block of the SNP

summary_statistics_dataframe

The summary_statistics dataframe contains information about which studies have summary statistics. From the API, column names are typically uppercase (SNP, CHR, BP, EA, OA, EAF, Z, BETA, SE, P, LBF_1, etc.). It has the following columns (names may be upper or lower case depending on source):

- SNP / variant_id: the id of the SNP
- CHR / chr: the chromosome of the SNP
- BP / bp: the base pair position of the SNP
- EA / ea: the effect allele
- OA / oa: the other allele
- EAF / eaf: the estimated allele frequency
- Z / z: the z-score
- BETA / beta: the beta value
- SE / se: the standard error
- P / p: the p-value
- imputed: whether the summary statistics are imputed
- LBF_* / lbf_*: all different finemapped log-bayes factors for each credible set. Each credible set is numbered from 1 to 10. If finemapped failed or only returned 1 credible set, the LBF_1 column is just converted directly from the z-score.

coloc_pairs_dataframe

The coloc_pairs dataframe contains information about which studies have coloc pairs. It has the following columns:

- study_extraction_a_id: the id of the study extraction associated with this coloc pair
- study_extraction_b_id: the id of the study extraction associated with this coloc pair
- ld_block_id: the id of the LD block
- h3: the h3 value for this coloc pair
- h4: the h4 value for this coloc pair
- spurious: whether this coloc pair is spurious

Index