Package 'gwas2crispr'

June 2, 2026

Type Package

Title GWAS-to-CRISPR Data Pipeline for High-Throughput SNP Target Extraction

Version 0.1.5

Description Provides a reproducible pipeline to conduct genome-wide association studies (GWAS) and extract single-nucleotide polymorphisms (SNPs) for a human trait or disease. Given aggregated GWAS dataset(s) and a user-defined significance threshold, the package retrieves significant SNPs from the GWAS Catalog using supported trait identifiers, annotates their gene context, and can write a harmonised metadata table in comma-separated values (CSV) format, genomic intervals in the Browser Extensible Data (BED) format, and sequences in the FASTA (text-based sequence) format with user-defined flanking regions for clustered regularly interspaced short palindromic repeats (CRISPR) guide design. The existing efo_id argument is retained for backward compatibility. The package prepares computational artifacts for downstream workflows; it does not perform biological causality testing, clinical interpretation, therapeutic design, or wet-lab validation. For details on the resources and methods see: Buniello et al. (2019) <doi:10.1093/nar/gky1120>; Sollis et al. (2023) <doi:10.1093/nar/gkac1010>; Jinek et al. (2012) <doi:10.1126/science.1225829>.

License MIT + file LICENSE

URL https://github.com/leopard0ly/gwas2crispr

BugReports https://github.com/leopard0ly/gwas2crispr/issues

Depends R (>= 4.1)

Imports httr, dplyr, purrr, tibble, tidyr, readr, stringr, tidyselect

Suggests Biostrings, BSgenome.Hsapiens.UCSC.hg38, GenomeInfoDb, optparse, testthat, knitr, rmarkdown

VignetteBuilder knitr, rmarkdown

Encoding UTF-8

Language en-US

RoxygenNote 7.3.3

biocViews Software, Genetics, VariantAnnotation, SNP, DataImport

NeedsCompilation no

Author Othman S. I. Mohammed [aut, cre],
LEOPARD.LY LTD [cph]

Maintainer Othman S. I. Mohammed <admin@leopard.ly>

Repository CRAN

Date/Publication 2026-06-02 06:50:07 UTC

Contents

`fetch_gwas` *Fetch significant GWAS associations for a GWAS Catalog trait identifier*

Description

Retrieves significant GWAS Catalog associations directly from the EMBL-EBI GWAS Catalog REST API v2. The function resolves the supplied GWAS Catalog trait identifier to direct identifier queries and trait labels, retrieves paginated association records, filters by p-value, and returns a list used by `run_gwas2crispr`.

Usage

```
fetch_gwas(efo_id = "EFO_0001663", p_cut = 5e-08, verbose = interactive())
```

Arguments

`efo_id`	character. GWAS Catalog trait identifier. The argument name is retained for backward compatibility. Examples include EFO_0001663, MONDO_0007254, and NCIT_C4872 when supported by the GWAS Catalog API.
`p_cut`	numeric. P-value threshold for significance.
`verbose`	logical. If `TRUE`, prints a compact progress line.

Details

This function performs network calls to the GWAS Catalog REST API v2 and may be affected by service availability or rate limits. Selected supported disease and cancer trait identifier prefixes include EFO, MONDO, and NCIT. HP, Orphanet, and ORPHA are accepted for compatibility. GO identifiers are not supported as primary GWAS Catalog trait identifiers in gwas2crispr 0.1.5.

Value

A list with:

- associations: tibble with association_id and pvalue.
- risk_alleles: tibble mapping association_id to variant_id.
- cache: internal tibble with variant metadata used downstream.

See Also

run_gwas2crispr

Examples

```
a <- fetch_gwas("EFO_0000707", p_cut = 1e-6, verbose = FALSE)
head(a$associations)
```

---

run_gwas2crispr *Run the GWAS-to-CRISPR export pipeline using GRCh38/hg38*

Description

Runs the complete computational preparation workflow: retrieves GWAS Catalog associations for a supported trait identifier through fetch_gwas, prepares SNP metadata, creates BED intervals, and optionally writes CSV, BED, and FASTA files for downstream CRISPR guide-design preparation.

Usage

```
run_gwas2crispr(
  efo_id,
  p_cut = 5e-08,
  flank_bp = 200,
  out_prefix = NULL,
  genome_pkg = "BSgenome.Hsapiens.UCSC.hg38",
  verbose = interactive()
)
```

Arguments

efo_id	character. GWAS Catalog trait identifier. The argument name is retained for backward compatibility. Examples include EFO_0001663, MONDO_0007254, and NCIT_C4872 when supported by the GWAS Catalog API.
p_cut	numeric. P-value threshold for significance.
flank_bp	integer. Number of flanking bases for FASTA sequence extraction.
out_prefix	character or NULL. Prefix for output files. If NULL, no files are written.
genome_pkg	character. BSgenome package name used for hg38 FASTA extraction.
verbose	logical. If TRUE, prints a compact progress line.

Details

Only GRCh38/hg38 is supported. CSV and BED outputs can be produced without genome packages. FASTA output is generated only when **BSgenome.Hsapiens.UCSC.hg38** and **Biostrings** are installed. If FASTA dependencies are unavailable, the function still writes CSV and BED. Selected supported disease and cancer trait identifier prefixes include EFO, MONDO, and NCIT. HP, Orphanet, and ORPHA are accepted for compatibility. GO identifiers are not supported as primary GWAS Catalog trait identifiers in gwas2crispr 0.1.5.

Value

Invisibly returns a list with:

- `summary`: one-row tibble with basic counts.
- `chr_freq`: chromosome frequency table.
- `snps_full`: harmonized SNP metadata.
- `bed`: BED-style interval table.
- `fasta`: DNAStringSet if FASTA was generated; otherwise `NULL`.
- `written`: character vector of written file paths.

See Also

`fetch_gwas`

Examples

```
res <- run_gwas2crispr(
  efo_id     = "EFO_0000707",
  p_cut      = 1e-6,
  flank_bp   = 300,
  out_prefix = file.path(tempdir(), "lung"),
  verbose    = FALSE
)
res$summary
res$written
```

Index