Package 'gwas2crispr'

May 10, 2026

Type Package

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

Version 0.1.4

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 and the Experimental Factor Ontology (EFO), 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. 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>; Malone et al. (2010) <doi:10.1093/bioinformatics/btq099>; Experimental Factor Ontology (EFO) <https://www.ebi.ac.uk/efo>.

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 2026-05-09 22:50:02 UTC

Contents

`fetch_gwas` *Fetch significant GWAS associations for an EFO trait*

Description

Retrieves significant GWAS Catalog associations directly from the EMBL-EBI GWAS Catalog REST API v2. The function resolves the supplied Experimental Factor Ontology (EFO) identifier to 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. EFO trait identifier, such as EFO_0001663.
`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.

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 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. EFO trait identifier, such as EFO_0001663.
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.

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