Package 'gwas2crispr'

August 22, 2025

Type Package

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

Version 0.1.2

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, methods

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

VignetteBuilder knitr, rmarkdown

Encoding UTF-8

Language en-US

RoxygenNote 7.3.2

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 2025-08-22 18:50:06 UTC

Contents

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

Description

Tries `gwasrapidd::get_associations()` first; if it returns no rows or fails, falls back to the EBI GWAS Summary Statistics REST API to retrieve significant associations up to the given p-value threshold.

Usage

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

Arguments

`efo_id`	character. Experimental Factor Ontology (EFO) trait identifier (e.g., "EFO_0001663").
`p_cut`	numeric. P-value threshold for significance (default 5e-8).

Details

This function performs network calls and may be rate-limited. Column names returned by the REST API may change; defensive checks are applied.

Value

An S4 object of class `"associations"` with slots:

- `associations`: data frame with `association_id` and `pvalue`.
- `risk_alleles`: data frame mapping `association_id` to `variant_id`.

See Also

run_gwas2crispr

Examples

```
# Network call; may be rate-limited, so we mark it as \donttest.
a <- try(fetch_gwas("EFO_0001663", p_cut = 5e-8), silent = TRUE)
if (!inherits(a, "try-error")) {
  head(a@associations)
}
```

---

run_gwas2crispr *Run the GWAS to CRISPR export pipeline (hg38)*

Description

End-to-end pipeline: fetch significant associations, annotate, and optionally write CSV/BED/FASTA outputs. By default no files are written; set `out_prefix` to write results.

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. Experimental Factor Ontology (EFO) identifier, e.g., "EFO_0001663".
`p_cut`	numeric. P-value threshold for significance (default 5e-8).
`flank_bp`	integer. Flanking bases for FASTA sequences (default 200).
`out_prefix`	character or NULL. File prefix (including path) for outputs. If `NULL` (default), nothing is written to disk and a result object is returned. To write files safely in examples/tests, use `file.path(tempdir(), "prefix")`.
`genome_pkg`	character. BSgenome package to use for FASTA (default "BSgenome.Hsapiens.UCSC.hg38"); FASTA step is skipped if not installed.
`verbose`	logical. If TRUE, emit progress via `message()`.

Details

Network I/O may occur when fetching data. Only GRCh38/hg38 is supported.

Value

(Invisibly) a list with elements:

- `summary`: tibble with basic counts.
- `snps_full`: tibble of SNP metadata.
- `bed`: tibble of BED intervals (if computed).
- `fasta`: `Biostrings::DNAStringSet` (if computed).
- `written`: character vector of file paths written (possibly empty).

See Also

`fetch_gwas`

Examples

```
# Write into a temporary directory so we don't touch the user's filespace:
tmp <- tempdir()
res <- run_gwas2crispr(
  efo_id     = "EFO_0001663",
  p_cut      = 5e-8,
  flank_bp   = 200,
  out_prefix = file.path(tmp, "prostate"),
  verbose    = FALSE
)

# If you omit 'out_prefix', nothing is written; an object is returned:
res2 <- run_gwas2crispr(
  efo_id     = "EFO_0001663",
  p_cut      = 5e-8,
  flank_bp   = 200,
  out_prefix = NULL,
  verbose    = FALSE
)
```

Index