Package 'annotaR'

January 27, 2026

Title Tidy, Integrated Gene Annotation

Version 0.1.1

Maintainer MD. Arshad `<arshad10867c@gmail.com>`

Description A framework for intuitive, multi-source gene and protein annotation, with a focus on integrating functional genomics with disease and drug data for translational insights. Methods used include g:Profiler (Raudvere et al. (2019) <doi:10.1093/nar/gkz369>), biomaRt (Durinck et al. (2009) <doi:10.1038/nprot.2009.97>), and the Open Targets Platform (Koscielny et al. (2017) <doi:10.1093/nar/gkw1055>).

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.3.3

VignetteBuilder knitr

Imports dplyr, gprofiler2, httr, jsonlite, purrr, tidyr, ggplot2, biomaRt, tibble, magrittr, later (>= 1.3.1), testthat (>= 3.0.0), knitr (>= 1.50), rmarkdown

Config/testthat/edition 3

NeedsCompilation no

Author MD. Arshad [aut, cre]

Repository CRAN

Date/Publication 2026-01-27 08:40:19 UTC

Contents

`add_disease_links` *Add disease association data*

Description

Augments an annotaR object with disease association data from the OpenTargets platform.

Usage

```
add_disease_links(annotaR_object, score_threshold = 0.5)
```

Arguments

`annotaR_object`	A tibble, typically from `annotaR()`, containing a 'gene' column with HGNC symbols.
`score_threshold`	Minimum association score (from 0 to 1) to include. Defaults to 0.5.

Value

A new tibble with the original data joined with disease association columns (disease_name, association_score).

Examples

```
annotaR(c("TP53", "EGFR")) %>%
  add_disease_links(score_threshold = 0.8)
```

`add_drug_links` *Add known drug association data*

Description

Augments an annotaR object with known drug/compound data from the OpenTargets platform. This includes the drug name, type, mechanism of action, and clinical trial phase.

Usage

```
add_drug_links(annotaR_object)
```

Arguments

`annotaR_object`	A tibble, typically from `annotaR()`, containing a 'gene' column with HGNC symbols.

Value

A new tibble with the original data joined with drug association columns (e.g., drug_name, drug_type, mechanism_of_action, phase).

Examples

```
annotaR(c("EGFR", "BRAF")) %>%
  add_drug_links()
```

`add_go_terms` *Add GO functional enrichment data*

Description

Augments an annotaR object with functional enrichment data from g:Profiler. It performs a Gene Ontology (GO) analysis on the gene list and joins the results.

Usage

```
add_go_terms(annotaR_object, organism = "hsapiens", sources = c("GO:BP"), ...)
```

Arguments

`annotaR_object`	A tibble, typically the output of `annotaR()`. Must contain a 'gene' column.
`organism`	The organism name to use for the query (e.g., "hsapiens"). Passed to `gprofiler2::gost`.
`sources`	A vector of data sources to query. Defaults to GO Biological Process. See `gprofiler2::gost` for options.
`...`	Additional parameters passed on to `gprofiler2::gost`.

Value

A new tibble with the original 'gene' column joined with functional annotation columns (e.g., term_id, term_name, p_value, source).

Examples

```
annotaR(c("TP53", "EGFR")) %>%
  add_go_terms()
```

`annotaR` *Create an annotaR object*

Description

Initializes the annotation pipeline by creating a tibble from a character vector of gene symbols. This is the entry point for a typical annotaR workflow.

Usage

```
annotaR(genes)
```

Arguments

`genes`	A character vector of HGNC gene symbols (e.g., c("TP53", "BRCA1")).

Value

A tibble with a single column 'gene', ready to be used in downstream annotation functions.

Examples

```
my_genes <- c("TP53", "EGFR", "BRCA1")
annotaR(my_genes)
```

`plot_enrichment_dotplot` *Plot GO Enrichment Results as a Dot Plot*

Description

Creates a publication-ready dot plot from the results of an `add_go_terms()` call. The plot shows the top enriched terms, with dot size representing the number of genes and color representing the p-value.

Usage

```
plot_enrichment_dotplot(
  annotaR_object,
  n_terms = 20,
  title = "Top GO Enrichment Results"
)
```

Arguments

`annotaR_object`	An object processed by `add_go_terms()`. Must contain `term_name`, `p_value`, and gene columns.
`n_terms`	The maximum number of top terms to display, ordered by p-value. Defaults to 20.
`title`	The title of the plot.

Value

A ggplot object.

Examples

```
# Create a dummy annotaR object with enrichment data
annotated_data <- tibble::tibble(
  gene = c("TP53", "TP53", "EGFR"),
  term_name = c("Cell cycle", "Apoptosis", "Cell cycle"),
  p_value = c(0.001, 0.005, 0.001),
  source = "GO:BP",
  intersection = "TP53,EGFR"
)

plot_enrichment_dotplot(annotated_data)
```

Index