Package 'ActivePathways'

May 6, 2026

Title Integrative Pathway Enrichment Analysis of Multivariate Omics Data

Version 2.0.6

Description
Framework for analysing multiple omics datasets in the context of molecular pathways, biological processes and other types of gene sets. The package uses p-value merging to combine gene- or protein-level signals, followed by ranked hypergeometric tests to determine enriched pathways and processes. Genes can be integrated using directional constraints that reflect how the input datasets are expected interact with one another. This approach allows researchers to interpret a series of omics datasets in the context of known biology and gene function, and discover associations that are only apparent when several datasets are combined. The recent version of the package is part of the following publication: Directional integration and pathway enrichment analysis for multi-omics data. Slobodyanyuk M^, Bahcheli AT^, Klein ZP, Bayati M, Strug LJ, Reimand J. Nature Communications (2024) <doi:10.1038/s41467-024-49986-4>.

Depends R (>= 3.6)

Imports data.table, ggplot2

License GPL-3

BugReports https://github.com/reimandlab/ActivePathways/issues

Encoding UTF-8

RoxygenNote 7.3.2

Suggests testthat, knitr, rmarkdown, RColorBrewer

VignetteBuilder knitr

NeedsCompilation no

Author Juri Reimand [aut, cre],
Jonathan Barenboim [ctb],
Mykhaylo Slobodyanyuk [aut]

Maintainer Juri Reimand <juri.reimand@utoronto.ca>

Repository CRAN

Date/Publication 2025-07-11 22:00:02 UTC

Contents

`ActivePathways` *ActivePathways*

Description

ActivePathways

Usage

```
ActivePathways(
  scores,
  gmt,
  background = makeBackground(gmt),
  geneset_filter = c(5, 1000),
  cutoff = 0.1,
  significant = 0.05,
  merge_method = c("Fisher", "Fisher_directional", "Brown", "DPM", "Stouffer",
    "Stouffer_directional", "Strube", "Strube_directional"),
 correction_method = c("holm", "fdr", "hochberg", "hommel", "bonferroni", "BH", "BY",
    "none"),
  cytoscape_file_tag = NA,
  color_palette = NULL,
  custom_colors = NULL,
  color_integrated_only = "#FFFFF0",
  scores_direction = NULL,
  constraints_vector = NULL
)
```

Arguments

scores	A numerical matrix of p-values where each row is a gene and each column represents an omics dataset (evidence). Rownames correspond to the genes and colnames to the datasets. All values must be 0<=p<=1. We recommend converting missing values to ones.
gmt	A GMT object to be used for enrichment analysis. If a filename, a GMT object will be read from the file.
background	A character vector of gene names to be used as a statistical background. By default, the background is all genes that appear in gmt.
geneset_filter	A numeric vector of length two giving the lower and upper limits for the size of the annotated geneset to pathways in gmt. Pathways with a geneset shorter than geneset_filter[1] or longer than geneset_filter[2] will be removed. Set either value to NA to not enforce a minimum or maximum value, or set geneset_filter to NULL to skip filtering.
cutoff	A maximum merged p-value for a gene to be used for analysis. Any genes with merged, unadjusted p > significant will be discarded before testing.
significant	Significance cutoff for selecting enriched pathways. Pathways with adjusted_p_val <= significant will be selected as results.
merge_method	Statistical method to merge p-values. See section on Merging P-Values
correction_method	Statistical method to correct p-values. See p.adjust for details.
cytoscape_file_tag	The directory and/or file prefix to which the output files for generating enrichment maps should be written. If NA, files will not be written.
color_palette	Color palette from RColorBrewer::brewer.pal to color each column in the scores matrix. If NULL grDevices::rainbow is used by default.
custom_colors	A character vector of custom colors for each column in the scores matrix.
color_integrated_only	A character vector of length 1 specifying the color of the "combined" pathway contribution.
scores_direction	A numerical matrix of log2 transformed fold-change values where each row is a gene and each column represents a dataset (evidence). Rownames correspond to the genes and colnames to the datasets. We recommend converting missing values to zero. Must contain the same dimensions as the scores parameter. Datasets without directional information should be set to 0.
constraints_vector	A numerical vector of +1 or -1 values corresponding to the user-defined directional relationship between columns in scores_direction. Datasets without directional information should be set to 0.

Value

A data.table of terms (enriched pathways) containing the following columns:

term_id The database ID of the term

term_name The full name of the term

adjusted_p_val The associated p-value, adjusted for multiple testing

term_size The number of genes annotated to the term

overlap A character vector of the genes enriched in the term

evidence Columns of `scores` (i.e., omics datasets) that contributed individually to the enrichment of the term. Each input column is evaluated separately for enrichments and added to the evidence if the term is found.

Merging P-values

To obtain a single p-value for each gene across the multiple omics datasets considered, the p-values in `scores` #' are merged row-wise using a data fusion approach of p-value merging. The eight available methods are:

Fisher Fisher's method assumes p-values are uniformly distributed and performs a chi-squared test on the statistic sum(-2 log(p)). This method is most appropriate when the columns in `scores` are independent.

Fisher_directional Fisher's method modification that allows for directional information to be incorporated with the `scores_direction` and `constraints_vector` parameters.

Brown Brown's method extends Fisher's method by accounting for the covariance in the columns of `scores`. It is more appropriate when the tests of significance used to create the columns in `scores` are not necessarily independent. The Brown's method is therefore recommended for many omics integration approaches.

DPM DPM extends Brown's method by incorporating directional information using the `scores_direction` and `constraints_vector` parameters.

Stouffer Stouffer's method assumes p-values are uniformly distributed and transforms p-values into a Z-score using the cumulative distribution function of a standard normal distribution. This method is appropriate when the columns in `scores` are independent.

Stouffer_directional Stouffer's method modification that allows for directional information to be incorporated with the `scores_direction` and `constraints_vector` parameters.

Strube Strube's method extends Stouffer's method by accounting for the covariance in the columns of `scores`.

Strube_directional Strube's method modification that allows for directional information to be incorporated with the `scores_direction` and `constraints_vector` parameters.

Cytoscape

To visualize and interpret enriched pathways, ActivePathways provides an option to further analyse results as enrichment maps in the Cytoscape software. If `!is.na(cytoscape_file_tag)`, four files will be written that can be used to build enrichment maps. This requires the EnrichmentMap and enhancedGraphics apps.

The four files written are:

pathways.txt A list of significant terms and the associated p-value. Only terms with `adjusted_p_val <= significant` are written to this file.

subgroups.txt A matrix indicating whether the significant terms (pathways) were also found to be significant when considering only one column from `scores`. A one indicates that term was found to be significant when only p-values in that column were used to select genes.

pathways.gmt A Shortened version of the supplied GMT file, containing only the significantly enriched terms in pathways.txt.

legend.pdf A legend with colours matching contributions from columns in `scores`.

How to use: Create an enrichment map in Cytoscape with the file of terms (pathways.txt) and the shortened gmt file (pathways.gmt). Upload the subgroups file (subgroups.txt) as a table using the menu File > Import > Table from File. To paint nodes according to the type of supporting evidence, use the 'style' panel, set image/Chart1 to use the column 'instruct' and the passthrough mapping type. Make sure the app enhancedGraphics is installed. Lastly, use the file legend.pdf as a reference for colors in the enrichment map.

Examples

```
fname_scores <- system.file("extdata", "Adenocarcinoma_scores_subset.tsv",
     package = "ActivePathways")
fname_GMT = system.file("extdata", "hsapiens_REAC_subset.gmt",
     package = "ActivePathways")

dat <- as.matrix(read.table(fname_scores, header = TRUE, row.names = 'Gene'))
dat[is.na(dat)] <- 1

ActivePathways(dat, fname_GMT)
```

`brownsMethod` *Merge p-values using the Brown's method.*

Description

Merge p-values using the Brown's method.

Usage

```
brownsMethod(p_values, data_matrix = NULL, cov_matrix = NULL)
```

Arguments

`p_values`	A matrix of m x n p-values.
`data_matrix`	An m x n matrix representing m tests and n samples. NA's are not allowed.
`cov_matrix`	A pre-calculated covariance matrix of `data_matrix`. This is more efficient when making many calls with the same data_matrix. Only one of `data_matrix` and `cov_matrix` must be given. If both are supplied, `data_matrix` is ignored.

Value

A p-value vector representing the merged significance of multiple p-values.

`columnSignificance` *Determine which terms are found to be significant using each column individually.*

Description

Determine which terms are found to be significant using each column individually.

Usage

```
columnSignificance(
  scores,
  gmt,
  background,
  cutoff,
  significant,
  correction_method,
  pvals
)
```

Arguments

`scores`	A numerical matrix of p-values where each row is a gene and each column represents an omics dataset (evidence). Rownames correspond to the genes and colnames to the datasets. All values must be 0<=p<=1. We recommend converting missing values to ones.
`gmt`	A GMT object to be used for enrichment analysis. If a filename, a GMT object will be read from the file.
`background`	A character vector of gene names to be used as a statistical background. By default, the background is all genes that appear in `gmt`.
`cutoff`	A maximum merged p-value for a gene to be used for analysis. Any genes with merged, unadjusted p > `significant` will be discarded before testing.
`significant`	Significance cutoff for selecting enriched pathways. Pathways with `adjusted_p_val` <= `significant` will be selected as results.
`correction_method`	Statistical method to correct p-values. See `p.adjust` for details.
`pvals`	p-value for the pathways calculated by ActivePathways

Value

a data.table with columns 'term_id' and a column for each column in `scores`, indicating whether each term (pathway) was found to be significant or not when considering only that column. For each term, either report the list of related genes if that term was significant, or NA if not.

DPM *Merge p-values using the DPM method.*

Description

Merge p-values using the DPM method.

Usage

```
DPM(
  p_values,
  data_matrix = NULL,
  cov_matrix = NULL,
  scores_direction,
  constraints_vector
)
```

Arguments

p_values	A matrix of m x n p-values.
data_matrix	An m x n matrix representing m tests and n samples. NA's are not allowed.
cov_matrix	A pre-calculated covariance matrix of data_matrix. This is more efficient when making many calls with the same data_matrix. Only one of data_matrix and cov_matrix must be given. If both are supplied, data_matrix is ignored.
scores_direction	A matrix of log2 fold-change values. Datasets without directional information should be set to 0.
constraints_vector	A numerical vector of +1 or -1 values corresponding to the user-defined directional relationship between columns in scores_direction. Datasets without directional information should be set to 0.

Value

A p-value vector representing the merged significance of multiple p-values.

---

export_as_CSV *Export the results from ActivePathways as a comma-separated values (CSV) file.*

Description

Export the results from ActivePathways as a comma-separated values (CSV) file.

Usage

```
export_as_CSV(res, file_name)
```

Arguments

res	the data.table object with ActivePathways results.
file_name	location and name of the CSV file to write to.

Examples

```
fname_scores <- system.file("extdata", "Adenocarcinoma_scores_subset.tsv",
     package = "ActivePathways")
fname_GMT = system.file("extdata", "hsapiens_REAC_subset.gmt",
     package = "ActivePathways")

dat <- as.matrix(read.table(fname_scores, header = TRUE, row.names = 'Gene'))
dat[is.na(dat)] <- 1

res <- ActivePathways(dat, fname_GMT)

export_as_CSV(res, "results_ActivePathways.csv")
```

GMT *Read and Write GMT files*

Description

Functions to read and write Gene Matrix Transposed (GMT) files and to test if an object inherits from GMT.

Usage

```
read.GMT(filename)

write.GMT(gmt, filename)

is.GMT(x)
```

Arguments

filename	Location of the gmt file.
gmt	A GMT object.
x	The object to test.

Format

A GMT object is a named list of terms, where each term is a list with the items:

id The term ID.

name The full name or description of the term.

genes A character vector of genes annotated to this term.

Details

A GMT file describes gene sets, such as biological terms and pathways. GMT files are tab delimited text files. Each row of a GMT file contains a single term with its database ID and a term name, followed by all the genes annotated to the term.

Value

`read.GMT` returns a GMT object.
`write.GMT` returns NULL.
`is.GMT` returns TRUE if `x` is a GMT object, else FALSE.

Examples

```
fname_GMT <- system.file("extdata", "hsapiens_REAC_subset.gmt", package = "ActivePathways")
gmt <- read.GMT(fname_GMT)
gmt[1:10]
gmt[[1]]
gmt[[1]]$id
gmt[[1]]$genes
gmt[[1]]$name
gmt$`REAC:1630316`
```

`hypergeometric` *Hypergeometric test*

Description

Perform a hypergeometric test, also known as the Fisher's exact test, on a 2x2 contingency table with the alternative hypothesis set to 'greater'. In this application, the test finds the probability that counts[1, 1] or more genes would be found to be annotated to a term (pathway), assuming the null hypothesis of genes being distributed randomly to terms.

Usage

```
hypergeometric(counts)
```

Arguments

`counts`	A 2x2 numerical matrix representing a contingency table.

Value

a p-value of enrichment of genes in a term or pathway.

`makeBackground` *Make a background list of genes (i.e., the statistical universe) based on all the terms (gene sets, pathways) considered.*

Description

Returns A character vector of all genes in a GMT object.

Usage

```
makeBackground(gmt)
```

Arguments

`gmt`	A GMT object.

Value

A character vector containing all genes in GMT.

Examples

```
fname_GMT <- system.file("extdata", "hsapiens_REAC_subset.gmt", package = "ActivePathways")
gmt <- read.GMT(fname_GMT)
makeBackground(gmt)[1:10]
```

`merge_gmt` *Merge and filter GMT files based on a list of term IDs*

Description

This function reads a GMT file and filters it to include only the specified term IDs. It's useful for creating a filtered GMT file for visualization in Cytoscape.

Usage

```
merge_gmt(gmt_file, term_ids)
```

Arguments

`gmt_file`	Path to the GMT file
`term_ids`	Character vector of term IDs to include

Value

A GMT object containing only the specified terms

Examples

```
## Not run:
# Get term IDs from merged results
merged_results <- merge_results(
  enriched_pathways, enriched_pathways_directional,
  gmt_file = fname_GMT2,
  output_prefix = "Aggregated",
  tests = c('rna', 'protein', 'combined'),
  col_colors = c("#FF0000", "#00FF00", "#FFFFF0")
)

# Merge and filter GMT file
merge_gmt(
  gmt_file = fname_GMT2,
  term_ids = merged_results$term_ids
)

## End(Not run)
```

---

merge_p_values *Merge a list or matrix of p-values*

Description

Merge a list or matrix of p-values

Usage

```
merge_p_values(
  scores,
  method = "Fisher",
  scores_direction = NULL,
  constraints_vector = NULL
)
```

Arguments

scores	Either a list/vector of p-values or a matrix where each column is a test.
method	Method to merge p-values. See 'methods' section below.
scores_direction	Either a vector of log2 transformed fold-change values or a matrix where each column is a test. Must contain the same dimensions as the scores parameter. Datasets without directional information should be set to 0.

`constraints_vector`
: A numerical vector of +1 or -1 values corresponding to the user-defined directional relationship between the columns in scores_direction. Datasets without directional information should be set to 0.

Value

If `scores` is a vector or list, returns a number. If `scores` is a matrix, returns a named list of p-values merged by row.

Methods

Eight methods are available to merge a list of p-values:

Fisher Fisher's method (default) assumes that p-values are uniformly distributed and performs a chi-squared test on the statistic sum(-2 log(p)). This method is most appropriate when the columns in `scores` are independent.

Fisher_directional Fisher's method modification that allows for directional information to be incorporated with the `scores_direction` and `constraints_vector` parameters.

Brown Brown's method extends Fisher's method by accounting for the covariance in the columns of `scores`. It is more appropriate when the tests of significance used to create the columns in `scores` are not necessarily independent. Note that the "Brown" method cannot be used with a single list of p-values. However, in this case Brown's method is identical to Fisher's method and should be used instead.

DPM DPM extends Brown's method by incorporating directional information using the `scores_direction` and `constraints_vector` parameters.

Stouffer Stouffer's method assumes p-values are uniformly distributed and transforms p-values into a Z-score using the cumulative distribution function of a standard normal distribution. This method is appropriate when the columns in `scores` are independent.

Stouffer_directional Stouffer's method modification that allows for directional information to be incorporated with the `scores_direction` and `constraints_vector` parameters.

Strube Strube's method extends Stouffer's method by accounting for the covariance in the columns of `scores`.

Strube_directional Strube's method modification that allows for directional information to be incorporated with the `scores_direction` and `constraints_vector` parameters.

Examples

```
merge_p_values(c(0.05, 0.09, 0.01))
merge_p_values(list(a=0.01, b=1, c=0.0015, d=0.025), method='Fisher')
merge_p_values(matrix(data=c(0.03, 0.061, 0.48, 0.052), nrow = 2), method='Brown')
```

merge_results *Merge results from multiple ActivePathways analyses*

Description

This function combines results from multiple ActivePathways analyses into a single set of files for visualization in Cytoscape. This is particularly useful for comparing results with and without directional penalties.

Usage

```
merge_results(
  enriched_pathways,
  enriched_pathways_directional,
  gmt_file,
  output_prefix = "Aggregated",
  col_colors = NULL,
 tests = c(gsub("^Genes_", "", grep("^Genes_", colnames(enriched_pathways), value =
    TRUE)), "combined")
)
```

Arguments

enriched_pathways	A data.table returned by ActivePathways
enriched_pathways_directional	A data.table returned by ActivePathways
gmt_file	Path to GMT file
output_prefix	A string prefix for output files
col_colors	A character vector of colors for each test (must match length of tests)
tests	A character vector of names for the data sources (e.g., c('rna', 'protein', 'combined')) or NULL

Value

A list containing the merged results

Examples

```
## Not run:
# Run two different ActivePathways analyses
enriched_pathways <- ActivePathways(
  pval_matrix, gmt = fname_GMT2, cytoscape_file_tag = "original_")

enriched_pathways_directional <- ActivePathways(
  pval_matrix, gmt = fname_GMT2, cytoscape_file_tag = "directional_",
  merge_method = "DPM", scores_direction = dir_matrix,
```

```
  constraints_vector = constraints_vector)

# Merge the results
merge_results(
  enriched_pathways, enriched_pathways_directional,
  gmt_file = fname_GMT2,
  output_prefix = "Aggregated",
  col_colors = c("#FF0000", "#00FF00", "#FFFFF0"),
  tests = c('rna', 'protein', 'combined')
)

## End(Not run)
```

`orderedHypergeometric` *Ordered Hypergeometric Test*

Description

Perform a series of hypergeometric tests (a.k.a. Fisher's Exact tests), on a ranked list of genes ordered by significance against a list of annotation genes. The hypergeometric tests are executed with increasingly larger numbers of genes representing the top genes in order of decreasing scores. The lowest p-value of the series is returned as the optimal enriched intersection of the ranked list of genes and the biological term (pathway).

Usage

```
orderedHypergeometric(genelist, background, annotations)
```

Arguments

`genelist`	Character vector of gene names, assumed to be ordered by decreasing importance. For example, the genes could be ranked by decreasing significance of differential expression.
`background`	Character vector of gene names. List of all genes used as a statistical background (i.e., the universe).
`annotations`	Character vector of gene names. A gene set representing a functional term, process or biological pathway.

Value

a list with the items:

p_val The lowest obtained p-value

ind The index of `genelist` such that `genelist[1:ind]` gives the lowest p-value

Examples

```
orderedHypergeometric(c('HERC2', 'SP100'), c('PHC2', 'BLM', 'XPC', 'SMC3', 'HERC2', 'SP100'),
                      c('HERC2', 'PHC2', 'BLM'))
```

prepareCytoscape *Prepare files for building an enrichment map network visualization in Cytoscape*

Description

This function writes four text files that are used to build an network using Cytoscape and the EnrichmentMap app. The files are prefixed with `cytoscape_file_tag`. The four files written are:

pathways.txt A list of significant terms and the associated p-value. Only terms with `adjusted_p_val <= significant` are written to this file

subgroups.txt A matrix indicating whether the significant pathways are found to be significant when considering only one column (i.e., type of omics evidence) from `scores`. A 1 indicates that that term is significant using only that column to test for enrichment analysis

pathways.gmt A shortened version of the supplied GMT file, containing only the terms in pathways.txt.

legend.pdf A legend with colours matching contributions from columns in `scores`

Usage

```
prepareCytoscape(
  terms,
  gmt,
  cytoscape_file_tag,
  col_significance,
  color_palette = NULL,
  custom_colors = NULL,
  color_integrated_only = "#FFFFF0"
)
```

Arguments

`terms`	A data.table object with the columns 'term_id', 'term_name', 'adjusted_p_val'.
`gmt`	An abridged GMT object containing only the pathways that were found to be significant in the ActivePathways analysis.
`cytoscape_file_tag`	The user-defined file prefix and/or directory defining the location of the files.
`col_significance`	A data.table object with a column 'term_id' and a column for each type of omics evidence indicating whether a term was also found to be significant or not when considering only the genes and p-values in the corresponding column of the `scores` matrix. If term was not found, NA's are shown in columns, otherwise the relevant lists of genes are shown.
`color_palette`	Color palette from RColorBrewer::brewer.pal to color each column in the scores matrix. If NULL grDevices::rainbow is used by default.

`custom_colors` A character vector of custom colors for each column in the scores matrix.

`color_integrated_only`
A character vector of length 1 specifying the color of the "combined" pathway contribution.

Value

None

Index