Package: DGEAR (via r-universe)

October 9, 2026

Type Package

Title Differential Gene Expression Analysis with R

Version 0.2.1

Description Analyses gene expression data derived from microarray experiments to detect differentially expressed genes (DEGs) by employing majority voting across five statistical models: Welch t-test, one-way ANOVA, Dunnett's test, Half's modified t-test, and the Wilcoxon-Mann-Whitney U-test. Combined p-values are computed with Fisher's method. Gene annotation is optional: users may supply a GEO SOFT annotation table or rely on row names directly. Boyer, R.S., Moore, J.S. (1991) <doi:10.1007/978-94-011-3488-0_5>.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Imports DescTools, metapod

RoxygenNote 7.3.2

Depends R (>= 3.5.0)

Suggests spelling, testthat (>= 3.0.0)

Config/testthat/edition 3

NeedsCompilation no

Language en-US

Author Koushik Bardhan [aut, cre, ctb] (ORCID: <https://orcid.org/0009-0002-8846-8347>), Chiranjib Sarkar [aut, ths] (ORCID: <https://orcid.org/0000-0003-1536-7449>)

Maintainer Koushik Bardhan <koushikbardhan2000@gmail.com>

Config/pak/sysreqs cmake make libuv1-dev libssl-dev libx11-dev zlib1g-dev

Repository https://koushikbardhan2000.r-universe.dev

Date/Publication 2026-07-03 08:50:35 UTC

RemoteUrl https://github.com/cran/DGEAR

RemoteRef HEAD

RemoteSha 87ec9971693ed49357a5aa2ee7619d1cd6e971d9

Contents

DGEAR 2
gene_exp_data 4
perform_anova 5
perform_dunnett_test 6
perform_h_test 7
perform_t_test 8
perform_wilcox_test 9

Index **11**

DGEAR *Differential Gene Expression Analysis with R*

Description

Main orchestration function that runs five statistical tests (Welch t-test, one-way ANOVA, Dunnett's test, Half's modified t-test, and Wilcoxon-Mann-Whitney U-test) on a gene expression matrix, combines their BH-adjusted p-values with Fisher's combined probability method, and identifies differentially expressed genes (DEGs) by majority voting.

Usage

```
DGEAR(
  dataframe,
  con1,
  con2,
  exp1,
  exp2,
  alpha = 0.05,
  votting_cutoff = 3,
  annot_df = NULL
)
```

Arguments

dataframe	A numeric matrix or data.frame of gene expression values (rows = genes, columns = samples). Raw intensity / count values are automatically log2-transformed when they appear to be on a linear scale.
con1	Integer. Index of the first control column.
con2	Integer. Index of the last control column.
exp1	Integer. Index of the first experiment column.

`exp2`	Integer. Index of the last experiment column.
`alpha`	Numeric significance threshold for BH-adjusted p-values (default `0.05`).
`votting_cutoff`	Integer. Minimum number of tests (out of 5) that must independently declare a gene significant for it to be included in the majority-vote DEG list (default `3`). Must be between 1 and 5.
`annot_df`	Optional annotation data.frame with columns `ID` and `Gene.Symbol` (or `Gene.symbol`) that maps probe/row identifiers to gene symbols. When `NULL` (default) row names of `dataframe` are used directly as gene identifiers. Typical source: a GEO SOFT family file parsed with `read.delim()`.

Details

The function internally calls:

- `perform_t_test` — Welch two-sample t-test
- `perform_anova` — one-way ANOVA
- `perform_dunnett_test` — Dunnett's test
- `perform_h_test` — Half's modified t-test
- `perform_wilcox_test` — Wilcoxon rank-sum test

Each test independently assigns an FDR flag (1 = significant, 0 = not). The five flags are summed per gene; genes whose sum meets or exceeds `votting_cutoff` are reported as DEGs (majority voting, Boyer & Moore 1991). Combined p-values across all five tests are computed with Fisher's method via `parallelFisher`.

Annotation via `annot_df` is entirely optional. When supplied, the first gene symbol listed for each probe (delimited by `///`) is used. When absent, row names serve as identifiers, making the function fully self-contained without GEO annotation files.

Value

A named list with four elements:

`DEGs` Data.frame of gene identifiers that passed majority voting.

`FDR_Table` Wide data.frame with BH-adjusted p-values from every test, the Fisher-combined FDR, the ensemble voting score, and log2 fold change for every gene.

`Results_Table` Concise data.frame with `G_Symbol`, `CombineFDR`, `log2FC`, and `Ensemble` score.

`IndividualTests` Named list of the raw output from each of the five test functions (each containing a `Table` and a `DEGs` element).

References

Boyer, R.S. and Moore, J.S. (1991). MJRTY — A Fast Majority Vote Algorithm. In *Automated Reasoning: Essays in Honor of Woody Bledsoe*, pp. 105–117. Springer, Dordrecht. doi:10.1007/9789401134880_5

Examples

```
library(DGEAR)
data("gene_exp_data")

## Basic usage - no annotation file needed
result <- DGEAR(dataframe    = gene_exp_data,
                con1         = 1,
                con2         = 10,
                exp1         = 11,
                exp2         = 20,
                alpha        = 0.05,
                votting_cutoff = 2)
result$DEGs
head(result$FDR_Table)

## With an optional annotation data.frame (GEO SOFT format)
## annot <- read.delim("GSExxxxx_family.soft")
## result <- DGEAR(dataframe = gene_exp_data,
##                 con1 = 1, con2 = 10, exp1 = 11, exp2 = 20,
##                 annot_df = annot)
```

---

gene_exp_data *A dataset containing gene expression data*

Description

This dataset contains statistically simulated gene expression data for ease of exercise.

Usage

```
gene_exp_data
```

Format

A data frame with 10 rows and 20 columns, the columns represents samples, say first 10 columns 1 to 10 being control and 11 to 20 being experiment. Whereas, the rows of the dataset contains genes. First 5 out of 10 genes, gene1-gene5 are the true DEGs as the expression values for the first 10 samples are ~13 times higher than the rest.

Examples

```
# Data will be loaded with lazy loading and can be accessible when needed.
data("gene_exp_data")
head(gene_exp_data)
```

`perform_anova` *One-Way ANOVA Test for Differential Gene Expression*

Description

Performs a one-way ANOVA (Welch correction via `oneway.test`) for every gene (row) in the expression matrix, applies BH correction, and returns a results table together with the list of significant DEGs.

Usage

```
perform_anova(dataframe, con1, con2, exp1, exp2, alpha = 0.05, annot_df = NULL)
```

Arguments

`dataframe`	A numeric matrix or data.frame of gene expression values (rows = genes, columns = samples). Values are automatically log2- transformed when they appear to be on a linear / intensity scale.
`con1`	Integer. Index of the first control column.
`con2`	Integer. Index of the last control column.
`exp1`	Integer. Index of the first experiment column.
`exp2`	Integer. Index of the last experiment column.
`alpha`	Numeric significance threshold for BH-adjusted p-values (default `0.05`).
`annot_df`	Optional annotation data.frame with columns `ID` and `Gene.Symbol` (or `Gene.symbol`). When `NULL` (default) row names of `dataframe` are used as gene identifiers.

Value

A named list:

Table Data.frame with columns `G_Symbol`, `log2FC`, `statistic.F`, `p.value`, `BH`, `fdr`.

DEGs Data.frame of significant gene identifiers.

Examples

```
library(DGEAR)
data("gene_exp_data")
result <- perform_anova(dataframe = gene_exp_data,
                        con1 = 1, con2 = 10,
                        exp1 = 11, exp2 = 20)
head(result$Table)
result$DEGs
```

perform_dunnett_test *Dunnett's Test for Differential Gene Expression*

Description

Performs Dunnett's multiple comparison test (control vs. treatment) for every gene (row) in the expression matrix using `DunnettTest`, applies BH correction, and returns a results table with the list of significant DEGs. Genes with insufficient variance, missing values, or other numerical issues are skipped gracefully and receive `NA` p-values.

Usage

```
perform_dunnett_test(
  dataframe,
  con1,
  con2,
  exp1,
  exp2,
  alpha = 0.05,
  annot_df = NULL
)
```

Arguments

`dataframe`	A numeric matrix or data.frame of gene expression values (rows = genes, columns = samples). Values are automatically log2- transformed when they appear to be on a linear / intensity scale.
`con1`	Integer. Index of the first control column.
`con2`	Integer. Index of the last control column.
`exp1`	Integer. Index of the first experiment column.
`exp2`	Integer. Index of the last experiment column.
`alpha`	Numeric significance threshold for BH-adjusted p-values (default `0.05`).
`annot_df`	Optional annotation data.frame with columns `ID` and `Gene.Symbol` (or `Gene.symbol`). When `NULL` (default) row names of `dataframe` are used as gene identifiers.

Value

A named list:

Table Data.frame with columns `G_Symbol`, `log2FC`, `p.value`, `BH`, `fdr`.

DEGs Data.frame of significant gene identifiers.

Examples

```
library(DGEAR)
data("gene_exp_data")
result <- perform_dunnett_test(dataframe = gene_exp_data,
                               con1 = 1, con2 = 10,
                               exp1 = 11, exp2 = 20)
head(result$Table)
result$DEGs
```

`perform_h_test` *Half's Modified t-Test for Differential Gene Expression*

Description

Computes a modified t-statistic (sometimes called "Half's t-test") that uses only the control standard deviation in its denominator, applies BH correction, and returns a results table together with the list of significant DEGs.

Usage

```
perform_h_test(
  dataframe,
  con1,
  con2,
  exp1,
  exp2,
  alpha = 0.05,
  annot_df = NULL
)
```

Arguments

`dataframe`	A numeric matrix or data.frame of gene expression values (rows = genes, columns = samples). Values are automatically log2- transformed when they appear to be on a linear / intensity scale.
`con1`	Integer. Index of the first control column.
`con2`	Integer. Index of the last control column.
`exp1`	Integer. Index of the first experiment column.
`exp2`	Integer. Index of the last experiment column.
`alpha`	Numeric significance threshold for BH-adjusted p-values (default `0.05`).
`annot_df`	Optional annotation data.frame with columns `ID` and `Gene.Symbol` (or `Gene.symbol`). When `NULL` (default) row names of `dataframe` are used as gene identifiers.

Value

A named list:

Table Data.frame with columns `G_Symbol`, `log2FC`, `statistic`, `p.value`, `BH`, `fdr`.

DEGs Data.frame of significant gene identifiers.

Examples

```
library(DGEAR)
data("gene_exp_data")
result <- perform_h_test(dataframe = gene_exp_data,
                         con1 = 1, con2 = 10,
                         exp1 = 11, exp2 = 20)
head(result$Table)
result$DEGs
```

`perform_t_test` *Welch Two-Sample t-Test for Differential Gene Expression*

Description

Performs an independent two-sample Welch t-test for every gene (row) in the expression matrix, applies Benjamini-Hochberg (BH) correction, and returns a results table together with the list of significant DEGs.

Usage

```
perform_t_test(
  dataframe,
  con1,
  con2,
  exp1,
  exp2,
  alpha = 0.05,
  annot_df = NULL
)
```

Arguments

`dataframe`	A numeric matrix or data.frame of gene expression values (rows = genes, columns = samples). Values are automatically log2- transformed when they appear to be on a linear / intensity scale.
`con1`	Integer. Index of the first control column.
`con2`	Integer. Index of the last control column.
`exp1`	Integer. Index of the first experiment column.
`exp2`	Integer. Index of the last experiment column.

alpha	Numeric significance threshold for BH-adjusted p-values (default 0.05).
annot_df	Optional annotation data.frame with columns ID and Gene.Symbol (or Gene.symbol). When NULL (default) row names of dataframe are used as gene identifiers.

Value

A named list:

Table Data.frame with columns G_Symbol, log2FC, statistic.t, p.value, BH, fdr.

DEGs Data.frame of gene identifiers whose BH-adjusted p-value is $\leq$ alpha.

Examples

```
library(DGEAR)
data("gene_exp_data")
result <- perform_t_test(dataframe = gene_exp_data,
                         con1 = 1, con2 = 10,
                         exp1 = 11, exp2 = 20)
head(result$Table)
result$DEGs
```

---

perform_wilcox_test *Wilcoxon-Mann-Whitney U-Test for Differential Gene Expression*

Description

Performs the Wilcoxon rank-sum (Mann-Whitney U) test for every gene (row) in the expression matrix, applies BH correction, and returns a results table together with the list of significant DEGs.

Usage

```
perform_wilcox_test(
  dataframe,
  con1,
  con2,
  exp1,
  exp2,
  alpha = 0.05,
  annot_df = NULL
)
```

Arguments

dataframe	A numeric matrix or data.frame of gene expression values (rows = genes, columns = samples). Values are automatically log2- transformed when they appear to be on a linear / intensity scale.
con1	Integer. Index of the first control column.

con2	Integer. Index of the last control column.
exp1	Integer. Index of the first experiment column.
exp2	Integer. Index of the last experiment column.
alpha	Numeric significance threshold for BH-adjusted p-values (default 0.05).
annot_df	Optional annotation data.frame with columns ID and Gene.Symbol (or Gene.symbol). When NULL (default) row names of dataframe are used as gene identifiers.

Value

A named list:

Table Data.frame with columns G_Symbol, log2FC, statistic.W, p.value, BH, fdr.

DEGs Data.frame of significant gene identifiers.

Examples

```
library(DGEAR)
data("gene_exp_data")
result <- perform_wilcox_test(dataframe = gene_exp_data,
                              con1 = 1, con2 = 10,
                              exp1 = 11, exp2 = 20)
head(result$Table)
result$DEGs
```

Index

∗ **datasets**
gene_exp_data, 4

DGEAR, 2
DunnettTest, *6*

gene_exp_data, 4

oneway.test, *5*

parallelFisher, *3*
perform_anova, *3*, 5
perform_dunnett_test, *3*, 6
perform_h_test, *3*, 7
perform_t_test, *3*, 8
perform_wilcox_test, *3*, 9