Package 'DEHOGT'

July 21, 2025

Type Package

Title Differentially Expressed Heterogeneous Overdispersion Gene Test for Count Data

Version 0.99.0

Description Implements a generalized linear model approach for detecting differentially expressed genes across treatment groups in count data. The package supports both quasi-Poisson and negative binomial models to handle over-dispersion, ensuring robust identification of differential expression. It allows for the inclusion of treatment effects and gene-wise covariates, as well as normalization factors for accurate scaling across samples. Additionally, it incorporates statistical significance testing with options for p-value adjustment and log2 fold range thresholds, making it suitable for RNA-seq analysis as described in by Xu et al., (2024) <doi:10.1371/journal.pone.0300565>.

License GPL-3

Encoding UTF-8

Depends R (>= 3.5.0)

Imports doParallel, foreach, MASS,

Suggests knitr, rmarkdown, BiocStyle

biocViews GeneExpression, DifferentialExpression, StatisticalMethod, Regression, Normalization

VignetteBuilder knitr

RoxygenNote 7.3.2

URL `https://github.com/ahshen26/DEHOGT`

BugReports `https://github.com/ahshen26/DEHOGT/issues`

NeedsCompilation no

Author Qi Xu [aut],
Arlina Shen [cre] (ORCID: <`https://orcid.org/0009-0008-5330-6659`>),
Yubai Yuan [ctb],
Annie Qu [ctb]

Maintainer Arlina Shen <`ahshen24@berkeley.edu`>

Repository CRAN

Date/Publication 2024-09-13 18:30:06 UTC

Contents

`dehogt_func` *Differentially Expressed Heterogeneous Overdispersion Genes Testing for Count Data This script implements the main function of the proposed method in the above paper*

Description

Differentially Expressed Heterogeneous Overdispersion Genes Testing for Count Data This script implements the main function of the proposed method in the above paper

Usage

```
dehogt_func(
  data,
  treatment,
  norm_factors = NULL,
  covariates = NULL,
  dist = "qpois",
  padj = TRUE,
  pval_thre = 0.05,
  l2fc = FALSE,
  l2fc_thre = 1,
  num_cores = 1
)
```

Arguments

`data`	A matrix of gene expression data where rows represent genes and columns represent samples.
`treatment`	A vector specifying the treatment conditions for each sample.
`norm_factors`	An optional vector of normalization factors for each sample. Default is NULL, which assumes equal normalization factors.
`covariates`	An optional matrix of gene-wise covariates. Default is NULL.
`dist`	The distribution family for the GLM. Can be "qpois" for quasi-Poisson or "negbin" for negative binomial. Default is "qpois".

`padj`	Logical value indicating whether to adjust p-values using the Benjamini-Hochberg (BH) procedure. Default is TRUE.
`pval_thre`	The threshold for identifying differentially expressed genes based on adjusted p-values. Default is 0.05.
`l2fc`	Logical value indicating whether to consider log2 fold change for identifying differentially expressed genes. Default is FALSE.
`l2fc_thre`	The threshold for log2 fold change in identifying differentially expressed genes. Default is 1.
`num_cores`	The number of CPU cores to use for parallel computing. Default is 1.

Value

A list containing:

`DE_idx`	A logical vector indicating differentially expressed genes.
`pvals`	A numeric vector of p-values for each gene.
`log2fc`	A numeric vector of log2 fold changes for each gene.

Examples

```
# simulate gene expression data
data <- matrix(rpois(1000, 10), nrow = 100, ncol = 10)
# simulate random treatment assignments
treatment <- sample(0:1, 10, replace = TRUE)
# Run main function with parallel computing using 2 cores
result <- dehogt_func(data, treatment, num_cores = 2)
```

Index