Package 'LimROTS'

July 23, 2025

Title LimROTS: A Hybrid Method Integrating Empirical Bayes and Reproducibility-Optimized Statistics for Robust Differential Expression Analysis

Version 1.1.4

Description Differential expression analysis is a prevalent method utilised in the examination of diverse biological data. The reproducibility-optimized test statistic (ROTS) modifies a t-statistic based on the data's intrinsic characteristics and ranks features according to their statistical significance for differential expression between two or more groups (f-statistic). Focussing on proteomics and metabolomics, the current ROTS implementation cannot account for technical or biological covariates such as MS batches or gender differences among the samples. Consequently, we developed LimROTS, which employs a reproducibility-optimized test statistic utilising the limma methodology to simulate complex experimental designs. LimROTS is a hybrid method integrating empirical bayes and reproducibility-optimized statistics for robust analysis of proteomics and metabolomics data.

License Artistic-2.0

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Depends R (>= 4.5.0), SummarizedExperiment

biocViews Software, GeneExpression, DifferentialExpression, Microarray, RNASeq, Proteomics, ImmunoOncology, Metabolomics, mRNAMicroarray

URL https://github.com/AliYoussef96/LimROTS, https://aliyoussef96.github.io/LimROTS/

BugReports https://github.com/AliYoussef96/LimROTS/issues

VignetteBuilder knitr

Imports limma, stringr, qvalue, utils, stats, BiocParallel, S4Vectors, dplyr

Suggests BiocStyle, ggplot2, magick, testthat (>= 3.0.0), knitr, rmarkdown, caret, ROTS

Config/testthat/edition 3

git_url https://git.bioconductor.org/packages/LimROTS

git_branch devel

git_last_commit 24550a1

git_last_commit_date 2025-06-16

Repository Bioconductor 3.22

Date/Publication 2025-07-22

Author Ali Mostafa Anwar [aut, cre] (ORCID:
<https://orcid.org/0000-0002-5201-387X>),
Leo Lahti [aut, ths] (ORCID: <https://orcid.org/0000-0001-5537-637X>),
Akewak Jeba [aut, ctb] (ORCID: <https://orcid.org/0009-0007-1347-7552>),
Eleanor Coffey [aut, ths] (ORCID:
<https://orcid.org/0000-0002-9717-5610>)

Maintainer Ali Mostafa Anwar <aliali.mostafa99@gmail.com>

Contents

`bootstrapS` *Generate Bootstrap Samples*

Description

This function generates bootstrap samples from the input metadata. It samples with replacement within each group defined in the metadata, and optionally adjusts for paired groups.

Usage

```
bootstrapS(niter, meta.info, group.name)
```

Arguments

`niter`	Integer. The number of bootstrap samples to generate.
`meta.info`	Data frame. Metadata containing sample information, where each row corresponds to a sample.
`group.name`	Character. The name of the column in `meta.info` that defines the grouping variable for the samples.

Details

The function works by resampling the row names of the metadata for each group separately.

Value

A matrix of dimension `niter` x `n`, where `n` is the number of samples. Each row corresponds to a bootstrap sample, and each entry is a resampled row name from the metadata.

`bootstrapSamples_limRots`

Generate Stratified Bootstrap Samples for limRots

Description

This function generates stratified bootstrap samples based on the groupings and additional factors in the metadata. The function ensures that samples are drawn proportionally based on strata defined by the interaction of factor columns in the metadata.

Usage

```
bootstrapSamples_limRots(niter, meta.info, group.name)
```

Arguments

`niter`	Integer. The number of bootstrap samples to generate.
`meta.info`	Data frame. Metadata containing sample information, where each row corresponds to a sample. Factor columns in `meta.info` are used to define strata for sampling.
`group.name`	Character. The name of the column in `meta.info` that defines the grouping variable for the samples.

Details

The function works by first identifying the factors in the `meta.info` data frame that are used to create strata for sampling. Within each group defined by `group.name`, the function samples according to the strata proportions, ensuring that samples are drawn from the correct groups and strata in a proportional manner.

Value

A matrix of dimension `niter` x `n`, where `n` is the number of samples. Each row corresponds to a bootstrap sample, and each entry is a resampled row name from the metadata, stratified by group and additional factors.

`Boot_parallel` *Parallel processing handling function*

Description

Parallel processing handling function

Usage

```
Boot_parallel(
  BPPARAM = NULL,
  samples,
  data,
  formula.str,
  group.name,
  groups,
  meta.info,
  a1,
  a2,
  pSamples
)
```

Arguments

`BPPARAM`	A parallel BPPARAM object for distributed computation.
`samples`	bootstrapped samples matrix
`data`	A `SummarizedExperiment` object or a matrix where rows represent features (e.g., genes, proteins) and columns represent samples. The values should be log-transformed.
`formula.str`	A formula string used when covariates are present in meta. info for modeling. It should include "~ 0 + ..." to exclude the intercept from the model.
`group.name`	A string specifying the column in `meta.info` that represents the groups or conditions for comparison.
`groups`	groups information from `meta.info`
`meta.info`	A data frame containing sample-level metadata, where each row corresponds to a sample. It should include the grouping variable specified in `group.name`. If `x` is a `SummarizedExperiment` object, `meta.info` must be a vector of the metadata needed for the model to run and can be retrieved using `colData()`.
`a1`	Optional numeric value used in the optimization process. If defined by the user, no optimization occurs.
`a2`	Optional numeric value used in the optimization process. If defined by the user, no optimization occurs.
`pSamples`	a permutated list of samples

Value

A list containing: `D, S, pD, pS` for bootstrapped data and for permuted data.

calculateFalseDiscoveryRate *Calculate False Discovery Rate (FDR) Using Permuted Values (Adjusted)*

Description

This function calculates the false discovery rate (FDR) by comparing observed values to permuted values.The function sorts observed values, compares them against permuted data, and computes FDR using the median of permutation results.

Usage

```
calculateFalseDiscoveryRate(observedValues, permutedValues)
```

Arguments

observedValues	Numeric vector. The observed test statistics or values to be evaluated for significance.
permutedValues	Numeric matrix. The permuted test statistics or values, with rows corresponding to the same values as in observedValues and columns representing different permutations.

Value

A numeric vector of the same length as observedValues, containing the estimated FDR for each observed value.

calOverlaps *Calculate Overlaps Between Observed and Permuted Data*

Description

This function calculates the overlap between observed and permuted data for two sets of comparisons. It computes the ratio of overlap between pairs of vectors (res1/res2 and pres1/pres2) after sorting the values.

Usage

```
calOverlaps(D, S, pD, pS, nrow, N, N_len, ssq, niter, overlaps, overlaps_P)
```

Arguments

D	Numeric vector. Observed data values (e.g., differences).
S	Numeric vector. Standard errors or related values associated with the observed data.
pD	Numeric vector. Permuted data values (e.g., differences).
pS	Numeric vector. Standard errors or related values associated with the permuted data.

nrow	Integer. Number of rows in each block of data.
N	Integer vector. Number of top values to consider for overlap calculation.
N_len	Integer. Length of the N vector.
ssq	Numeric. A small constant added to standard errors for stability.
niter	Integer. Number of bootstrap samples or resampling iterations.
overlaps	Numeric matrix. Matrix to store overlap results for observed data.
overlaps_P	Numeric matrix. Matrix to store overlap results for permuted data.

Details

The function calculates overlaps for two sets of comparisons: one for observed data (res1/res2) and one for permuted data (pres1/pres2).For each bootstrap sample, the function orders the two vectors being compared, then calculates the proportion of overlap for the top N values.

Value

A list containing two matrices: overlaps for observed data and overlaps_P for permuted data.

---

calOverlaps_slr *Calculate Overlaps for Single-Label Replicates (SLR)*

Description

This function computes the overlap between two sets of observed and permuted 'values for single-label replicates (SLR). It calculates the proportion of overlap between pairs of vectors (res1/res2 and pres1/pres2) after sorting them.

Usage

```
calOverlaps_slr(D, pD, nrow, N, N_len, niter, overlaps, overlaps_P)
```

Arguments

D	Numeric vector. Observed data values (e.g., differences).
pD	Numeric vector. Permuted data values.
nrow	Integer. Number of rows in each block of data.
N	Integer vector. Number of top values to consider for overlap calculation.
N_len	Integer. Length of the N vector.
niter	Integer. Number of bootstrap samples or resampling iterations.
overlaps	Numeric matrix. Matrix to store overlap results for observed data.
overlaps_P	Numeric matrix. Matrix to store overlap results for permuted data.

Details

The function calculates the overlap for two sets of comparisons: one for observed data (res1/res2) and one for permuted data (pres1/pres2). For each bootstrap sample, the function orders the two vectors being compared, then computes the proportion of overlap for the top N values.

Value

A list containing two matrices: overlaps for observed data and overlaps_P for permuted data.

Check_meta_info *Check if meta info is correct*

Description

Check if meta info is correct

Usage

```
Check_meta_info(meta.info, data, log)
```

Arguments

meta.info	Data frame. Metadata associated with the samples (columns of data.exp). If data.exp is a SummarizedExperiment,
data	A matrix-like object or a SummarizedExperiment containing the data to be analyzed.
log	Logical, indicating whether the data is already log-transformed. Default is TRUE.

Value

Logical

Check_SummarizedExperiment *Check if SummarizedExperiment or data is correct*

Description

Check if SummarizedExperiment or data is correct

Usage

```
Check_SummarizedExperiment(data.exp, meta.info, group.name)
```

Arguments

data.exp	A matrix-like object or a SummarizedExperiment containing the data to be analyzed.
meta.info	Data frame. Metadata associated with the samples (columns of data.exp). If data.exp is a SummarizedExperiment,
group.name	Character. Column name in meta.info that defines the groups or conditions for comparison.

Value

a list of data , groups and meta.info

countLargerThan *Count Larger Permuted Values (Modified)*

Description

This helper function compares observed values against permuted values and counts the number of permuted values that are greater than or equal to each observed value.

Usage

```
countLargerThan(observedVec, permutedVec)
```

Arguments

observedVec	Numeric vector. The observed values.
permutedVec	Numeric vector. The permuted values to compare against the observed values.

Value

A numeric vector containing the counts of permuted values greater than or equal to the corresponding observed values.

Limma_bootstrap *Perform Linear Modeling with Covariates using Limma*

Description

This function performs linear modeling using the Limma package while accounting for covariates specified in the meta.info. It supports two-group comparisons and multi-group analysis, incorporating covariates through a design matrix.

Usage

```
Limma_bootstrap(x, group.name, meta.info, formula.str)
```

Arguments

x	A list containing two or more data matrices where rows represent features (e.g., genes, proteins) and columns represent samples. The list should contain at least two matrices for pairwise group comparison.
group.name	A character string indicating the name of the group variable in meta.info to be used in the analysis.
meta.info	A data frame containing the metadata for the samples. This includes sample grouping and any covariates to be included in the model.
formula.str	A string specifying the formula to be used in model fitting. It should follow the standard R formula syntax (e.g., ~ covariate1 + covariate2).

Details

This function first combines the data matrices from different groups and prepares a design matrix based on the covariates specified in `meta.info` using the provided formula. It fits a linear model using Limma, computes contrasts between groups, and applies empirical Bayes moderation. For two-group comparisons, the function returns log-fold changes and associated statistics. In multi-group settings with a single covariate, it calculates pairwise contrasts and moderated F-statistics.

Value

A list containing the following elements:

`d`	A vector of the test statistics (log-fold changes or F-statistics) for each feature.
`s`	A vector of the standard deviations for each feature, adjusted by the empirical Bayes procedure.

See Also

`lmFit`, `eBayes`, `topTable`, `makeContrasts`

`Limma_fit` *Perform Linear Modeling with Covariates using Limma*

Description

This function performs linear modeling using the Limma package, incorporating covariates in the model fitting process. It is designed to handle both two-group comparisons and multi-group settings with covariates.

Usage

```
Limma_fit(x, group.name, meta.info, formula.str, trend, robust)
```

Arguments

`x`	A list containing two or more data matrices where rows represent features (e.g., genes, proteins) and columns represent samples. The list should contain at least two matrices for pairwise group comparison.
`group.name`	A character string indicating the name of the group variable in`meta.info` to be used in the analysis.
`meta.info`	A data frame containing the metadata for the samples. This includes sample grouping and any covariates to be included in the model.
`formula.str`	A string specifying the formula to be used in model fitting. It should follow the standard R formula syntax (e.g., `~ covariate1 + covariate2`).
`trend`	A logical value indicating whether to allow for an intensity-dependent trend in the prior variance.
`robust`	A logical value indicating whether to use a robust fitting procedure to protect against outliers.

Details

This function combines the data matrices from different groups and fits a linear model using covariates provided in the `meta.info`. For two-group comparisons, the function computes contrasts between the two groups and applies empirical Bayes moderation. For multi-group analysis with a single covariate, pairwise contrasts are computed, and the moderated F-statistic is calculated for each feature.

Value

A list containing the following elements:

`d`	A vector of the test statistics (log-fold changes or F-statistics) for each feature.
`s`	A vector of the standard deviations for each feature, adjusted by t he empirical Bayes procedure.
`corrected.logfc`	The log-fold changes for each feature after fitting the model.

See Also

`lmFit`, `eBayes`, `topTable`, `makeContrasts`

`Limma_permutating` *Perform Permutation-Based Linear Modeling with Covariates using Limma*

Description

This function performs linear modeling using the Limma package with permutation of the covariates to evaluate the test statistics under random assignments. It handles two-group comparisons and multi-group settings.

Usage

```
Limma_permutating(x, group.name, meta.info, formula.str)
```

Arguments

`x`	A data matrices where rows represent features (e.g., genes, proteins) and columns represent samples. The list should contain at least two matrices for pairwise group comparison.
`group.name`	A character string indicating the name of the group variable in `meta.info` to be used in the analysis.
`meta.info`	A data frame containing the metadata for the samples. This includes sample grouping and any covariates to be included in the model.
`formula.str`	A string specifying the formula to be used in model fitting. It should follow the standard R formula syntax (e.g., `~ covariate1 + covariate2`).

Details

This function combines the data matrices from different groups and permutes the covariates from `meta.info`before fitting a linear model using Limma. Permutation helps assess how the covariates behave under random conditions, providing a null distribution of the test statistics. For two-group comparisons, the function computes contrasts between the two groups and applies empirical Bayes moderation. For multi-group analysis with a single covariate, pairwise contrasts are computed, and the moderated F-statistic is calculated for each feature.

Value

A list containing the following elements:

`d`	A vector of the test statistics (log-fold changes or F-statistics) for each feature.
`s`	A vector of the standard deviations for each feature, adjusted by the empirical Bayes procedure.

See Also

`lmFit`, `eBayes`, `topTable`, `makeContrasts`

`LimROTS` `LimROTS`*: A Hybrid Method Integrating Empirical Bayes and Reproducibility-Optimized Statistics for Robust Differential Expression Analysis*

Description

`LimROTS`: A Hybrid Method Integrating Empirical Bayes and Reproducibility-Optimized Statistics for Robust Differential Expression Analysis

Usage

```
LimROTS(
  x,
  niter = 1000,
  K = NULL,
  a1 = NULL,
  a2 = NULL,
  log = TRUE,
  verbose = TRUE,
  meta.info,
  BPPARAM = NULL,
  group.name,
  formula.str,
  robust = TRUE,
  trend = TRUE,
  permutating.group = FALSE
)
```

Arguments

x	A SummarizedExperiment object, where rows represent features (e.g., proteins, metabolites) and columns represent samples. The values should be log-transformed.
niter	An integer representing the amount of bootstrap iterations. Default is 1000.
K	An optional integer representing the top list size for ranking. If not specified, it is set to one-fourth of the number of features.
a1	Optional numeric value used in the optimization process. If defined by the user, no optimization occurs.
a2	Optional numeric value used in the optimization process. If defined by the user, no optimization occurs.
log	Logical, indicating whether the data is already log-transformed. Default is TRUE.
verbose	Logical, indicating whether to display messages during the function's execution. Default is TRUE.
meta.info	a character vector of the metadata needed for the model to run and can be retrieved using colData().
BPPARAM	A BiocParallelParam object specifying the parallelization backend (e.g., MulticoreParam, SnowParam). The default depends on the operating system: if the user is on Windows, SnowParam(workers = 2) is used; otherwise, MulticoreParam(workers = 2).
group.name	A string specifying the column in meta.info that represents the groups or conditions for comparison. group.name should be retrieved using colData() as factor.
formula.str	A formula string for modeling. It should include "~ 0 + ..." to exclude the intercept from the model. All the model parameters must be present in meta.info.
robust	indicating whether robust fitting should be used. Default is TRUE, see eBayes.
trend	indicating whether to include trend fitting in the differential expression analysis. Default is TRUE. see eBayes.
permutating.group	Logical, If TRUE, the permutation for calculating the null distribution is performed by permuting the target group only specified in group.name Preserving all the other sample information. If FALSE, the entire sample information retrieved from meta.info will be permuted (recommended to be set to FALSE).

Details

The **LimROTS** approach initially uses **limma** package functionality to simulate the intensity data of proteins and metabolites. A linear model is subsequently fitted using the design matrix. Empirical Bayes variance shrinking is then implemented. To obtain the moderated t-statistics, the adjusted standard error $SE_{post} = \sqrt{s^2_{\text{post}}} \times unscaledSD$ for each feature is computed, along with the regression coefficient for each feature (indicating the impact of variations in the experimental settings). Then, by adapting a reproducibility-optimized technique known as ROTS to establish an optimality based on the largest overlap of top-ranked features within group-preserving bootstrap datasets, Finally based on the optimized parameters $\alpha 1$ and $\alpha 2$ this equation used to calculates the final statistics:

$$t_{\alpha_{(p)}} = \frac{\beta_{(p)}}{\alpha 1 + \alpha 2 \times SEpost_{(p)}}$$

where $t_{\alpha_{(p)}}$ is the final statistics for each feature, $\beta_{(p)}$ is the coefficient, and $SEpost_{(p)}$ is the the adjusted standard error. LimROTS generates p-values from permutation samples using the implementation available in qvalue package, along with internal implementation of FDR adapted from ROTS package. Additionally, the qvalue package is used to calculate q-values, were the proportion of true null p-values is set to the bootstrap method pi0est. We recommend using permutation-derived p-values and qvalues.

This function processes a dataset using parallel computation. It leverages the **BiocParallel** framework to distribute tasks across multiple workers, which can significantly reduce runtime for large datasets.

Value

An object of class `"SummarizedExperiment"` with the following elements:

`data`	The original data matrix.
`niter`	The number of bootstrap samples used.
`statistics`	The optimized statistics for each feature.
`logfc`	Log-fold change values between groups.
`pvalue`	P-values computed based on the permutation samples.
`FDR`	False discovery rate estimates.
`a1`	Optimized parameter used in differential expression ranking.
`a2`	Optimized parameter used in differential expression ranking.
`k`	Top list size used for ranking.
`corrected.logfc`	estimate of the log2-fold-change corresponding to the effect corrected by the s model see topTable.
`q_values`	Estimated q-values using the `qvalue` package.
`BH.pvalue`	Benjamini-Hochberg adjusted p-values.
`null.statistics`	The optimized null statistics for each feature.

References

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research 43(7), e47

Suomi T, Seyednasrollah F, Jaakkola M, Faux T, Elo L (2017). "ROTS: An R package for reproducibility-optimized statistical testing. " *PLoS computational biology*, *13*(5), e1005562. doi:10.1371/journal.pcbi.1005562 https://doi.org/10.1371/journal.pcbi.1005562, http://www.ncbi.nlm.nih.gov/pubmed/28542205

Elo LL, Filen S, Lahesmaa R, Aittokallio T. Reproducibility-optimized test statistic for ranking genes in microarray studies. IEEE/ACM Trans Comput Biol Bioinform. 2008;5(3):423-431. doi:10.1109/tcbb.2007.1078

Examples

```
# Example usage:

data <- data.frame(matrix(rnorm(500), nrow = 100, ncol = 10))
# Simulated data
meta.info <- data.frame(
    group = factor(rep(1:2, each = 5)),
    row.names = colnames(data)
)
formula.str <- "~ 0 + group"
result <- LimROTS(data,
    meta.info = meta.info, group.name = "group",
    formula.str = formula.str, niter = 10
)
```

`Optimizing` *Optimize Parameters Based on Overlap Calculations*

Description

This function optimizes parameters by calculating overlaps between observed and permuted data for multiple values of a smoothing constant (`ssq`) and a single-label replicate (SLR) comparison.

Usage

```
Optimizing(niter, ssq, N, D, S, pD, pS, verbose)
```

Arguments

`niter`	Integer. Number of bootstrap samples or resampling iterations.
`ssq`	Numeric vector. Smoothing constants to be evaluated.
`N`	Integer vector. Number of top values to consider for overlap calculation.
`D`	Numeric matrix. Observed data values.
`S`	Numeric matrix. Standard errors or related values for observed data.
`pD`	Numeric matrix. Permuted data values.
`pS`	Numeric matrix. Standard errors or related values for permuted data.
`verbose`	Logical. If `TRUE`, progress messages will be displayed.

Details

The function calculates overlaps for a range of smoothing constants and identifies the optimal set of parameters by maximizing a z-score-based metric, which compares the overlap of observed data to permuted data. It computes overlap matrices for both observed (`D` and `S`) and permuted (`pD` and `pS`) data and returns the optimal parameters based on the highest z-score.

Value

A list containing the optimal parameters:

- a1: Optimal smoothing constant or 1 for SLR.
- a2: SLR flag (1 if smoothing constant is optimal, 0 if SLR is optimal).
- k: Optimal number of top values to consider for overlap.
- R: Optimal overlap value.
- Z: Optimal z-score.
- ztable: Matrix of z-scores for all evaluated parameters.

SanityChecK *Sanity Check for Input Data and Parameters*

Description

This function performs a series of checks and initial setups for input data, metadata, and parameters, ensuring everything is correctly formatted for downstream analysis.

Usage

```
SanityChecK(
  x,
  niter = 1000,
  K = NULL,
  meta.info,
  group.name,
  formula.str,
  verbose = TRUE,
  log = TRUE
)
```

Arguments

x	A matrix-like object or a SummarizedExperiment containing the data to be analyzed.
niter	Integer. Number of bootstrap samples or resampling iterations. Default is 1000.
K	Integer. Top list size. If NULL, it will be set to a quarter of the number of rows in the data matrix. Default is NULL.
meta.info	Data frame. Metadata associated with the samples (columns of data.exp). If data.exp is a SummarizedExperiment, meta.info can be a vector of colData column names to use.
group.name	Character. Column name in meta.info that defines the groups or conditions for comparison.
formula.str	Optional character string representing the formula for the model.
verbose	Logical, indicating whether to display messages during the function's execution. Default is TRUE.
log	Logical, indicating whether the data is already log-transformed. Default is TRUE.

Details

This function checks whether the input data and metadata are in the correct format, processes metadata from a `SummarizedExperiment` object if provided, and ensures that group information is correctly specified. If no top list size (`K`) is provided, it defaults to a quarter of the number of rows in the data.

Value

A list containing:

- `meta.info`: Processed metadata.
- `data`: Processed data matrix.
- `groups`: Numeric or factor vector indicating group assignments.
- `K`: Top list size to be used in the analysis.

`UPS1.Case4` *Spectronaut and ScaffoldDIA UPS1 Spiked Dataset case 4*

Description

A `SummarizedExperiment` object containing DIA proteomics data from a UPS1-spiked E. coli proteins, processed using both Spectronaut and ScaffoldDIA separately, then merged.

Format

An instance of the `SummarizedExperiment` class with the following assays:

norm This assay includes log2 protein intensities calculated by averaging the peptides derived from the same protein

The object also contains colData and rowData:

colData A DataFrame with metadata for samples.

rowData A DataFrame with metadata for proteins.

The `colData` contains the following columns:

SampleID Unique identifier for each sample.

Conc Experimental condition or group for each sample , representing different conc. of UPS1-spiked proteins.

tool software tool used, Spectronaut or ScaffoldDIA.

fake.batch A fake digestion batch.

Source

Generated with both spectronaut and ScaffoldDIA separately. using a mixed mode acquisition method and FASTA mode for demonstration purposes.

References

Gotti, C., Roux-Dalvai, F., Joly-Beauparlant, C., Mangnier, L., Leclercq, M., & Droit, A. (2022). DIA proteomics data from a UPS1-spiked E.coli protein mixture processed with six software tools. In Data in Brief (Vol. 41, p. 107829). Elsevier BV. https://doi.org/10.1016/j.dib.2022.107829

Index