Package ‘mitch’

October 24, 2025

Title Multi-Contrast Gene Set Enrichment Analysis

Version 1.21.3

Description mitch is an R package for multi-contrast enrichment analysis. At it’s heart, it uses a rank-MANOVA based statistical approach to detect sets of genes that exhibit enrichment in the multidimensional space as compared to the background. The rank-MANOVA concept dates to work by Cox and Mann (https://doi.org/10.1186/1471-2105-13-S16-S12). mitch is useful for pathway analysis of profiling studies with one, two or more contrasts, or in studies with multiple omics profiling, for example proteomic, transcriptomic, epigenomic analysis of the same samples. mitch is perfectly suited for pathway level differential analysis of scRNA-seq data.
We have an established routine for pathway enrichment of Infinium Methylation Array data (see vignette).
The main strengths of mitch are that it can import datasets easily from many upstream tools and has advanced plotting features to visualise these enrichments.

Depends R (>= 4.4)

Suggests stringi, testthat (>= 2.1.0), HGNChelper, IlluminaHumanMethylation450kanno.ilmn12.hg19, IlluminaHumanMethylationEPICanno.ilm10b4.hg19

Imports stats, grDevices, graphics, utils, MASS, plyr, reshape2, parallel, GGally, grid, gridExtra, knitr, rmarkdown, ggplot2, gplots, beeswarm, echarts4r, kableExtra, dplyr, network

License CC BY-SA 4.0 + file LICENSE

Encoding UTF-8

URL `https://github.com/markziemann/mitch`

BugReports `https://github.com/markziemann/mitch`

LazyData false

RoxygenNote 7.3.2

biocViews GeneExpression, GeneSetEnrichment, SingleCell, Transcriptomics, Epigenetics, Proteomics, DifferentialExpression, Reactome, DNAMethylation, MethylationArray, DataImport

VignetteBuilder knitr

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

git_branch devel

git_last_commit 369a425

git_last_commit_date 2025-04-15

Repository Bioconductor 3.22

Date/Publication 2025-10-23

Author Mark Ziemann [aut, cre, cph] (ORCID:
<`https://orcid.org/0000-0002-7688-6974`>),
Antony Kaspi [aut, cph]

Maintainer Mark Ziemann <`mark.ziemann@gmail.com`>

Contents

`genesetsExample` *Reactome gene sets*

Description

Genesets from Reactome database suitable for enrichment analysis. Acquired August 2019. The structure of this data is a named list of vectors, containing human gene names (character strings). This is a sample of 200 gene sets from the approximately 2000 present in the full dataset.

Usage

```
data(genesetsExample)
```

Format

A list of gene sets

Source

Reactome website: https://reactome.org/

References

Fabregat et al. (2017) BMC Bioinformatics volume 18, Article number: 142, https://www.ncbi.nlm.nih.gov/pubmed/282

Examples

```
data(genesetsExample)
```

`gmt_import` *gmt_import*

Description

This function imports GMT files into a list of character vectors for mitch analysis. GMT files are a commonly used format for lists of genes used in pathway enrichment analysis. GMT files can be obtained from Reactome, MSigDB, etc.

Usage

```
gmt_import(gmtfile)
```

Arguments

`gmtfile`	a gmt file.

Value

a list of gene sets.

Examples

```
# Import some gene sets
genesetsExample<-gmt_import(system.file('extdata/sample_genesets.gmt',
package = 'mitch'))
```

`k36a` *H3K36ac profile*

Description

Example edgeR result of differential ChIP-seq H3K36ac. This is a dataframe which contains columns for log fold change, log counts per million, p-value and FDR adjusted p-value. These columns consist of numerical values. The row names represent human gene names. This is a sample of 1000 gene of an original dataset that contains measurements of ~30000 genes.

Usage

```
data(k36a)
```

Format

data frame

Examples

```
data(k36a)
```

`k9a` *H3K9ac profile*

Description

Example edgeR result of differential ChIP-seq H3K9ac. This is a dataframe which contains columns for log fold change, log counts per million, p-value and FDR adjusted p-value. These columns consist of numerical values. The row names represent human gene names. This is a sample of 1000 gene of an original dataset that contains measurements of ~30000 genes.

Usage

```
data(k9a)
```

Format

data frame

Examples

```
data(k9a)
```

`mitch` *mitch: An R package for multi-dimensional pathway enrichment analysis*

Description

mitch is an R package for multi-dimensional enrichment analysis. At it's heart, it uses a rank-MANOVA based statistical approach to detect sets of genes that exhibit enrichment in the multidimensional space as compared to the background. mitch is useful for pathway analysis of profiling studies with two to or more contrasts, or in studies with multiple omics profiling, for example proteomic, transcriptomic, epigenomic analysis of the same samples. mitch is perfectly suited for pathway level differential analysis of scRNA-seq data.

Details

A typical mitch workflow consists of: 1) Import gene sets with gmt_import() 2) Import profiling data with mitch_import() 3) Calculate enrichments with mitch_calc() 4) And generate plots and reports with mitch_plots() and mitch_report()

More documentation on the github page https://github.com/markziemann/mitch or with ?<function>, eg: ?mitch_import

Author(s)

Maintainer: Mark Ziemann `<mark.ziemann@gmail.com>` (ORCID) [copyright holder]

Authors:

- Antony Kaspi [copyright holder]

See Also

Useful links:

- `https://github.com/markziemann/mitch`
- Report bugs at `https://github.com/markziemann/mitch`

Examples

```
# Example workflow
# Import some gene sets
genesetsExample<-gmt_import(system.file('extdata/sample_genesets.gmt',
package = 'mitch'))
# Load some edgeR tables (rna, k9a, k36a).
data(rna,k9a,k36a)
# Create a list of differential profiles
myList<-list('rna'=rna,'k9a'=k9a,'k36a'=k36a)
# Import as edgeR table
myImportedData<-mitch_import(myList,DEtype='edger')
# Calculate enrichment using MANOVA
resExample<-mitch_calc(myImportedData,genesetsExample,priority='effect',
resrows=5,cores=2)
# Generate some high res plots in PDF format
mitch_plots(resExample,outfile='outres.pdf')
#' Generate a report of the analysis in HTML format
mitch_report(resExample,'outres.html')
```

`mitch_calc` *mitch_calc*

Description

This function performs multivariate gene set enrichment analysis.

Usage

```
mitch_calc(
  x,
  genesets,
  minsetsize = 10,
  cores = 1,
  resrows = 50,
  priority = NULL
)
```

Arguments

`x`	a multicolumn numerical table with each column containing differential expression scores for a contrast. Rownames must match genesets.
`genesets`	lists of genes imported by the gmt_imprt function or similar.
`minsetsize`	the minimum number of genes required in a set for it to be included in the statistical analysis. Default is 10.
`cores`	the number of parallel threads for computation. Defaults to 1.
`resrows`	an integer value representing the number of top genesets for which a detailed report is to be generated. Default is 50.
`priority`	the prioritisation metric used to selecting top gene sets. Valid options are 'significance', 'effect' and 'SD'.

Value

mitch res object with the following parts: $input_profile: the supplied input differential profile $input_genesets: the supplied input gene sets $ranked_profile: the differential profile after ranking $enrichment_result: the table of MANOVA/ANOVA enrichment results for each gene set $analysis_metrics: several metrics that are important to the interpretation of the results $detailed_sets: a list of dataframes containing ranks of members of prioritised gene sets.

Examples

```
# Example using mitch to calculate multivariate enrichments and
# prioritise based on effect size
data(myImportedData,genesetsExample)
resExample<-mitch_calc(myImportedData,genesetsExample,priority='effect',
minsetsize=5,cores=2)
```

`mitch_import` *mitch_import*

Description

This function imports differential omics data from common differential tools like edgeR, limma and DESeq2. It calculates a summarised differential expression metric by multiplying the sign of the log fold change by the -log10 of the p-value. If this behaviour is not desired, mitch_import can be bypassed in favour of another scoring metric.

Usage

```
mitch_import(x, DEtype, geneIDcol = NULL, geneTable = NULL, joinType = NULL)
```

Arguments

`x`	a list of differential expression tables.
`DEtype`	the program that generated the differential expression table Valid options are 'edgeR', 'DESeq2', 'limma', 'ABSSeq', 'Sleuth', 'Seurat', 'topConfects', 'muscat', 'Swish', 'scDE', 'MAST', 'DEsingle', 'ballgown', 'NOIseq', 'TCC', 'DEDS', 'cuffdiff', 'fishpond', 'missMethyl', 'DMRcate', 'DEP', 'msmsTests', 'plgem', 'SDAMS', 'DEqMS', 'DiffBind' and 'prescored'. Where 'prescored' is a dataframe

	containing the test statistic and gene ID (either in rowname or separate column) and nothing else. 'preranked' is an alias for 'prescored'.
`geneIDcol`	the column containing gene names. If gene names are specified as row names, then geneIDcol=NULL.
`geneTable`	a 2 column table mapping gene identifiers in the profile to gene identifiers in the gene sets.
`joinType`	the type of join to perform, either 'inner' or 'full'. By default, joins are 'inner' except for Seurat and muscat where full is used.

Value

a multi-column table compatible with mitch_calc analysis.

Examples

```
# first step is to create a list of differential profiles
data(rna,k9a,k36a)
x<-list('rna'=rna,'k9a'=k9a,'k36a'=k36a)
# import as edgeR table
imported<-mitch_import(x,DEtype='edger')
```

`mitch_plots` *mitch_plots*

Description

This function generates several plots of multivariate gene set enrichment in high resolution PDF format. The number of detailed sets to generate is dictated by the resrows set in the mitch_calc command.

Usage

```
mitch_plots(res, outfile = "Rplots.pdf")
```

Arguments

`res`	a mitch results object.
`outfile`	the destination file for the plots in PDF format. should contain 'pdf' suffix. Defaults to 'Rplots.pdf'

Value

generates a PDF file containing enrichment plots.

Examples

```
data(resExample)
mitch_plots(resExample,outfile='outres.pdf')
```

`mitch_report` *mitch_report*

Description

This function generates an R markdown based html report containing tables and several plots of mitch results The plots are in png format, so are not as high in resolution as compared to the PDF generated by mitch_plots function. The number of detailed sets to generate is dictated by the resrows set in the mitch_calc command.

Usage

```
mitch_report(res, outfile, overwrite = FALSE)
```

Arguments

`res`	a mitch results object.
`outfile`	the destination file for the html report. should contain 'html' suffix. Defaults to 'Rplots.pdf'
`overwrite`	should overwrite the report file if it already exists?

Value

generates a HTML file containing enrichment plots.

Examples

```
data(resExample)
mitch_report(resExample,'outres2.html')
```

`myImportedData` *myImportedData: Example imported profiles*

Description

Example of three edgeR profiles imported using mitch. The structure of this data is a dataframe where each column represents one of the following profiling datasets after scoring: RNA, H3K9ac and H3K36ac. Each row represents one gene and this dataset contains just 1000 rows to keep the example dataset small.

Usage

```
data(myImportedData)
```

Format

data frame

Examples

```
data(myImportedData)
```

`myList` *myList: A list of three edgeR results*

Description

Example edgeR results of differential RNA, H3K9ac and H3K36ac profiling. The structure of this data is a list of three dataframes. Each data frame is 1000 lines only.

Usage

```
data(myList)
```

Format

data frame

Examples

```
data(myList)
```

`networkplot` *networkplot*

Description

This function generates gene set network diagrams to complement the enrichment analysis. It produces separate charts for up and down-regulated sets. This only includes genes ranked in the top and bottom tertiles. Genes in the set which do not meet this criterion are discarded. The intensity of the colour is proportional to the s.dist value (enrichment score). Circle size is proportional to the number of genes in the set. Line thickness is proportional to the Jaccard similarity value. This function works best after prioritisation with "effect" when running mitch_calc(). Note that the circle size and the line width shown in the legend is approximate, although the values shown are exactly the smallest and largest respectively. Note that this chart works best when the width is double the height, otherwise many of the long gene set names could be cut off. There is an element of stochasticity with regard to the network projection, so it could be a good idea to repeat it a few times until you get a nice layout.

Usage

```
networkplot(eres, FDR = 0.05, n_sets = 20)
```

Arguments

`eres`	a mitch results object.
`FDR`	the significance threshold for inclusion. By default, it is 0.05
`n_sets`	the number of sets to include. This type of graph can get messy, so the default is 20, but it can be customised to your needs.

Value

generates network diagrams.

Examples

```
data(resExample)
networkplot(resExample)
```

`network_genes` *network_genes*

Description

This is a companion function for the network diagrams to display the genes which contribute to the enrichment and are shared between gene sets.

Usage

```
network_genes(eres, FDR = 0.05, n_sets = 20)
```

Arguments

`eres`	a mitch results object.
`FDR`	the significance threshold for inclusion. By default, it is 0.05
`n_sets`	the number of sets to include.

Value

a list of up- and down-regulated gene sets

Examples

```
data(resExample)
network_genes(resExample)
```

`resExample` *resExample: Example mitch result*

Description

Example of mitch results. Enrichment of the Reactome gene sets in the RNA, H3K9ac and H3K36ac datasets. The structure of this data set is a list where the 1st element is "input_profile" that has been imported (data frame), 2nd element is the "input_genesets" (names list of gene names [character vectors]), 3rd is "ranked_profile" which is the input profiling data after ranking (data frame), 4th is "enrichment_result" which is a data frame which provides enrichment information on each gene set in the profiling data including s scores, p-values and FDR adjusted p-values. 5th is "analysis_metrics" (list). The 6th slot is "detailed_sets" which is a list of 5 matrices which details the enrichment of members of selected gene sets.

Usage

```
data(resExample)
```

Format

list of mixed data types

Examples

```
data(resExample)
```

`rna` *RNA profile*

Description

Example edgeR result of differential RNA expression. This is a dataframe which contains columns for log fold change, log counts per million, p-value and FDR adjusted p-value. These columns consist of numerical values. The row names represent human gene names. This is a sample of 1000 gene of an original dataset that contains measurements of ~15000 genes.

Usage

```
data(rna)
```

Format

data frame

Examples

```
data(rna)
```

Index