

Package ‘HetSeq’

February 3, 2025

Title Identifying Modulators of Cellular Responses Leveraging Intercellular Heterogeneity

Version 0.1.0

Author Kevin Berg [aut, cre],
Florian Erhard [aut] (<<https://orcid.org/0000-0002-3574-6983>>),
Lygeri Sakellaridi [aut]

Maintainer Kevin Berg <Kevin.Berg@informatik.uni-regensburg.de>

Description Cellular responses to perturbations are highly heterogeneous and depend largely on the initial state of cells. Connecting post-perturbation cells via cellular trajectories to untreated cells (e.g. by leveraging metabolic labeling information) enables exploitation of intercellular heterogeneity as a combined knock-down and overexpression screen to identify pathway modulators, termed Heterogeneity-seq (see 'Berg et al' <[doi:10.1101/2024.10.28.620481](https://doi.org/10.1101/2024.10.28.620481)>). This package contains functions to generate cellular trajectories based on scSLAM-seq (single-cell, thiol-(SH)-linked alkylation of RNA for metabolic labelling sequencing) time courses, functions to identify pathway modulators and to visualize the results.

License Apache License (>= 2)

URL <https://github.com/erhard-lab/HetSeq>

BugReports <https://github.com/erhard-lab/HetSeq/issues>

Depends R (>= 4.1.0)

Imports cowplot, doParallel, foreach, DoubleML, e1071, igraph, ggplot2, ggrastr, ggrepel, grandR, lpSolve, mlr3, pROC, reshape2, scales, Seurat, stats

Encoding UTF-8

RoxygenNote 7.2.3

NeedsCompilation no

Repository CRAN

Date/Publication 2025-02-03 18:00:05 UTC

Contents

distmat	2
Hetseq	3
HetseqClassify	4
HetseqDoubleML	5
HetseqTest	7
mincostflow	7
PlotClassify	8
PlotDoubleML	9
prune	10
Index	11

distmat	<i>Calculate scSLAM-seq based distance matrices</i>
---------	---

Description

This function calculates distances matrices between cells of different time points based on metabolic labeling RNA profiles.

Usage

```
distmat(prev.t, next.t, prevAssay, nextAssay, gene_subset = NULL)
```

Arguments

prev.t	Cells to be used from the previous time point in distance matrix calculation.
next.t	Cells to be used from the next time point in distance matrix calculation.
prevAssay	Name of the expression assay of cells from the previous time point.
nextAssay	Name of the expression assay of cells from the next time point.
gene_subset	Set a subset of genes on which trajectories should be calculated. Other genes will be disregarded.

Value

Distance matrix between cells from two time points.

Examples

```
# Full vignette available on https://grandr.erhard-lab.de/articles/web/hetseq.html

obj.list <- SplitObject(seuratObject, split.by = "time")
D.list=list(
  distmat(treatment.list[["0h"]],treatment.list[["2h"]], "RNA", "prevRNA"),
  distmat(treatment.list[["2h"]],treatment.list[["4h"]], "RNA", "prevRNA")
)
```

)

Hetseq*Heterogeneity-seq wrapper*

Description

Wrapper function for all Heterogeneity-seq functions.

Usage

```
Hetseq(method = c("test", "classify", "DoubleML"), ...)
```

Arguments

method	The method to run Heterogeneity-seq. Calls <code>hetseq.test</code> , <code>hetseq.classify</code> or <code>hetseq.DoubleML</code> .
...	Parameters given to the chosen Hetseq method. See respective help pages.

Details

Heterogeneity-seq uses intercellular heterogeneity to identify modulators of cellular response to perturbations. Three approaches to identify these factors are currently available:

- test: Differential Gene Expression testing between cells from two response groups.
- classify: Predicting cellular outcome by expression of single genes (+informative features) reveals genes with strong predictive capabilities (high AUC) as potential pathway modulators.
- DoubleML: A strict predictor that utilizes Causal Inference (DoubleML) to distinguish causal factors from simply correlating genes. Genes with high estimated effects on the outcome and high significance are identified as potential pathway modulators.

Value

Table of Heterogeneity-seq results.

Examples

```
# Full vignette available on https://grandr.erhard-lab.de/articles/web/hetseq.html  
  
tab <- Hetseq(method="classify", data, trajectories, score.name = "score")
```

HetseqClassify	<i>Heterogeneity-seq: Classifying cellular response by gene expression values</i>
----------------	---

Description

Classifying the cellular response of control cells using single gene expression (+ informative features) to identify features with the strongest predictive capabilities.

Usage

```
HetseqClassify(
  object,
  trajectories,
  score.group = NULL,
  score.name = NULL,
  quantiles = c(0.25, 0.75),
  compareGroups = c("Low", "High"),
  posClass = NULL,
  basefeatures = NULL,
  genes = NULL,
  assay = NULL,
  split = NULL,
  kernel = "radial",
  cross = 10,
  num_cores = 1
)
```

Arguments

object	Seurat object
trajectories	Matrix of cell-cell trajectories. Columns represent time points, rows represent trajectories of connected cells over time points.
score.group	A named vector of response groups. Names represent cells, the values represent the score groups. If no score.group is set, use score.name and quantiles parameters must be set to define score groups.
score.name	The name of a numeric Seurat meta data column, which will be used to calculate score groups. Only used if no score.group is given.
quantiles	Thresholds of the score.name meta data to define 3 response groups. Low, Middle, High.
compareGroups	Which score groups to test. Default: Low vs. High
posClass	Define the positive Class for classification.
basefeatures	Additional informative features to include in the classification. Must be meta data available in the Seurat object.
genes	Vector of genes to test.

assay	The name of the Seurat assay to perform Heterogeneity-seq on. If NULL, the default assay will be used.
split	Set a training-test data split. Must be in [0,1]
kernel	The kernel for the SVM. linear, polynomial, radial or sigmoid. Default: radial.
cross	Number of cross-validations.
num_cores	The number of cores used in parallel processing.

Value

Table of log2FC and AUC values for each gene and an additional AUC value for the baseline features.

Examples

```
# Full vignette available on https://grandr.erhard-lab.de/articles/web/hetseq.html

t <- HetseqClassify(data, trajectories, score.name = "score")

t <- HetseqClassify(data, trajectories, score.group = groups,
  compareGroups = c("Weak", "Strong"))
```

HetseqDoubleML	<i>Heterogeneity-seq: Classifying cellular response by gene expression values including causal inference by DoubleML</i>
----------------	--

Description

Classifying the cellular response of control cells using single gene expression (+ informative features) to identify features with the strongest predictive capabilities and applying causal inference by a DoubleML approach.

Usage

```
HetseqDoubleML(
  object,
  trajectories,
  score.group = NULL,
  score.name = NULL,
  quantiles = c(0.25, 0.75),
  compareGroups = c("Low", "High"),
  posClass = NULL,
  basefeatures = NULL,
  genes = NULL,
  background = NULL,
  assay = NULL,
```

```

    split = NULL,
    cross = 10,
    num_cores = 1
  )

```

Arguments

object	Seurat object
trajectories	Matrix of cell-cell trajectories. Columns represent time points, rows represent trajectories of connected cells over time points.
score.group	A named vector of response groups. Names represent cells, the values represent the score groups. If no score.group is set, use score.name and quantiles parameters must be set to define score groups.
score.name	The name of a numeric Seurat meta data column, which will be used to calculate score groups. Only used if no score.group is given.
quantiles	Thresholds of the score.name meta data to define 3 response groups. Low, Middle, High.
compareGroups	Which score groups to test. Default: Low vs. High
posClass	Define the positive Class for classification.
basefeatures	Additional informative features to include in the classification. Must be meta data available in the Seurat object.
genes	Vector of genes to test.
background	A set of genes that will be considered as potential confounding factors in the DoubleML analysis. Must contain all genes set in the genes parameter. By default, all genes are used.
assay	The name of the Seurat assay to perform Heterogeneity-seq on. If NULL, the default assay will be used.
split	Set a training-test data split. Must be in [0,1]
cross	Number of cross-validations.
num_cores	The number of cores used in parallel processing.

Value

Table of log2FC and AUC values for each gene and an additional AUC value for the baseline features.

Examples

```

# Full vignette available on https://grandr.erhard-lab.de/articles/web/hetseq.html

t <- HetseqDoubleML(data, trajectories, score.name = "score")

t <- HetseqDoubleML(data, trajectories, score.group = group_vector,
  compareGroups = c("Weak", "Strong"))

```

HetseqTest	<i>Heterogeneity-seq: Testing for differential gene expression</i>
------------	--

Description

Testing for differential gene expression in predecessors of treated cells

Usage

```
HetseqTest(mat, A, B)
```

Arguments

mat	Gene expression matrix of control cells
A	Vector of cells (columns) in positive class
B	Vector of cells (columns) in negative class

Value

Table of log2FC, p-values and adjusted p-values of differentially expressed genes.

mincostflow	<i>Min-Cost-Max-Flow for cellular trajectories</i>
-------------	--

Description

Applies Min-Cost-Max-Flow to calculate optimal trajectories from distance matrices.

Usage

```
mincostflow(D.list, verbose = TRUE)
```

Arguments

D.list	List of (pruned) distance matrices. Can be generated by the distmat function.
verbose	Show verbose output.

Value

Matrix of cell-cell trajectories spanning all given timepoints.

Examples

```
# Full vignette available on https://grandr.erhard-lab.de/articles/web/hetseq.html
```

```
mincostflow(D.list)
```

Description

This plotting functions creates AUC Scatter plots to visualize Classifier Results.

Usage

```
PlotClassify(
  table,
  highlights = NULL,
  highlights.color = NULL,
  auc.cutoff = NULL,
  plot.baseline = TRUE,
  density.color = TRUE,
  density.n = 500,
  point.scale = 0.5,
  xlab = "AUC",
  ylab = bquote(log[2] ~ FC ~ (`0h`)),
  linetype = "dashed"
)
```

Arguments

<code>table</code>	Table from HetSeq using the Classifier method.
<code>highlights</code>	A vector of genes to highlight in the plot.
<code>highlights.color</code>	A vector of colors for gene highlights.
<code>auc.cutoff</code>	Inserts a vertical line at the cutoff AUC value.
<code>plot.baseline</code>	Inserts a vertical line at the baseline AUC value (= AUC of the classifier with basefeatures but no further gene info).
<code>density.color</code>	Color data points by density. Default is TRUE.
<code>density.n</code>	Set granularity of 2d density color.
<code>point.scale</code>	Set point size.
<code>xlab</code>	Set label of the x-axis.
<code>ylab</code>	Set label of the y-axis.
<code>linetype</code>	Set the linetype of the baseline AUC line.

Value

ggplot object.

Examples

```
tab <- HetseqClassify(data, trajectories, score.name = "score")
PlotClassify(tab, highlights=c("MYC", "GAPDH", "ISG15"))
```

PlotDoubleML

Plot Heterogeneity-seq DoubleML Results

Description

This plotting functions creates a Volcano Plot to visualize DoubleML Results.

Usage

```
PlotDoubleML(
  table,
  highlights = NULL,
  p.cutoff = 0.05,
  est.cutoff = NULL,
  highlights.color = NULL,
  label.repulsion = 1,
  density.color = TRUE,
  density.n = 500,
  point.scale = 0.5,
  xlab = "Estimate",
  ylab = bquote("-" ~ log[10] ~ FDR),
  linetype = "dashed"
)
```

Arguments

table	Table from the Hetseq using the doubleML method.
highlights	A vector of genes to highlight in the plot.
p.cutoff	Adds a dashed horizontal line at the given adjusted p-value cutoff.
est.cutoff	Adds two dashed vertical lines (+/-) at the given estimate cutoff.
highlights.color	A vector of colors for gene highlights.
label.repulsion	Represents the force parameter of the ggrepel::geom_label_repel() function. Higher values reduce label overlap.
density.color	Color data points by density. Default is TRUE.
density.n	Set granularity of 2d density color.
point.scale	Set point size.
xlab	Set label of the x-axis.
ylab	Set label of the y-axis.
linetype	Set the linetype of the p-value and estimate cutoff line.

Value

ggplot object.

Examples

```
tab <- HetseqDoubleML(data, trajectories, score.name = "score")
PlotDoubleML(tab, highlights=c("MYC", "GAPDH", "ISG15"))
```

prune

Prune trajectories

Description

Prune trajectories down to top.n candidates to reduce runtime of subsequent mincostflow function.

Usage

```
prune(D.list, top.n = 10)
```

Arguments

D.list	List of distance matrices. Can be generated by the distmat function.
top.n	Prune trajectories to only top.n possible connections to optimize subsequent application of the mincostflow function.

Value

Pruned distance matrix between cells from multiple timepoints.

Examples

```
# Full vignette available on https://grandr.erhard-lab.de/articles/web/hetseq.html
prune(D.list, top.n = 10)
```

Index

`distmat`, [2](#)

`Hetseq`, [3](#)

`HetseqClassify`, [4](#)

`HetseqDoubleML`, [5](#)

`HetseqTest`, [7](#)

`mincostflow`, [7](#)

`PlotClassify`, [8](#)

`PlotDoubleML`, [9](#)

`prune`, [10](#)