Package ‘anamiR’

March 22, 2017

Type Package

Title An integrated analysis package of miRNA and mRNA expression data

Description This package is intended to identify potential interactions of miRNA-target gene interactions from miRNA and mRNA expression data. It contains functions for statistical test, databases of miRNA-target gene interaction and functional analysis.

biocViews Software, AssayDomain, GeneExpression, BiologicalQuestion, GeneSetEnrichment, GeneTarget, Normalization, Pathways, DifferentialExpression, GeneRegulation, ResearchField, Genetics, Technology, Microarray, Sequencing, miRNA, WorkflowStep

License GPL-2


LazyData TRUE

Imports stats, DBI, limma, lumi, agricolae, RMySQL, DESeq2, SummarizedExperiment, gplots

Depends R (>= 3.3.1), SummarizedExperiment(>= 1.1.6)

RoxygenNote 5.0.1

Author Ti-Tai Wang [aut, cre], Tzu-Pin Lu [aut], Chien-Yueh Lee[ctb,] Eric Y. Chuang [aut]

Maintainer Ti-Tai Wang <r04945002@ntu.edu.tw>

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

R topics documented:

 anamiR .............................................................. 2
database_support .............................................. 3
differExp_continuous ........................................... 4
differExp_discrete ............................................. 5
anamiR: An integrated analysis package of miRNA and mRNA expression data.

Description

The anamiR package is used to identify miRNA-target genes interactions. The anamiR package provides a whole workflow, which contains important functions: ‘normalization’, ‘differExp_discrete’, ‘negative_cor’, ‘miR_converter’, ‘database_support’, ‘enrichment’.

normalization

The normalization function is used to normalize the expression data with one of three methods, including normal, quantile, rank.invariant.

differExp_discrete

The differExp_discrete function is used to find the differential genes or miRNAs from given expression data with one of three statistical methods, including t.test, wilcox.test.limma and DESeq. The miRNA would remain if its p-value lower than the cutoff value.

miR_converter

The miR_annotation function is used to convert the older miRNA annotation to the miRBase 21 version.

negative_cor

The negative_cor function is used to identify the possible miRNA-target gene interactions from given miRNA and mRNA expression data by calculating the correlation coefficient between each miRNA and gene. interaction would remain if its correlation coefficient is negative and lower than cutoff value.

database_support

The database_support function would search information about miRNA-target gene interactions from an integrated database, which contains 8 algorithm predicted databases and 2 experiment validated databases. Eventually return a big table, which is in data.frame format and contains extra 10 columns for those 10 databases to count if interactions were predicted or validated by these databases.
enrichment

The enrichment function is used to do the functional analysis from the output of `database_support`. Not only p-value from hypergeometric test but empirical p-value from 10000 times of permutation would be provided by this function.

**database_support**

**Intersect with databases for potential miRNA-target gene interactions**

**Description**

This function will intersect potential miRNA-target gene interactions from the input matrix, which is generated by `negative_cor` or `miR_converter`, with 8 predict databases and 2 validate databases about miRNA-target gene interactions. If the input contains hundreds of interactions, it would take a few minutes to intersect all of them.

**Usage**

database_support(cor_data, org = "hsa", Sum.cutoff = 2)

**Arguments**

- **cor_data**: matrix format generated from `negative_cor` or `miR_converter`, including miRNA, gene, correlation coefficient for column names.
- **org**: species of genes and miRNAs, only support "hsa", "mmu"
- **Sum.cutoff**: a Threshold for total hits by predict databases. This one should not be greater than 8. Default is 2.

**Value**

data.frame format. Each row represent one potential interaction. The first four columns are information about interactions: miRNA, gene symbol, Ensembl ID, gene ID, as for column 5 to 12 represent the predict databases, while column 13 to 14 are validate databases. If databases truly hit this interactions, the number in it would be 1. The column ‘Sum’ means total hits by 8 databases, and column ‘Validate’ would be TRUE if at least one validate database hit the interaction. Furthermore, ‘Fold-Change’ and ‘P-adjust’ can also be found in this output, and if the ‘de novo’ column contains 1 means that row is not supported by any databases. The column ‘evidence’ represents if the experiment for validation is strong or limited, considering http://mirtarbase.mbc.nctu.edu.tw/.

**Examples**

```r
## Use the internal dataset
data("mirna", package = "anamiR", envir = environment())
data("pheno.mirna", package = "anamiR", envir = environment())
data("mrna", package = "anamiR", envir = environment())
data("pheno.mrna", package = "anamiR", envir = environment())
## SummarizedExperiment class
require(SummarizedExperiment)
mirna_se <- SummarizedExperiment(
  assays = SimpleList(counts=mirna),
  colData = pheno.mirna)
```
differExp_continuous

Find differential expression genes or miRNAs from given expression data

Description

This function will apply linear regression model to find differential expression genes or miRNAs with continuous phenotype data, and then filter the genes or miRNAs (rows) which have bigger p-value than cutoff.

Usage

differExp_continuous(se, class, log2 = FALSE, p_value.cutoff = 0.05)

Arguments

- **se**: `SummarizedExperiment` for input format.
- **class**: string. Choose one features from all rows of phenotype data.
- **log2**: logical, if this data hasn't been log2 transformed yet, this one should be TRUE. Default is FALSE.
- **p_value.cutoff**: an numeric value indicating a threshold of p-value for every genes or miRNAs (rows). Default is 0.05.
**differExp_discrete**

**Value**

data expression data in matrix format, with sample name in columns and gene symbol or miRNA name in rows.

**See Also**

`lm` for fitting linear models.

**Examples**

```r
## Use the internal dataset
data("mirna", package = "anamiR", envir = environment())
data("pheno.mirna", package = "anamiR", envir = environment())

## SummarizedExperiment class
require(SummarizedExperiment)
mirna_se <- SummarizedExperiment(
  assays = SimpleList(counts=mirna),
  colData = pheno.mirna)

## Finding differential miRNA from miRNA expression data with lm
differExp_continuous(
  se = mirna_se, class = "Survival"
)
```

`differExp_discrete` Find differential expression genes or miRNAs from given expression data

**Description**

This function will apply one of three statistical methods, including t.test, wilcox.test and limma, to find differential expression genes or miRNAs with discrete phenotype data, and then filter the genes or miRNAs (rows) which have bigger p-value than cutoff.

**Usage**

differExp_discrete(se, class, method = c("t.test", "limma", "wilcox.test", "DESeq"), limma.trend = FALSE, t.test.var = FALSE, log2 = FALSE, p_value.cutoff = 0.05, p_adjust.method = "BH", foldchange = 0.5)

**Arguments**

- `se`: `SummarizedExperiment` for input format.
- `class`: string. Choose one features from all rows of phenotype data.
- `method`: statistical method for finding differential genes or miRNAs, including "t.test", "wilcox.test", "limma". Default is "t.test".
- `limma.trend`: logical, only matter when limma is chosen to be the method. From function `eBayes`.
enrichment

t_test.var logical, only matter when limma is chosen to be the method. Whether to treat the two variances as being equal. From function t.test

log2 logical, if this data hasn’t been log2 transformed yet, this one should be TRUE. Default is FALSE.

p_value.cutoff an numeric value indicating a threshold of p-value for every genes or miRNAs (rows). Default is 0.05.

p_adjust.method Correction method for multiple testing. (If you are using DESeq for method, this param would not affect the result) From function p.adjust. Default is "BH".

foldchange an numeric value indicating a threshold of foldchange (log2) for every genes or miRNAs (rows). Default is 0.5.

Value
data expression data in matrix format, with sample name in columns and gene symbol or miRNA name in rows.

See Also
t.test for Student’s t-Test; wilcox.test for Wilcoxon Rank Sum and Signed Rank Tests.

Examples

```r
## Use the internal dataset
data("mirna", package = "anamiR", envir = environment())
data("pheno.mirna", package = "anamiR", envir = environment())

## SummarizedExperiment class
require(SummarizedExperiment)
mirna_se <- SummarizedExperiment(
  assays = SimpleList(counts=mirna),
  colData = pheno.mirna)

## Finding differential miRNA from miRNA expression data with t.test
mirna_d <- differExp_discrete(
  se = mirna_se,
  class = "ER",
  method = "t.test"
)
```

Description

This function will do function analysis with genes from potential miRNA-target gene interactions in the input data.frame, which is generated by database_support, with total 4 kinds of pathway databases, including mouse and human two species, besides, this function will permute 5000 times for each pathway to show an empirical p_value to avoid the bias from hypergeometric p-value, this indicating that it would take a few minutes to do functional analysis.
enrichment

Usage

enrichment(data_support, org = "hsa", per_time = 5000)

Arguments

data_support  matrix format generated from database_support.
org  species of genes and miRNAs, only support "hsa", "mmu"
per_time  Times of permutation about each enriched pathways, higher times, more precise empirical p-value user can obtain, meanwhile, this function would cost more time. Default is 5000.

Value

matrix format. There are 7 columns in it, including database, term, total genes of the term, targets in the term, targets in total genes of the term (p-value).

See Also

phyper for details.

Examples

## Use the internal dataset
data("mirna", package = "anamiR", envir = environment())
data("pheno.mirna", package = "anamiR", envir = environment())
data("mrna", package = "anamiR", envir = environment())
data("pheno.mrna", package = "anamiR", envir = environment())

## SummarizedExperiment class
require(SummarizedExperiment)
mirna_se <- SummarizedExperiment(  
  assays = SimpleList(counts=mirna),  
  colData = pheno.mirna)

## SummarizedExperiment class
require(SummarizedExperiment)
mrna_se <- SummarizedExperiment(  
  assays = SimpleList(counts=mrna),  
  colData = pheno.mrna)

## Finding differential miRNA from miRNA expression data with t.test
mirna_d <- differExp_discrete(  
  se = mirna_se,  
  class = "ER",  
  method = "t.test"
)

## Finding differential mRNA from mRNA expression data with t.test
mrna_d <- differExp_discrete(  
  se = mrna_se,  
  class = "ER",  
  method = "t.test"
)

## Convert annotation to miRBse 21
heat_vis

```r
mirna_21 <- miR_converter(data = mirna_d, original_version = 17)

## Correlation
cor <- negative_cor(mrna_data = mrna_d, mirna_data = mirna_21)

## Intersect with known databases
sup <- database_support(cor_data = cor)

## Functional analysis
pat <- enrichment(data_support = sup, org = "hsa", per_time = 1000)
```

---

**heat_vis**

*Using correlation information to draw a heatmaps*

**Description**

This function would base on Fold-Change information from the output of `negative_cor`, `differExp_discrete` and show heatmaps to users. Note that if miRNA-gene interactions (row) from input are larger than 100, the label in plot would be unclear.

**Usage**

```r
heat_vis(cor_data, mrna_d, mirna_21)
```

**Arguments**

- `cor_data` matrix format generated from `negative_cor`.
- `mrna_d` differential expressed data in data.frame format, with sample name in columns and gene symbol in rows, which is generated by `differExp_discrete` or `differExp_continuous`.
- `mirna_21` differential expressed data in data.frame format, with sample name in columns and miRNA in rows, which is generated by `differExp_discrete` or `differExp_continuous`, miRNA should be miRBase 21 version now.

**Value**

Heatmap plots of miRNA and gene.

**See Also**

`heatmap.2` for plot.

**Examples**

```r
## Use the internal dataset
data("mirna", package = "anamiR", envir = environment())
data("pheno.mirna", package = "anamiR", envir = environment())
data("mrna", package = "anamiR", envir = environment())
data("pheno.mrna", package = "anamiR", envir = environment())

## SummarizedExperiment class
require(SummarizedExperiment)
```
mirna se <- SummarizedExperiment(
    assays = SimpleList(counts=mirna),
    colData = pheno.mirna)

## SummarizedExperiment class
require(SummarizedExperiment)
mrna_se <- SummarizedExperiment(
    assays = SimpleList(counts=mrna),
    colData = pheno.mrna)

## Finding differential miRNA from miRNA expression data with t.test
mirna_d <- differExp_discrete(
    se = mirna_se,
    class = "ER",
    method = "t.test"
)

## Finding differential mRNA from mRNA expression data with t.test
mrna_d <- differExp_discrete(
    se = mrna_se,
    class = "ER",
    method = "t.test"
)

## Convert annotation to miRBse 21
mirna_21 <- miR_converter(data = mirna_d, original_version = 17)

## Correlation
cor <- negative_cor(mrna_data = mrna_d, mirna_data = mirna_21)

## Draw heatmap
heat_vis(cor, mrna_d, mirna_21)

---

mirna

miRNA expression data about breast cancer

Description

This miRNA expression dataset is originally from GSE19536. To make dataset smaller, we have selected 30 samples in columns and 489 miRNAs in rows.

Usage

mirna

Format

A data frame with 489 obs (miRNAs) and 30 variables:

BC.M.014 sample name, in column
BC.M.015 sample name, in column
BC.M.017 sample name, in column
BC.M.019  sample name, in column
BC.M.023  sample name, in column
BC.M.031  sample name, in column
BC.M.053  sample name, in column
BC.M.083  sample name, in column
BC.M.088  sample name, in column
BC.M.112  sample name, in column
BC.M.119  sample name, in column
BC.M.144  sample name, in column
BC.M.148  sample name, in column
BC.M.150  sample name, in column
BC.M.209  sample name, in column
BC.M.220  sample name, in column
BC.M.221  sample name, in column
BC.M.300  sample name, in column
BC.M.308  sample name, in column
BC.M.309  sample name, in column
BC.M.318  sample name, in column
BC.M.357  sample name, in column
BC.M.381  sample name, in column
BC.M.388  sample name, in column
BC.M.406  sample name, in column
BC.M.451  sample name, in column
BC.M.457  sample name, in column
BC.M.493  sample name, in column
BC.M.512  sample name, in column
BC.M.709  sample name, in column

Value
data.frame

Source
miR_converter

Convert miRNA annotation to the miRBase 21 version

Description
This function will convert the miRNA names from the data frame, which is produced by differ-
Exp_discrete, to the miRBase 21 version of miRNA annotation. If the input contains hundreds of
miRNAs, it would take a few minutes to convert all of them.

Usage
miR_converter(data, remove_old = TRUE, original_version,
latest_version = 21)

Arguments
data      expression data in data.frame format, with sample name in columns and miRNA
name in rows.
remove_old logical value, if the miRNA is deleted in miRBase 21, should it be removed from
row? Default is TRUE.
original_version   the original version of miRNA in input matrix. This one is necessary.
latest_version    choose an integer under 21, and this function would convert miRNA annotation
to that version. Default is 21.

Value
expression data in data.frame format, with sample name in columns and miRNA name for miRBase
version 21 in rows.

Examples
## Use the internal dataset
data("mirna", package = "anamiR", envir = environment())
data("pheno.mirna", package = "anamiR", envir = environment())

## SummarizedExperiment class
require(SummarizedExperiment)
mirna_se <- SummarizedExperiment(
  assays = SimpleList(counts=mirna),
colData = pheno.mirna)

## Finding differential miRNA from miRNA expression data with t.test
mirna_d <- differExp_discrete(
  se = mirna_se,
class = "ER",
method = "t.test"
)

## Convert annotation to miRBase 21
mirna_21 <- miR_converter(data = mirna_d, original_version = 17)
mRNA expression data about breast cancer

Description

This mRNA expression dataset is originally from GSE19536. To make dataset smaller, we have selected 30 samples in columns and 19210 genes in rows.

Usage

Usage

Format

A data frame with 15000 rows (genes) and 30 variables:

- BC.M.014 sample name, in column
- BC.M.015 sample name, in column
- BC.M.017 sample name, in column
- BC.M.019 sample name, in column
- BC.M.023 sample name, in column
- BC.M.031 sample name, in column
- BC.M.053 sample name, in column
- BC.M.083 sample name, in column
- BC.M.088 sample name, in column
- BC.M.112 sample name, in column
- BC.M.119 sample name, in column
- BC.M.144 sample name, in column
- BC.M.148 sample name, in column
- BC.M.150 sample name, in column
- BC.M.209 sample name, in column
- BC.M.220 sample name, in column
- BC.M.221 sample name, in column
- BC.M.300 sample name, in column
- BC.M.308 sample name, in column
- BC.M.309 sample name, in column
- BC.M.318 sample name, in column
- BC.M.357 sample name, in column
- BC.M.381 sample name, in column
- BC.M.388 sample name, in column
- BC.M.406 sample name, in column
- BC.M.451 sample name, in column
- BC.M.457 sample name, in column
- BC.M.493 sample name, in column
- BC.M.512 sample name, in column
- BC.M.709 sample name, in column
multi_Differ

Description
This function will apply anova, a statistical method, for each gene or miRNA (row) to find not only whether expression data of multiple groups is differentially expressed or not, but also to tell specifically which two groups among all are differentially expressed.

Usage
multi_Differ(se, class, anova_p_value = 0.05, post_hoc = c("scheffe.test", "duncan.test", "HSD.test"), post_hoc_p_value = 0.05)

Arguments
- se: SummarizedExperiment for input format.
- class: string. Choose one feature from all rows of phenotype data.
- anova_p_value: an numeric value indicating a threshold of p-value from anova for every genes or miRNAs (rows). Default is 0.05.
- post_hoc: post hoc test for anova, including "scheffe.test", "HSD.test", "duncan.test".
- post_hoc_p_value: an numeric value indicating a threshold of p-value from post hoc test for every genes or miRNAs (rows). Default is 0.05.

Value
data.frame format with extra columns containing information about differential expressed groups among all.

See Also
- aov for fit an analysis of variance model.

Examples
```r
## Use the internal dataset
data("mirna", package = "anamiR", envir = environment())
data("pheno.mirna", package = "anamiR", envir = environment())

## SummarizedExperiment class
require(SummarizedExperiment)
mirna_se <- SummarizedExperiment(
```
negative_cor

Find the correlation coefficient between each gene and miRNA.

Description

This function will calculate the correlation coefficient between each gene and miRNA from differential expressed data, which are produced by differExp_discrete or differExp_continuous. After filtering the positive and higher than cutoff value of correlation, this function would return a matrix with seven columns, including miRNA, gene, correlation coefficients and Fold change, P-adjust value for miRNA and gene. Each row represents one potential miRNA-target gene interaction.

Usage

negative_cor(mrna_data, mirna_data, method = c("pearson", "kendall", "spearman"), cut.off = -0.5)

Arguments

- **mrna_data**: differential expressed data in data.frame format, with sample name in columns and gene symbol in rows, which is generated by differExp_discrete or differExp_continuous.
- **mirna_data**: differential expressed data in data.frame format, with sample name in columns and miRNAl in rows, which is generated by differExp_discrete or differExp_continuous, miRNA should be miRBase 21 version now.
- **method**: methods for calculating correlation coefficient, including "pearson", "spearman", "kendall". Default is "pearson". From function cor.
- **cut.off**: an numeric value indicating a threshold of correlation coefficient for every potential miRNA-genes interactions. Default is -0.5, however, if no interaction pass the threshold, this function would add 0.2 value in threshold until at least one interaction passed the threshold.

Value

matrix format with each row indicating one potential miRNA-target gene interaction and seven columns are miRNA, gene, correlation coefficient and Fold change, P-adjust value for miRNA and gene.

See Also

cor for calculation of correlation.
normalization

Examples

```r
## Use the internal dataset
data("mirna", package = "anamiR", envir = environment())
data("pheno.mirna", package = "anamiR", envir = environment())
data("mrna", package = "anamiR", envir = environment())
data("pheno.mrna", package = "anamiR", envir = environment())

## SummarizedExperiment class
require(SummarizedExperiment)
mirna_se <- SummarizedExperiment(
  assays = SimpleList(counts=mirna),
  colData = pheno.mirna)

## SummarizedExperiment class
require(SummarizedExperiment)
mrna_se <- SummarizedExperiment(
  assays = SimpleList(counts=mrna),
  colData = pheno.mrna)

## Finding differential miRNA from miRNA expression data with t.test
mirna_d <- differExp_discrete(
  se = mirna_se,
  class = "ER",
  method = "t.test"
)

## Finding differential mRNA from mRNA expression data with t.test
mrna_d <- differExp_discrete(
  se = mrna_se,
  class = "ER",
  method = "t.test"
)

## Correlation
cor <- negative_cor(mrna_data = mrna_d, mirna_data = mirna_d,
  method = "pearson"
)
```

description

This function will normalize the given expression data and return it in the same data format.

Usage

```r
normalization(data, method = c("quantile", "normal", "rank.invariant"))
```

Arguments

data expression data in matrix format, with sample name in columns and gene symbol or miRNA name in rows.
method: normalization methods, including "quantile", "normal", "rank.invariant". Default is "quantile". As for method "normal", we trim the extreme value and calculate the mean in the data.

Value

SummarizedExperiment for return object.

See Also

normalizeQuantiles for quantile normalization; rankinvariant for rank invariant normalization.

Examples

```r
## Use the internal dataset
data("mirna", package = "anamiR", envir = environment())

## Normalize miRNA expression data
normalization(data = mirna, method = "quantile")
```

<table>
<thead>
<tr>
<th>pheno.mirna</th>
<th>phenotype data of mirna about breast cancer</th>
</tr>
</thead>
</table>

Description

This phenotype dataset is originally from GSE19536. It contains 3 features in row and 30 samples in column.

Usage

pheno.mirna

Format

A data frame with 30 obs and 3 variables:

- **ER** estrogen receptor status
- **Subtype** breast cancer subtype
- **Survival** disease free survival time (months)

Value

data.frame

Source

Description

This phenotype dataset is originally from GSE19536. It contains 3 features in row and 30 samples in column.

Usage

pheno.mrna

Format

A data frame with 30 obs and 3 variables:

- **ER**  estrogen receptor status
- **Subtype**  breast cancer subtype
- **Survival**  disease free survival time (months)

Value

data.frame

Source

Index

Topic datasets
   mirna, 9
   mrna, 12
   pheno.mirna, 16
   pheno.mrna, 17

   anamiR, 2
   anamiR-package (anamiR), 2
   aov, 13
   cor, 14
   database_support, 3, 6, 7
   differExp_continuous, 4, 8, 14
   differExp_discrete, 5, 8, 11, 14
   eBayes, 5
   enrichment, 6
   heat_vis, 8
   heatmap.2, 8
   lm, 5
   miR_converter, 3, 11
   mirna, 9, 16
   mrna, 12, 17
   multi_Differ, 13
   negative_cor, 3, 8, 14
   normalization, 15
   normalizeQuantiles, 16
   p.adjust, 6
   pheno.mirna, 16
   pheno.mrna, 17
   phyper, 7
   rankinvariant, 16
   SummarizedExperiment, 4, 5, 13, 16
   t.test, 6
   wilcox.test, 6