Package 'miRSM'

April 16, 2019

Type Package

Title Inferring miRNA sponge modules by integrating expression data and miRNA-target binding information

Version 1.0.0

Description
The package aims to identify miRNA sponge modules by integrating expression data and miRNA-target binding information.
It provides several functions to study miRNA sponge modules, including popular methods for inferring gene modules
(candidate miRNA sponge modules), and a function to identify miRNA sponge modules, as well as a function to conduct
functional analysis of miRNA sponge modules.

Depends R (>= 3.5.0)

License GPL-3

URL https://github.com/zhangjunpeng411/miRSM

Encoding UTF-8

biocViews GeneExpression, BiomedicalInformatics, Clustering, GeneSetEnrichment, Microarray, Software, GeneRegulation, GeneTarget

RoxygenNote 6.1.0

Imports WGCNA, flashClust, dynamicTreeCut, GFA, igraph, linkcomm, MCL, NMF, biclust, runibic, iBBiG, fabia, BicARE, isa2, s4vd, BiBitR, rqubic, Biobase, PMA, stats, miRsponge, Rcpp, utils, SummarizedExperiment, GSEABase

Suggests BiocStyle, knitr, rmarkdown, testthat, org.Hs.eg.db

VignetteBuilder knitr

BugReports https://github.com/zhangjunpeng411/miRSM/issues

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

git_branch RELEASE_3_8

git_last_commit 54b5bd5

git_last_commit_date 2018-10-30

Date/Publication 2019-04-15

Author Junpeng Zhang [aut, cre]

Maintainer Junpeng Zhang <zhangjunpeng_411@yahoo.com>

R topics documented:

`ceRExp` *ceRNA expression data*

Description

ceRNA expression data

Format

ceRExp: A SummarizedExperiment object with 72 BRCA and 72 normal samples (rows) and 305 lncRNAs (columns).

Details

The matched breast invasive carcinoma (BRCA) miRNA, lncRNA and mRNA expression data is obtained from TCGA (http://cancergenome.nih.gov/). lncRNA expression data is regarded as ceRNA expression data. The data focuses on 72 individuals for which the complete sets of tumor and matched normal (i.e., normal tissue taken from the same patient) profiles are available. A lncRNA which has missing values in more than 10 are imputed using the k-nearest neighbours (KNN) algorithm from the impute R package. We use the limma R package to infer differentially expressed lncRNAs between tumour and normal samples. After the analysis, we select top 305 lncRNAs which are differentially expressed at a significant level (adjusted p-value < 1E-02, adjusted by Benjamini & Hochberg method).

cor_binary *cor_binary*

Description

Generation of positively correlated binary matrix between ceRNAs and mRNAs

Usage

```
cor_binary(ceRExp, mRExp, cor.method = "pearson",
  pos.p.value.cutoff = 0.01)
```

Arguments

ceRExp	A SummarizedExperiment object. ceRNA expression data: rows are samples and columns are ceRNAs.
mRExp	A SummarizedExperiment object. mRNA expression data: rows are samples and columns are mRNAs.
cor.method	The method of calculating correlation selected, including 'pearson' (default), 'kendall', 'spearman'.
pos.p.value.cutoff	The significant p-value cutoff of positive correlation.

Value

A binary matrix.

Author(s)

Junpeng Zhang (https://www.researchgate.net/profile/Junpeng_Zhang3)

References

Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008, 9:559.

Examples

```
data(BRCASampleData)
cor_binary_matrix <- cor_binary(ceRExp, mRExp)
```

miRExp *miRNA expression data*

Description

miRNA expression data

Format

miRExp: A SummarizedExperiment object with 72 BRCA and 72 normal samples (rows) and 226 miRNAs (columns).

Details

The matched breast invasive carcinoma (BRCA) miRNA, lncRNA and mRNA expression data is obtained from TCGA (http://cancergenome.nih.gov/). The data focuses on 72 individuals for which the complete sets of tumor and matched normal (i.e., normal tissue taken from the same patient) profiles are available. A miRNA which has missing values in more than 10 are imputed using the k-nearest neighbours (KNN) algorithm from the impute R package. We use the limma R package to infer differentially expressed miRNAs, ceRNAs and mRNAs between tumour and normal samples. After the analysis, we select top 226 miRNAs which are differentially expressed at a significant level (adjusted p-value < 1E-02, adjusted by Benjamini & Hochberg method).

miRSM *miRSM*

Description

Identify miRNA sponge modules using cannonical correlation (CC) and sensitivity cannonical correlation (SCC) methods

Usage

```
miRSM(miRExp, ceRExp, mRExp, miRTarget, CandidateModulegenes,
  typex = "standard", typez = "standard", nperms = 100,
  method = c("CC", "SCC"), num_shared_miRNAs = 3,
  pvalue.cutoff = 0.05, CC.cutoff = 0.8, SCC.cutoff = 0.3)
```

Arguments

`miRExp`	A SummarizedExperiment object. miRNA expression data: rows are samples and columns are miRNAs.
`ceRExp`	A SummarizedExperiment object. ceRNA expression data: rows are samples and columns are ceRNAs.
`mRExp`	A SummarizedExperiment object. mRNA expression data: rows are samples and columns are mRNAs.
`miRTarget`	A SummarizedExperiment object. Putative miRNA-target binding information.
`CandidateModulegenes`	List object: a list of candidate miRNA sponge modules.

typex	The columns of x unordered (type='standard') or ordered (type='ordered').
typez	The columns of z unordered (type='standard') or ordered (type='ordered').
nperms	The number of permutations.
method	The method selected to identify miRNA sponge modules, including 'CC' and 'SCC'.
num_shared_miRNAs	The number of common miRNAs shared by a group of ceRNAs and mRNAs.
pvalue.cutoff	The p-value cutoff of significant sharing of common miRNAs by a group of ceRNAs and mRNAs.
CC.cutoff	The cutoff of cannonical correlation for 'CC' method.
SCC.cutoff	The cutoff of sensitivity cannonical correlation for 'SCC' method.

Value

List object: Cannonical correlation or sensitivity cannonical correlation, and genes of miRNA sponge modules.

Author(s)

Junpeng Zhang (https://www.researchgate.net/profile/Junpeng_Zhang3)

References

Witten DM, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. Biostatistics. 2009, 10(3):515-34.

Examples

```
data(BRCASampleData)
modulegenes_igraph <- module_igraph(ceRExp[, seq_len(10)],
    mRExp[, seq_len(10)])
# Identify miRNA sponge modules using cannonical correlation (CC)
miRSM_igraph_CC <- miRSM(miRExp, ceRExp, mRExp, miRTarget,
                         modulegenes_igraph, nperms = 5,
                         method = 'CC')
```

---

miRTarget *miRNA-target ineractions*

Description

miRNA-target ineractions

Format

miRTarget: A SummarizedExperiment object with 29901 miRNA-target interactions.

Details

The miRNA-target binding information is from miRTarBase v7.0 (http://mirtarbase.mbc.nctu.edu.tw/php/index.php), and LncBase v2.0 (http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=lncbasev2/index). Among 226 miRNAs, 305 lncRNAs and 500 mRNAs which are differentially expressed, we obtain 29901 miRNA-target interactions (including miRNA-lncRNA and miRNA-mRNA interactions).

References

Hastie T, Tibshirani R, Narasimhan B, Chu G. impute: Imputation for microarray data. R package version 1.54.0. doi: 10.18129/B9.bioc.impute.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43(7):e47.

`module_biclust` *module_biclust*

Description

Identification of gene modules from matched ceRNA and mRNA expression data using a series of biclustering packages, including biclust, runibic, iBBiG, fabia, BicARE, isa2, s4vd, BiBitR and rqubic

Usage

```
module_biclust(ceRExp, mRExp, BCmethod = "fabia", num.modules = 10,
  num.ModuleceRs = 2, num.ModulemRs = 2)
```

Arguments

`ceRExp`	A SummarizedExperiment object. ceRNA expression data: rows are samples and columns are ceRNAs.
`mRExp`	A SummarizedExperiment object. mRNA expression data: rows are samples and columns are mRNAs.
`BCmethod`	Specification of the biclustering method, including 'BCBimax', 'BCCC', 'BCPlaid' (default), 'BCQuest', 'BCSpectral', 'BCXmotifs', 'BCUnibic', iBBiG', 'fabia', 'fabiap', 'fabias', 'mfsc', 'nmfdiv', 'nmfeu', 'nmfsc', 'FLOC', 'isa', 'BCs4vd', 'BCssvd', 'bibit' and 'quBicluster'.
`num.modules`	The number of modules to be identified.
`num.ModuleceRs`	The minimum number of ceRNAs in each module.
`num.ModulemRs`	The minimum number of mRNAs in each module.

Value

GeneSetCollection object: a list of module genes.

Author(s)

Junpeng Zhang (https://www.researchgate.net/profile/Junpeng_Zhang3)

References

Preli\'c A, Bleuler S, Zimmermann P, Wille A, B\'uhlmann P, Gruissem W, Hennig L, Thiele L, Zitzler E. A systematic comparison and evaluation of biclustering methods for gene expression data. Bioinformatics. 2006, 22(9):1122-9.

Cheng Y, Church GM. Biclustering of expression data. Proc Int Conf Intell Syst Mol Biol. 2000, 8:93-103.

Turner H, Bailey T, Krzanowski W. Improved biclustering of microarray data demonstrated through systematic performance tests. Comput Stat Data Anal. 2003, 48(2): 235-254.

Murali TM, Kasif S. Extracting conserved gene expression motifs from gene expression data. Pac Symp Biocomput. 2003:77-88.

Kluger Y, Basri R, Chang JT, Gerstein M. Spectral biclustering of microarray data: coclustering genes and conditions. Genome Res. 2003, 13(4):703-16.

Wang Z, Li G, Robinson RW, Huang X. UniBic: Sequential row-based biclustering algorithm for analysis of gene expression data. Sci Rep. 2016, 6:23466.

Gusenleitner D, Howe EA, Bentink S, Quackenbush J, Culhane AC. iBBiG: iterative binary bi-clustering of gene sets. Bioinformatics. 2012, 28(19):2484-92.

Hochreiter S, Bodenhofer U, Heusel M, Mayr A, Mitterecker A, Kasim A, Khamiakova T, Van Sanden S, Lin D, Talloen W, Bijnens L, G\'ohlmann HW, Shkedy Z, Clevert DA. FABIA: factor analysis for bicluster acquisition. Bioinformatics. 2010, 26(12):1520-7.

Yang J, Wang H, Wang W, Yu, PS. An improved biclustering method for analyzing gene expression. Int J Artif Intell Tools. 2005, 14(5): 771-789.

Bergmann S, Ihmels J, Barkai N. Iterative signature algorithm for the analysis of large-scale gene expression data. Phys Rev E Stat Nonlin Soft Matter Phys. 2003, 67(3 Pt 1):031902.

Sill M, Kaiser S, Benner A, Kopp-Schneider A. Robust biclustering by sparse singular value decomposition incorporating stability selection. Bioinformatics. 2011, 27(15):2089-97.

Lee M, Shen H, Huang JZ, Marron JS. Biclustering via sparse singular value decomposition. Biometrics. 2010, 66(4):1087-95.

Rodriguez-Baena DS, Perez-Pulido AJ, Aguilar-Ruiz JS. A biclustering algorithm for extracting bit-patterns from binary datasets. Bioinformatics. 2011, 27(19):2738-45.

Li G, Ma Q, Tang H, Paterson AH, Xu Y. QUBIC: a qualitative biclustering algorithm for analyses of gene expression data. Nucleic Acids Res. 2009, 37(15):e101.

Examples

```
data(BRCASampleData)
modulegenes_biclust <- module_biclust(ceRExp[, seq_len(30)],
    mRExp[, seq_len(30)])
```

`module_FA` *module_FA*

Description

Functional analysis of miRNA sponge modules, including functional enrichment and disease enrichment analysis

Usage

```
module_FA(Modulelist, GOont = "BP", Diseaseont = "DO",
  KEGGorganism = "hsa", Reactomeorganism = "human",
  OrgDb = "org.Hs.eg.db", padjustvaluecutoff = 0.05,
  padjustedmethod = "BH", Analysis.type = c("FEA", "DEA"))
```

Arguments

`Modulelist`	List object: a list of miRNA sponge modules.
`GOont`	One of 'MF', 'BP', and 'CC' subontologies.
`Diseaseont`	One of 'DO', and 'DOLite' subontologies.
`KEGGorganism`	Organism, supported organism listed in http://www.genome.jp/kegg/catalog/org_list.html.
`Reactomeorganism`	Organism, one of 'human', 'rat', ' mouse', 'celegans', 'yeast', 'zebrafish', 'fly'.
`OrgDb`	OrgDb
`padjustvaluecutoff`	A cutoff value of adjusted p-values.
`padjustedmethod`	Adjusted method of p-values, can select one of 'holm', 'hochberg', 'hommel', 'bonferroni', 'BH', 'BY', 'fdr', 'none'.
`Analysis.type`	The type of functional analysis selected, including 'FEA' (functional enrichment analysis) and 'DEA' (disease enrichment analysis).

Value

List object: a list of enrichment analysis results.

Author(s)

Junpeng Zhang (https://www.researchgate.net/profile/Junpeng_Zhang3)

References

Zhang J (2017). miRsponge: Identification and analysis of miRNA sponge interaction networks and modules. R package version 1.2.0, (https://github.com/zhangjunpeng411/miRsponge).

Examples

```
## Not run:
data(BRCASampleData)
modulegenes_WGCNA <- module_WGCNA(ceRExp, mRExp)
# Identify miRNA sponge modules using cannonical correlation (CC)
miRSM_WGCNA_CC <- miRSM(miRExp, ceRExp, mRExp, miRTarget,
                        modulegenes_WGCNA, nperms = 10, method = 'CC')
miRSM_WGCNA_CC_genes <- miRSM_WGCNA_CC[[2]]
miRSM_WGCNA_CC_FEA <- module_FA(miRSM_WGCNA_CC_genes, Analysis.type = 'FEA')
miRSM_WGCNA_CC_DEA <- module_FA(miRSM_WGCNA_CC_genes, Analysis.type = 'DEA')

## End(Not run)
```

module_GFA *module_GFA*

Description

Identification of gene modules from matched ceRNA and mRNA expression data using GFA package

Usage

```
module_GFA(ceRExp, mRExp, StrengthCut = 0.9, iter.max = 5000,
  num.ModuleceRs = 2, num.ModulemRs = 2)
```

Arguments

ceRExp	A SummarizedExperiment object. ceRNA expression data: rows are samples and columns are ceRNAs.
mRExp	A SummarizedExperiment object. mRNA expression data: rows are samples and columns are mRNAs.
StrengthCut	Desired minimum strength (absolute value of association with interval [0 1]) for each bicluster.
iter.max	The total number of Gibbs sampling steps (default 1000).
num.ModuleceRs	The minimum number of ceRNAs in each module.
num.ModulemRs	The minimum number of mRNAs in each module.

Value

GeneSetCollection object: a list of module genes.

Author(s)

Junpeng Zhang (https://www.researchgate.net/profile/Junpeng_Zhang3)

References

Bunte K, Lepp\'aaho E, Saarinen I, Kaski S. Sparse group factor analysis for biclustering of multiple data sources. Bioinformatics. 2016, 32(16):2457-63.

Lepp\'aaho E, Ammad-ud-din M, Kaski S. GFA: exploratory analysis of multiple data sources with group factor analysis. J Mach Learn Res. 2017, 18(39):1-5.

Examples

```
data(BRCASampleData)
modulegenes_GFA <- module_GFA(ceRExp[seq_len(20), seq_len(15)],
    mRExp[seq_len(20), seq_len(15)], iter.max = 2600)
```

module_igraph *module_igraph*

Description

Identification of gene modules from matched ceRNA and mRNA expression data using igraph package

Usage

```
module_igraph(ceRExp, mRExp, cor.method = "pearson",
  pos.p.value.cutoff = 0.01, cluster.method = "greedy",
  num.ModuleceRs = 2, num.ModulemRs = 2)
```

Arguments

ceRExp	A SummarizedExperiment object. ceRNA expression data: rows are samples and columns are ceRNAs.
mRExp	A SummarizedExperiment object. mRNA expression data: rows are samples and columns are mRNAs.
cor.method	The method of calculating correlation selected, including 'pearson' (default), 'kendall', 'spearman'.
pos.p.value.cutoff	The significant p-value cutoff of positive correlation.
cluster.method	The clustering method selected in **igraph** package, including 'betweenness', 'greedy' (default), 'infomap', 'prop', 'eigen', 'louvain', 'walktrap'.
num.ModuleceRs	The minimum number of ceRNAs in each module.
num.ModulemRs	The minimum number of mRNAs in each module.

Value

GeneSetCollection object: a list of module genes.

Author(s)

Junpeng Zhang (https://www.researchgate.net/profile/Junpeng_Zhang3)

References

Csardi G, Nepusz T. The igraph software package for complex network research, InterJournal, Complex Systems. 2006:1695.

Examples

```
data(BRCASampleData)
modulegenes_igraph <- module_igraph(ceRExp[, seq_len(10)],
    mRExp[, seq_len(10)])
```

module_NMF *module_NMF*

Description

Identification of gene modules from matched ceRNA and mRNA expression data using NMF package

Usage

```
module_NMF(ceRExp, mRExp, NMF.algorithm = "brunet", num.modules = 10,
  num.ModuleceRs = 2, num.ModulemRs = 2)
```

Arguments

ceRExp	A SummarizedExperiment object. ceRNA expression data: rows are samples and columns are ceRNAs.
mRExp	A SummarizedExperiment object. mRNA expression data: rows are samples and columns are mRNAs.
NMF.algorithm	Specification of the NMF algorithm, including 'brunet' (default), 'Frobenius', 'KL', 'lee', 'nsNMF', 'offset', 'siNMF', 'snmf/l', 'snmf/r'.
num.modules	The number of modules to be identified.
num.ModuleceRs	The minimum number of ceRNAs in each module.
num.ModulemRs	The minimum number of mRNAs in each module.

Value

GeneSetCollection object: a list of module genes.

Author(s)

Junpeng Zhang (https://www.researchgate.net/profile/Junpeng_Zhang3)

References

Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics. 2010, 11:367.

Examples

```
data(BRCASampleData)
# Reimport NMF package to avoid conflicts with DelayedArray package
library(NMF)
modulegenes_NMF <- module_NMF(ceRExp[, seq_len(10)],
    mRExp[, seq_len(10)])
```

module_ProNet *module_ProNet*

Description

Identification of gene modules from matched ceRNA and mRNA expression data using ProNet package

Usage

```
module_ProNet(ceRExp, mRExp, cor.method = "pearson",
  pos.p.value.cutoff = 0.01, cluster.method = "MCL",
  num.ModuleceRs = 2, num.ModulemRs = 2)
```

Arguments

ceRExp	A SummarizedExperiment object. ceRNA expression data: rows are samples and columns are ceRNAs.
mRExp	A SummarizedExperiment object. mRNA expression data: rows are samples and columns are mRNAs.
cor.method	The method of calculating correlation selected, including 'pearson' (default), 'kendall', 'spearman'.
pos.p.value.cutoff	The significant p-value cutoff of positive correlation
cluster.method	The clustering method selected in **ProNet** package, including 'FN', 'MCL' (default), 'LINKCOMM', 'MCODE'.
num.ModuleceRs	The minimum number of ceRNAs in each module.
num.ModulemRs	The minimum number of mRNAs in each module.

Value

GeneSetCollection object: a list of module genes.

Author(s)

Junpeng Zhang (https://www.researchgate.net/profile/Junpeng_Zhang3)

References

Clauset A, Newman ME, Moore C. Finding community structure in very large networks. Phys Rev E Stat Nonlin Soft Matter Phys., 2004, 70(6 Pt 2):066111.

Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res., 2002, 30(7):1575-84.

Kalinka AT, Tomancak P. linkcomm: an R package for the generation, visualization, and analysis of link communities in networks of arbitrary size and type. Bioinformatics, 2011, 27(14):2011-2.

Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics, 2003, 4:2.

Examples

```
data(BRCASampleData)
modulegenes_ProNet <- module_ProNet(ceRExp[, seq_len(10)],
    mRExp[, seq_len(10)])
```

`module_WGCNA` *module_WGCNA*

Description

Identification of co-expressed gene modules from matched ceRNA and mRNA expression data using WGCNA package

Usage

```
module_WGCNA(ceRExp, mRExp, RsquaredCut = 0.9, num.ModuleceRs = 2,
  num.ModulemRs = 2)
```

Arguments

`ceRExp`	A SummarizedExperiment object. ceRNA expression data: rows are samples and columns are ceRNAs.
`mRExp`	A SummarizedExperiment object. mRNA expression data: rows are samples and columns are mRNAs.
`RsquaredCut`	Desired minimum scale free topology fitting index R^2 with interval [0 1].
`num.ModuleceRs`	The minimum number of ceRNAs in each module.
`num.ModulemRs`	The minimum number of mRNAs in each module.

Value

GeneSetCollection object: a list of module genes.

Author(s)

Junpeng Zhang (https://www.researchgate.net/profile/Junpeng_Zhang3)

References

Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008, 9:559.#'

Examples

```
data(BRCASampleData)
modulegenes_WGCNA <- module_WGCNA(ceRExp[, seq_len(80)],
    mRExp[, seq_len(80)])
```

mRExp *mRNA expression data*

Description

mRNA expression data

Format

mRExp: A SummarizedExperiment object with 72 BRCA and 72 normal samples (rows) and 226 miRNAs (columns).

Details

The matched breast invasive carcinoma (BRCA) miRNA, lncRNA and mRNA expression data is obtained from TCGA (http://cancergenome.nih.gov/). The data focuses on 72 individuals for which the complete sets of tumor and matched normal (i.e., normal tissue taken from the same patient) profiles are available. A mRNA which has missing values in more than 10 are imputed using the k-nearest neighbours (KNN) algorithm from the impute R package. We use the limma R package to infer differentially expressed mRNAs between tumour and normal samples. After the analysis, we select top 500 mRNAs which are differentially expressed at a significant level (adjusted p-value < 1E-02, adjusted by Benjamini & Hochberg method).

Index