Package ‘diffcoexp’

May 29, 2024

Title Differential Co-expression Analysis
Version 1.24.0
Date 2023-08-09
Author Wenbin Wei, Sandeep Amberkar, Winston Hide
Description A tool for the identification of differentially coexpressed links (DCLs) and differentially coexpressed genes (DCGs). DCLs are gene pairs with significantly different correlation coefficients under two conditions. DCGs are genes with significantly more DCLs than by chance.
Maintainer Wenbin Wei <wenbin.wei2@durham.ac.uk>
Depends R (>= 3.5), WGCNA, SummarizedExperiment
Imports stats, DiffCorr, psych, igraph, BiocGenerics
Suggests GEOquery, RUnit
URL https://github.com/hidelab/diffcoexp
biocViews GeneExpression, DifferentialExpression, Transcription,
  Microarray, OneChannel, TwoChannel, RNASeq, Sequencing,
  Coverage, ImmunoOncology
License GPL (>=2)
Encoding UTF-8
LazyData true
git_url https://git.bioconductor.org/packages/diffcoexp
git_branch RELEASE_3_19
git_last_commit 60ff62d
git_last_commit_date 2024-04-30
Repository Bioconductor 3.19
Date/Publication 2024-05-29

Contents

coexpr ................................................................. 2
comparecor ......................................................... 3
Identification of gene pairs coexpressed in at least one of two conditions

This function identifies gene pairs coexpressed in at least one of two conditions.

Usage

```r
coexpr(exprs.1, exprs.2, r.method = c("pearson", "spearman")[1],
q.method = c("BH", "holm", "hochberg", "hommel", "bonferroni", "BY", "fdr",
"none")[1], rth = 0.5, qth = 0.1)
```

Arguments

- `exprs.1`: a SummarizedExperiment, data frame or matrix for condition 1, with gene IDs as rownames and sample IDs as column names.
- `exprs.2`: a SummarizedExperiment, data frame or matrix for condition 2, with gene IDs as rownames and sample IDs as column names.
- `r.method`: a character string specifying the method to be used to calculate correlation coefficients. It is passed to the `cor` function of the WGCNA package.
- `q.method`: a character string specifying the method for adjusting p values. It is passed to the `p.adjust` function of the stats package.
- `rth`: the cutoff of absolute value of correlation coefficients; must be within [0,1].
- `qth`: the cutoff of q-value (adjusted p value); must be within [0,1].

Value

A data frame containing gene pairs that are coexpressed in at least one of the conditions with the criteria that absolute value of correlation coefficient is greater than `rth` and q value less than `qth`. It has the following columns:

- `Gene.1`: Gene ID
- `Gene.2`: Gene ID
- `cor.1`: correlation coefficients under condition 1
- `cor.2`: correlation coefficients under condition 2
- `cor.diff`: difference between correlation coefficients under condition 2 and condition 1
- `p.1`: p value under null hypothesis that correlation coefficient under condition 1 equals to zero
**comparecor**

- **p.2**  
  p value under null hypothesis that correlation coefficient under condition 2 equals to zero

- **p.diffcor**  
  p value under null hypothesis that difference between two correlation coefficients under two conditions equals to zero using Fisher’s r-to-Z transformation

- **q.1**  
  adjusted p value under null hypothesis that correlation coefficient under condition 1 equals to zero

- **q.2**  
  adjusted p value under null hypothesis that correlation coefficient under condition 2 equals to zero

- **q.diffcor**  
  adjusted p value under null hypothesis that the difference between two correlation coefficients under two conditions equals to zero using Fisher’s r-to-Z transformation

**Examples**

```r
data(gse4158part)
allowWGCNAThreads()
res=coexpr(exprs.1 = exprs.1, exprs.2 = exprs.2, r.method = "spearman")
#The result is a data frames.
str(res)
```

---

**comparecor**  
Compare gene-gene correlation coefficients under two conditions

**Description**

This function calculates correlation coefficients of all gene pairs under two conditions and compare them using Fisher’s Z-transformation.

**Usage**

```r
comparecor(exprs.1, exprs.2, r.method = c("pearson", "spearman")[1],
q.method = c("BH", "holm", "hochberg", "hommel", "bonferroni", "BY", "fdr",
"none")[1])
```

**Arguments**

- **exprs.1**  
  a SummarizedExperiment, data frame or matrix for condition 1, with gene IDs as rownames and sample IDs as column names.

- **exprs.2**  
  a SummarizedExperiment, data frame or matrix for condition 2, with gene IDs as rownames and sample IDs as column names.

- **r.method**  
  a character string specifying the method to be used to calculate correlation coefficients. It is passed to the cor function of the WGCNA package.

- **q.method**  
  a character string specifying the method for adjusting p values. It is passed to the p.adjust function of the stats package.
diffcoexp

Value

a data frame containing the differences between the correlation coefficients under two conditions and their p values. It has the following columns:

<table>
<thead>
<tr>
<th>Gene.1</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene.2</td>
<td>Gene ID</td>
</tr>
<tr>
<td>cor.1</td>
<td>correlation coefficients under condition 1</td>
</tr>
<tr>
<td>cor.2</td>
<td>correlation coefficients under condition 2</td>
</tr>
<tr>
<td>cor.diff</td>
<td>difference between correlation coefficients under condition 2 and condition 1</td>
</tr>
<tr>
<td>p.1</td>
<td>p value under null hypothesis that correlation coefficient under condition 1 equals to zero</td>
</tr>
<tr>
<td>p.2</td>
<td>p value under null hypothesis that correlation coefficient under condition 2 equals to zero</td>
</tr>
<tr>
<td>p.diffcor</td>
<td>p value under null hypothesis that difference between two correlation coefficients under two conditions equals to zero using Fisher’s r-to-Z transformation</td>
</tr>
<tr>
<td>q.1</td>
<td>adjusted p value under null hypothesis that correlation coefficient under condition 1 equals to zero</td>
</tr>
<tr>
<td>q.2</td>
<td>adjusted p value under null hypothesis that correlation coefficient under condition 2 equals to zero</td>
</tr>
<tr>
<td>q.diffcor</td>
<td>adjusted p value under null hypothesis that the difference between two correlation coefficients under two conditions equals to zero using Fisher’s r-to-Z transformation</td>
</tr>
</tbody>
</table>

Examples

data(gse4158part)
allowWGCNAThreads()
res=comparecor(exprs.1 = exprs.1, exprs.2 = exprs.2, r.method = "spearman")
#The result is a data frames.
str(res)

diffcoexp

Differential co-expression analysis

Description

This function identifies differentially coexpressed links (DCLs) and differentially coexpressed genes (DCGs).

Usage

diffcoexp(exprs.1, exprs.2, r.method = c("pearson", "kendall", "spearman")[1],
q.method = c("BH", "holm", "hochberg", "hommel", "bonferroni", "BY", "fdr",
"none")[1], rth = 0.5, qth = 0.1, r.diffth = 0.5, q.diffth = 0.1,
q.dcgth = 0.1)
**Arguments**

- **exprs.1**: a SummarizedExperiment, data frame or matrix for condition 1, with gene IDs as rownames and sample IDs as column names.
- **exprs.2**: a SummarizedExperiment, data frame or matrix for condition 2, with gene IDs as rownames and sample IDs as column names.
- **r.method**: a character string specifying the method to be used to calculate correlation coefficients. It is passed to the cor function of the WGCNA package.
- **q.method**: a character string specifying the method for adjusting p values. It is passed to the p.adjust function of the stats package.
- **rth**: the cutoff of absolute value of correlation coefficients; must be within \([0,1]\).
- **qth**: the cutoff of q-value (adjusted p value); must be within \([0,1]\).
- **r.diffth**: the cutoff of absolute value of the difference between the correlation coefficients of the two conditions; must be within \([0,1]\).
- **q.diffth**: the cutoff of q-value (adjusted p value) of the difference between the correlation coefficients of the two conditions; must be within \([0,1]\).
- **q.dcgth**: the cutoff of q-value (adjusted p value) of the genes enriched in the differentially correlated gene pairs between the two conditions; must be within \([0,1]\).

**Details**

diffcoexp function identifies differentially coexpressed links (DCLs) and differentially coexpressed genes (DCGs). DCLs are gene pairs with significantly different correlation coefficients under two conditions (de la Fuente 2010, Jiang et al., 2016). DCGs are genes with significantly more DCLs than by chance (Yu et al., 2011, Jiang et al., 2016). It takes two gene expression matrices or data frames under two conditions as input, calculates gene-gene correlations under two conditions and compare them with Fisher’s Z transformation, filter the correlation with the rth and qth and the correlation changes with r.diffth and q.diffth. It identifies DCGs using binomial probability model (Jiang et al., 2016).

The main steps are as follows:

- **a)**. Correlation coefficients and p values of all gene pairs under two conditions are calculated.
- **b)**. The difference between the correlation coefficients under two conditions are calculated and the p value is calculated using Fisher’s Z-transformation.
- **c)**. p values are adjusted.
- **d)**. Gene pairs (links) coexpressed in at least one condition are identified using the criteria that at least one of the correlation coefficients under two conditions has absolute value greater than the threshold rth and adjusted p value less than the threshold qth. The links that meet the criteria are included in CLs.
- **e)**. Differentially coexpressed gene pairs (links) are identified from CLs using the criteria that the absolute value of the difference between the two correlation coefficients is greater the threshold r.diffth and adjusted p value is less than the threshold q.diffth. The links that meet the criteria are included in DCLs.
- **f)**. The DCLs are classified into three categories: "same signed", "diff signed", or "switched opposites". "same signed" indicates that the gene pair has same signed correlation coefficients under both
conditions. "diff signed" indicates that the gene pair has oppositely signed correlation coefficients under two conditions and only one of them meets the criteria that absolute correlation coefficient is greater than the threshold \( r_{th} \) and adjusted \( p \) value less than the threshold \( q_{th} \). "switched opposites" indicates that the gene pair has oppositely signed correlation coefficients under two conditions and both of them meet the criteria that absolute correlation coefficient is greater than the threshold \( r_{th} \) and adjusted \( p \) value less than the threshold \( q_{th} \).

g). All the genes in DCLs are tested for their enrichment of DCLs, i.e., whether they have more DCLs than by chance using binomial probability model (Jiang et al., 2016). Those with adjusted \( p \) value less than the threshold \( q_{dcgth} \) are included in DCGs.

Value

a list of two data frames.

The DCGs data frame contains genes that contribute to differentially correlated links (gene pairs) with \( q \) value less than \( q_{dcgth} \). It has the following columns:

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLs</td>
<td>Number of links with absolute correlation coefficient greater than ( r_{th} ) and ( q ) value less than ( q_{th} ) in at least one condition</td>
</tr>
<tr>
<td>DCLs</td>
<td>Number of links that meet the criteria for CLs and the criteria that absolute difference between the correlation coefficients of the two condition is greater than ( r_{diffth} ) and ( q ) value less than ( q_{diffth} )</td>
</tr>
<tr>
<td>DCL.same</td>
<td>Number of subset of DCLs with same signed correlation coefficients in both conditions</td>
</tr>
<tr>
<td>DCL.diff</td>
<td>Number of subset of DCLs with oppositely signed correlation coefficients under two conditions but only one of them has absolute correlation coefficient greater than ( r_{th} ) and ( q ) value less than ( q_{th} )</td>
</tr>
<tr>
<td>DCL.switch</td>
<td>Number of subset of DCLs with oppositely signed correlation coefficients under two conditions and both of them have absolute correlation coefficient greater than ( r_{th} ) and ( q ) value less than ( q_{th} )</td>
</tr>
<tr>
<td>p</td>
<td>( p ) value of having &gt;=DCLs given CLs</td>
</tr>
<tr>
<td>q</td>
<td>adjusted ( p ) value</td>
</tr>
</tbody>
</table>

The DCLs data frame contains the differentially correlated links (gene pairs) that meet the criteria that at least one of their correlation coefficients (cor.1 and/or cor.2) is greater than \( r_{th} \) with \( q \) value (\( q_{1} \) and/or \( q_{2} \)) less than \( q_{th} \) and the absolute value of the difference between the correlation coefficients under two conditions (cor.diff) is greater than \( r_{diffth} \) with \( q_{diffcor} \) less than \( q_{diffth} \). It has the following columns:

<table>
<thead>
<tr>
<th>Gene.1</th>
<th>Gene ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene.2</td>
<td>Gene ID</td>
</tr>
<tr>
<td>cor.1</td>
<td>correlation coefficients under condition 1</td>
</tr>
<tr>
<td>cor.2</td>
<td>correlation coefficients under condition 2</td>
</tr>
<tr>
<td>cor.diff</td>
<td>difference between correlation coefficients under condition 2 and condition 1</td>
</tr>
<tr>
<td>p.1</td>
<td>( p ) value under null hypothesis that correlation coefficient under condition 1 equals to zero</td>
</tr>
</tbody>
</table>
p.2  p value under null hypothesis that correlation coefficient under condition 2 equals to zero

p.diffcor  p value under null hypothesis that difference between two correlation coefficients under two conditions equals to zero using Fisher's r-to-Z transformation

q.1  adjusted p value under null hypothesis that correlation coefficient under condition 1 equals to zero

q.2  adjusted p value under null hypothesis that correlation coefficient under condition 2 equals to zero

q.diffcor  adjusted p value under null hypothesis that the difference between two correlation coefficients under two conditions equals to zero using Fisher's r-to-Z transformation

type  can have value "same signed", "diff signed", or "switched opposites". "same signed" indicates that the gene pair has same signed correlation coefficients under both conditions. "diff signed" indicates that the gene pair has oppositely signed correlation coefficients under two conditions and only one of them meets the criteria that absolute correlation coefficient is greater than rth and q value less than qth. "switched opposites" indicates that the gene pair has oppositely signed correlation coefficients under two conditions and both of them meet the criteria that absolute correlation coefficient is greater than rth and q value less than qth.

Author(s)

Wenbin Wei

References


Examples

data(gse4158part)
allowWGCNAThreads()
res=diffcoexp(exprs.1 = exprs.1, exprs.2 = exprs.2, r.method = "spearman")
#The results are a list of two data frames, one for differentially co-expressed links (DCLs, gene pairs) and one for differentially co-expressed genes (DCGs).
str(res)
<table>
<thead>
<tr>
<th>exprs.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
</tr>
<tr>
<td>Usage</td>
</tr>
<tr>
<td>exprs.1</td>
</tr>
<tr>
<td>Format</td>
</tr>
<tr>
<td>A matrix with 400 genes and 14 samples.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>exprs.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
</tr>
<tr>
<td>expression of 400 genes in 14 samples (GSM94988 to GSM95001) of yeast after pulses 0.2 g/l glucose, <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE4158">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE4158</a>.</td>
</tr>
<tr>
<td>Usage</td>
</tr>
<tr>
<td>exprs.2</td>
</tr>
<tr>
<td>Format</td>
</tr>
<tr>
<td>A matrix with 400 genes and 12 samples.</td>
</tr>
</tbody>
</table>
Index

* **coexpression**
  - coexpr, 2
  - comparecor, 3
  - diffcoexp, 4

* **datasets**
  - exprs.1, 8
  - exprs.2, 8

  coexpr, 2
  comparecor, 3
  diffcoexp, 4

  exprs.1, 8
  exprs.2, 8