About *diffcoexp*

Wenbin Wei, Sandeep Amberkar, Winston Hide

Sheffield Institute of Translational Neuroscience, University of Sheffield, Sheffield, United Kingdom

November 23, 2020

Contents

1 Description **1**

2 Installation and removal **2**

3 Input and output of *diffcoexp* function **3**

4 Analysis and interpretation of DCGs and DCLs **3**

5 Example **3**

6 References **4**

1 Description

This package 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 compares them with Fisher's Z transformation(Fisher 1915 and Fisher 1921). It filters gene pairs with the thresholds for correlation coefficients and their adjusted p value as well as the thresholds for the difference between the two correlation coefficients and its adjusted p value. 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 differences between the correlation coefficients under two conditions are calculated and their p values are 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 co-expressed links(CLs).

e). Differentially coexpressed links (gene pairs) are identified from CLs using the criteria that the absolute value of the difference between the two correlation coefficients is greater than the threshold *r.diffth* and adjusted p value is less than the threshold *q.diffth*. The links that meet the criteria are included in differentially coexpressed links (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 *rth* and adjusted p value less than the threshold *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 the threshold *rth* and adjusted p value less than the threshold *qth*.

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.

2 Installation and removal

This package is available from Bioconductor and can be installed within R as follows:

```
## try http:// if https:// URLs are not supported
if (!requireNamespace("BiocManager", quietly=TRUE))
install.packages("BiocManager")
BiocManager::install("diffcoexp")
```

To install this package from GitHub, start R and enter:

```
library(devtools)
devtools::install_git("git://github.com/hidelab/diffcoexp.git",
branch = "master")
```

The above method does not build and install vignette. To install the package with vignette, enter the following from command line:

```
git clone https://github.com/hidelab/diffcoexp.git
R CMD build diffcoexp
R CMD check diffcoexp_1.11.1.tar.gz
```

R CMD INSTALL diffcoexp_1.11.1.tar.gz

To remove this package, start R and enter:

remove.packages("diffcoexp")

3 Input and output of *diffcoexp* function

The main function of this package is *diffcoexp* function. The first two arguments, *exprs.1* and *exprs.2*, are normalized gene expression data under two conditions with rows as genes and columns as samples. They should be objects of classes *SummarizedExperiment*, *data.frame* or *matrix*. Both should have the same number of genes in the same order. The third argument *r.method* is passed to the *cor* function of the *WGCNA* package as argument *method*, details of which can be found by typing

```
> help(cor, WGCNA)
```

The fourth argument *q.method* is passed to the *p.adjust* function of the *stats* package as argument *method*, details of which can be found by typing

```
> help(p.adjust, stats)
```

Details of other arguments of *diffcoexp* function can be found by typing

```
> help(diffcoexp, diffcoexp)
```

The output of *diffcoexp* function is a list of two data frames, one for differentially co-expressed links (DCLs), the other for differentially co-expressed genes (DCGs). Further details of the output can be seen on the help page.

4 Analysis and interpretation of DCGs and DCLs

DCGs are a list of genes and therefore can be further analysed using other tools such as FGNet (https://bioconductor.org/packages/release/bioc/html/FGNet.html), clusterProfiler (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) and enrichr (http://amp.pharm.mssm.edu/Enrichr/). DCLs are a list of differentially co-expressed gene pairs and can be assembled into a differential co-expression network. The network is scale-free but not smallworld (Hsu et al., 2017). The network can be visualized and analyzed using igraph (https://cran.r-project.org/web/packages/igraph/index.html). DCLs can also be further analyzed to identify upstream causal regulators using other tools such as DCGL v2.0 (Yang et al., 2013).

5 Example

This example illustrates the identification of differentially coexpressed links (gene pairs) and differentially coexpressed genes of yeast after pulses of 2 g/l and

0.2 g/l glucose separately. The data were downloaded from GEO (GSE4158). Only 400 genes were analysed in this example. Analysis of all the genes (6104) will take about 20 minutes on a computer with 8 cores and 16GB RAM.

```
> library(diffcoexp)
> data(gse4158part)

> allowWGCNAThreads()

Allowing multi-threading with up to 72 threads.

> 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), the other for differentially co-expressed genes (DCGs).

```
> str(res)

List of 2
 $ DCGs:'data.frame':        15 obs. of  8 variables:
  ..$ Gene      : chr [1:15] "YBR010W" "YCLX04W" "YBR110W" "YBL107W-A" ...
  ..$ CLs       : num [1:15] 100 66 102 47 148 33 37 41 70 78 ...
  ..$ DCLs      : num [1:15] 40 17 20 12 17 7 7 7 9 9 ...
  ..$ DCL.same  : num [1:15] 0 0 0 0 0 0 0 0 1 0 ...
  ..$ DCL.diff  : num [1:15] 32 17 20 9 17 7 7 7 8 6 ...
  ..$ DCL.switch: num [1:15] 8 0 0 3 0 0 0 0 0 3 ...
  ..$ p         : num [1:15] 1.07e-35 2.10e-12 5.97e-12 4.06e-09 8.29e-07 ...
  ..$ q         : num [1:15] 4.28e-33 4.19e-10 7.96e-10 4.06e-07 6.63e-05 ...
 $ DCLs:'data.frame':        363 obs. of  12 variables:
  ..$ Gene.1   : chr [1:363] "YBR290W" "YCR016W" "YCR005C" "YCL064C" ...
  ..$ Gene.2   : chr [1:363] "YAL019W" "YAL055W" "YAR031W" "YBL009W" ...
  ..$ cor.1    : num [1:363] 0.9473 -0.956 0.2352 0.0857 0.9912 ...
  ..$ cor.2    : num [1:363] 0.21 -0.287 0.979 0.937 0.49 ...
  ..$ cor.diff : num [1:363] -0.737 0.669 0.744 0.851 -0.502 ...
  ..$ p.1      : num [1:363] 2.77e-07 9.47e-08 4.18e-01 7.71e-01 6.54e-12 ...
  ..$ p.2      : num [1:363] 5.13e-01 3.66e-01 3.09e-08 6.99e-06 1.06e-01 ...
  ..$ p.diffcor: num [1:363] 3.99e-04 3.63e-04 6.04e-06 2.94e-04 2.64e-06 ...
  ..$ q.1      : num [1:363] 1.50e-04 9.10e-05 6.91e-01 9.02e-01 2.61e-07 ...
  ..$ q.2      : num [1:363] 0.794531 0.69894 0.000616 0.004811 0.415126 ...
  ..$ q.diffcor: num [1:363] 0.0921 0.0921 0.0153 0.0841 0.01 ...
  ..$ type     : chr [1:363] "same signed" "same signed" "same signed" "same signed" ...
```

6 References

de la Fuente A (2010). From "differential expression" to "differential networking" – identification of dysfunctional regulatory networks in diseases. *Trends in Genetics*, 26(7):326-33.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika*, 10 (4): 507–521.

Fisher, R. A. (1921). On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron*, 1: 3–32.

Hsu C-L, Juan H-F, Huang H-C (2015). Functional analysis and characterization of differential coexpression networks. *Scientific Reports*, 5: 13295

Jiang Z, Dong X, Li Z-G, He F, Zhang Z (2016). Differential coexpression analysis reveals extensive rewiring of Arabidopsis gene coexpression in response to Pseudomonas syringae infection. *Scientific Reports*, 6(1):35064.

Yang J, Yu H, Liu B-H, Zhao Z, Liu L, Ma L-X, et al. (2013) DCGL v2.0: An R package for unveiling differential regulation from differential co-expression. *PLoS ONE*, 8(11):e79729.

Yu H, Liu B-H, Ye Z-Q, Li C, Li Y-X, Li Y-Y (2011). Link-based quantitative methods to identify differentially coexpressed genes and gene pairs. *BMC bioinformatics*, 12(1):315.