

# Package ‘disco’

July 2, 2018

**Title** Discordance and Concordance of Transcriptomic Responses

**Version** 0.6

**Author** Teresa Domaszewska [aut, cre],  
January Weiner [aut]

**Description** Concordance and discordance of homologous gene regulation allows comparing reaction to stimuli in different organisms, for example human patients and animal models of a disease. The package contains functions to calculate discordance and concordance score for homologous gene pairs, identify concordantly or discordantly regulated transcriptional modules and visualize the results. It is intended for analysis of transcriptional data.

**License** GPL (>= 2.0)

**Depends** R (>= 2.10)

**Encoding** UTF-8

**LazyData** true

**Type** Package

**Maintainer** Teresa Domaszewska <domaszewska@mpiib-berlin.mpg.de>

**RoxygenNote** 6.0.1.9000

**Imports** methods,ggplot2,tmod,RColorBrewer

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2018-07-02 14:20:03 UTC

## R topics documented:

corOrt . . . . .	2
disco.score . . . . .	3
makeMatchedOrtholog . . . . .	4
makeSummary . . . . .	5
matchedOrtholog-class . . . . .	5
modCor . . . . .	6

orthologs . . . . .	7
percentGenes . . . . .	8
plotDisco . . . . .	8

<b>Index</b>	<b>10</b>
--------------	-----------

---

corOrt	<i>Calculate ortholog gene correlation coefficients</i>
--------	---

---

## Description

Calculate correlation between sets of matched orthologous genes using either method from Takao et al. (2015) or Seok et al. (2013)

## Usage

```
corOrt(x, g = NULL, pval = 0.05, lfc = 0)
```

## Arguments

x	an object of class matchedOrtholog
g	either a vector of gene names of genes of interest, or an object of class "tmod" with definitions of modules. If NULL, all genes will be used
pval	p value threshold
lfc	log <sub>2</sub> fold change threshold

## Details

This function calculates Pearson Correlation, squared Pearson correlation and Spearman correlation for mouse and human logFC values for genes with significant p-values for differential regulation. As in PNAS 2013 110 (9) 3507-3512, for the calculation of squared Pearson correlation coefficient all the genes that are significantly regulated in at least one specie are taken into account, while for the calculation of Spearman's correlation coefficient only the genes significantly regulated in both species as in PNAS 2015 112 (4) 1167-1172.

The parameters can be calculated either for all genes in the orthologs object, for an explicit list of genes as a character vector parameter g, or for a set of tmod modules, if the parameter g is an object of class tmod.

## Value

corOrt returns a data frame with the following columns:

**r** Pearson correlation coefficient r

**Seok** correlation measure described by Seok et al. – squared Pearson r correlation coefficient for genes significantly up- or downregulated in one of the two orthologous sets

**Seok.N** number of gene pairs that were used to calculate the Seok coefficient

**rho** Spearman correlation coefficient rho

**Takao** correlation measure described by Takao et al. – Spearman correlation coefficient for genes significantly up- or downregulated in both of the two gene orthologous sets

**Takao.N** number of gene pairs that were used to calculate the Takao coefficient; smaller or equal to Seok.N

**N** total number of gene pairs in the given gene set

## Bibliography

Seok, Junhee, et al. "Genomic responses in mouse models poorly mimic human inflammatory diseases." *Proceedings of the National Academy of Sciences* 110.9 (2013): 3507-3512.

Takao, Keizo, and Tsuyoshi Miyakawa. "Genomic responses in mouse models greatly mimic human inflammatory diseases." *Proceedings of the National Academy of Sciences* 112.4 (2015): 1167-1172.

## See Also

modCor

## Examples

```
library(tmod)
data(tmod)
data(orthologs)
a <- corOrt(orthologs)
disco <- disco.score(orthologs)
ord <- order(disco, decreasing = TRUE)
concordant <- tmodCERNOtest(toupper(orthologs$genes)[ord])
corOrt(orthologs, g=tmod[concordant$ID])
```

---

disco.score	<i>Calculate concordance score for differentially expressed genes in two data sets</i>
-------------	--

---

## Description

For each pair of matched (orthologous) genes from an object of class `mtchedOrtholog` calculates the concordance / discordance score. The rationale is to formulate a metric allowing to compare heterologous data sets, e.g. data sets from different sources, technical platforms or even organisms.

## Usage

```
disco.score(x)
```

## Arguments

x                      object of class `matchedOrtholog`

**Details**

$$\text{disco.score} := |\log FC_1| \cdot |\log FC_2| \cdot (|\log_{10}(p_1)| + |\log_{10}(p_2)|) \cdot \text{sign}(\log FC_1) \cdot \text{sign}(\log FC_2)$$

Where

$\log FC_1$  log fold change of the expression change of the gene from the first data set

$\log FC_2$  log fold change of the expression change of the gene from the second data set

$p_1$  differential regulation p-value for the gene from the first data set

$p_2$  differential regulation p-value for the gene from the second data set

**Value**

A numerical vector of the disco scores

**Examples**

```
data(orthologs)
orthologs$disco.score <- disco.score(orthologs)
```

---

makeMatchedOrtholog    *Create an object of class matchedOrtholog*

---

**Description**

Create an object representing regulation in two sets of genes (e.g. from two organisms).

**Usage**

```
makeMatchedOrtholog(names, genes, lfc1, lfc2, pval1, pval2, row.names = NULL,
  extra = NULL)
```

**Arguments**

names	character vector with names of the matched sets
genes	names associated with the ortholog pairs
lfc1, lfc2	log fold changes or effect sizes
pval1, pval2	vectors of p values for the two data sets
row.names	character vector to be used as row names for the data set
extra	a data frame or a vector with the same length or number of rows as "genes"

**Details**

The object contains a mapping between two sets of genes and the associated effect sizes (i.e. log fold changes) and p-values used to calculate disco score and other statistics.

**See Also**

matchedOrtholog-class

---

`makeSummary`*Common modules in specified comparisons*

---

**Description**

Summarize numbers of enriched modules in an unrestricted number of comparisons and list all the modules that belong to intersection of specified comparisons.

**Usage**

```
makeSummary(...)
```

**Arguments**

... any number of data frames created as a result of tmodCERNOtest functions

**Value**

list of two elements: RowN - list, each element of it contains number of rows in every subsequent comparison, int character vector with IDs of all modules belonging to the intersection of comparisons provided as the input

**Examples**

```
library(tmod)
data(tmod)
data(orthologs)
disco <- disco.score(orthologs)
g <- toupper(orthologs$genes)
concordant <- tmodCERNOtest(g[order(disco, decreasing = TRUE)])
discordant <- tmodCERNOtest(g[order(disco)])
sum1 <- makeSummary(concordant, discordant)
```

---

`matchedOrtholog-class` *S4 class for disco*

---

**Description**

S4 class for disco

**Usage**

```
## S4 method for signature 'matchedOrtholog'
show(object)
```

**Arguments**

object            An object of the matchedOrtholog class

**Details**

Class for holding matched orthologs and associated data. See makeMatchedOrtholog for details.

**See Also**

makeMatchedOrtholog

---

modCor	<i>Calculate correlation between regulation of orthologous genes in a chosen gene set</i>
--------	---

---

**Description**

Calculates correlation between regulation of group of genes in two data sets. Takes as arguments an object of class matchedOrtholog, a vector of genes of interest (for example, genes belonging to an expression module) and method of correlation calculation. The function also checks how many of the genes in the vector are present in the matchedOrtholog object. frame.

**Usage**

```
modCor(x, g = NULL, method = "pearson")
```

**Arguments**

x                    an object of class matchedOrtholog

g                    either a vector of gene names of genes of interest, or an object of class "tmod" with definitions of modules. If NULL, all genes will be used

method              method for calculation of correlation (default: pearson)

**Value**

data frame with two columns: Cor - calculated correlation between the genes, P.Value - p value from test of significance of the correlation, NGenes - number of genes included in correlation calculation. Row names of the data frame correspond to the names of the modules.

**Examples**

```
library(tmod)
data(tmod)
data(orthologs)

genes <- tmod$MODULES2GENES[["LI.M0"]]
a <- modCor(orthologs, genes)
```

```
# Using tmod objects directly
a <- modCor(orthologs, tmod[c("LI.M0", "LI.M1.0")], "spearman")
```

---

orthologs

*Example ortholog data set*

---

## Description

Example object of class `matchedOrtholog` containing mouse and human data sets

## Details

The `matchedOrtholog` object `orthologs` contains matched orthologous genes from mouse and human samples, and the results of differential expression analysis as  $\log_2$  fold changes and p-values.

The human data set contains blood transcriptional profiles of 46 TB patients and 62 healthy individuals and is available on the Gene Expression Omnibus database under the accession number GSE28623 (Maertzdorf et al., 2011).

The mouse data set was derived from 129S2 mice infected with TB for 24h (5 mice) and before infection (5 mice) and is available under the accession number GSE89392. The data sets have been analyzed with `limma` R package for differential expression analysis (Ritchie et al., 2015). The data sets were background corrected using the `normexp` method and quantile normalized between arrays. `Limma lmFit` function was used to fit linear models which included the factors: stimulus type and time point. The p-values were calculated based on the moderated t-statistic and most differentially regulated genes were retrieved with `topTable` function. Orthologous genes were assigned to each other between corresponding human and mouse data sets used in each comparison. Probe names specific to the microarray used were assigned an ENSEMBL identifier with use of `mapIds` function from the `biomaRt` package (version 2.24.1, Durinck et al., 2005; Durinck, Spellman, Birney, & Huber, 2009).

Multiple repeating probes were averaged by applying `limma avereps` function. Then, orthologous human and mouse genes were identified with `biomaRt getLDS` function based on homology mapping between different species interlinked in Ensembl data base (with attributes and filters defined as `ensembl_gene_id`). Only the putative orthologs with a 1:1 mapping (no potential in-paralogs) were included in the further analysis.

The gene names in the object correspond to the human gene names.

## Examples

```
data(orthologs)
# view the object as a data frame
head(as(orthologs, "data.frame"))
# calculate the disco score
ds <- disco.score(orthologs)
```

---

percentGenes	<i>Calculates percentage of present genes from a specified superset of expression modules</i>
--------------	---

---

### Description

This function counts the number of genes in modules specified in data.frame created by tmodCERNOtest function and returns percentage of those genes which are present in data.frame on the basis of which the enrichment was calculated, as well as original number of genes present in the module.

### Usage

```
percentGenes(orthologs, mset = NULL)
```

### Arguments

orthologs	object of class matchedOrtholog
mset	a tmod module set (object of class tmod). If NULL, the default tmod object will be used.

### Value

data frame with columns specifying percentages of genes present in subsequent modules "p\_genes" and the overall amount of genes in the modules "all\_genes"

### Examples

```
library(tmod)
data(tmod)
data(orthologs)
disco <- disco.score(orthologs)
ord <- order(disco, decreasing = TRUE)
concordant <- tmodCERNOtest(toupper(orthologs$genes)[ord])
modtable <- percentGenes(orthologs, tmod[concordant$ID])
cbind(concordant, modtable)
```

---

plotDisco	<i>plots log2FC of the corresponding genes against each other</i>
-----------	---

---

### Description

Creates a plot of log2FC of the chosen genes in one data set against log2FC of corresponding genes in the other data set. Each dot represents a gene and is colored according to disco.score calculated for this pair of heterologous genes: the stronger is red color the more concordantly regulated is the gene pair, and the stronger is the blue color, the more discordantly regulated is the gene pair.



**Usage**

```
plotDisco(x, disco.score, g = NULL)
```

**Arguments**

<code>x</code>	an object of class <code>matchedOrtholog</code>
<code>disco.score</code>	a numerical vector of values of <code>disco.score</code> assigned to each orthologous gene pair, ordered according to the gene order in <code>matchedOrtholog</code> object
<code>g</code>	either a vector of gene names of genes of interest

**Value**

plot

**Examples**

```
library(tmod)
data(tmod)
data(orthologs)
ds <- disco.score(orthologs)
plotDisco(orthologs, ds, tmod$MODULES2GENES[["LI.M1.0"]])
```

# Index

`as (matchedOrtholog-class)`, 5

`corOrt`, 2

`disco.score`, 3

`makeMatchedOrtholog`, 4

`makeSummary`, 5

`matchedOrtholog-class`, 5

`modCor`, 6

`orthologs`, 7

`percentGenes`, 8

`plotDisco`, 8

`show, matchedOrtholog-method`  
`(matchedOrtholog-class)`, 5