# Package 'GOFunction'

March 26, 2013

**Type** Package

**Title** GO-function: deriving biologcially relevant functions from statistically significant functions

**Version** 1.4.0

**Date** 2011-06-28

**Author** Jing Wang <jingwang.uestc@gmail.com>

**Maintainer** Zheng Guo <guoz@ems.hrbmu.edu.cn>

**Description** The GO-function package provides a tool to address the redundancy that result from the GO structure or multiple annotation genes and derive biologically relevant functions from the statistically significant functions based on some intuitive assumption and statistical testing.

**License** GPL (>= 2)

**LazyLoad** yes

**Depends** R (>= 2.11.0), methods, Biobase (>= 2.8.0), graph (>= 1.26.0),Rgraphviz (>= 1.26.0), GO.db (>= 2.4.1), AnnotationDbi (>= 1.10.2), SparseM (>= 0.85)

**Imports** methods, Biobase, graph, Rgraphviz, GO.db, AnnotationDbi,SparseM

**Collate** zzz.R GOFunction.R enrichmentFunction.R localRedundancy.R globalRedundancy.R createGODAG.R showSigNodes.R

**biocViews** GO, Pathways, Microarray, GeneSetEnrichment

## R topics documented:

---

GOFunction-package       *GO-function: deriving biologically relevant functions from statistically significant functions*

---

## Description

The GO-function package is an enrichment analysis tool for Gene Ontology (GO). According to some explicit rules, it is designed for treating the redundancy resulting from the GO structure or multiple annotation genes. Different from current redundancy treatment tools simply based on some numerical considerations, GO-function can find terms which are both statistically interpretable and biologically meaningful.

## Details

|           |            |
|-----------|------------|
| Package:  | GOFunction |
| Type:     | Package    |
| Version:  | 1.0.1      |
| Date:     | 2011-06-14 |
| License:  | GPL (>=2)  |
| LazyLoad: | yes        |

## Author(s)

Jing Wang Maintainer: Zheng Guo <guoz@ems.hrbmu.edu.cn>

## References

Jing Wang et al. GO-function: deriving biologically relevant functions from statistically significant functions, Briefings in Bioinformatics, doi: 10.1093/bib/BBR041.

## See Also

GOFunction

---

createGODAG       *Creation of GO DAG stucture for statistically significant GO terms*

---

## Description

To plot the relationship between statistically significant GO terms, this function creates a GO DAG structure for these terms.

## Arguments

| | |
|---|---|
| sigNodes | sigNodes is the statistically significant GO terms found by "enrichmentFunction" function. |
| ontology | The default ontology is "BP" (Biological Process). The "CC" (Cellular Component) and "MF" (Molecular Function) ontologies can also be used. |

## Value

This function returns a object of 'graphNEL' class.

## Note

This function simulates the related program in TopGO (Alexa, A. et al. (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics, 22, 1600-1607).

## Author(s)

Jing Wang

## See Also

GOFunction enrichmentFunction

---

| enrichmentFunction | *Identification of statistically significant GO terms* |
|---|---|

---

## Description

Using the hypergenomic distribution test and FDR correction procedure, this function can identify a set of statistically significant GO terms.

## Arguments

| | |
|---|---|
| annRef | annRef is the annotated reference genes and their annotated terms. |
| annInterest | annInterest is the annotated interesting genes and their annotated terms. |
| method | method is the fdr method. GO-function provides three p value correction methods: "bonferroni", "BH" and "BY". The default is "BY". |
| fdrth | fdrth is the fdr cutoff to identify statistically significant GO terms. The default is 0.05. |

## Value

The function returns a list contained two values: The first value is a table contained the information of all statistically significant terms under a FDR control level and the second value is a table contained the information of all annotated GO terms.

## Author(s)

Jing Wang

**See Also**

GOFunction

---

| globalRedundancy | *Treating for glocal redundancy between GO terms which share multiple function genes but have no ancient-offspring relationship* |

---

**Description**

For a pair of terms with overlapping genes, this function can identify whether their significance might be simply introduced by the overlapping genes (see details).

**Arguments**

| | |
|---|---|
| generalAnn | generalAnn is the mapping between gene and GO terms. |
| sigTermRelation | sigTermRelation is the GO DAG relationship between the statistically significant terms. |
| annRef | annRef is the annotated reference genes and their annotated terms. |
| annInterest | annInterest is the annotated interesting genes and their annotated terms. |
| sigTermRedun | sigTermRedun is the remained statistically signifcant terms after treating local redundancy. |
| poth | poth is the significant level to test whether the overlapping genes of one term is significantly different from the non-overlapping genes of the term. The default is 0.05. |
| peth | peth is the significant level to test whether the non-overlapping genes of one term is enriched with interesting genes. The default is 0.05. |

**Details**

This function uses the following principles to treat for glocal redundancy between GO terms which share multiple function genes but have no ancient-offspring relationship. For a pair of overlapping terms, GO-function remains one term only if (1) there is additional evidence that the non-overlapping genes of a term may be related to the disease if the frequency of interesting genes in these genes is not significantly different from that in the overlapping genes or is significantly higher than the random background; and (2) there is no such evidence for the non-overlapping genes of another term. In all the other situations, GO-function remains both terms.

**Value**

The function returns a set of remained terms with evidence that their significance should not be simply due to the overlapping genes.

**Author(s)**

Jing Wang

**See Also**

GOFunction localRedundancy

---

GOFunction                    *main function of the GO-function package*

---

**Description**

The GOFunction function is the main function of the GO-function package and can generate a set of biologically relevant GO terms.

**Usage**

GOFunction(interestGenes, refGenes, organism = "org.Hs.eg.db", ontology = "BP", fdrmethod = "BY", fdrth =

**Arguments**

| | |
|---|---|
| interestGenes | interestGenes is a set of interesting genes (e.g. differential expressed genes), which should be denoted using the Entrez gene ID. |
| refGenes | refGenes is the background genes corresponding to the interesting genes, which should be denoted using the Entrez gene ID. |
| organism | The GO-function package can be currently applied to analyse data for 18 organisms and the user should install the corresponding gene annotation package when analysing data for these organisms. The 18 organisms and the corresponding packages are as follows: Anopheles "org.Ag.eg.db", Bovine "org.Bt.eg.db", Canine "org.Cf.eg.db", Chicken "org.Gg.eg.db", Chimp "org.Pt.eg.db", E coli strain K12 "org.EcK12.eg.db", E coli strain Sakai "org.EcSakai.eg.db", Fly "org.Dm.eg.db", Human "org.Hs.eg.db", Mouse "org.Mm.eg.db", Pig "org.Ss.eg.db", Rat "org.Rn.eg.db", Rhesus "org.Mmu.eg.db", Streptomyces coelicolor "org.Sco.eg.db", Worm "org.Ce.eg.db", Xenopus "org.Xl.eg.db", Yeast "org.Sc.sgd.db", Zebrafish "org.Dr.eg.db". The default organism is "org.Hs.eg.db" (Human). |
| ontology | The default ontology is "BP" (Biological Process). The "CC" (Cellular Component) and "MF" (Molecular Function) ontologies can also be used. |
| fdrmethod | GO-function provides three p value correction methods: "bonferroni", "BH" and "BY". The default fdrmethod is "BY". |
| fdrth | fdrth is the fdr cutoff to identify statistically significant GO terms. The default is 0.05. |
| ppth | ppth is the significant level to test whether the remaining genes of the ancestor term are enriched with interesting genes after removing the genes in its significant offspring terms. The default is 0.05. |
| pcth | pcth is the significant level to test whether the frequency of interesting genes in the offspring terms are significantly different from that in the ancestor term. The default is 0.05. |
| poth | poth is the significant level to test whether the overlapping genes of one term is significantly different from the non-overlapping genes of the term. The default is 0.05. |
| peth | peth is the significant level to test whether the non-overlapping genes of one term is enriched with interesting genes. The default is 0.05. |
| bmpSize | bmpSize is the width and height of the plot of GO DAG for all statistically significant terms. GO-function set the default width and height of the plot as 2000 pixels in order to clearly show the GO DAG structure. If the GO DAG is |

very complexity, the user should increase $\mathrm{bmpSize}$. Note: If there is an error at the step of "bmp(filename, width = 2000, ..." when running GO-function, the user should decrease $\mathrm{bmpSize}$.

filename        filename is the name of the files saving the table and the GO DAG of all statistically significant terms.

**Value**

There are two types of result output of GO-function. The first type is that GO-function saves a table contained all statistically significant terms to a CSV file (e.g. "sigTerm.csv") in the current working folder. This table contains seven columns: goid, name, refnum (the number of the reference genes in a GO term), interestnum (the number of the interesting genes in a GO term), pvalue, adjustp (the corrected p value by the fdr control), FinalResults. The "FinalResults" contains three types: (1) "Local" represents terms removed after treating for local redundancy; (2) "Global" represents terms removed after treating for global redundancy; (3) "Final" represents the remained terms with evidence that their significance should not be simply due to the overlapping genes. GO-function also saves the structure of GO DAG for all statistic significant terms into a plot (e.g. "sigTerm.bmp") in the current working folder. In this plot, "circle", "box" and "rectangle" represent "Local", "Global" and "Final" terms in the table, respectively. The different color shades represent the adjusted p values of the terms.

**Note**

GO-function use the GO data and annotation data from Bioconductor, so the user does not need to update the data manually.

**Author(s)**

Jing Wang

**Examples**

```
data(exampledata)
sigTerm <- GOFunction(interestGenes, refGenes, organism = "org.Hs.eg.db", ontology= "BP", fdrmethod = "BY",
fdrth = 0.05, ppth = 0.05, pcth = 0.05, poth = 0.05, peth = 0.05, bmpSize = 2000, filename="sigTerm")
```

---

interestGenes                    *Colon cancer differentially expressed genes*

---

**Description**

The interestGenes contain 9201 differentially expressed genes extracted from Sabates-Bellver dataset.

**Usage**

```
data(exampledata)
```

**Format**

vector

**Details**

The differentially expressed genes are selected using the Significant Analysis of Microarray (SAM) with an FDR of 1

**References**

Sabates-Bellver J. et al. Transcriptome pro

le of human colorectal adenomas. Mol Cancer Res, 5(12):1263-1275, 2007.

---

| localRedundancy | *Treating for local redundancy between GO terms with ancient-offspring relationship* |
|---|---|

---

**Description**

This function uses some explicit rules to treat local redundancy (see details).

**Arguments**

| | |
|---|---|
| sigTerm | sigTerm is the statistically significant terms found by the enrichmentFunction function. |
| generalAnn | generalAnn is the mapping between gene and GO terms. |
| sigTermRelation | sigTermRelation is the GO DAG relationship between the statistically significant terms. |
| annRef | annRef is the annotated reference genes and their annotated terms. |
| annInterest | annInterest is the annotated interesting genes and their annotated terms. |
| ppth | ppth is the significant level to test whether the remaining genes of the ancestor term are enriched with interesting genes after removing the genes in its significant offspring terms. The default is 0.05. |
| pcth | pcth is the significant level to test whether the frequency of interesting genes in the offspring terms are significantly different from that in the ancestor term. The default is 0.05. |

**Details**

This function uses three principles to treat local redundancy between GO terms with ancient-offspring relationship: 1. if the frequency of interesting genes in the remaining genes of the ancestor term is not significantly higher than the random background, and at the same time, it is lower than the frequency of its significant offspring terms, GO-function only remains the offspring terms; Otherwise, GO-function remains ancient term. In this situation, 2. if the offspring terms are no different with the ancient term, GO-function only remains the ancestor term; otherwise, 3. GO-function remains both ancient term and its offspring terms.

**Value**

The function returns a set of remained terms without local redundancy.

**Author(s)**

Jing Wang

**See Also**

GOFunction enrichmentFunction

---

refGenes                    *Colon cancer background genes*

---

**Description**

The refGenes contain 19950 background genes.

**Usage**

data(exampledata)

**Format**

vector

**References**

Sabates-Bellver J. et al. Transcriptome pro

le of human colorectal adenomas. Mol Cancer Res, 5(12):1263-1275, 2007.

---

showSigNodes                *Drawing the GO DAG structure of statistically significant terms*

---

**Description**

According to the GO DAG structure created by the "createGODAG" function, this function draws the corresponding plot.

**Arguments**

| | |
|---|---|
| DAG | DAG is a object of 'graphNEL' class which contains the GO DAG structure of statistically significant GO terms. |
| sigTerm | sigTerm is the GO terms found by enrichmentFunction function. |
| sigTerm_Local | sigTerm_Local is the remained statistically signifcant terms after treating local redundancy. |
| sigTerm_Global | sigTerm_Glocal is the remained statistically signifcant terms after treating global redundancy. |
| dagTermInfo | dagTermInfo is the information of all related GO terms in GO DAG. |
| bmpSize | bmpSize is the width and height of the plot of GO DAG for all statistically significant terms. GO-function set the default width and height of the plot as 2000 pixels in order to clearly show the GO DAG structure. If the GO DAG is very complexity, the user should increase bmpSize. Note: If there is an error at the step of "bmp(filename, width = 2000, ..." when running GO-function, the user should decrease bmpSize. |
| filename | filename is the name of the files saving the table and the GO DAG of all statistically significant terms. |

## Value

This function saves the structure of GO DAG for all statistic significant terms into a plot (e.g. "sigTerm.bmp") in the current working folder. In this plot, "circle" represents terms removed after local redundancy treatment, "box" represents terms removed after treating global redundancy and "rectangle" represents the remained terms with evidence that their significance should not be simply due to the overlapping genes. The dfferent color shades represent the adjusted p values of the terms.

## Author(s)

Jing Wang

## See Also

GOFunction enrichmentFunction localRedundancy globalRedundancy createGODAG

# Index