

# Package ‘GOpro’

October 3, 2021

**Type** Package

**Title** Find the most characteristic gene ontology terms for groups of human genes

**Version** 1.18.0

**Author** Lidia Chrabaszcz

**Maintainer** Lidia Chrabaszcz <chrabaszcz.lidia@gmail.com>

**Description** Find the most characteristic gene ontology terms for groups of human genes.  
This package was created as a part of the thesis which was developed under the auspices of MI<sup>2</sup> Group (<http://mi2.mini.pw.edu.pl/>, <https://github.com/geneticsMiNIng>).

**License** GPL-3

**Depends** R (>= 3.4)

**Imports** AnnotationDbi, dendextend, doParallel, foreach, parallel, org.Hs.eg.db, GO.db, Rcpp, stats, graphics, MultiAssayExperiment, IRanges, S4Vectors

**LinkingTo** Rcpp, BH

**RoxygenNote** 6.0.1

**Suggests** knitr, rmarkdown, RTCGA.PANCAN12, BiocStyle, testthat

**VignetteBuilder** knitr

**LazyData** true

**biocViews** Annotation, Clustering, GO, GeneExpression, GeneSetEnrichment, MultipleComparison

**URL** <https://github.com/mi2-warsaw/GOpro>

**BugReports** <https://github.com/mi2-warsaw/GOpro/issues>

**git\_url** <https://git.bioconductor.org/packages/GOpro>

**git\_branch** RELEASE\_3\_13

**git\_last\_commit** b5688fa

**git\_last\_commit\_date** 2021-05-19

**Date/Publication** 2021-10-03

R topics documented:

|                   |   |
|-------------------|---|
| exrtcga . . . . . | 2 |
| findGO . . . . .  | 2 |
| GOpro . . . . .   | 4 |

|       |   |
|-------|---|
| Index | 5 |
|-------|---|

---

|         |                                    |
|---------|------------------------------------|
| exrtcga | <i>Expressions of human genes.</i> |
|---------|------------------------------------|

---

Description

A dataset containing gene expressions of 300 human genes randomly chosen from a list returned by function prepareData(RTCGA = TRUE,cohorts = c('leukemia','colon','bladder')).

Usage

exrtcga

Format

A MultiAssayExperiment object of 3 listed experiments with user-defined names and respective classes. Containing an ExperimentList class object of length 3: [1] leukemia: matrix with 300 rows and 173 columns [2] colon: matrix with 300 rows and 190 columns [3] bladder: matrix with 300 rows and 122 columns.

Value

data

---

|        |                                                      |
|--------|------------------------------------------------------|
| findGO | <i>Find top Gene Ontology terms for given genes.</i> |
|--------|------------------------------------------------------|

---

Description

Find top Gene Ontology terms for given genes.

Usage

```
findGO(groups, topAOV = 50, sig.levelAOV = 0.05, parallel = FALSE,
grouped = "tukey", sig.levelGO = 0.05, minGO = 5, maxGO = 500,
clust.metric = NULL, clust.method = NULL, dist.matrix = NULL,
topGO = 3, sig.levelTUK = 0.05, onto = c("MF", "BP", "CC"),
extend = FALSE, over.rep = FALSE)
```

**Arguments**

|              |                                                                                                                                                                                                                                                               |
|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| groups       | a MultiAssayExperiment object containing an ExperimentList class object representing gene expressions for at least 3 cohorts. Rows must be named with genes' aliases. The order of samples and genes has to be the same for each ExperimentList class object. |
| topAOV       | A numeric value, a number of most significantly differentiated genes to be returned.                                                                                                                                                                          |
| sig.levelAOV | a numeric value, a significance level used in BH correction for multiple testing (aovTopTest).                                                                                                                                                                |
| parallel     | A logical value indicating if a task should be run on more than one core.                                                                                                                                                                                     |
| grouped      | A method of grouping genes, one of 'tukey' and 'clustering'.                                                                                                                                                                                                  |
| sig.levelGO  | A numeric value, a significance level used in BH correction for multiple testing (findTopGOs).                                                                                                                                                                |
| minGO        | A minimum number of functions that a gene needs to represent to be considered as frequent.                                                                                                                                                                    |
| maxGO        | A maximum number of functions that a gene needs to represent to be considered as frequent.                                                                                                                                                                    |
| clust.metric | The method to calculate a distance measure used in hierarchical clustering, possible names: "euclidean", "maximum", "manhattan", "canberra", "binary" or "minkowski".                                                                                         |
| clust.method | The agglomeration method used to cluster genes. This should be #'one of "ward.D", "ward.D2", "single", "complete", "average" (= UPGMA), "mcquitty" (= WPGMA), "median" (= WPGMC) or "centroid" (= UPGMC).                                                     |
| dist.matrix  | A matrix with calculated distances to be used as a metric by hclust function.                                                                                                                                                                                 |
| topGO        | A number of the most characteristic functions of groups of genes to be returned.                                                                                                                                                                              |
| sig.levelTUK | A numeric value, a significance level used in Tukey's all pairwise comparison (groupByTukey).                                                                                                                                                                 |
| onto         | An ontology or ontologies to be searched for significant GO terms, at least one of 'MF' (molecular function), 'BP' (biological process), and 'CC' (cellular component).                                                                                       |
| extend       | A logical value indicating if an extended version of the output should be presented.                                                                                                                                                                          |
| over.rep     | A logical value indicating if an over represented GO terms should be presented in the plot.                                                                                                                                                                   |

**Value**

A data frame containing the top gene ontology terms for each group of genes and the gene aliases.

**Examples**

```
findGO(exrtcga, grouped = 'clustering', topGO = 10, onto = 'MF')
findGO(exrtcga, grouped = 'tukey', topGO = 2, extend = TRUE)
```

---

GOpro*GOpro: find the most characteristic gene ontology terms for groups of genes*

---

**Description**

Based on the gene expressions find the structure somewhat comparable to gene signature. From all given genes, determine which are significantly different between sets. These sets may relate to different health conditions of patients, i.e. different types of cancer. Then divide interesting genes into subsets. Genes belong to a particular subset if they share the same feature. There are two implemented methods that can be used to create genes' subsets. The first method is so-called all pairwise comparisons by Tukey's procedure. Genes that have the same profile (a result of all comparisons) are assigned to one subset. The second way of determining subsets is a method of hierarchical clustering. When all genes are divided into subsets, then for each subset all relevant GO terms are searched for in org.Hs.eg.db database. Each found GO terms is tested using Fisher's test to find out which of them are the most characteristic for the given subset of genes.

**See Also**[findGO](#)

# Index

## \* **datasets**

exrtcga, [2](#)

exrtcga, [2](#)

findGO, [2](#), [4](#)

G0pro, [4](#)

G0pro-package (G0pro), [4](#)