Package 'pathwayko'

February 1, 2022

Version 0.1.1

Title PathwayKO: An R package for knock-out pathway enrichment analysis

Description PathwayKO is an R package for knock-out pathway enrichment analysis, excavating the true positive KO KEGG pathways that contain and are impacted by a KO gene at the system-level. It enables ROC curve-based statistics analysis for assessing the performance of methods in terms of AUC (area under ROC curve), partial AUCs, Youden's best p-value threshold, specificity, sensitivity, accuracy, precision and recall. It is flexible to incorporate custom methods and currently integrates the state-of-the-art SPIA, ROntoTools (PE and pDIS), PADOG, GSA, GESA and SAFE. A benchmark dataset of mouse 10 KO GEO datasets is embedded for demo. The PathwayKO package provides a novel solution to the knock-out pathway enrichment analysis.

Maintainer Hannan Ai <allenai94@icloud.com>

Depends R (>= 4.0.0)

Imports stats, methods, graphics, graph, Biobase (>= 2.48.0), org.Mm.eg.db (>= 3.11.4), KEGGgraph (>= 1.48.0), Rgraphviz (>= 2.32.0), AnnotationDbi (>= 1.50.0), limma (>= 3.44.0), annotate (>= 1.66.0), oligo (>= 1.52.0), ggplot2 (>= 3.3.0), gridExtra (>= 2.3) , SPIA (>= 2.40.0), magick (>= 2.4.0), pROC (>= 1.16.0), RColorBrewer (>= 1.1-0), stringr (>= 1.4.0), changepoint (>= 2.2.0), igraph (>= 1.2.5), ROntoTools (>= 2.16.0), GSA (>= 1.03.0), PADOG (>= 1.30.0), fgsea (>= 1.14.0), safe (>= 3.28.0)

Suggests pathview (>= 1.28.0), GEOquery (>= 2.56.0), knitr, rmarkdown

License CC BY-NC-ND 4.0 + file LICENSE

URL https://github.com/allenaigit/pathwayko

BugReports https://github.com/allenaigit/pathwayko/issues

LazyData TRUE

Encoding UTF-8

RoxygenNote 7.1.2

VignetteBuilder knitr

R topics documented:

`combineresult` *combining and formatting the individual PA results*

Description

It yields _STATS.RData, summarizing and formatting the intermediate outputs such as true positive rates (TPR), true negative rates (TNR), false positive rates (FPR) and false negative rates (FNR), plus the Youden's best p-value threshold, sensitivity (TPR), specificity (TNR), accuracy, precision and recall.

Usage

```
combineresult()
```

Details

This function takes "*_SUM.RData" generated by pathwayko, combines them into a single object which can be further analyzed or plotted. All "_SUM" data should be analyzed by the same sets of PA methods, if not by identical PA methods under identical parameters. (i.e. with same KEGG xmls, same criteria for choosing DE genes, etc)

Value

TRUE when successful, FALSE otherwise

evidenceplot *generating two-dimensional evidence plot*

Description

It generates two-dimensional evidence plot for methods based on or derived from SPIA

Usage

```
evidenceplot()
```

Details

Specifically, this function looks for two columns in raw pathway analysis results: "pPERT" and "pNDE", to plot the evidence plot. Reference: Tarca A. L. et al. (2009) A novel signaling pathway impact analysis. Bioinformatics, 25, 75–82.)

filtertrue *extracting the list of true positive KO KEGG pathways*

Description

It extracts the list of true positive KO KEGG pathways that contain and are impacted by the KO gene from the results of each method

Usage

```
filtertrue()
```

Details

This function scans user provided directory for any .csv files under the current directory recursively and reads in all .csv files not tagged with control keywords like "AUC" or "pathway". Hence this function should only be used on directories with no .csv files from other sources. After data input, this function filters out from all PA results entries of true positive KEGG pathways in that analysis and saves each of filtered pathway analysis result in a separate .csv file.

highedges *calculating edge scores, determining the change-point and generating the KO-gene associated sub-graph*

Description

It is not executed itself, instead, it is invoked by pathwayko.

Usage

```
highedges(
  KEGGgraph = NULL,
  DataObject = NULL,
  condition_FC = NULL,
  condition_pVal = NULL,
  mc.cores = 1
)
```

Arguments

KEGGgraph	graphical representation of pathways in KEGG
DataObject	Object generated by calling preprocess.R
condition_FC	named list of logFC of all genes in DataObject
condition_pVal	named list of adjusted pval of all genes in DataObject
mc.cores	number of cores to be used in parallel lapply

Details

Reference: Hanoudi S., et al. (2017) Identifying biologically relevant putative mechanisms in a given phenotype comparison. PLoS ONE 12, e0176950.

Value

various information grouped together as a list

Author(s)

Original implementation by Samer Hanoudi and Sorin Draghici, modification by Hannan Ai

makeSPIAdata *parsing KEGG xmls ready for pathway enrichment analysis*

Description

Reference: Tarca A. L. et al. (2009) A novel signaling pathway impact analysis. Bioinformatics, 25, 75–82.

Usage

```
makeSPIAdata(kgml.path=NULL,organism="mmu",out.path=".")
```

Arguments

kgml.path	path to KEGG xml files
organism	three letter organism code from KEGG database
out.path	path to which generated RData file will be placed

Value

TRUE on success, FALSE otherwise

Author(s)

Original implementation by Adi Laurentiu Tarca <atarca AT med.wayne.edu>, Purvesh Kathri <purvesh AT cs.wayne.edu> and Sorin Draghici <sorin AT wayne.edu>, source code shown at the bottom. Modification by Hannan Ai

pathwayko *performing the KO pathway enrichment analysis on single dataset*

Description

It invokes utilities, highedges, functions and methods to calculate the scores of all edges (Hanoudi et al., 2017) in a known global KEGG graph (Zhang and Wiemann, 2009) to yield the distribution of edge scores, and thus automatically determines a change point on the distribution of edge scores by the change-point analysis method (Killick and Eckley, 2014). Outputs include the KO gene-associated sub-graph, the list of differential expression (DE) genes, and the list of true positive KO KEGG pathways. References: Hanoudi S., et al. (2017) Identifying biologically relevant putative mechanisms in a given phenotype comparison. PLoS ONE 12, e0176950; Zhang J. D. and Wiemann S. (2009) KEGGgraph: a graph approach to KEGG pathway in R and bioconductor. Bioinformatics, 25, 1470–1471; Killick R. and Eckley I. (2014) changepoint: An R package for changepoint analysis. J. Statistic Software. 58,1–19.

Usage

```
pathwayko()
```

Arguments

batch	boolean, should only be set to TRUE by pathwayko_batch function
batch_DataObject	should only be set by pathwayko_batch function
batch_uip	should only be set by pathwayko_batch function

Value

NULL if encountered error, TRUE when batch is false and an internal object with results if batch is true

---

pathwayko_batch *batch-performing the KO pathway enrichment analysis on multiple datasets*

Description

It conducts the KO pathway enrichment analysis on multiple datasets in a pipeline manner.

Usage

```
pathwayko_batch()
```

Details

This function scans all directory under the working directory to identify data object generated by preprocess function and apply pathwayko to all said objects using same sets of parameters obtained from the user

Value

TRUE if all process completed successfully, FALSE otherwise

`pathwayko_demo` *for pathwayko_demo on 10 KO GEO datasets*

Description

It generates demo results for testing on the 10 KO GEO datasets

Usage

```
pathwayko_demo()
```

Value

TRUE if all process completed successfully, FALSE otherwise

`pathwayview` *rendering the true positive KO KEGG pathways by the up- and down-regulated DE genes*

Description

It renders the true positive KO KEGG pathways by the up- and down-regulated DE genes, which contain and are impacted by the KO genes. Reference: Luo, W. and Brouwer C., Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics, 2013, 29(14): 1830-1831, doi: 10.1093/bioinformatics/btt285

Usage

```
pathwayview()
```

Value

TRUE on success, FALSE otherwise

preprocess *preprocessing GSE data from CEL files and series matrix file*

Description

This function reads in relevant CEL files (from *_RAW.tar) file and series matrix file (*_series_matrix.txt.gz) typically obtained from NCBI database for a given GSE. The preprocessing of these data relies on the user to interactively provide necessary information as such information is difficult to obtain prior to execution and hence to automate. Reference: Carvalho B.S. and Irizarry R.A. (2010) A framework for oligonucleotide microarray preprocessing. Bioinformatics, 16, 2363–2367; Ritchie M.E., et al. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res, 43, e47.

Usage

```
preprocess()
```

Details

To begin, the user is expected to have obtained '*_RAW.tar' and '*_series_matrix.txt.gz' files of a given GSE. The user will first be asked to choose from files under the working directory with matching suffix as input files. Then the user will be asked to provide relevant information such as keywords for case/control and KO gene names to build experiment design for packages like 'oligo' and 'limma'. The final resault will be saved as an compressed R object to be read in and used by other parts of the package.

Value

TRUE when successful, FALSE otherwise

roctest *conducting ROC-curve based statistics analysis*

Description

It performs roctest on two ROC curves

Usage

```
roctest()
```

Details

Reference: Robin X. et al. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics, 12, 77.

`violinplot` *generating violinplots based on the combined PA results*

Description

It displays 12 key metrics including the Youden's best p-value threshold, specificity (i.e., TNR), sensitivity (i.e., TPR), FDR, FPR, FNR, accuracy, precision, recall, AUC, pAUC_SP and pAUC_SE across methods under study when benchmarked on a set of KO GEO datasets.

Usage

```
violinplot()
```

`wilcoxtest` *performing pairwise comparisons on the PA results*

Description

It conducts the "paired", pairwise comparisons between each pair of two methods under study when benchmarked on the same set of KO datasets. pairwise.wilcox.test is from the R 'stats' package, with corrections (BH FDR adjustment of p-value) for multiple testing under the two-sided mode.

Usage

```
wilcoxtest()
```

Index