Package ‘TPEA’

January 20, 2025

Type Package

Title A Novel Topology-Based Pathway Enrichment Analysis Approach

Version 3.1.0

Date 2017-6-25

Author Wei Jiang

Maintainer Wei Jiang <`jiangwei@hrbmu.edu.cn`>

Description We described a novel Topology-based pathway enrichment analysis, which integrated the global position of the nodes and the topological property of the pathways in Kyoto Encyclopedia of Genes and Genomes Database.
We also provide some functions to obtain the latest information about pathways to finish pathway enrichment analysis using this method.

License GPL-2

Depends R (>= 2.10), MESS, Matrix, foreach

Imports XML, RCurl, utils, igraph

Suggests geeM, geepack

LazyData true

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2017-06-25 15:42:32 UTC

Contents

`TPEA-package` *TPEA: A Novel Pathway enrichment analysis approach based on topological structure and updated annotation of pathway*

Description

This package descirbed A Novel Pathway enrichment analysis approach based on topological structure and updated annotation of pathway which integrated the topological property of the pathway and the global position of nodes in pathways.Additionally,it also provided the update functions which could obtain the latest pathway information from KEGG database and users can use the latest information to do the pathway enrichment analysis.

Details

The function AUEC is to calculate the area under the cumulative enrichment curve. The function TPEA is to measure the significance of pathways. The function UPDATE is to online download the latest KEGG pathway information. The viewpathway function is to visualize the pathway in the result based on the genes you input, such as differentially expressed genes. Several other functions are the update related functions, including ViewUpdateTime,UpdateKGML,PathNetwork,NodeGeneData,NodeGene,importUpd The functions involved in relationship between nodes and genes were provided by Chunquan Li. If you want to use the latest information of KEGG database,please run "UPDATE()" functions first, and then run the pathway enrichment analysis functions AUEC and TPEA.

Author(s)

Wei Jiang

`all_genes` *All human protein coding genes*

Description

Human protein coding genes from NCBI Database. We use this set as background gene set.

`AUEC` *Calculate the area under the cumulative enrichment curve (AUEC) based on the interested gene set.*

Description

The interested gene set may be the differentially expressed genes or any other gene set. The function calculate the AUEC based on the interested genes. AUEC is the area under the cumulative enrichment curve in a coordinate system. X-axis displays the nodes by the scores from maximum to minimum. Y-axis displays the cumulative enrichment curve.

Usage

```
AUEC(DEGs)
```

Arguments

`DEGs`	The interested genes you input and the format must be "Entrez ID". If not,translate the interested genes into Entrez ID.

Details

The function only identifies Entrez ID of genes. The nodes are sorted by their AUEC in the pathway. If genes locates on the upstream or the nodes with high degree in a certain pathway, the AUEC of this pathway is high.

Value

The AUEC of 109 pathways based on the interested gene set.

Author(s)

Wei Jiang

Examples

```
##Randomly generated interested genes
DEGs<-sample(100:100000,15)
DEG<-as.matrix(DEGs);
## The function is used to calculate the observed statistic
area<-AUEC(DEG);
```

`DownloadKGML` *Download the latest KGML files*

Description

Download the latest KGML files from KEGG database if you want the latest KGML files from KEGG database.

Details

Download the latest KGML files from KEGG database before pathway enrichment analysis.

Value

The latest KGML files from KEGG database.

Author(s)

Wei Jiang

`filterNode` *Filter the nodes in pathways*

Description

Filter the nodes in pathways.

Author(s)

Wei Jiang

`gene2ec` *The relationship of genes and EC*

Description

The relationship of genes and EC.

`gene2ko` *The relationship of genes and KO*

Description

The relationship of genes and KO.

`getEntry` *Obtain the nodes*

Description

Processe the pathways

`getGeneFromEnzyme` *Obtain the genes from enzymes*

Description

Processe the pathways

`getGeneFromKGene` *Obtain the genes from KGenes*

Description

Processe the pathways

`getGeneFromKO` *Obtain the genes from KO*

Description

Processe the pathways

`getGraphics` *Recontructe the network based on pathways*

Description

Processe the pathways

`getKGeneFromEnzyme` *Obtain genes from KGnenes*

Description

Processe the pathways

`getKGeneFromKO` *Obtain the genes from KO*

Description

Processe the pathways

`getNonMetabolicGraph` *Convert the non-metaboloc pathway to network*

Description

Processe the pathways

`getOrgAndIdType` *Get the type names of nodes*

Description

Processe the pathways

`getPathway` *Get the pathway from KEGG database.*

Description

Processe the pathways

`getProduct` *Get the products*

Description

Processe the pathways

`getReaction` *Get the reaction of nodes in pathways*

Description

Processe the pathways

`getRelation` *Get the relation of nodes in pathways*

Description

Processe the pathways

`getSimpleGraph` *Obtain the graph of pathways*

Description

Processe the pathways

`getSubstrate` *Obtain the information about nodes in KEGG database*

Description

Processe the pathways

`getSubtype` *Get the type of nodes*

Description

Processe the pathways

`getUGraph` *Obtain the graph of pathways*

Description

Obtain the graph of pathways.

Usage

```
getUGraph(graphList, simpleGraph = TRUE)
```

Arguments

`graphList`	Get the list.
`simpleGraph`	Convert the network.

Value

The graphList relationship.

Author(s)

Wei Jiang

`getUnknowProduct` *Get the products*

Description

Processe the pathways

`getUnknowReaction` *Get the reaction of nodes in pathways*

Description

Processe the pathways

`getUnknowRelation` *Get the relation of nodes in pathways*

Description

Processe the pathways

`getUnknowSubstrate` *Obtain the information about nodes in KEGG database*

Description

Processe the pathways

`getUnknowSubtype` *Obtain the types of genes in pathways*

Description

Processe the pathways

Author(s)

Wei Jiang

`importLatesData` *Import the latest relationship information.*

Description

Import the latest relationship information about node,gene and score.

Usage

```
importLatesData()
```

Details

Import the latest relationship information about nodes,genes and their scores based on KGML files.

Value

Import the latest relationship information about node,gene and score.

Author(s)

Wei Jiang

`keggGene2gene` *KeggGene to genes*

Description

Processe the pathways

`mapNode` *Obtain the relationship of nodes and genes*

Description

Processe the pathways

`NodeGene` *Restract the relationship between nodes and genes.*

Description

Restract the relationship between nodes and genes from KGML files.

Usage

```
NodeGene()
```

Details

This function must be used behind the function NodeGeneData.

Value

Restract the relationship between nodes and genes in each network based on the information of KGML files.

Author(s)

Wei Jiang

`NodeGeneData` *Intergate list of node,gene and the score of node.*

Description

Intergate list of node,gene and the score of node based on latest KGML files from KEGG database.

Usage

```
NodeGeneData()
```

Details

Intergate list of node,gene and the score of node based on latest KGML files from KEGG database.

Value

List contains the relationship of node,gene and the score of node based on latest KGML files.

Author(s)

Wei Jiang

`node_gene` *The relationship between nodes and genes*

Description

The relationship between nodes and genes in each pathway in KEGG Database

---

`num_node_gene_score` *The score of each node in a certain pathway*

Description

The dataset includes 109 list and each list contains four columns (the order of node, node, gene and the score).

`PathNetwork` *Reconstruct pathways to networks*

Description

Reconstruct pathways to networks based on KGML files from KEGG database.

Usage

```
PathNetwork()
```

Details

Reconstruct pathways to networks based on KGML files from KEGG database.

Value

The relationship of edges in network.

Author(s)

Wei Jiang

`pathway_names` *Pathway names in KEGG Database*

Description

All pathway names we used in this method

`simplifyGraph` *Recontructe the network based on pathways*

Description

Processe the pathways

`TPEA` *Statistical test and calculate the significance*

Description

Comparing with the AUEC_R which the interested gene set extract from the background gene set randomly and the corresponding AUEC based on interested gene set you input. The last step is to calculate the significance.

Usage

```
TPEA(DEGs, scores, n, FDR_method)
```

Arguments

`DEGs`	Interested gene set such as differentailly expressed gene set.
`scores`	The "AUEC" based on the interested gene set of 109 pathways.
`n`	Randomly number,e.g. 1000, 5000.
`FDR_method`	The methods of calculating FDR value,e.g. "fdr","BH","BY" ,"bonferroni" and etc..

Details

To calculate the significance of the result, you can set "n" as "1000" or any other number you want.

Value

The ultimately result of this topology-based enrichment analysis method.

Author(s)

Wei Jiang

Examples

```
##Randomly generated interested gene set
ViewLatestTime()
##If you want to use the latest information,please run "UPDATE()".
DEGs<-sample(100:10000,10);
DEG<-as.matrix(DEGs);
##Set the times of perturbation
number<-50;
##Calculate the observed statistic
scores<-AUEC(DEG);
##Significant computational
FDR_method<-"fdr";
results<-TPEA(DEG,scores,number,FDR_method);
```

`UPDATE` *Update the latest data from KEGG database*

Description

Updating the latest information of pathways in KEGG database and the time of this process is about 1-2 minutes.

`ViewLatestTime` *Check up the latest date of KGML files*

Description

Check up the latest date of KGML files from KEGG database.

Usage

```
ViewLatestTime()
```

Value

The latest date of KGML files from KEGG database.

Author(s)

Wei Jiang

`viewpathway` *The visualization of interested pathway based on the genes you input, such as differentially expressed genes.*

Description

Input the number of the interested pathway in KEGG Database and genes you interested in, such as differentially expressed genes.

Usage

```
viewpathway(pathwayID, DEGs)
```

Arguments

`pathwayID`	The number of interested pathway ID in KEGG Database, such as "hsa05210".
`DEGs`	The genes you interested in, such as differentially expressed genes.

Details

The "DEGs" must be Entrez ID. If not, please translate them into Entrez ID.

Value

The interface link to KEGG Database to visualize the pathway you input.

Author(s)

Wei Jiang

Examples

```
DEGs<-c(836,842,5594,595);
DEG<-as.data.frame(DEGs);
pathwayID<-"hsa05210";
viewpathway(pathwayID,DEG);
```

Index