Package 'PANDA'

December 5, 2016

Type Package

Title Preferential Attachment Based Common Neighbor Distribution
Derived Functional Associations

Version 0.9.9

Date 2014-11-07

Author Hua Li and Pan Tong

Maintainer Hua Li <kaixinsjtu@hotmail.com>

Description PANDA (Preferential Attachment based common Neighbor Distribution derived Associations) is designed to perform the following tasks in PPI networks: (1) identify significantly functionally associated protein pairs, (2) predict GO terms and KEGG pathways for proteins, (3) make a cluster of proteins based on the significant protein pairs, (4) identify subclusters whose members are enriched in KEGG pathways. For other types of biological networks, (1) and (3) can still be performed.

License Artistic-2.0

LazyData true

Imports GO.db, cluster

NeedsCompilation no

Repository CRAN

Date/Publication 2016-12-05 18:28:46

Depends R (>= 2.10)

R topics documented:

---

dfPPI   *An example of protein-protein interaction data*

Description

This is an example of protein-protein interaction data.

Usage

```
data(dfPPI)
```

Details

A two-column data frame consisting of binary interactions where each row represents an undirected edge (interaction) between two nodes (proteins) from two columns.

GENE2GOtopLite   *An example of GO annotation data for proteins*

Description

This is only an example of GO annotations for proteins.

Usage

```
data(GENE2GOtopLite)
```

Details

A list that maps GO ID to genes. The names should be gene symbols and the elements should be GO IDs. The list have the following format:

$SHC1

[1] "GO:0005158" "GO:0005068" "GO:0005159" "GO:0070435"

$POU5F1

[1] "GO:0035413" "GO:0003130" "GO:0060391" "GO:0090308" "GO:0060965" "GO:0035198"

$FGF12

[1] "GO:0008201"

......

GENE2KEGG *An example of KEGG annotation data for proteins*

Description

This is only an example of KEGG pathway annotations for proteins.

Usage

```
data(GENE2KEGG)
```

Details

A list that maps KEGG pathway ID to genes. The names should be gene symbols and the elements should be KEGG pathways. The list have the following format:

$IGSF5

[1] "hsa04530" "hsa05120"

$OR2T8

[1] "hsa04740"

$hCG_1776980

[1] "hsa00020" "hsa00190" "hsa01100" "hsa05010" "hsa05012" "hsa05016"

......

GOpredict *Predict GO annotations for proteins*

Description

This function uses a direct annotation scheme to predict GO annotations for proteins in the network derived with the PAND algorithm

Usage

```
GOpredict(Pfile, PPIdb, Gene2Annotation, p_value=0.001)
```

Arguments

Pfile   A data frame returned from the function "SignificantPairs"

PPIdb   A 2-column data frame consisting of binary interactions where each row i.e. c(A, B) represents an undirected edge (interaction) between gene A and gene B.

Gene2Annotation

> A list that maps GO ID to genes. The names should be gene symbols and the elements should be GO IDs. i.e.
>
> $SHC1
>
> [1] "GO:0005158" "GO:0005068" "GO:0005159" "GO:0070435"
>
> $POU5F1
>
> [1] "GO:0035413" "GO:0003130" "GO:0060391" "GO:0090308" "GO:0060965" "GO:0035198"
>
> $FGF12
>
> [1] "GO:0008201"
>
> ......

p_value   A cut-off for p-values from Fishers exact test when predicting GO annotations

Value

This function returns a data frame with column names: "Symbol", "GOID", "GOterm", "Ratio" and "Pvalue". "Symbol" is the name of the node this function predicts GO annotation for. "GOID" and "GOterm" are the predicted GO ID and the GO term, respectively. "Ratio" is the proportion of neighboring nodes that have the predicted GO annotation. "Pvalue" is calculated from the "Ratio" by Fishers exact test.

See Also

[SignificantPairs](), [ProteinCluster](), [KEGGpredict](), [SignificantSubcluster]()

KEGGID2NAME   *KEGG pathway ID to KEGG pathway name*

Description

This is a data frame mapping between KEGG pathway ID and KEGG pathway name.

Usage

```
data(KEGGID2NAME)
```

KEGGpredict	*Predict KEGG pathway annotations for proteins*

Description

This function uses a direct annotation scheme to predict KEGG pathway annotations for proteins in the network derived with the PAND algorithm.

Usage

```
KEGGpredict(Pfile, PPIdb, Gene2Annotation, p_value=0.001, IDtoNAME)
```

Arguments

Pfile   A data frame returned from the function "SignificantPairs"

PPIdb   A 2-column data frame consisting of binary interactions where each row i.e. c(A, B) represents an undirected edge (interaction) between gene A and gene B.

Gene2Annotation

A list that maps KEGG pathway ID to genes. The names should be gene symbols and the elements should be KEGG pathways. i.e.

$IGSF5

[1] "hsa04530" "hsa05120"

$OR2T8

[1] "hsa04740"

$hCG_1776980

[1] "hsa00020" "hsa00190" "hsa01100" "hsa05010" "hsa05012" "hsa05016"

......

p_value   A cut-off for p-values from Fishers exact test when predicting KEGG pathway annotations

IDtoNAME   A table that maps KEGG pathway ID to KEGG pathway names.

Value

This function returns a data frame with column names: "Symbol", "KEGGID", "PathName", "Ratio" and "Pvalue". "Symbol" is the name of the node this function predicts KEGG annotation for. "KEGGID" and "PathName" are the predicted KEGG pathway identifier and the pathway name, respectively. "Ratio" is the proportion of neighboring nodes that have the predicted KEGG annotation. "Pvalue" is calculated from the "Ratio" by Fishers exact test.

See Also

SignificantPairs, ProteinCluster, GOpredict, SignificantSubcluster

---

ProteinCluster   *Cluster proteins based on significant protein pairs*

Description

This function uses the p-values (or probabilities) derived from the PAND algorithm to perform agglomerative hierarchical clustering (using the unweighted group average) for proteins that form significant protein pairs.

Usage

```
ProteinCluster(Pfile, Plot=FALSE, TextScaler=50, height=10, width)
```

Arguments

Pfile	A data frame returned from the function SignificantPairs()
Plot	If FALSE, a dendrogram will NOT be generated
TextScaler	Scale the size of the label in the generated PDF file
height	The height of the generated PDF file
width	The width of the generated PDF file

Value

This function returns an object in the class "dendrogram". If the argument "Plot" is "TRUE", it will also plot the dendrogram.

See Also

[SignificantPairs](), [KEGGpredict](), [GOpredict](), [SignificantSubcluster]()

Examples

```
## not run
## data(dfPPI)
## OrderAll=SignificantPairs(dfPPI)
## dendMap=ProteinCluster(Pfile=OrderAll, Plot=TRUE, TextScaler=30)
```

SignificantPairs *Identify functionally associated protein pairs*

Description

This function uses the PAND distribution to calculate p-values (or probabilities) for each pair of proteins with at least one common neighbor in a protein-protein interaction network. It returns protein pairs with significant p-values (or probabilities).

Usage

```
SignificantPairs(PPIdb, Lambda=2, pvalue=FALSE)
```

Arguments

PPIdb   A two-column data frame consisting of binary interactions where each row represents an undirected edge (interaction) between two nodes (proteins) from two columns.

Lambda   Weight of direct interactions in the PAND algorithm. Lamda has different biological meanings with different values: "0" indicates that a direct link gives no information on the functional association; "1" indicates that a direct link is as informative as sharing one common neighbor (defined as an indirect link) on the functional association; "2" (or greater integer) indicates that a direct link is more informative than an indirect link. Since direct links should represent stronger evidence of functional associations than indirect links, we recommend using "2" as Lamda.

pvalue   logical; if TRUE, p-values for protein pairs will be calculated using PAND; if FALSE, probabilities will be calculated.

Value

This function returns a data frame with column names: "Sym_A", "Sym_B", "p_value" and "CommonNeighbor". "Sym_A" and "Sym_B" are a pair of nodes that share a significant functional linkage. "p_value" or "Probability" (calculated by the PAND algorithm) measures the significance of the linkage. "CommonNeighbor" is the number of shared nodes.

See Also

[ProteinCluster](), [KEGGpredict](), [GOpredict](), [SignificantSubcluster]()

Examples

```
## not run
## data(dfPPI)
## OrderAll=SignificantPairs(dfPPI)
```

SignificantSubcluster   *Subclusters with KEGG annotations significantly enriched*

Description

This function identifies subclusters whose members are significantly enrichment in certain KEGG pathways.

Usage

```
SignificantSubcluster(Dendrogram, Gene2Annotation, PPIdb, KGremove=TRUE,
  SPoint=1, EPoint=9.7, p_value=0.001, ini.p_value=0.05)
```

Arguments

Dendrogram	An object in the class "dendrogram" returned from the function "ProteinCluster"
Gene2Annotation	
	A list that maps KEGG pathway ID to genes. The names should be gene symbols and the elements should be KEGG pathways. i.e.
	\$IGSF5
	[1] "hsa04530" "hsa05120"
	\$OR2T8
	[1] "hsa04740"
	\$hCG_1776980
	[1] "hsa00020" "hsa00190" "hsa01100" "hsa05010" "hsa05012" "hsa05016"
	......
PPIdb	A 2-column data frame consisting of binary interactions where each row i.e. c(A, B) represents an undirected edge (interaction) between gene A and gene B.
KGremove	If "TRUE", "hsa05200" and "hsa01100" will be excluded from KEGG-pathway based enrichment analysis as they are too broad.
SPoint	The starting point for searching for the KEGG-pathway enriched subclusters
EPoint	The endpoint for searching for the KEGG-pathway enriched subclusters
p_value	Description for p_value
ini.p_value	Description for ini.p_value

Value

This function identifies subclusters whose members are significantly enrichment in KEGG pathways and generate a dendrogram for these subclusters.

See Also

[SignificantPairs](#), [ProteinCluster](#), [KEGGpredict](#), [GOpredict](#)

Index