Package ‘KEGGprofile’

March 26, 2013

Type Package

Title An annotation and visualization package for multi-types and multi-groups expression data in KEGG pathway

Version 1.0.0

Date 2012-6-29

Author Shilin Zhao

Maintainer Shilin Zhao <zhaoshilin@gmail.com>

Description KEGGprofile is an annotation and visualization tool which integrated the expression profiles and the function annotation in KEGG pathway maps. The multi-types and multi-groups expression data can be visualized in one pathway map. KEGGprofile facilitated more detailed analysis about the specific function changes inner pathway or temporal correlations in different genes and samples.

License GPL (>= 2)

LazyLoad yes

Depends XML, png, TeachingDemos, KEGG.db

Imports AnnotationDbi

biocViews Pathways

Collate ’KEGG_function.R’

R topics documented:

---

col_by_value *col_by_value*

Description

The function will transfer a numeric matrix into a matrix of colors, in which the colors represent the values of numeric matrix

Usage

```
col_by_value(x, col, range = NA, breaks = NA)
```

Arguments

x	a numeric matrix
col	colors used to represent the values. (See also 'Details')
range	values out of the range will be modified to in the range.
breaks	a numeric vector of three or more cut points giving the number of intervals into which x is to be cut. See also 'Details'

Details

A colorbar would also be ploted. The returned colors of the function can be used in function plot_profile. if breaks not equal to NA, col must have the same length with breaks-1.

Value

a matrix equal to x, but the values were instead by colors.

Examples

```
data(pho_sites_count)
col<-col_by_value(pho_sites_count,col=colorRampPalette(c('white','khaki2'))(4),breaks=c(0,1,4,10,Inf))
```

---

download_KEGGfile *download_KEGGfile*

Description

The function download XML files and png files from KEGG website to local disk

Usage

```
download_KEGGfile(pathway_id = "00010", specis = "hsa",
  target_dir = getwd())
```

Arguments

pathway_id	the KEGG pathway id, such as '00010'
specis	the specis id in KEGG database, 'hsa' means human, 'mmu' means mouse, 'rno' means rat, etc
target_dir	the local directory where the downloaded files are saved

Details

If pathway_id is set as 'all', all KEGG pathway ids in KEGG.db package will be used and downloaded from KEGG website

Examples

```
download_KEGGfile(pathway_id="00010",specis='hsa')
```

---

find_enriched_pathway *find_enriched_pathway*

Description

The function will map the genes in KEGG pathway database, and then hypergegeometric tests would be used to estimate the significance of enrichment for each pathway

Usage

```
find_enriched_pathway(gene, specis = "hsa",
  returned_pvalue = 0.05, returned_genenumber = 5)
```

Arguments

gene	a numeric matrix
returned_pvalue	the minimum p value for enriched pathways
returned_genenumber	the minimum number of annotated genes for enriched pathways
specis	the specis id in KEGG database, 'hsa' means human, 'mmu' means mouse, 'rno' means rat, etc

Details

Only the pathways with p value <= returned_pvalue in hypergegeometric tests and number of annotated genes >= returned_genenumber would be taken as enriched and returned.

Value

a list with two parts

name stastic	description a matirx containing the pathway IDs of enriched pathways, and their names, p values, number of annotated genes
name detail	description a list with the genes annotated for each pathway

Examples

```
data(pho_sites_count)
#the 300 genes with most phospholation sites quantified
genes<-names(rev(sort(pho_sites_count[,1]))[1:300])
pho_KEGGresult<-find_enriched_pathway(genes,specis=’hsa’)
```

---

parse_XMLfile *parse_XMLfile*

Description

The function parses KEGG XML (KGML) files

Usage

```
parse_XMLfile(pathway_id, specis, database_dir = getwd())
```

Arguments

database_dir	the directory where the XML files and png files are located
pathway_id	the KEGG pathway id, such as ’00010’
specis	the specis id in KEGG database, ’hsa’ means human, ’mmu’ means mouse, ’rno’ means rat, etc

Details

This function will parse the KEGG XML (KGML) file. Then a matrix with genes in this pathway and related infomations will be returned. This matrix can be used for plot the expression profiles on the pathway figure.

Value

a matrix containing genes in this pathway, and their names, locations etc, which could be used in the function plot_profile as param KEGG_database

Examples

```
XML2database<-parse_XMLfile(pathway_id="04110",specis="hsa",database_dir=system.file("extdata",package="K
```

---

pho_sites_count *number of phosphorylation sites quantified for each gene*

Description

This data set is a data.frame with number of phosphorylation sites quantified for each gene in the analysis.

Source

Olsen, J.V., et al. (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis, Sci Signal, 3, ra3.

plot_pathway *plot_pathway*

Description

A wrapper for function download_KEGGfile, parse_XMLfile and plot_profile

Usage

```
plot_pathway(gene_expr, line_col, groups,
  pathway_id = "00010", specis = "hsa", pathway_min = 5,
  database_dir = getwd(), ...)
```

Arguments

...	any other Arguments for function plot_profile
gene_expr	the matrix for gene expression, row.names should be NCBI gene ID, such as 67040, 93683
line_col	line color for expression in different samples in the pathway map, valid when type='lines'
groups	a character used to indicate expression values from different samples
specis	the specis id in KEGG database, 'hsa' means human, 'mmu' means mouse, 'rno' means rat, etc
pathway_min	The pathways with number of annotated genes less than pathway_min would be ignored
database_dir	the directory where the XML files and png files are located
pathway_id	the KEGG pathway id, such as '00010'

Details

This wrapper function is developed to make the visualization process more easier. Firstly the existence of XML file and png file would be checked, if not, the download_KEGGfile function would be used to download the files. Then the parse_XMLfile function would be used to parse the XML file. At last the plot_profile function would be used to generate the pathway map.

See Also

download_KEGGfile, parse_XMLfile, plot_profile

Examples

```
data(pro_pho_expr)
data(pho_sites_count)
#type='lines'
col<-col_by_value(pho_sites_count,col=colorRampPalette(c('white','khaki2'))(4),breaks=c(0,1,4,10,Inf))
temp<-plot_pathway(pro_pho_expr,bg_col=col,line_col=c("brown1","seagreen3"),groups=c(rep("Proteome ",6),rep
#type='bg'
pho_expr<-pro_pho_expr[,7:12]
temp<-apply(pho_expr,1,function(x) length(which(is.na(x))))
pho_expr<-pho_expr[which(temp==0),]
col<-col_by_value(pho_expr,col=colorRampPalette(c('green','black','red'))(1024),range=c(-6,6))
temp<-plot_pathway(pho_expr,type="bg",bg_col=col,text_col="white",magnify=1.2,specis='hsa',database_dir=sy
```

---

plot_profile *plot_profile*

Description

The function plot gene expression profiles on KEGG pathway maps

Usage

```
plot_profile(gene_expr, pathway_name, KEGG_database,
  groups, bg_col = "white", text_col = "black", line_col,
  border_col = "grey", text_cex = 0.25, magnify = 1,
  type = c("lines", "bg"), pathway_min = 5,
  genes_kept = c("foldchange", "first", "random", "var", "abs"),
  specis = "hsa", database_dir = getwd(), max_dist,
  lwd = 1.2)
```

Arguments

gene_expr	the matrix for gene expression, row.names should be NCBI gene ID, such as 67040, 93683
pathway_name	the specis id and KEGG pathway id, such as 'hsa00010'
KEGG_database	the matrix returned by function parse_XMLfile, which contains genes in this pathway, and their names, locations etc
groups	a character used to indicate expression values from different samples
bg_col	background color for gene rectangles in the pathway map
line_col	line color for expression in different samples in the pathway map, valid when type='lines'
text_col	the colors for text in the pathway map. A color matrix generated by function col_by_value can be used here
border_col	border color for gene rectangles in the pathway map. A color matrix generated by function col_by_value can be used here
text_cex	cex for text in the pathway map. A color matrix generated by function col_by_value can be used here
magnify	the coefficient used to magnify the gene rectangles
type	the type of pathway map visulization, could be 'bg' or 'lines'. Default is 'bg'. See also 'Details'
pathway_min	The pathways with number of annotated genes less than pathway_min would be ignored
genes_kept	methods used for choosing genes when several genes corresponding to one location in pathway map. Default is 'foldchange', which kept the gene with largest fold changes. 'first' kept the first gene. 'random' chosed gene random. 'var' kept the gene with largest variation. 'abs' kept the gene with largest absolute value

max_dist	The expression changes that represented by the distance from the bottom to the top of gene rectangle, valid when type='lines'. This param is used to ensure the dynamic changes of lines in different gene polygon represent equal variation. It would be calculated from the maximum changes of genes in this pathway by default. If max_dist=NA, then the lines would be plotted from top to bottom in each gene rectangle
lwd	The line width when type='lines'
specis	the specis id in KEGG database, 'hsa' means human, 'mmu' means mouse, 'rno' means rat, etc
database_dir	the directory where the XML files and png files are located

Details

There are two visualization methods to represent gene expression profiles: 'background' and 'lines'. The first one is applicable for analysis with only one sample or one type of data, which divides the gene polygon into several sub-polygons to represent different time points. And each sub-polygon has a specific background color to represent expression changes in that time point. The second method plots lines with different colors in the gene polygon to represent different samples or different types of data. The dynamic changes of lines mean the profiles of genes in different time points.

Value

a matrix containing genes maped in this pathway, and their names, expressions

Examples

```
XML2database<-parse_XMLfile(pathway_id="04110",specis="hsa",database_dir=system.file("extdata",package="
data(pro_pho_expr)
temp<-plot_profile(pro_pho_expr,pathway_name="hsa04110",KEGG_database=XML2database,line_col=c("brow
```

---

pro_pho_expr *expression profiles in proteome and phosphoproteome*

Description

This data set is from a previously published data of proteome and phosphoproteome analysis in different cell phase. The column 1-6 are proteome data and column 7-12 are phosphoproteome data in this data.frame. The 6 time points are G1, G1/S, Early S, Late S, G2, Mitosis.

Source

Olsen, J.V., et al. (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis, Sci Signal, 3, ra3.

Index