Package ‘domainsignatures’

March 22, 2017

Type Package

Title Geneset enrichment based on InterPro domain signatures

Version 1.34.0

Author Florian Hahne, Tim Beissbarth

Maintainer Florian Hahne <florian.hahne@novartis.com>

Description Find significantly enriched gene classifications in a list of functionally unde-
scribed genes based on their InterPro domain structure.

Depends R (>= 2.4.0), KEGG.db, prada, biomaRt, methods

Imports AnnotationDbi

License Artistic-2.0


LazyLoad yes

biocViews Annotation, Pathways, GeneSetEnrichment

NeedsCompilation no

R topics documented:

domainsignatures-package ................................................. 2
dataSource ............................................................ 2
getKEGGdata ............................................................ 3
getKEGGdescription ..................................................... 4
gseDomain .............................................................. 5
ipDataSource-class ...................................................... 6

Index 7
Description

Compute similarities to pathways for a set of entrezgene identifiers based on the InterPro domain signature.

Details

<table>
<thead>
<tr>
<th>Package</th>
<th>domainsignatures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>1.0</td>
</tr>
<tr>
<td>Date</td>
<td>2007-07-02</td>
</tr>
<tr>
<td>License</td>
<td>LGPL?</td>
</tr>
</tbody>
</table>

see help for `gseDomain` for details

Author(s)

Florian Hahne, Tim Beissbarth
Maintainer: Florian Hahne <f.hahne@dkfz.de>

dataSource

Constructor for annotation object

Description

This function creates the necessary annotation object of class `ipDataSource` containing the pathway and InterPro ID mappings.

Usage

dataSource(mapping, type = "generic")

Arguments

- `mapping`: A named list providing a mapping between entrezgene identifiers and arbitrary groupings of genes or pathways.
- `type`: The type of pathway. A character scalar.

Details

For genes without pathway membership, `NA` list items need to be included in `mapping`. The names of the list comprise the gene universe to test against. The function will access the ensembl biomart database in order to retrieve the necessary InterPro domain information.
getKEGGdata

Value
Object of class ipDataSource

Author(s)
Florian Hahne

See Also
gseDomain, getKEGGdata

Examples
if(interactive())
{
  grouping <- list("653361"="pw1", "pw2", "729230"="pw1",
  "415117"="pw3")
dataSource(grouping)
}

getKEGGdata
 Fetch KEGG annotations and InterPro domains

Description
Get all available KEGG annotations and InterPro domains for a set of entrezgene identifiers from
the KEGG annotation package and from the ensembl biomaRt.

Usage
getKEGGdata(universe=NULL, pathways=NULL, ensemblMart="hsapiens_gene_ensembl")

Arguments
universe Character vector of entrezgene identifiers. This is the global universe of genes
to test against.
pathways Optional character vector of KEGG pathway identifiers. This can be used in
order to test for over-representation of only a subset of all the available KEGG
pathways.
ensemblMart Character giving the type of Biomart to use. Defaults to Human Ensembl.

Details
This function is a wrapper around the KEGG annotation package and a customized query of the
ensembl biomaRt database. For the gene identifiers in universe and all or a selection of KEGG
pathways it will fetch the necessary information and create an object of class ipDataSource which
can later be used as input to gseDomain

Value
An object of class ipDataSource.
getKEGGdescription

Author(s)
Florian Hahne

See Also
gseDomain

Examples
if(interactive())
getKEGGdata()

description

Description
Get description of KEGG pathways from a list of KEGG identifiers.

Usage
getKEGGdescription(ids)

Arguments
ids Character vector of KEGG identifiers

Value
Character vector of KEGG descriptions

Author(s)
Florian Hahne

Examples
getKEGGdescription("hsa03050")
gseDomain

Geneset enrichment based on InterPro domain signatures

Description
Compute the similarity to pathways specified through dataSource for a set of entrezgene identifiers.

Usage
gseDomain(dataSource, geneset, n=10000, verbose=TRUE, samples=FALSE)

Arguments
- **dataSource**: Object of class `ipDataSource` containing pathway and InterPro domain mappings
- **geneset**: Character vector of entrezgene identifiers
- **n**: Number of subsampling iterations
- **verbose**: Toggle progress report
- **samples**: Logical indicating whether to return the similarity measures for all the resamples.

Details
Use this function to compute p-values for similarity of the domain signature of a gene set to all signatures of the pathways defined in dataSource. You should have created dataSource using either function `dataSource` or `getKEGGdata`.

Value
A list with items
- **similarity**: Named vector of similarity measures for each pathway
- **pvalue**: The p-values of similarity to each pathway. A named vector.

and optional item (if `samples=TRUE`)
- **dist**: A named list containing similarity measures for all the resamples

Author(s)
Florian Hahne

See Also
gseDomain

Examples
```r
## see Vignette of this package for examples how to use this function
```
ipDataSource-class

A class to store mapping information between genes, pathways and interPro domains

Description

This class represents the data necessary to run gseDomain.

Details

You should always create these objects using either one of the functions dataSource or getKEGGdata.

Creating Objects

Objects can be created using

```r
new('ipDataSource',
genes = ...., # Object of class character
pathways = .... # Object of class character
domains = .... # Object of class character
gene2Domains = .... # Object of class environment
path2Domains = .... # Object of class environment
type = ...., # Object of class character
)
```

or the functions dataSource or getKEGGdata.

Slots

- `genes`: Vector of unique entrezgene identifiers
- `pathways`: Vector of unique pathway identifiers
- `domains`: Vector of unique InterPro identifiers
- `gene2Domains`: Hash table mapping entrezgene IDs to Interpro IDs
- `path2Domains`: Hash table mapping pathway IDs to InterPro IDs
- `type`: Type of pathway. A character skalar
- `dims`: Numeric vector of dimensions

Methods

- `show` display summary.

Author(s)

Florian Hahne

See Also

dataSource or getKEGGdata
Index

*Topic IO
  dataSource, 2
  getKEGGdata, 3
  getKEGGdescription, 4

*Topic classes
  ipDataSource-class, 6

*Topic manip
  gseDomain, 5

*Topic package
  domainsignatures-package, 2
    dataSource, 2, 5, 6
    domainsignatures
      (domainsignatures-package), 2
      domainsignatures-package, 2
    getKEGGdata, 3, 3, 5, 6
    getKEGGdescription, 4
    gseDomain, 2–5, 5, 6
  ipDataSource (ipDataSource-class), 6
  ipDataSource-class, 6

  show, ipDataSource-method
    (ipDataSource-class), 6