

Package ‘AtlasRDF’

September 9, 2013

Type Package

Title Gene Expression Atlas query and gene set enrichment package.

Version 1.0

Date 2013-09-09

Author James Malone, Simon Jupp, Maryam Soleimani

Maintainer James Malone <malone@ebi.ac.uk>

Depends hash, SPARQL, methods

Description Query the Gene Expression Atlas RDF data at the European Bioinformatics Institute using genes, experimental factors (such as disease, cell type, compound treatments), pathways and proteins. Also contains a function to perform an enrichment of your gene list across Experimental Factor Ontology (EFO) using the Atlas background set.

License Apache 2.0

R topics documented:

AtlasRDF-package	2
doFishersEnrichment	3
doFishersEnrichmentForEnsemblIds	4
doFishersEnrichmentForGeneNames	6
drawHeatMapForAtlasExperiment	7
getAllEnsemblGenesForExFactor	8
getClassLabel	9
getExperimentsByDescription	9
getGenesForExperiment	10
getGeneUriFromEnsemblId	11
getGeneUriFromName	11
getLabel	12
getPathwaysFromGenesAndCondition	12
getSpeciesSpecificEnsemblGenesForExFactor	13
getTaxonURI	14
vizPvalues	14
Index	16

AtlasRDF-package *Gene Expression Atlas query and gene set enrichment package.*

Description

Query the Gene Expression Atlas RDF data at the European Bioinformatics Institute using genes, experimental factors (such as disease, cell type, compound treatments), pathways and proteins. Also contains a function to perform an enrichment of your gene list across Experimental Factor Ontology (EFO) using the Atlas background set.

Details

Package: AtlasRDF
Type: Package
Version: 1.0
Date: 2013-09-09
License: Apache 2.0

Author(s)

James Malone, Simon Jupp, Maryam Soleimani
Maintainer: James Malone <malone@ebi.ac.uk>

References

Gene Expression Atlas RDF - <http://wwwdev.ebi.ac.uk/rdf/services/atlas/> Gene Expression Atlas
- <http://www.ebi.ac.uk/gxa> Experimental Factor Ontology - <http://www.ebi.ac.uk/efo> Kapushesky
M, Adamusiak T, Burdett T, Culhane A, Farne A, Filippov A, Holloway E, Klebanov A, Kryvykh
N, Kurbatova N, Kurnosov P, Malone J, Melnichuk O, Petryszak R, Pultsin N, Rustici
G, Tikhonov A, Travillian RS, Williams E, Zorin A, Parkinson H, Brazma A. (2012) Gene Expres-
sion Atlas update—a value-added database of microarray and sequencing-based functional genomics
experiments. Nucleic Acids Res. 2012 Jan;40(Database issue):D1077-81.

Examples

```
###find experiments where the description contains a given search term
#experimentlist <- getExperimentsByDescription("breast carcinoma")

###get gene list for given experiment ID
#genelist <- getGenesForExperiment("E-GEOD-1561")

###get list of pathways associated with a gene for which the genes are differentially
#pathwayinfo <- getPathwaysFromGenesAndCondition("efo:EFO_0001359")
#pathwayinfo$pathwayname

###perform gene set enrichment using Atlas as background to find factors genes are en
#load the background files for the species of interest (requires downloading)
#load("human_bg/human_gene_list.RData")      #human_genelist_bg
```

```
#load("human_bg/human_factor_counts.RData")      #human_factor_counts

###creat or load your genelist of interest
#genelist <- c("GSTT1", "PLCE1", "XIST", "ENOSF-1", "AHR", "FSTL1", "YIPF5", "RBM5",

###get id for species 'human'
#taxon <-getTaxonURI("human")  #obo:NCBITaxon_9606

###do enrichment
#results <- doFishersEnrichmentForGeneNames(genelist, taxon, human_genelist_bg, human

###visualize results, filtering by p-value
#vizPvalues(results, "0.005")
```

doFishersEnrichment

Perform a gene set enrichment using gene list based on gene URIs.

Description

Perform an enrichment of an input gene list using the Atlas as a background set for a specified species. The genes are given as a set of gene URIs (e.g. <<http://identifiers.org/ensembl/ENSG00000229807>>). The results are a list of experimental factors for which these genes are enriched.

Usage

```
doFishersEnrichment(genelist, genelist_bg, genecounts)
```

Arguments

genelist	Variable genelist a vector of genes using gene URIs
genelist_bg	genelist_bg background list appropriate for the species of interest should be obtained from svn
genecounts	genecounts counts list appropriate for the species of interest should be obtained from svn

Value

Returns a list of enrichmentresult objects which are described as follows:

factoruri	the URI of the experimental factor
label	the human readable label for the experimental factor
p.value	the p.value for the Fisher's exact test for the gene list with this factor. This can be seen as evidence for this factor being highly enriched (low p value) or not (high p value) for the input gene list
estimate	an estimate of the odds ratio. Note that the conditional Maximum Likelihood Estimate (MLE) rather than the unconditional MLE (the sample odds ratio) is used. Only present in the 2 by 2 case.
alternative	a character string describing the alternative hypothesis.
null.value	the odds ratio under the null, or. Only present in the 2 by 2 case.

method the character string "Fisher's Exact Test for Count Data".

enrichedgenes list of the genes from the input gene list that were enriched for this factor as a list of Gene URIs. To get the gene names use the function getClassLabel

Author(s)

James Malone, European Bioinformatics Institute Simon Jupp, European Bioinformatics Institute
Maryam Soleimani, European Bioinformatics Institute

References

Agresti, A. (1990) Categorical data analysis. New York: Wiley. Pages 59-66.

Examples

```
###perform gene set enrichment using Atlas as background to find factors genes are enriched
#load the background files for the species of interest (requires downloading)
#load("human_gene_list.RData")    #human_genelist_bg
#load("human_factor_counts.RData")    #human_factor_counts

###creat or load your genelist of interest
genelist <- c("<http://identifiers.org/ensembl/ENSG00000184674>",
"<http://identifiers.org/ensembl/ENSG00000138193>",
"<http://identifiers.org/ensembl/ENSG00000229807>",
"<http://identifiers.org/ensembl/ENSG00000106546>",
"<http://identifiers.org/ensembl/ENSG00000163430>", "<http://identifiers.org/ensembl/ENSG00000177731>")

###do enrichment
#results <- doFishersEnrichment(genelist, human_genelist_bg, human_factor_counts)
```

doFishersEnrichmentForEnsemblIds

Perform a gene set enrichment using gene list based on common ENSEMBL gene IDs.

Description

Perform an enrichment of an input gene list using the Atlas as a background set for a specified species. The genes are given as a set of ENSEMBL gene IDs (e.g. "ENSG00000188554"). The results are a list of experimental factors for which these genes are enriched.

Usage

```
doFishersEnrichmentForEnsemblIds(geneids, taxon, genelist_bg, genecounts)
```

Arguments

geneids	Variable geneids a vector of genes using ENSEMBL gene IDs
taxon	The taxon is an ontology URI which can be found using the function <code>getTaxonURI()</code>
genelist_bg	genelist_bg background list appropriate for the species of interest should be obtained from svn
genecounts	genecounts counts list appropriate for the species of interest should be obtained from svn

Value

Returns a list of enrichmentresult objects which are described as follows:

factoruri	the URI of the experimental factor
label	the human readable label for the experimental factor
p.value	the p.value for the Fisher's exact test for the gene list with this factor. This can be seen as evidence for this factor being highly enriched (low p value) or not (high p value) for the input gene list
estimate	an estimate of the odds ratio. Note that the conditional Maximum Likelihood Estimate (MLE) rather than the unconditional MLE (the sample odds ratio) is used. Only present in the 2 by 2 case.
alternative	a character string describing the alternative hypothesis.
null.value	the odds ratio under the null, or. Only present in the 2 by 2 case.
method	the character string "Fisher's Exact Test for Count Data".
enrichedgenes	list of the genes from the input gene list that were enriched for this factor as a list of Gene URIs. To get the gene names use the function <code>getClassLabel</code>

Author(s)

James Malone, European Bioinformatics Institute Simon Jupp, European Bioinformatics Institute
Maryam Soleimani, European Bioinformatics Institute

References

Agresti, A. (1990) Categorical data analysis. New York: Wiley. Pages 59-66.

Examples

```
###perform gene set enrichment using Atlas as background to find factors genes are enriched
#load the background files for the species of interest (requires downloading)
#load("human_bg/human_gene_list.RData") #human_genelist_bg
#load("human_bg/human_factor_counts.RData") #human_factor_counts

###creat or load your genelist of interest
genelist <- c("ENSG00000184674", "ENSG00000138193", "ENSG00000229807", "ENSG00000254212",
              "ENSG00000163430", "ENSG00000118260", "ENSG00000086991")

###get id for species 'human'
#taxon <-getTaxonURI("human") #obo:NCBITaxon_9606
```

```
###do enrichment
#results <- doFishersEnrichmentForEnsemblIds(genelist, taxon, human_genelist_bg, huma
```

```
doFishersEnrichmentForGeneNames
```

Perform a gene set enrichment using gene list based on common gene names.

Description

Perform an enrichment of an input gene list using the Atlas as a background set for a specified species. The genes are given as a set of common gene names (e.g. Brca1). The results are a list of experimental factors for which these genes are enriched.

Usage

```
doFishersEnrichmentForGeneNames(genenames, taxon, genelist_bg, genecounts)
```

Arguments

genenames	Variable genenames a vector of genes using common gene names
taxon	The taxon is an ontology URI which can be found using the function <code>getTaxonURI()</code>
genelist_bg	genelist_bg background list appropriate for the species of interest should be obtained from svn
genecounts	genecounts counts list appropriate for the species of interest should be obtained from svn

Value

Returns a list of enrichmentresult objects which are described as follows:

factoruri	the URI of the experimental factor
label	the human readable label for the experimental factor
p.value	the p.value for the Fisher's exact test for the gene list with this factor. This can be seen as evidence for this factor being highly enriched (low p value) or not (high p value) for the input gene list
estimate	an estimate of the odds ratio. Note that the conditional Maximum Likelihood Estimate (MLE) rather than the unconditional MLE (the sample odds ratio) is used. Only present in the 2 by 2 case.
alternative	a character string describing the alternative hypothesis.
null.value	the odds ratio under the null, or. Only present in the 2 by 2 case.
method	the character string "Fisher's Exact Test for Count Data".
enrichedgenes	list of the genes from the input gene list that were enriched for this factor as a list of Gene URIs. To get the gene names use the function <code>getClassLabel</code>

Author(s)

James Malone, European Bioinformatics Institute Simon Jupp, European Bioinformatics Institute
Maryam Soleimani, European Bioinformatics Institute

References

Agresti, A. (1990) Categorical data analysis. New York: Wiley. Pages 59-66.

Examples

```
###perform gene set enrichment using Atlas as background to find factors genes are en
#load the background files for the species of interest (requires downloading)
#load("human_gene_list.RData")      #human_genelist_bg
#load("human_factor_counts.RData")   #human_factor_counts

###creat or load your genelist of interest
#genelist <- c("GSTT1", "PLCE1", "XIST", "ENOSF-1", "AHR", "FSTL1", "YIPF5", "RBM5",

###get id for species 'human'
taxon <-getTaxonURI("human")      #obo:NCBITaxon_9606

###do enrichment
#results <- doFishersEnrichmentForGeneNames(genelist, taxon, human_genelist_bg, human
```

drawHeatMapForAtlasExperiment

Produce a heatmap of differentially expressed genes vs conditions for an Atlas experiment.

Description

Draws a heatmap of genes that are differentially expressed beyond a given t stat significance against conditions they are diff expressed for for a given Atlas experiment.

Usage

```
drawHeatMapForAtlasExperiment(experimentid, tstatsignificance, endpoint)
```

Arguments

experimentid Atlas experiment ID which is to be visualized

tstatsignificance

The significance beyond which a gene should be included for visualisation. A significance value of '5' will only include genes with a t-stat of ≥ 5 or ≤ -5 .

endpoint

SPARQL endpoint to query data from. Default value is the live Atlas SPARQL endpoint but this can be overridden.

Value

matrix of gene names with t-statistics for level of differential expression for each factor (one factor per column)

Author(s)

Simon Jupp, James Malone

Examples

```
#data <- drawHeatMapForAtlasExperiment("E-MTAB-1670", 12)
```

```
getAllEnsemblGenesForExFactor
```

Get all ENSEMBL genes reported as differentially expressed for a given experimental factor.

Description

Get all ENSEMBL genes reported as differentially expressed for a given experimental factor. Returns the gene URI, ENSEMBL Id and common gene name.

Usage

```
getAllEnsemblGenesForExFactor(exfactor, limit, endpoint)
```

Arguments

exfactor	The URI of the experimental factor e.g. "efo:EFO_0001420"
limit	Limit the number of genes returned by the query to this, no limit specified or value of 0 returns all.
endpoint	SPARQL endpoint to query data from. Default value is the live Atlas SPARQL endpoint but this can be overridden.

Value

Data frame describing genes for the specified experimental factor with the following slots:

dbXref	the URI of the gene
geneName	the common name of the gene
ensemblid	the ENSEMBL id of the gene
propertyValue	the property value for this gene corresponding to the experimental factor in the query

Author(s)

James Malone, Simon Jupp

Examples

```
###get the genes for the experimental factor 'sepsis' (efo:EFO_0001420)
#results <- getAllEnsemblGenesForExFactor("efo:EFO_0001420")
#results[1,] #first gene result
```

getClassLabel	<i>Get label of a class based on the URI.</i>
---------------	---

Description

Get human readable label for any class given a URI.

Usage

```
getClassLabel(classURI, endpoint = "http://www.ebi.ac.uk/rdf/services/atlas/
```

Arguments

classURI	
endpoint	The location of the SPARQL endpoint - default value is live Atlas endpoint

Author(s)

James Malone, Simon Jupp

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==> Define data, use random,
##--      or do help(data=index) for the standard data sets.
```

getExperimentsByDescription	<i>Search for experiments in Atlas based on a string, e.g. liver, cancer etc.</i>
-----------------------------	---

Description

Find experiments in Atlas based on a search string such as a "liver" or "cancer". Returns the experiment IDs plus the text in which the search string was found.

Usage

```
getExperimentsByDescription(searchterm, limit = 0, endpoint = "http://www.ebi.ac
```

Arguments

searchterm
 limit
 endpoint The location of the SPARQL endpoint - default value is live Atlas endpoint

Author(s)

James Malone, Simon Jupp

Examples

```
##----- Should be DIRECTLY executable !! -----
##-- ==> Define data, use random,
##--      or do help(data=index) for the standard data sets.
```

```
getGenesForExperiment
```

Get all of the genes reported in an experiment, speicified by the Atlas experiment ID.

Description

Get all of the gene as reported within an experiment. The experiment is specified as an Atlas experiment ID e.g. E-GEOD-3865

Usage

```
getGenesForExperiment(experiment, endpoint = "http://www.ebi.ac.uk/rdf/services/
```

Arguments

experiment Atlas experiment ID.
 endpoint The location of the SPARQL endpoint - default value is live Atlas endpoint

Author(s)

James Malone, Simon Jupp

Examples

```
##----- Should be DIRECTLY executable !! -----
##-- ==> Define data, use random,
##--      or do help(data=index) for the standard data sets.
```

```
getGeneUriFromEnsemblId
```

Get gene URI from an ENSEMBL ID.

Description

Get the URI for an ENSEMBL gene based on the ENSEMBL ID.

Usage

```
getGeneUriFromEnsemblId(id, taxon, endpoint = "http://www.ebi.ac.uk/rdf/services
```

Arguments

id	ENSEMBL gene ID
taxon	The URI of the species of interest
endpoint	The location of the SPARQL endpoint - default value is live Atlas endpoint

Value

List of URIs for a given ENSEMBL ID - this is almost always one but it is feasible there are two so be aware.

Author(s)

James Malone, Simon Jupp

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==> Define data, use random,
##--    or do help(data=index) for the standard data sets.
```

```
getGeneUriFromName
```

Get URI of a gene based on the common gene name.

Description

Get the URI of an ENSEMBL gene based on a search by the common gene name.

Usage

```
getGeneUriFromName(genename, taxon, endpoint = "http://www.ebi.ac.uk/rdf/ser
```

Arguments

genename	
taxon	
endpoint	The location of the SPARQL endpoint - default value is live Atlas endpoint

Author(s)

James Malone, Simon Jupp

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==> Define data, use random,
##--      or do help(data=index) for the standard data sets.
```

getLabel

Get label of an entity based on the URI.

Description

Get label of an entity base on the URI

Usage

```
getLabel(uri, endpoint = "http://www.ebi.ac.uk/rdf/services/atlas/sparql")
```

Arguments

uri	URI of the class for which the label is required.
endpoint	The location of the SPARQL endpoint - default value is live Atlas endpoint

Author(s)

James Malone, Simon Jupp

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==> Define data, use random,
##--      or do help(data=index) for the standard data sets.
```

getPathwaysFromGenesAndCondition

Get pathways connected to genes which are differentially expressed for a user specified condition.

Description

Get list of pathway names that are connected to genes (via proteins which are encoded for by those genes) and which are reported as differentially expressed under a specified condition. The condition is specified as an ontology URI.

Usage

```
getPathwaysFromGenesAndCondition(condition, endpoint = "http://www.ebi.ac.uk/rdf")
```

Arguments

condition URI of a condition based on the EFO ontology class.
 endpoint The location of the SPARQL endpoint - default value is live Atlas endpoint

Author(s)

James Malone, Simon Jupp

Examples

```
##----- Should be DIRECTLY executable !! -----
##-- ==> Define data, use random,
##--      or do help(data=index) for the standard data sets.
```

```
getSpeciesSpecificEnsemblGenesForExFactor
```

Get ENSEMBL genes which have been differentially expressed for a given condition and species.

Description

Get the ENSEMBL gene IDs and names which are reported as differentially expressed in Atlas for a specified conditions and species.

Usage

```
getSpeciesSpecificEnsemblGenesForExFactor(exfactor, taxon, limit = 0, endpoint="")
```

Arguments

exfactor
 taxon
 limit
 endpoint The location of the SPARQL endpoint - default value is live Atlas endpoint

Author(s)

James Malone, Simon Jupp

Examples

```
##----- Should be DIRECTLY executable !! -----
##-- ==> Define data, use random,
##--      or do help(data=index) for the standard data sets.
```

getTaxonURI	<i>Get URI of a taxon class based on the common or taxonomic name.</i>
-------------	--

Description

Convenience method to the ontology URI, used in other queries, based on an input common or taxonomic name such as "human" or "Homo sapiens"

Usage

```
getTaxonURI (taxonName)
```

Arguments

taxonName	Common or taxonomic name
-----------	--------------------------

Author(s)

James Malone, Simon Jupp

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==> Define data, use random,
##--      or do help(data=index) for the standard data sets.
```

vizPvalues	<i>Visualize the results of an enrichment filtering for a specified p-value.</i>
------------	--

Description

Produce a plot of the most enriched experimental factors following an Atlas gene set enrichment. The cutoff allows a filter to be placed by p-value.

Usage

```
vizPvalues(resultset, cutoff = "0.05")
```

Arguments

resultset	A set of results following an AtlasRDF gene set enrichment.
cutoff	The p-value threshold at which to include a gene for visualization

Value

Plot of results

Author(s)

James Malone

Examples

```
##---- Should be DIRECTLY executable !! ----  
##-- ==> Define data, use random,  
##--      or do help(data=index) for the standard data sets.
```

Index

*Topic \textasciitildekw1

vizPvalues, 14

*Topic \textasciitildekw2

vizPvalues, 14

*Topic **package**

AtlasRDF-package, 2

AtlasRDF (*AtlasRDF-package*), 2

AtlasRDF-package, **2**

doFishersEnrichment, **3**

doFishersEnrichmentForEnsemblIds,

4

doFishersEnrichmentForGeneNames,

6

drawHeatMapForAtlasExperiment, **7**

getAllEnsemblGenesForExFactor, **8**

getClassLabel, **9**

getExperimentsByDescription, **9**

getGenesForExperiment, **10**

getGeneUriFromEnsemblId, **11**

getGeneUriFromName, **11**

getLabel, **12**

getPathwaysFromGenesAndCondition,

12

getSpeciesSpecificEnsemblGenesForExFactor,

13

getTaxonURI, **14**

vizPvalues, **14**