Package 'LEANR'

October 12, 2022

Type Package

Title Finds ``Local Subnetworks'' Within an Interaction Network which Show Enrichment for Differentially Expressed Genes

Version 1.4.9

Date 2016-11-11

Author Frederik Gwinner

Maintainer Frederik Gwinner <`frederik.gwinner@gmail.com`>

Description Implements the method described in ``Network-based analysis of omics data: The LEAN method'' [Gwinner Boulday (2016) <DOI:10.1093/bioinformatics/btw676>]
Given a protein interaction network and a list of p-values describing a measure of interest (as e.g. differential gene expression) this method computes an enrichment p-value for the protein neighborhood of each gene and compares it to a background distribution of randomly drawn p-values.
The resulting scores are corrected for multiple testing and significant hits are returned in tabular format.

License GPL-3

Depends R (>= 2.14), igraph(>= 0.7.1), foreach(>= 1.4.2)

Suggests knitr, doMC, rmarkdown, ROCR, testthat

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2016-11-12 15:47:01

R topics documented:

`LEANR-package` *Finds "local subnetworks" within an interaction network which show enrichment for differentially expressed genes*

Description

Implements the method described in "Network-based analysis of omics data: The LEAN method". Given a protein interaction network and a list of p-values describing a measure of interest (as e.g. differential gene expression) this method computes an enrichment p-value for the protein neighborhood of each gene and compares it to a background distribution of randomly drawn p-values. The resulting scores are corrected for multiple testing and significant hits are returned in tabular format.

Details

Package:	LEANR
Type:	Package
Version:	1.4.8
Date:	2016-11-11
License:	GPL-3

See help page of run.lean for a more detailed description of how to use this package. Type vignette("CCM-data") for an example showing the application of LEAN to the CCM knockout data set discussed in the paper. Type vignette("subnet-sim") for an example showing the application of LEAN to simulated subnetwork data discussed in the paper.

Author(s)

Frederik Gwinner

Maintainer: Frederik Gwinner <frederik.gwinner@gmail.com>

References

Gwinner et al., Network-based analysis of omics data: The LEAN method, Bioinformatics 2016

See Also

`run.lean` vignette("CCM-data") vignette("subnet-sim")

`CCM.pvals` *Gene p-value list derived from knock-out experiments of the three CCM genes*

Description

Gene p-value list derived from knock-out experiments of the three CCM genes CCM1, CCM2 and CCM3. Contains p-values obtained from a limma differential expression analysis of knock-out samples versus control samples (each done in triplicate).

Usage

```
data("CCM.pvals")
```

Format

Named list (CCM1,2,3) of named numericals (names = gene ids, values = limma p-values)

`g2` *igraph graph object used in examples for function run.lean.fromdata*

Description

igraph graph object used in examples for function run.lean.fromdata. Obtained by parsing the STRING v.91 murine interaction network and restricting it to proteins mappable to genes contained on the Affymetrix MouseGene v1.0 ST chip.

Usage

```
data("g2")
```

Format

The format is: IGRAPH UNW- 7342 63617 –

Source

STRING v9.1 Mouse filtered for confidence scores >= 0.9

`gene.annots` *Annotation for STRING protein Ids*

Description

Annotation table giving gene names and descriptions for each protein contained in the STRING network

Usage

```
data("gene.annots")
```

Format

A data frame with 7342 observations on the following 4 variables.

`ensembl_gene_id` a character vector

`mgi_symbol` a character vector

`entrezgene` a character vector

`description` a character vector

Details

Row.names of the data.frame are STRING protein Ids

Examples

```
data(gene.annots)
str(gene.annots)
```

`gene.list.scores` *Gene p-value list used in examples for function run.lean.fromdata*

Description

Gene p-value list used in examples for function run.lean.fromdata Contains p-values obtained from a limma differential expression analysis

Usage

```
data("gene.list.scores")
```

Format

The format is: Named num [1:7342] 0.772 0.813 0.979 0.841 0.607 ... - attr(*, "names")= chr [1:7342] "10090.ENSMUSP00000000001" "10090.ENSMUSP00000010205" "10090.ENSMUSP00000053818" "10090.ENSMUSP00000000153" ...

`get.ls.info` *Extract the genes of a "local subnetwork"" around a given protein*

Description

Extract the genes of a "local subnetwork"" around a given protein and present in tabular format

Usage

```
get.ls.info(prot_id, LEANres)
```

Arguments

`prot_id`	Protein id compatible with node names used in graph.
`LEANres`	LEAN result object (list) returned by <run.lean> or <run.lean.fromdata>

Author(s)

Frederik Gwinner

See Also

`run.lean`

`g_red` *igraph graph object used in unit tests*

Description

igraph graph object used in unit tests. Obtained by restricting the graph <g2> to the graph induced by randomly selecting 1500 genes.

Usage

```
data("g_red")
```

Format

The format is: IGRAPH UNW- 1500 2818 –

Source

STRING v9.1 Mouse filtered for confidence scores >= 0.9; radnomly reduced to 1500 genes and all interactions between them

---

`pvals_red` *Gene p-value list used in unit tests*

Description

Gene p-value list used in unit tests Contains p-values obtained on the CCM2 data for a radnom subselection of 1500 genes. To be used in conjunction with the network contained in <g_red>.

Usage

```
data("pvals_red")
```

Format

The format is: Named num [1:1500] 0.5091 0.4833 0.0454 0.0814 0.0324 ... - attr(*, "names")= chr [1:1500] "10090.ENSMUSP00000079341" "10090.ENSMUSP00000106951" "10090.ENSMUSP00000045284" "10090.ENSMUSP00000077744" ...

`run.lean` *Run the LEAN approach*

Description

Apply the LEAN approach to a given network and a list of pvalues

Usage

```
run.lean(ranking, network, ranked = F,
    add.scored.genes = F, keep.nodes.without.scores = F,
    verbose = F, n_reps = 10000, bootstrap = F, ncores = NULL)
```

Arguments

`ranking`	Either a file containing gene p-values or a named numerical vector of p-values with names matching node names used in the network
`network`	Either a file containing the network in sif format or an igraph graph object representing the network
`ranked`	whether to transform input p-values into a uniformly distributed list of p-values based on the genes' rank before p* calculation
`add.scored.genes`	whether to create one singleton node for each gene with a score but not occurring in the graph

`keep.nodes.without.scores`	whether to keep nodes of the graph that have no recorded score. For those nodes it is still possible to compute enrichment scores if at least one of their network neighbors has a recorded score.
`verbose`	whether to print additional status messages
`n_reps`	the number of samples each background distribution should consist of. Largely influences the run-time, but higher values needed for meaningful empirical pvalues !
`bootstrap`	whether to draw the pvalues of the background distributions with or without replacement
`ncores`	number of cores to be used in parallel computation. Default (NULL) leads to automatic guessing of max number of cores to be used (depending on operating system).

Value

A list object containing the results of the LEAN run. The list encompasses the following elements:

`restab`	Result table of applying LEAN to the real data
`randtab`	Result table of applying LEAN to a permuted p-value list
`indGraph`	igraph graph representing the input network after adapting it according to parameters <add.scored.genes>, <keep.nodes.without.scores> and the presence of gene scores in the input scores
`nhs`	The extracted local subnetworks. Encoded as a named (by protein/gene ids) list of igraph node indices detailing each evaluated local subnetwork
`gene.scores`	The gene p-values extracted from the input scores. Encoded as a numeric vector named with protein/gene ids

Author(s)

Frederik Gwinner

References

Gwinner et al., Network-based analysis of omics data: The LEAN method, MS submitted to Bioinformatics

See Also

LEANR-package

Examples

```
## Simple use case starting from a test network and p-value list
## Not run:
# compute LEAN p-values starting from a p-value file and a network file
rank_file<-system.file('extdata/pvals_red.txt.gz', package='LEANR')
net_file<-system.file('extdata/g_red.sif.gz', package='LEANR')
```

```
system.time(res<-run.lean(ranking=rank_file, network=net_file,
  add.scored.genes=T, verbose=T, n_reps=1000, ncores=3))

# compute LEAN p-values starting from a list of gene scores and a graph
data(pvals_red)
data(g_red)
system.time(res2<-run.lean(ranking=pvals_red, network=g_red, verbose=T,
   n_reps=1000, ncores=3))


## End(Not run)
```

`subnet.simulation` *Simulate subnetworks*

Description

Simulate subnetworks (also called modules) and gene p-values to be then used in a ROC performance evaluation study.

Usage

```
subnet.simulation(g, nmods=10, mod_lims=c(10,50), pval_scaling=0.1, mod_enrich_perc=0.5,
          spec='', prob_function=function(degs){degs/sum(degs)}, create.files=T)
```

Arguments

`g`	igraph graph representing the network in which subnetworks are supposed to be simulated
`nmods`	number of subnetworks/modules to simulate
`mod_lims`	minimum and maximum size (number of genes) of each module
`pval_scaling`	parameter value for <p_scale>
`mod_enrich_perc`	parameter value for <p_enrich>
`spec`	string, specifier appended to the created pvalue files (if create.files=T)
`prob_function`	probability function used for picking attachment point in iterative construction of subnetworks. defaults to preferential attachment based on node degree. To disable preferential attachment, use prob_function=function(degs)rep(1/length(degs),length(degs))
`create.files`	whether to write subnetwork simulation results to file so external approaches can be run and evaluated on them

Value

A list object containing the simulated subnetworks. The list encompasses the following elements:

`mods`	List of simulated modules/subnetworks. Each module is given by the igraph indices of the nodes contained in it.
`pvals`	Result table containing for each gene in the graph its simulated pvalue (column P.Value) and its association to subnetworks or background (column NodeType)
`pvalfile`	String containing the name of the file containing the equivalent information to <pvals> created in this run if create.files=T

Author(s)

Frederik Gwinner

References

Gwinner et al., Network-based analysis of omics data: The LEAN method, MS submitted to Bioinformatics

Examples

```
### See vignette("subnet-sim") for a use case.
```

`write.ls.to.sif` *Extract the "local subnetwork" around a given protein*

Description

Extract the "local subnetwork" around a given protein and write it to a Cytoscape-readable .sif file

Usage

```
write.ls.to.sif(prot_id, LEANres, outfile)
```

Arguments

`prot_id`	protein id compatible with node names used in graph g
`LEANres`	LEAN result object (list) returned by <run.lean> or <run.lean.fromdata>
`outfile`	character string describing the location of an output file. Should end in .sif to be able to load it in Cytoscape.

Author(s)

Frederik Gwinner

See Also

`run.lean`

Index