BUS

April 19, 2010

`BUS-package` *For network reconstruction.*

Description

This package can be used to compute associations among genes (gene-networks) or between genes and some external traits (i.e. clinical). [Function: BUS]

Both associations can be computed via correlation or mutual information (MI). [Functions: gene.similarity (for adjacency matrix for genes) and gene.trait.similarity (for association between genes and traits)]

Statistical significance of the association is computed for single and multiple hypotheses testing, using (i) random permutations method and (ii) MM-correction. [Functions: gene.pvalue, gene.trait.pvalue and MM.correction]

The package can handle data with missing values using bootstrapping methods to fill NAs. [Arguments: na.replica]

Details

Package:	BUS
Type:	Package
Version:	1.0.0
Date:	2008-12-09
License:	GPL-3

Author(s)

Yin Jin, Hesen Peng, Lei Wang and Christine Nardini

Maintainer: Yin Jin<golden.jinyin@gmail.com>

References

C. Nardini, H. Peng, L. Wang, L. Benini, M.D. Kuo, MM-Correction: Meta-analysis-Based Multiple Hypotheses Correction in Omic Studies, Springer CCIS, Vol 25, pp 242-255, 2008.

BUS *A wrapper function for matrices of p-value and predicted network*

Description

A wrapper function to calculate the computation of two types of similarities (correlation and mutual information) with two different goals: (i) identification of the statistically significant similarities among the activity of molecules sampled across different experiments (option Unsupervised, U), (ii) identification of the statistically significant similarities between such molecules and other types of information (clinical etc., option supervised, S) .

Usage

```
BUS(EXP,trait=NULL,measure,method.permut=2,method.correct,n.replica=400,net.trim
```

Arguments

EXP	Gene expression data in form of a matrix. Row stands for genes and column for experiments.
trait	Trait data in form of a matrix. The row stands for traits and column for experiments.
measure	Metric used to calculate similarity: "corr" for correlation, "MI" for mutual information.
method.permut	A flag to indicate which method is used to correct permutation p-values, default as 2. See gene.pvalue for details.
method.correct	Option for method to calculate p-value corrected for multiple hypothesis: "MM-correction" for MM-correction, "permutation" for corrected permutation, "both" for both MM-correction and permutation methods.
n.replica	Number of permutations; default value is 400.
net.trim	Method used to trim the network: "mrnet", "clr", "aracne" and "none" for no trim. Option "mrnet" infers a network using the maximum relevance/minimum redundancy feature selection method; option "clr" use the CLR algorithm; option "aracne" applies the data processing inequality to all triplets of nodes in order to remove the least significant edge in each triplet. These options come from the package minet. As these methods are only for mutual information, option "none" should be chosen if correlation is set to be the .
thresh	Threshold for significance of the corrected p-value, in option Unsupervised is used to trim the adjacency matrix (contains the results of the gene-gene association based on the chosen metric) and obtain a predicted gene interaction network. In the Supervised option, the network is not output, hence default is NULL.
nflag	A flag to indicate a gene-gene interaction case (Unsupervised) or a gene-trait interaction case (Supervised); 1 for Unsupervised and 2 for Supervised.

Value

`single.perm.p.value`
: A matrix of single p-values obtained with permutation method

`multi.perm.p.value`
: A matrix of corrected p-values obtained with permutation method

`MMcorrected.p.value`
: A matrix of multiple hypothesis corrected p-values obtained with MM-correction

`net.pred.MMcorrected`
: Predicted network obtained trimming non significant values, as they are computed with the MM-correction.

`net.pred.permut`
: Predicted network obtained trimming non significant values, as they are computed with the permutation based multiple-hypothesis correction.

Author(s)

Yin Jin, Hesen Peng, Lei Wang and Christine Nardini

See Also

`gene.pvalue`,`gene.trait.pvalue`,`pred.network`

Examples

```
data(copasi)
mat=as.matrix(copasi)[1:5,]
BUS(mat,measure="corr",method.correct="both",net.trim="none",thresh=0.05,nflag=1)
```

---

copasi *copasi data*

Description

This dataset is taken from Copasi2 (Complex Pathway Simulator), a software for simulation and analysis of biochemical networks. The system generates random artificial gene networks according to well-defined topological and kinetic properties. These are used to run in silico experiments simulating real laboratory micro-array experiments. Noise with controlled properties is added to the simulation results several times emulating measurement replicates, before expression ratios are calculated. This series consists of 150 artificial gene networks. Each network consists of 100 genes with a total of 200 gene interactions (on average each gene has 2 modulators).

Format

A data frame is size of 100x200, the 100 rows represent 100 genes and 20 columns for 200 experiments.

References

See http://www.comp-sys-bio.org/AGN/data.html for detailed information.

`gene.pvalue` *Calculates p-value for gene-gene interaction*

Description

Calculates p-value for the null hypothesis that there is no gene-gene interaction. For gene data with M genes, a p-value matrix under MxM single null hypotheses (each two genes have no interaction) is computed; besides, matrices with correct the p-values are output: corrected permutation method using a distribution of MxMxP (P number of permutations) null hypotheses tests (multi.perm.p.value) and using a meta-analysis based method (MMcorrected.p.value). p-values are calculated based for the adjacency matrix for gene-gene interaction computed by function "gene.similarity".

Usage

```
gene.pvalue(EXP,measure,net.trim,n.replica=400)
```

Arguments

`EXP`	Gene expression data in form of a matrix. Row stands for genes and column for experiments.
`measure`	Metric used to calculate similarity between genes: "corr" for correlation, "MI" for mutual information.
`net.trim`	Method used to trim the adjacency matrix: "mrnet", "clr", "aracne" and "none" for no trim. Option "mrnet" infers a network using the maximum relevance/minimum redundancy feature selection method; option "clr" use the CLR algorithm; option "aracne" applies the data processing inequality to all triplets of nodes in order to remove the least significant edge in each triplet, these options come from the package minet. As the above methods are only for mutual information, option "none" should be chosen if correlation is set to be the metric.
`n.replica`	Number of permutations; default value is 400.

Details

Normally, in a permutation method, we use the empirical distribution of some statistics to estimate the p-value. To get a simple p-value for no interaction between gene i and j, empirical distribution of a vector with length of P (number of replicates) is used; to correct for multiple hypothesis with permutations, an empirical distribution of a vector with length of PxM (M being the number of hypotheses tested) is used.

Value

`single.perm.p.value`
: A matrix of single p-values obtained with permutation method + beta distribution for extreme values (for MI) or obtained with the exact distribution computed directly by cor.test (for correlation)

`multi.perm.p.value`
: A matrix of corrected p-values obtained with permutation method

`MMcorrected.p.value`
: A matrix of multiple hypothesis corrected p-values obtained with MM-correction

Author(s)

Yin Jin, Hesen Peng, Lei Wang and Christine Nardini

See Also

gene.similarity

Examples

```
data(copasi)
mat=as.matrix(copasi)[1:5,]
gene.pvalue(mat,measure="MI",net.trim="mrnet")
```

---

gene.similarity *Calculate adjacency matrix for gene-gene interaction*

Description

Calculate an adjacency matrix for gene-gene interaction (using correlation/mutual information metric). For gene expression data with M genes and N experiments, the adjacency matrix is in size of MxM. It is optional to get a trimmed adjacency matrix according to the net.trim argument mrnet(),clrnet() or aracnenet() (from the minet package).

Usage

```
gene.similarity(EXP, measure,net.trim,na.replica = 50)
```

Arguments

EXP	Gene expression data in form of a matrix. Row stands for genes and column for experiments.
measure	Metric used to calculate similarity between genes: "corr" for correlation, "MI" for mutual information.
net.trim	Method used to trim the adjacency matrix: "mrnet", "clr", "aracne" and "none" for no trim. Option "mrnet" infers a network using the maximum relevance/minimum redundancy feature selection method; option "clr" use the CLR algorithm; option "aracne" applies the data processing inequality to all triplets of nodes in order to remove the least significant edge in each triplet, options from minet package. As the above methods are only for mutual information, option "none" should be chosen if correlation is set to be the metric.
na.replica	Impute method. Times of replicate for filling NANs; default value is 50. The (smooth) bootstrapping approach is used to give an estimation to missing value in the data.

Value

An adjacency matrix in size of MxM with rows and columns both standing for genes. Element in row i and column j indicates the similarity between gene i and gene j.

Author(s)

Yin Jin, Hesen Peng, Lei Wang and Christine Nardini.

Examples

```
data(copasi)
mat=as.matrix(copasi)[1:5,]
gene.similarity(mat,measure="corr",net.trim="none")
```

---

gene.trait.pvalue *Calculate p-value for gene-trait interaction*

Description

Calculates p-value for null hypothesis that there is no interaction between gene and trait. There are MxT interactions between M genes and T traits. Results are given with 4 possibilities 1 for single p-value, and 3 for different types of correction. p-values are calculated based for the adjacency matrix for gene-gene interaction computed by function "gene.trait.similarity".

Usage

```
gene.trait.pvalue(EXP,trait,measure,method.permut=2,n.replica=400)
```

Arguments

EXP	Gene expression data in form of a matrix. Row stands for genes and column for experiments.
trait	Trait data in form of matrix. The row stands for traits and column for experiments.
measure	Metric used to calculate similarity: "corr" for correlation, "MI" for mutual information.
method.permut	A flag to indicate which dimension is considered when a correcting for multiple hypotheses. 1 for traits correction, 2 for genes and 3 for both genes and traits. The default value is 2.
n.replica	Number of permutations; default value is 400.

Details

According to a permutation method, we use the empirical distribution of some statistics to estimate the p-value. For single p-value the empirical distribution is a vector of P (number of random replicates for each test) test values. It is then possible to correct p-value in different ways: method.permut = 1, it is the empirical distribution of a vector with length of TxP, corrects for the multiple traits tested; method.permut = 2, it is the empirical distribution of a vector with length of MxP, corrects for the multiple genes tested; method.permut = 3, it is empirical distribution of a vector with length of MxTxP, corrects for the multiple traits and genes tested.

Value

`single.perm.p.value`
: A matrix of single p-values obtained with permutation method + beta distribution for extreme values (for MI) or obtained with the exact distribution computed directly by cor.test (for correlation)

`multi.perm.p.value`
: A matrix of corrected p-values obtained with permutation method

`MMcorrected.p.value`
: A matrix of multiple hypothesis corrected p-values obtained with MM-correction

Author(s)

Yin Jin, Hesen Peng, Lei Wang and Christine Nardini

See Also

`gene.trait.similarity`

Examples

```
data(copasi)
mat=as.matrix(copasi)[1:5,]
gt<-matrix(rnorm(200),2,100)
gene.trait.pvalue(EXP=mat,trait=gt,measure="corr")
```

---

gene.trait.similarity

Calculate similarity for gene-trait interaction

Description

Calculate similarity for gene-trait interaction (using correlation/mutual information metric).

Usage

```
gene.trait.similarity(EXP, trait, measure, data.type=c("continuous","continuous"
```

Arguments

`EXP`	Gene expression data in form of a matrix. Row stands for genes and column for experiments.
`trait`	Trait data in form of matrix. The row stands for traits and column for experiments.
`measure`	Metric used to calculate similarity: "corr" for correlation, "MI" for mutual information.
`data.type`	A vector for the type of gene expression data and trait data. Each element in the vector has two options, "discrete" and "countinuous". The default value is c("continuous","continuous").
`na.replica`	Impute method. Times of replicate for filling NANs; default value is 50. The (smooth) bootstrapping approach is used to give an estimation to missing value in the data.

Value

A matrix, row stands for gene and column for trait. Element in row i and column j stands for the link between the gene i and trait j.

Author(s)

Yin Jin, Hesen Peng, Lei Wang and Christine Nardini

Examples

```
data(copasi)
mat=as.matrix(copasi)[1:5,]
gt<-matrix(rnorm(200),2,100)
gene.trait.similarity(EXP=mat,trait=gt,measure="MI")
```

MM.correction *MM-correction method*

Description

Correct single p-value for Multiple Hypothesis Testing

Usage

```
MM.correction(data)
```

Arguments

data	A square matrix of single p-values; each element is a p-value to be corrected for multi-hypothesis testing.

Value

A square matrix with the same size of data, elements are corrected p-values.

Author(s)

Yin Jin, Hesen Peng, Lei Wang and Christine Nardini

References

C. Nardini, H. Peng, L. Wang, L. Benini, M.D. Kuo, MM-Correction: Meta-analysis-Based Multiple Hypotheses Correction in Omic Studies, Springer CCIS, Vol 25, pp 242-255, 2008.

Examples

```
pmat=matrix(c(0.0000,0.6675,0.5800,0.7375,0.3800,0.6675,0.0000,0.1975,0.3075,0.1400,0.580
MM.correction(pmat)
```

`pred.network` *Predict the network*

Description

Predict the matrix of gene network, based on the similarity matrix and filtered according to a corrected p-value matrix.

Usage

```
pred.network(pM,similarity,thresh)
```

Arguments

`pM`	A corrected p-value matrix, a MxM matrix for significance of similarity among M genes.
`similarity`	A MxM matrix for similarity between genes.
`thresh`	Threshold for significance of the p-value.

Value

A MxM matrix of the predicted network, cell ij infers a link between gene i and j and set 0 when the p-value is not significant (no link).

Author(s)

Yin Jin, Hesen Peng, Lei Wang and Christine Nardini

Examples

```
data(copasi)
mat=as.matrix(copasi)[1:5,]
similarity=gene.similarity(mat,measure="MI",net.trim="mrnet")
pM=gene.pvalue(mat,measure="MI",net.trim="mrnet")$single.perm.p.value
pred.network(pM,similarity,thresh=0.05)
```

Index