Package 'MetaPath'

October 3, 2015

Type Package

Title Perform the Meta-Analysis for Pathway Enrichment Analysis (MAPE)

Version 1.0

Date 2015-09-28

Author Kui Shen and Geroge Tseng

Maintainer Kui Shen <kuishen@alumni.pitt.edu>

Depends R (>= 3.0.0),Biobase,GSEABase,genefilter,impute

Description Perform the Meta-analysis for Pathway Enrichment (MAPE) methods introduced by Shen and Tseng (2010). It includes functions to automatically perform MAPE_G (integrating multiple studies at gene level), MAPE_P (integrating multiple studies at pathway level) and MAPE_I (a hybrid method integrating MAEP_G and MAPE_P methods). In the simulation and real data analyses in the paper, MAPE_G and MAPE_P have complementary advantages and detection power depending on the data structure. In general, the integrative form of MAPE_I is recommended to use. In the case that MAPE_G (or MAPE_P) detects almost none pathway, the integrative MAPE_I does not improve performance and MAPE_P (or MAPE_G) should be used. Reference: Shen, Kui, and George C Tseng. Meta-analysis for pathway enrichment analysis when combining multiple microarray studies.Bioinformatics (Oxford, England) 26, no. 10 (April 2010): 1316-1323. doi:10.1093/bioinformatics/btq148. http://www.ncbi.nlm.nih.gov/pubmed/20410053.

License GPL (>= 2.0)

LazyLoad yes

NeedsCompilation no

Repository CRAN

Date/Publication 2015-10-03 07:57:54

R topics documented:

`MetaPath-package` *Perform the Meta-Analysis for Pathway Enrichment (MAPE) analysis by combining multiple genomic studies*

Description

Description: This R package was implemented to perform the Meta-analysis for Pathway Enrichment (MAPE) methods introduced by Shen and Tseng (2010). It includes functions to automatically perform MAPE_G (integrating multiple studies at gene level), MAPE_P (integrating multiple studies at pathway level) and MAPE_I (a hybrid method integrating MAEP_G and MAPE_P methods).

In the simulation and real data analyses in the paper, MAPE_G and MAPE_P have complementary advantages and detection power depending on the data structure. In general, the integrative form of MAPE_I is recommended to use. In the case that MAPE_G (or MAPE_P) detects almost none pathway, the integrative MAPE_I does not improve performance and MAPE_P (or MAPE_G) should be used.

References

Shen, Kui, and George C Tseng. Meta-analysis for pathway enrichment analysis when combining multiple microarray studies.Bioinformatics (Oxford, England) 26, no. 10 (April 2010): 1316-1323. doi:10.1093/bioinformatics/btq148. http://www.ncbi.nlm.nih.gov/pubmed/20410053.

`cor.func` *internal functions from Dr. Tibshirani's software package GSA*

Description

internal functions from Dr. Tibshirani's software package GSA

References

Bair, Eric, Trevor Hastie, Debashis Paul, and Robert Tibshirani. Prediction by Supervised Principal Components.Journal of the American Statistical Association 101, no. 473 (March 2006): 119-137. doi:10.1198/016214505000000628.

`cox.perm.sample` *internal functions from Dr. Tibshirani's software package GSA*

Description

internal functions from Dr. Tibshirani's software package GSA

References

Bair, Eric, Trevor Hastie, Debashis Paul, and Robert Tibshirani. Prediction by Supervised Principal Components.Journal of the American Statistical Association 101, no. 473 (March 2006): 119-137. doi:10.1198/016214505000000628.

`coxfunc` *internal functions from Dr. Tibshirani's software package GSA*

Description

internal functions from Dr. Tibshirani's software package GSA

References

Bair, Eric, Trevor Hastie, Debashis Paul, and Robert Tibshirani. Prediction by Supervised Principal Components.Journal of the American Statistical Association 101, no. 473 (March 2006): 119-137. doi:10.1198/016214505000000628.

coxscor *internal functions from Dr. Tibshirani's software package GSA*

Description

internal functions from Dr. Tibshirani's software package GSA

References

Bair, Eric, Trevor Hastie, Debashis Paul, and Robert Tibshirani. Prediction by Supervised Principal Components.Journal of the American Statistical Association 101, no. 473 (March 2006): 119-137. doi:10.1198/016214505000000628.

coxstuff *internal functions from Dr. Tibshirani's software package GSA*

Description

internal functions from Dr. Tibshirani's software package GSA

References

Bair, Eric, Trevor Hastie, Debashis Paul, and Robert Tibshirani. Prediction by Supervised Principal Components.Journal of the American Statistical Association 101, no. 473 (March 2006): 119-137. doi:10.1198/016214505000000628.

coxvar *internal functions from Dr. Tibshirani's software package GSA*

Description

internal functions from Dr. Tibshirani's software package GSA

References

Bair, Eric, Trevor Hastie, Debashis Paul, and Robert Tibshirani. Prediction by Supervised Principal Components.Journal of the American Statistical Association 101, no. 473 (March 2006): 119-137. doi:10.1198/016214505000000628.

`Enrichment_KS_gene` *internal functions*

Description

internal functions

`Enrichment_KS_sample` *internal functions*

Description

internal functions

`F.perm.sample` *internal functions*

Description

internal functions

`MAPE` *perform the Meta-Analysis for Pathway Enrichment (MAPE) analysis.*

Description

Description: This is the major function in the MetaPath package to implement the Meta-analysis for Pathway Enrichment (MAPE) methods introduced by Shen and Tseng (2010). The function automatically performs MAPE_G (integrating multiple studies at gene level), MAPE_P (integrating multiple studies at pathway level) and MAPE_I (a hybrid method integrating MAEP_G and MAPE_P methods).

In the simulation and real data analyses in the paper, MAPE_G and MAPE_P have complementary advantages and detection power depending on the data structure. In general, the integrative form of MAPE_I is recommended to use. In the case that MAPE_G (or MAPE_P) detects almost none pathway, the integrative MAPE_I does not improve performance and MAPE_P (or MAPE_G) should be used.

Usage

```
MAPE(arraydata,pathway.DB,resp.type=c('twoclass','multiclass','continuous','survival'),
  stat=c('maxP','minP','rth','Fisher'),rth.value=NULL, permutation=c('sample','gene'),
  nperm=500,size.min=15,size.max=500,knn.neighbors=10,qvalue.cal=c('permute','estimate'))
```

Arguments

arraydata	The arraydata is a list of microarray data sets. Each microarray data set can be either an ExpressionSet or a list. If the microarray data set is a list, then it includes five elements as follows: 1)x–exprs data 2)y– the phenotype of interests 3)z– censoring.status if applicable. 1 stands for the event occurred and 0 stands for censored. 4)geneid 5)samplename If the microarray data set is in an ExpressionSet format, the users need to 1) store the phenotype of interests in the slot 'label'. 2) store the censor data is the slot 'censoring.status' if applicable
pathway.DB	The pathway database in a GeneSetCollection format defined by GSEABase. The pathway database can be downloaded from Broad institue (http://www.broadinstitute.org/gsea). PLEASE use the function 'getGmt' provided in the GSEABase package to load the pathway database.
resp.type	The phenotype of interest.It is one of the four values: 'twoclass','multiclass','continuous', 'survival'.
stat	The meta-analysis statistic to be used to combine two studies. It is one of the four values: 'minP','maxP','rth','Fisher'.
rth.value	The value of the rth statistic if the meta-anlaysis statistic is 'rth'. For example, rth.value=0.6.
permutation	The options for using sample permutation or gene permutation when performing enrichment analysis. it is one of the two values: 'gene' and 'sample'. The default option is sample permutation.
nperm	Number of permutations to be performed.
size.min	The minimum size of pathways to be considered. The default value is 15.
size.max	The maximum size of pathways to be considered. The default value is 500.
knn.neighbors	Number of neighbors to be used in the knn imputation method(default=10)
qvalue.cal	The method to calculate the q-values. The default method is to calcuate the q-values based on the permutation method. If qvalue.cal='estimate', the q-values were estimated based on the Storey's method.

Value

The qvalue and pvalue of each pathway.

Author(s)

Kui Shen and George C Tseng.

References

Shen, Kui, and George C Tseng. Meta-analysis for pathway enrichment analysis when combining multiple microarray studies. Bioinformatics (Oxford, England) 26, no. 10 (April 2010): 1316-1323. doi:10.1093/bioinformatics/btq148. http://www.ncbi.nlm.nih.gov/pubmed/20410053.

Examples

```
## Not run:
library(MetaPath)
data(MAQC)
data(pathway.DB)
## Supposed we are interested in the ER related pathways, we first store the ER
information in the slot 'label'. Then perform MAPE on this data set.
MAQC[[1]]$label=MAQC[[1]]$ER_status
MAQC[[2]]$label=MAQC[[2]]$ER_status
nperm=10 ## nperm was set to 10 to save the computational time. The default value is 500.
MAPE.sample.obj=MAPE(arraydata=MAQC,pathway.DB=pathway.DB,resp.type="twoclass",stat='maxP',
rth.value=NULL,nperm=nperm,permutation='gene',size.min=15,size.max=500)
cutoff=.1
subset(MAPE.sample.obj$qvalue,MAPE_I<=cutoff)
plotMAPE(MAPE.sample.obj,cutoff,MAPE.method='MAPE_I')

## End(Not run)
```

`MAPE_G_gene_KS` *internal functions*

Description

internal functions

`MAPE_G_sample_KS` *internal functions*

Description

internal functions

`MAPE_I_KS` *internal functions*

Description

internal functions

MAPE_P_gene_KS *internal functions*

Description

internal functions

MAPE_P_sample_KS *internal functions*

Description

internal functions

MAQC *The data sets from MAQC project.*

Description

This is the microarray data sets from MAQC project.

References

Popovici, Vlad, Weijie Chen, Brandon G Gallas, Christos Hatzis, Weiwei Shi, Frank W Samuelson, Yuri Nikolsky, et al. Effect of training-sample size and classification difficulty on the accuracy of genomic predictors. Breast cancer research : BCR 12, no. 1 (January 2010): R5. doi:10.1186/bcr2468. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2880423&tool=pmcentrez&rendertype=al

Examples

```
data(MAQC)
```

pathway.DB *An example of pathway database.*

Description

This data set is an example of gene set database in a GeneSetCollection format defined by GSEABase. This database is the C2 collection of Molecular Signatures Database provided by Broad institute(http://www.broadinstitute.org/gsea).

Usage

```
data(pathway.DB)
```

plotMAPE *Plot MAPE outcomes*

Description

This function will plot two figures. The first figure is the Venn diagram to show the overlapped enriched pathways identified by MAPE_G, MAPE_P and MAPE_I. The second figure is the heatmap of the q-values of enriched pathways.

Usage

```
plotMAPE(MAPE.obj, cutoff, MAPE.method = c("MAPE_I", "MAPE_P", "MAPE_G"))
```

Arguments

MAPE.obj	The output of MAPE.
cutoff	The q-value cutoff.
MAPE.method	The MAPE method of interest.

Value

A heatmap of q-values of enriched pathways will be plotted. When plot the heatmap, if the MAPE.method is MAPE_I, it will plot the q-values of enriched pathways for each individual study and q-values computed by three MAPE methods. if the MAPE.method is MAPE_P, it will plot the q-values of enriched pathways for each individual study and q-values computed by the MAPE_P method. if the MAPE.method is MAPE_G, it will plot the q-values of enriched pathways for each individual study and q-values computed by the MAPE_G method.

Examples

```
## Not run:
plot.MAPE(MAPE.obj, cutoff=0.05, MAPE.method = "MAPE_I")

## End(Not run)
```

pqvalues.compute *internal functions*

Description

internal functions

`reg.perm.sample` *internal functions*

Description

internal functions

`Tperm.sample` *internal functions*

Description

internal functions

Index