How To Use MPINet

Yanjun Xu, Chunquan Li and Xia Li

July 27, 2013

Contents

1 Overview

This vignette demonstrates how to easily use the MPINet package. This package can identify pathways related with studied conditon (e.g. dysregulated pathways related with a specific disease) via global weighted human metabolite network, which considers both the global non-equivalence of metabolites in pathway and the bias existing in metabonomic experiment technology.

2 Identifying biological pathways via global weighted human metabolite network

The section introduces our pathway analysis based on metabolite set via global weighted human metabolite network for identifying biological pathways associated with studied condition. MPINet uses a network-based approach to identify pathways by considering both the global non-equivalence of metabolites in pathway and the bias existing in metabonomic experiment technology. Firstly, we construct a global edge weighted human metabolite network. We calculate the strength of connection value between each metabolite pair in the network. Then, we evaluate the CGNB score of each metabolite in pathway(see the section 2.1). Finally, the scores are used to calculate pathway weight which is used in the Wallenius' noncentral hypergeometric test to evaluate the significance of the pathway by using the function `identifypathway`(see the section 2.2).

2.1 Calculating the CGNB scores of metabolites in pathway

The the monotonic spline model was used to evaluate the score values of metabolites by integrating the global non-equivalence scores and the initial bias scores of metabolites. The function `getPSS` can calculate the CGNB score values of metabolites based on the inputting interest metabolites. The following commands can calculate the CGNB score.

```
> #example 1
> ##########get example data
> risk<-GetExampleData(dataset="prostate")
> ###########calculate the CGNB score
> pss<-getPSS(risk ,plot=F)
> CGNBscore<-pss[,"CGNB"]
> names(CGNBscore)<-rownames(pss)
> ##########print the CGNB score of some metabolites to screen
> head(CGNBscore)

 11953968   9548588    420804   5281997    440744    124148
0.9956771 0.9956771 0.9956771 0.9956771 0.9956771 0.9956771

> #example 2
> #get example data from file
> risk<-read.table(paste(system.file(package="MPINet"),"/localdata/prostate.txt",sep=""),
+ header=F,sep="\t","\"")
> ####convert the data to a character vector
> risk<-as.character(risk[[1]])
> ###########calculate the CGNB score
> pss<-getPSS(risk ,plot=F)
> CGNBscore<-pss[,"CGNB"]
> names(CGNBscore)<-rownames(pss)
> ##########print the CGNB score of some metabolites to screen
> head(CGNBscore)

 11953968   9548588    420804   5281997    440744    124148
0.9956771 0.9956771 0.9956771 0.9956771 0.9956771 0.9956771
```

2.2 Identifying pathways related with studied condition

The function `identifypathway` can identify pathways associated with studied condition. The result is a list. (i)If the argument `method` is 'MPINet', it includes the following elements: 'pathwayName', 'annComponentList', 'annComponentNumber', 'annBgComponentList', 'annBgNumber', 'componentNumber', 'bgNumber', 'pvalue', 'fdr', 'InWeight', 'weight', 'anncompinNetworkNum', 'anncompinNetworkList', 'riskcompinNetworkNum', 'riskcompinNetworkList'. They correspond to pathway name, the submitted metabolites annotated to a pathway, numbers of submitted metabolites annotated to a pathway, the background metabolites annotated to a pathway, numbers of background metabolites annotated to a pathway, numbers of submitted metabolites, numbers of background metabolitess, p-value of the Wallenius' noncentral hypergeometric test, Benjamini-Hochberg fdr values, the mean score value of metabolites in pathway, the final weight of pathway, numbers of the submitted metabolites annotated to a pathway and in the global human metabolite network, the submitted metabolites annotated to a pathway and in the global human metabolite network, numbers of submitted metabolites in the global human metabolite network, submitted metabolites in the global human metabolite network. When the argument `pathType` is 'KEGG', the 'pathwayId' element is also included, which is the pathway identifier in KEGG. When the argument `pathType` is not 'KEGG', the 'pathsource' element is also included, which stands for the source of pathway. (ii)If the argument `method` is 'Hyper', it includes the same elements as (i), but not includes the following elements: 'InWeight', 'weight', 'anncompinNetworkNum', 'anncompinNetworkList', 'riskcompinNetworkNum', 'riskcompinNetworkList'. To save the results, the `list` can be converted to the `data.frame` by the function `printGraph`(see the section 2.3).

```
> #example 1
> #### get the metastatic prostate cancer interesting metabolite data set
```

```
> risk<-GetExampleData(dataset="prostate")
> #### integrate the non-equivalence of metabolites and the character of
> #### differential metabolites by the monotonic spline model
> pss<-getPSS(risk,plot=F)
> #identify dysregulated pathways
> anncpdpre<-identifypathway(risk,pss,pathType="KEGG",method="MPINet",annlim=1,bglim=6)
> #convert ann to data.frame
> result<-printGraph(anncpdpre,pathType="KEGG",method="MPINet")
> head(result)

   pathwayId                                          pathwayName
1 path:00330                      Arginine and proline metabolism
2 path:00232                                 Caffeine metabolism
3 path:00380                               Tryptophan metabolism
4 path:01040              Biosynthesis of unsaturated fatty acids
5 path:00120                       Primary bile acid biosynthesis
6 path:00130 Ubiquinone and other terpenoid-quinone biosynthesis
  annComponentRatio annBgRatio      weight       pvalue          fdr
1             10/92    89/4994 0.220476474 3.528808e-12 2.117285e-10
2              3/92    21/4994 0.005586403 1.457222e-09 4.371667e-08
3              3/92    80/4994 0.002731896 1.089850e-08 2.179699e-07
4              3/92    49/4994 0.005168286 1.622847e-08 2.434271e-07
5              2/92    47/4994 0.001598571 9.615872e-07 1.153905e-05
6              2/92    74/4994 0.001180029 1.323536e-06 1.287355e-05
  annComponentinNetRatio
1                  10/85
2                   3/85
3                   2/85
4                   2/85
5                   2/85
6                   2/85

> #example 2
> ########get example data from file
> risk<-read.table(paste(system.file(package="MPINet"),"/localdata/prostate.txt",sep=""),
+ header=F,sep="\t","\"")
> ####convert the data to a character vector
> risk<-as.character(risk[[1]])
> pss<-getPSS(risk,plot=F)
> #identify dysregulated pathways
> anncpdpre<-identifypathway(risk,pss,pathType="KEGG",method="MPINet",annlim=1,bglim=6)
> #convert ann to data.frame
> result<-printGraph(anncpdpre,pathType="KEGG",method="MPINet")
> head(result)

   pathwayId                                          pathwayName
1 path:00330                      Arginine and proline metabolism
2 path:00232                                 Caffeine metabolism
3 path:00380                               Tryptophan metabolism
4 path:01040              Biosynthesis of unsaturated fatty acids
5 path:00120                       Primary bile acid biosynthesis
6 path:00130 Ubiquinone and other terpenoid-quinone biosynthesis
```

```
  annComponentRatio annBgRatio      weight       pvalue          fdr
1             10/92    89/4994 0.220476474 3.528808e-12 2.117285e-10
2              3/92    21/4994 0.005586403 1.457222e-09 4.371667e-08
3              3/92    80/4994 0.002731896 1.089850e-08 2.179699e-07
4              3/92    49/4994 0.005168286 1.622847e-08 2.434271e-07
5              2/92    47/4994 0.001598571 9.615872e-07 1.153905e-05
6              2/92    74/4994 0.001180029 1.323536e-06 1.287355e-05
  annComponentinNetRatio
1                  10/85
2                   3/85
3                   2/85
4                   2/85
5                   2/85
6                   2/85
```

```
> #example 3
> #### get the metastatic prostate cancer interesting metabolite data set
> risk<-GetExampleData(dataset="prostate")
> pss<-getPSS(risk,plot=F)
> #identify dysregulated Reactome and KEGG pathways
> anncpdpre<-identifypathway(risk,pss,pathType=c("KEGG","Reactome"),
+                 method="MPINet",annlim=1,bglim=6)
> #convert ann to data.frame
> result<-printGraph(anncpdpre,pathType=c("KEGG","Reactome"),method="MPINet")
> head(result)
```

```
                                          pathwayName pathsource
1             Metabolic pathways - Homo sapiens (human)       KEGG
2                                          Metabolism   Reactome
3                                Biological oxidations   Reactome
4 Arginine and proline metabolism - Homo sapiens (human)       KEGG
5               ABC transporters - Homo sapiens (human)       KEGG
6                 Metabolism of lipids and lipoproteins   Reactome
  annComponentRatio annBgRatio     weight       pvalue          fdr
1             49/92  1040/4994 0.01211964 4.943871e-97 1.354621e-94
2             48/92   683/4994 0.04884447 2.033639e-76 2.786086e-74
3             19/92   219/4994 0.13108141 4.496308e-24 4.106628e-22
4             10/92    89/4994 0.01532967 1.258226e-23 8.618848e-22
5             17/92    79/4994 0.28284487 2.650334e-23 1.452383e-21
6             13/92   255/4994 0.03580393 7.804007e-21 3.563830e-19
  annComponentinNetRatio
1                  48/85
2                  46/85
3                  19/85
4                  10/85
5                  17/85
6                  12/85
```

2.3 Print the results of annotation and identification

The function `printGraph` can convert the result list of the function `identifypathway` to the `data.frame`. A `data.frame` of the identification results. (i)If the argument `method` is 'MPINet', it includes the following

elements: 'pathwayName', 'annComponentRatio', 'annBgRatio', 'weight', 'pvalue', 'fdr', 'annComponentList', 'annBgComponentList', 'annComponentinNetRatio', 'anncompinNetworkList', 'riskcompinNetworkList'. The 'annComponentRatio' is the ratio of the annotated metabolites. For example, 30/1000 means that 30 metabolites in 1000 interesting metabolites are annotated to this pathway. The 'annBgRatio' is the ratio of background metabolites. For example, 10/4994 means that 10 of the 4994 background metabolites are annotated to this pathway. The 'annComponentinNetRatio' indicates the ratio of annotated metabolites in the global human metabolite network. The 'annComponentList' and 'annBgComponentList' are the annotated metabolites and the annotated background metabolites. The 'anncompinNetworkList' and 'riskcompinNetworkList' are the annotated metabolites in network and the interesting metabolites in network. (ii)If the argument `method` is 'Hyper', it includes the following elements: 'pathwayName', 'annComponentRatio', 'annBgRatio', 'pvalue', 'fdr', 'annComponentList', 'annBgComponentList'. When the argument `pathType` is 'KEGG', the 'pathwayID' is included. When the the argument `pathType` is not 'KEGG', the 'pathsource' is included. Detailed information is provided in the function `identifypathway`. The following commands use the function `printGraph` to convert the result to `data.frame`.

```
> #example 1
> #### get the metastatic prostate cancer interesting metabolite data set
> risk<-GetExampleData(dataset="prostate")
> #### integrate the global non-equivalence of metabolites and the character of
> ####differential metabolites by the monotonic spline model
> pss<-getPSS(risk,plot=F)
> #identify dysregulated pathways
> anncpdpre<-identifypathway(risk,pss,pathType="KEGG",method="MPINet",annlim=1,bglim=6)
> #convert ann to data.frame
> result<-printGraph(anncpdpre,pathType="KEGG",method="MPINet")
> #print part of the results to screen
> head(result)
```

```
   pathwayId                                        pathwayName
1 path:00330                     Arginine and proline metabolism
2 path:00232                                Caffeine metabolism
3 path:00380                              Tryptophan metabolism
4 path:01040             Biosynthesis of unsaturated fatty acids
5 path:00120                      Primary bile acid biosynthesis
6 path:00130 Ubiquinone and other terpenoid-quinone biosynthesis
  annComponentRatio annBgRatio      weight       pvalue          fdr
1             10/92    89/4994 0.220476474 3.528808e-12 2.117285e-10
2              3/92    21/4994 0.005586403 1.457222e-09 4.371667e-08
3              3/92    80/4994 0.002731896 1.089850e-08 2.179699e-07
4              3/92    49/4994 0.005168286 1.622847e-08 2.434271e-07
5              2/92    47/4994 0.001598571 9.615872e-07 1.153905e-05
6              2/92    74/4994 0.001180029 1.323536e-06 1.287355e-05
  annComponentinNetRatio
1                  10/85
2                   3/85
3                   2/85
4                   2/85
5                   2/85
6                   2/85
```

```
> result1<-printGraph(anncpdpre,pathType="KEGG",method="MPINet",detail=TRUE)
> #example 2
```

```
> #### get the metastatic prostate cancer interesting metabolite data set
> risk<-GetExampleData(dataset="prostate")
> pss<-getPSS(risk,plot=F)
> #identify dysregulated pathways
> anncpdpre<-identifypathway(risk,pss,pathType="Reactome",method="MPINet",annlim=1,bglim=6)
> #convert ann to data.frame
> result<-printGraph(anncpdpre,pathType="Reactome",method="MPINet")
> #print part of the results to screen
> head(result)

                                pathwayName pathsource annComponentRatio
1                                Metabolism   Reactome             48/92
2                      Biological oxidations   Reactome             19/92
3 Transmembrane transport of small molecules   Reactome             34/92
4       SLC-mediated transmembrane transport   Reactome             31/92
5      Metabolism of lipids and lipoproteins   Reactome             13/92
6  Metabolism of amino acids and derivatives   Reactome             21/92
  annBgRatio     weight       pvalue          fdr annComponentinNetRatio
1   683/4994 0.02801184 6.974538e-88 1.422806e-85                  46/85
2   219/4994 0.07517393 1.443495e-28 1.472365e-26                  19/85
3   169/4994 1.00440559 1.370296e-27 9.318011e-26                  33/85
4   150/4994 1.00632502 2.117426e-25 1.079887e-23                  30/85
5   255/4994 0.02053321 6.073538e-24 2.478003e-22                  12/85
6   187/4994 0.25080347 2.732173e-23 9.289388e-22                  21/85

> result1<-printGraph(anncpdpre,pathType="Reactome",method="MPINet",detail=TRUE)
```

3 Session Info

The script runs within the following session:

```
R version 3.0.1 (2013-05-16)
Platform: i386-w64-mingw32/i386 (32-bit)

locale:
[1] LC_COLLATE=C
[2] LC_CTYPE=Chinese_People's Republic of China.936
[3] LC_MONETARY=Chinese_People's Republic of China.936
[4] LC_NUMERIC=C
[5] LC_TIME=Chinese_People's Republic of China.936

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] MPINet_1.0     mgcv_1.7-24    BiasedUrn_1.05

loaded via a namespace (and not attached):
[1] Matrix_1.0-12   grid_3.0.1      lattice_0.20-15 nlme_3.1-109
[5] tools_3.0.1
```

References

[Li *et al.*, 2009] Li, C., et al. (2009) Subpathwayminer: A Software Package for Flexible Identification of Pathways. Nucleic Acids Res, 37, e131.

[Young *et al.*, 2005] Young, M.D., Wakefield, M.J., Smyth, G.K. and Oshlack, A. (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol, 11, R14.