How To Use SubpathwayGMir

Li Feng, Chunquan Li and Xia Li

May 20, 2015

Contents

1 Overview

This vignette demonstrates how to easily use the SubpathwayGMir package. This package can implement the identification of Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic subpathways mediated by microRNAs (miRNAs), by topologically locating miRNAs and genes within reconstructed KEGG metabolic pathway graphs, which emmbedded by miRNAs through integrating miRNA-target interactions verified by low-throughput experiments. (1) This package provides the `GetK2riData` to return verified miRNA-target interactions, which collected from four databases, such as TarBase(v5.0), miRecords(v4.0), miR2Disease and miRTarBase. (see the section 2). (2) This package provides the `getInteGraphList` function to reconstruct KEGG metabolic pathways by embedding miRNAs into direct and/or undirect KEGG metabolic pathway graphs, these embedded miRNAs have verified targets within pathways.(see the section 3). (3) This package provides the `getLocSubGraph` function to locate miRNA-mediated metabolic subpathways by topologically analyzing the "lenient distance" of miRNAs and genes, based on reconstructed pathways.(see the section 4). (4) This package provides the `identifyGraph` function to identify the significantly enriched metabolic subpathways, based on located subpathways.(see the section 5). (5) This package provides the `GetK2riData` function to get variable data

in current environment.(see the section 6). (6) This package provides the `updateOrgEnvir` function to updata the organism-specific environment variables.(see the section 7).

2 The experimentally verified miRNA-target interactions

We can use function `GetK2riData` to return verified miRNA-gene interactions, which are collected from four databases, namely TarBase(v5.0), miRecords(v4.0), miR2Disease and miRTarBase. We process these datasets into a uniform format. The final dataset contains seven columns, such as "SourceDB", "Species", "miRNA", "target", "LowHTExps", "Experiments" and "PMID". The value of column "SourceDB" is one of "TarBase(v5.0)", "miRecords(v4.0)", "miR2Disease" and "miRTarBase", which represents where this relation pair was derived from. Besides, this dataset can supports six organisms, such as cel(caenorhabditis elegans), dre(Danio rerio), dme(Drosophila melanogaster), hsa(Homo sapiens), mmu(Mus musculus) and rno(Rattus norvegicus). The column "Experiments" describes what kind of experiments valided this relations and the column "LowTHExps" represents whether this relation have been valided by low-throughput experiments or not. The column "PMID" provides PubMed identifers for the references of relations.

```
> # get verified miRNA-target interactions
> expMir2Tar <- GetK2riData("expMir2Tar")
> # view first six rows of data
> expMir2Tar[1:6,]
```

```
      SourceDB         miRNA   Gene Species LowTHExps
279  miRecords   cel-miR-273  die-1     cel       YES
286 miRTarBase cel-miR-35-3p lin-23     cel       YES
287 miRTarBase cel-miR-35-3p  gld-1     cel       YES
303 miRTarBase cel-miR-84-5p let-60     cel       YES
304 miRTarBase  cel-lin-4-5p lin-14     cel       YES
305 miRTarBase  cel-lin-4-5p lin-14     cel       YES
                              Experiments     PMID
279                      GFP Activity Assay 15306811
286      Luciferase reporter assay//qRT-PCR 21691303
287      Luciferase reporter assay//qRT-PCR 21691303
303                      GFP reporter assay 15766527
304 Immunofluorescence//LacZ reporter assay  8252622
305    qRT-PCR//Western blot//Northern blot 19155321
```

3 Reconstruct KEGG metabolic pathways

We can use function `getInteGraphList` to return the integrated KEGG metabolic pathway graph list. We first convert KEGG metabolic pathways to direct/undirect graphs with genes as nodes, then reconstructed pathways by linking miRNAs to targets within it.

3.1 Embed miRNAs to direct KEGG metabolic pathway graphs

The function `getInteGraphList` can integrate miRNAs to direct KEGG metabolic pathway graphs. With integrated graph list, we can offer the additional interested miRNAs and/or genes sets to identify the condition-specific metabolic pathways mediated by miRNAs.

```
> # get hsa-specificd miRNA-target interactions
> expMir2Tar <- GetK2riData("expMir2Tar")
```

```
> row1 <- which(expMir2Tar[["LowTHExps"]]=="YES")
> row2 <- which(expMir2Tar[["Species"]]=="hsa")
> relations <- unique(expMir2Tar[intersect(row1,row2),c(2:3)])
> # get direct metabolic pathway graphs
>  DirectGraphList <- GetK2riData("MetabolicGEGEEMGraph")
> # get reconstructed direct pathway graph list
>  DirectInteGraphList <- getInteGraphList(DirectGraphList, relations)
```

The following commands can show the embedded pathways with genes and miRNAs as nodes.

```
> # visualize the reconstructed direct pathway
> plotGraph(DirectInteGraphList[[1]],layout=layout.random)
```

Figure 1 shows the reconstructed direct Glycolysis / Gluconeogenesis metabolic pathway.

3.2 Embed miRNAs to undirect KEGG metabolic pathway graphs

The function `getInteGraphList` can integrate miRNAs into undirect KEGG metabolic pathway graphs with genes as nodes. With integrated graph list, we can offer the additional interested miRNAs and/or genes sets to identify the condition-specific pathways mediated by miRNAs.

```
> # get undirect metabolic pathway graphs
> UndirectGraphList <- GetK2riData("MetabolicGEGEUEMGraph")
> # get reconstructed undirect pathway graph list
> UndirectInteGraphList <- getInteGraphList(UndirectGraphList, relations)
```

The following commands can show the reconstructed pathway graph with genes and miRNAs as nodes.

```
> # visualize the reconstructed undirect pathway
> plotGraph(UndirectInteGraphList[[1]],layout=layout.random)
```

Figure 2 shows the reconstructed undirect Glycolysis / Gluconeogenesis metabolic pathway.

4 Locate KEGG metabolic subpathways

We can use function `getLocSubGraph` to locate metabolic subpathways by topologically analyzing the "lenient distance" of miRNAs and/or genes based on reconstructed pathways.

4.1 Based on reconstructed direct KEGG metabolic pathways

The function `getLocSubGraph` can locate metabolic subpathways based on reconstructed direct KEGG metabolic pathways.

```
> # get user-interested miRNAs and genes
> moleculeList <- c(getBackground(type="gene")[1:1000],
+                getBackground(type="miRNA")[1:2000])
> # get located direct subpathways
> DirectSubGraphList <- getLocSubGraph(moleculeList,DirectInteGraphList,
+                       type="gene_miRNA",n=1,s=10)
```

The following commands can show the located subpathway graph with genes and miRNAs as nodes.

```
> # visualize the located direct pathway
> plotGraph(DirectSubGraphList[[1]],layout=layout.random)
```

Figure 3 shows the located direct purine metabolic subpathway.

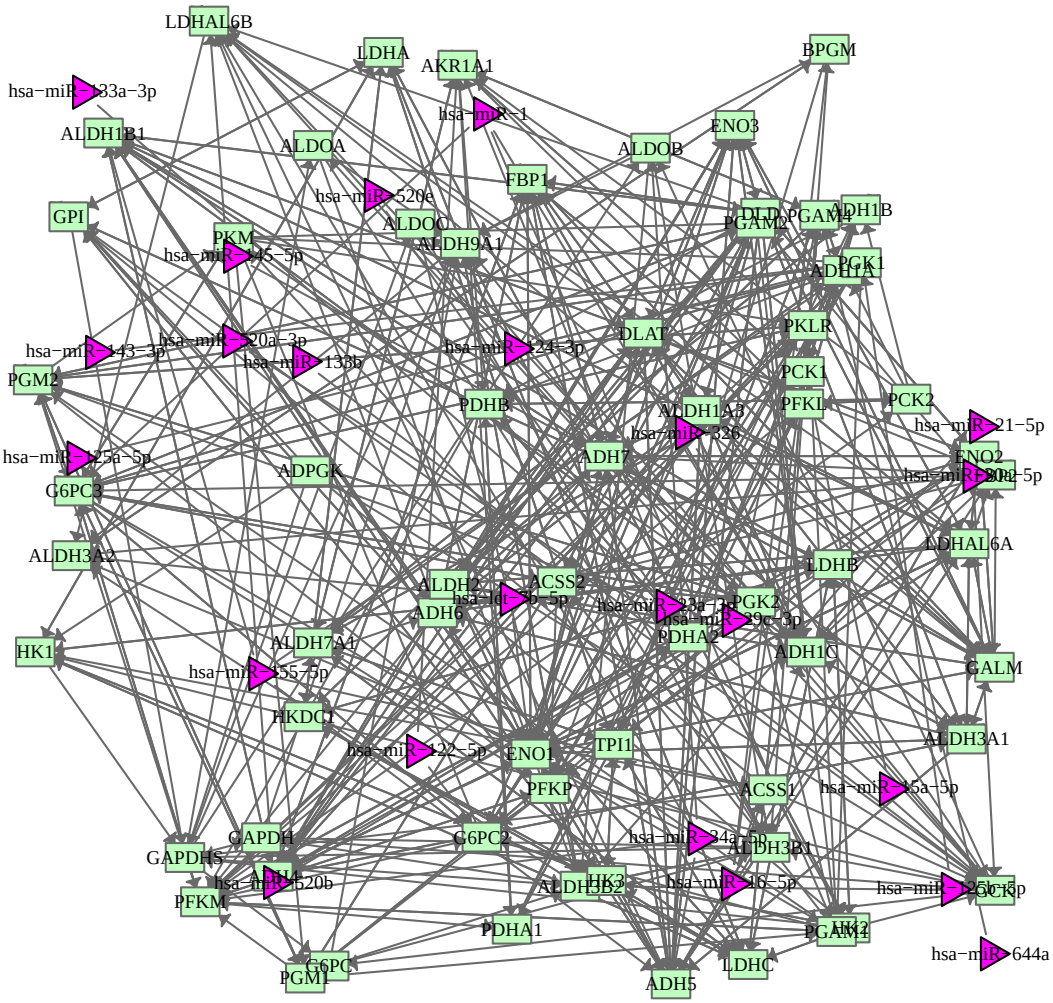

Figure 1: The visualization of reconstructed direct Glycolysis / Gluconeogenesis metabolic pathway.

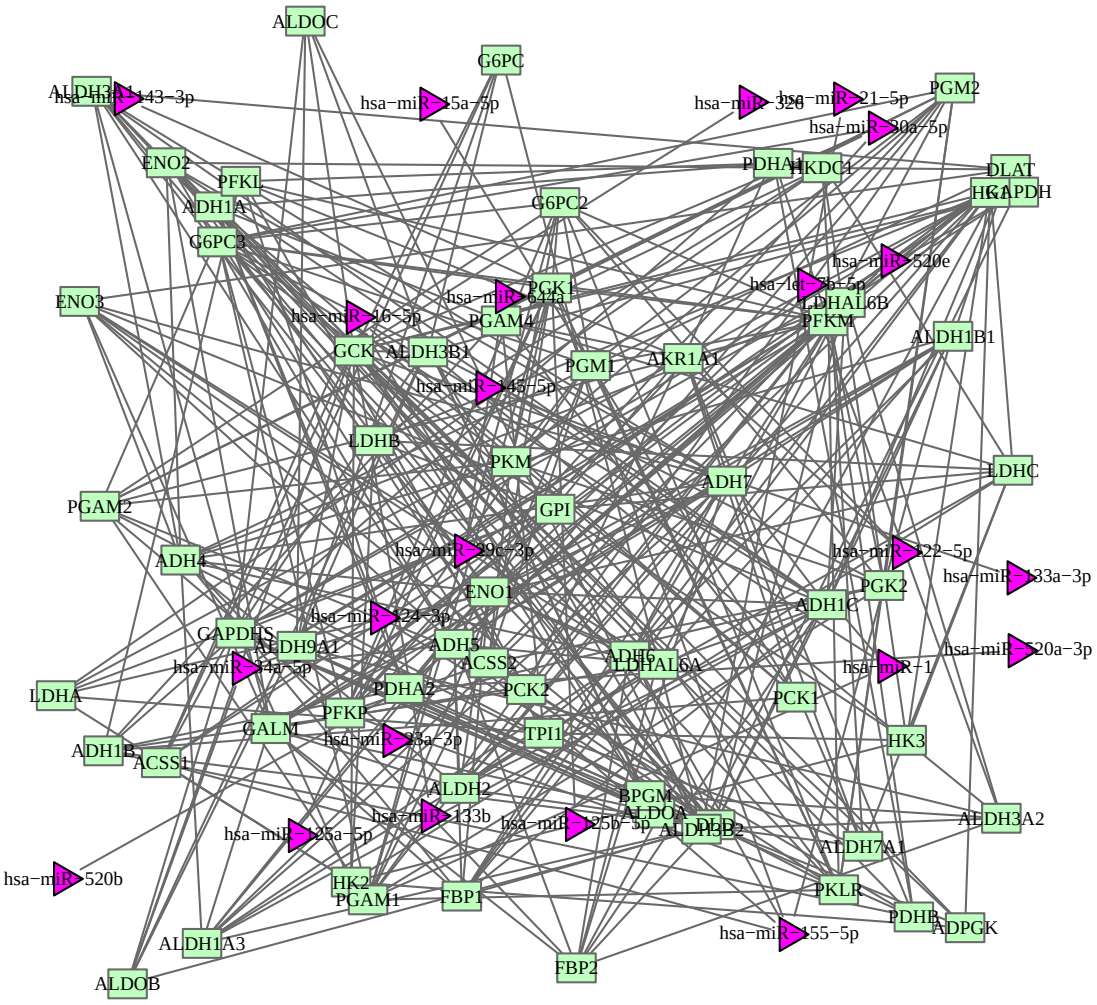

Figure 2: The visualization of reconstructed undirect Glycolysis / Gluconeogenesis metabolic pathway.

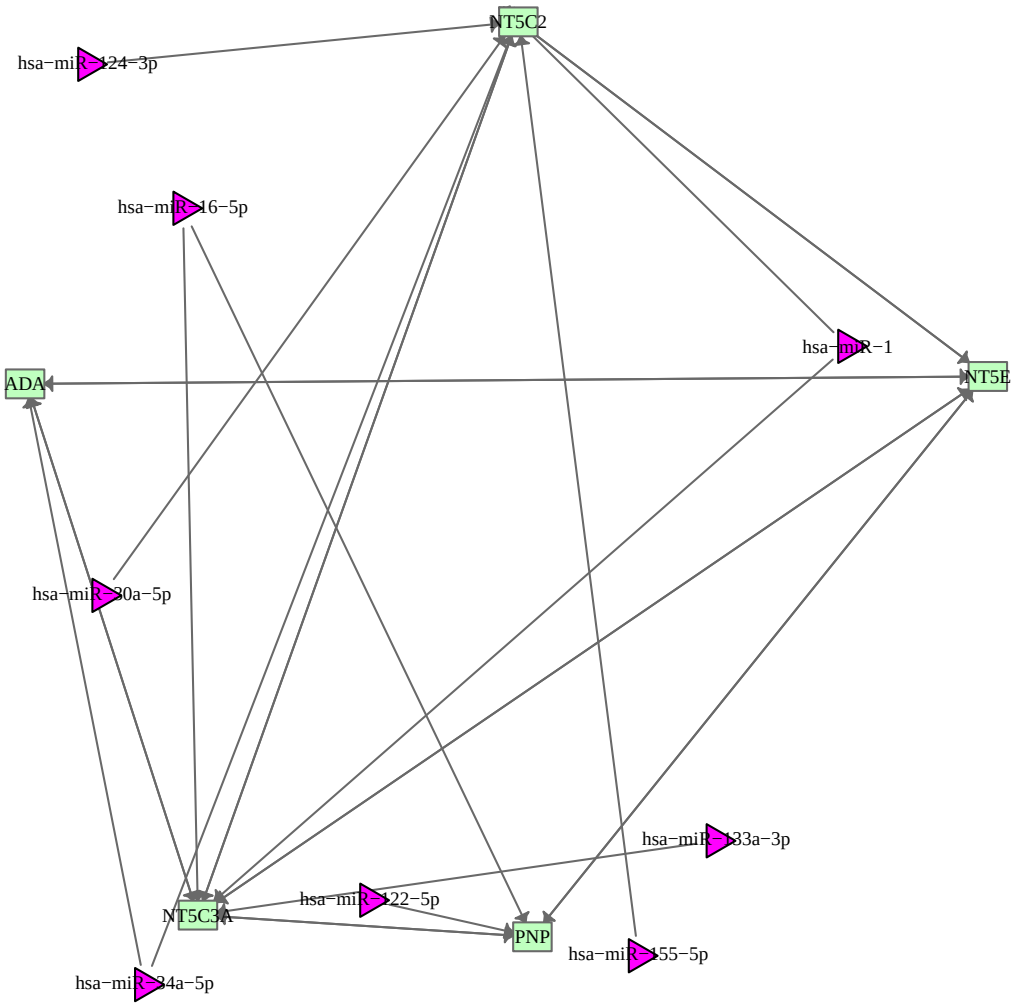

Figure 3: The visualization of located direct purine metabolic subpathway.

4.2 Based on reconstructed undirect KEGG metabolic pathways

The function `getLocSubGraph` can locate subpathways based on reconstructed undirect pathways.

```
> # get located undirect subpathways
> UnDirectSubGraphList <- getLocSubGraph(moleculeList,UndirectInteGraphList,
+                       type="gene_miRNA",n=1,s=10)
```

The following commands can show the located subpathway graph with genes and miRNAs as nodes.

```
> # visualize the located undirect pathway
> plotGraph(UnDirectSubGraphList[[6]],layout=layout.random)
```

Figure 4 shows the located undirect purine metabolic subpathway.

5 Identify the significantly enriched subpathways

We can use function `identifyGraph` to identify the significantly enriched subpathways based on located direct/undirect metabolic subpathways.

5.1 Based on located direct KEGG metabolic subpathways

The function `identifyGraph` can identify the significantly enriched subpathways based on located direct metabolic subpathways.

```
> # identify significant direct subpathways
> ann <- identifyGraph(moleculeList,DirectSubGraphList,type="gene_miRNA")
> result <- printGraph(ann,detail=TRUE)
> # view the result
> head(result[,c(1:2,5:6)])
```

```
     pathwayId                                    pathwayName       pvalue
1 path:00230_1                             Purine metabolism 1.992183e-07
2 path:00520_1 Amino sugar and nucleotide sugar metabolism 1.992183e-07
           fdr
1 1.992183e-07
2 1.992183e-07
```

5.2 Based on located undirect KEGG metabolic subpathways

The function `getLocSubGraph` can identify the significantly enriched subpathways based on located undirect metabolic subpathways.

```
> # identify significant undirect subpathways
> ann <- identifyGraph(moleculeList,UnDirectSubGraphList,type="gene_miRNA")
> result <- printGraph(ann,detail=TRUE)
> # view the result
> head(result[,c(1:2,5:6)])
```

```
     pathwayId                        pathwayName       pvalue          fdr
1 path:00562_2      Inositol phosphate metabolism 0.000000e+00 0.000000e+00
2 path:00230_1                  Purine metabolism 9.683032e-12 1.113549e-10
3 path:00270_2 Cysteine and methionine metabolism 9.957679e-11 7.634221e-10
```

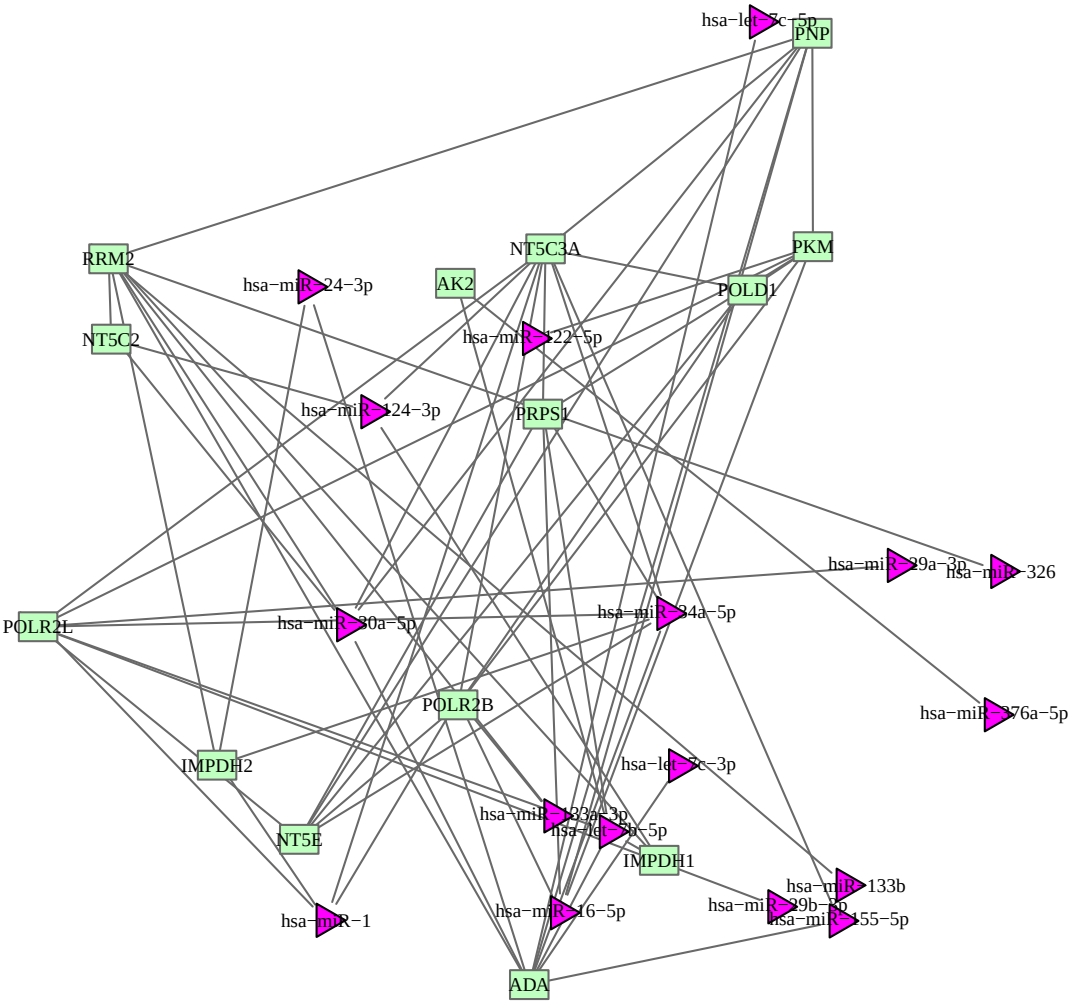

Figure 4: The visualization of located undirect purine metabolic subpathway.

```
4 path:00010_3       Glycolysis / Gluconeogenesis 2.587495e-09 1.487810e-08
5 path:00480_1             Glutathione metabolism 7.775946e-09 3.576935e-08
6 path:00240_1              Pyrimidine metabolism 1.223046e-08 4.688342e-08
```

```
> # save the result
> write.table(head(result),"result.txt",sep="\t",col.names=TRUE,row.names=FALSE)
```

6 Get the current environment variables

We can use function `GetK2riData` to obtain variable datas in current environment.

```
> # get verified miRNA-target interactions
> expMir2Tar <- GetK2riData(K2riData="expMir2Tar")
> # get the background of miRNAs
> BGMiRNA <- GetK2riData(K2riData="BGMiRNA")
> # get the background of genes
> BGGene <- GetK2riData(K2riData="BGGene")
>
```

7 Update the organism-specific environment variables

We can use function `updateOrgEnvir` to update the organism-specific environment variables.

```
> # update the cel-specific environment variables
>  updateOrgEnvir("cel")
```

```
[1] "Update the current organism : cel"
[1] "Note that the programming may be time consumming!"
[1] "Download relations between gene and symbol."
[1] "Download relations between KEGG gene and pathway"
[1] "Download background of miRNAs"
[1] "Download background of direct KEGG metabolic pathways"
[1] "Download background of undirect KEGG metabolic pathways"
```

```
> # show the current environment variables
>  ls(k2ri)
```

```
 [1] "BGGene"                    "BGMiRNA"
 [3] "CEL_MetabolicGEGEEMGraph"  "CEL_MetabolicGEGEUEMGraph"
 [5] "DME_MetabolicGEGEEMGraph"  "DME_MetabolicGEGEUEMGraph"
 [7] "DRE_MetabolicGEGEEMGraph"  "DRE_MetabolicGEGEUEMGraph"
 [9] "HSA_MetabolicGEGEEMGraph"  "HSA_MetabolicGEGEUEMGraph"
[11] "MMU_MetabolicGEGEEMGraph"  "MMU_MetabolicGEGEUEMGraph"
[13] "MetabolicGEGEEMGraph"      "MetabolicGEGEUEMGraph"
[15] "RNO_MetabolicGEGEEMGraph"  "RNO_MetabolicGEGEUEMGraph"
[17] "expMir2Tar"                "gene2path"
[19] "gene2symbol"               "miRNA2Org"
```

```
> # show the background of miRNAs
>  k2ri$BGMiRNA[1:3]
```

```
[1] "cel-let-7-5p" "cel-let-7-3p" "cel-lin-4-5p"
```

```
>
```

8 Session Info

The script runs within the following session:

```
R version 3.0.2 (2013-09-25)
Platform: x86_64-pc-linux-gnu (64-bit)

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=zh_CN.UTF-8        LC_COLLATE=C
 [5] LC_MONETARY=zh_CN.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=zh_CN.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=zh_CN.UTF-8 LC_IDENTIFICATION=C

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

other attached packages:
[1] SubpathwayGMir_1.0 igraph_0.7.1       XML_3.98-1.1

loaded via a namespace (and not attached):
[1] tools_3.0.2
```

References

[Antonov *et al.*, 2008] Antonov, A.V., et al. (2008) Kegg Spider: Interpretation of Genomics Data in the Context of the Global Gene Metabolic Network. Genome Biol, 9, R179.

[Csardi and Nepusz, 2006] Csardi, G. and Nepusz, T. (2006) The igraph software package for complex network research. InterJournal, Complex Systems, 1695.

[Draghici *et al.*, 2007] Draghici, S., et al. (2007) A Systems Biology Approach for Pathway Level Analysis. Genome Res, 17, 1537-1545.

[Guimera and Nunes Amaral, 2005] Guimera, R. and Nunes Amaral, L.A. (2005) Functional Cartography of Complex Metabolic Networks. Nature, 433, 895-900.

[Huber *et al.*, 2007] Huber, W., et al. (2007) Graphs in Molecular Biology. BMC Bioinformatics, 8 Suppl 6, S8.

[Hung *et al.*, 2010] Hung, J.H., et al. (2010) Identification of Functional Modules That Correlate with Phenotypic Difference: The Influence of Network Topology. Genome Biol, 11, R23.

[Kanehisa *et al.*, 2006] Kanehisa, M., et al. (2006) From Genomics to Chemical Genomics: New Developments in Kegg. Nucleic Acids Res, 34, D354-357.

[Koyuturk *et al.*, 2004] Koyuturk, M., et al. (2004) An Efficient Algorithm for Detecting Frequent Subgraphs in Biological Networks. Bioinformatics, 20 Suppl 1, i200-207.

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

[Li *et al.*, 2013] Li, C., et al. (2013) Subpathway-GM: identification of metabolic subpathways via joint power of interesting genes and metabolites and their topologies within pathways. Nucleic acids research, 41, e101.

[Smart *et al.*, 2008] Smart, A.G., et al. (2008) Cascading Failure and Robustness in Metabolic Networks. Proc Natl Acad Sci U S A, 105, 13223-13228.

[Strimmer, 2008] Strimmer, K. (2008) fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics, 24, 1461-1462.

[Vergoulis *et al.*, 2012] Vergoulis, T., et al. (2012) TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucleic acids research, 40, D222-229.

[Xiao *et al.*, 2009] Xiao, F., et al. (2009) miRecords: an integrated resource for microRNA-target interactions. Nucleic acids research, 37, D105-110.

[Hsu *et al.*, 2011] Hsu, S.D., et al. (2011) miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic acids research, 39, D163-169.

[Jiang *et al.*, 2009] Jiang, Q., et al. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic acids research, 37, D98-104.