How To Use SubpathwayLNCE

Xinrui Shi, Chunquan Li, Xia Li

November 20, 2024

Contents

1 Overview

This vignette demonstrates how to easily use the SubpathwayLNCE package. This package can implement the identification of Kyoto Encyclopedia of Genes and Genomes (KEGG) signal subpathways competitively regulated by long nocoding RNAs (lncRNAs), by topologically locating lncRNAs and genes within reconstructed KEGG signal pathway graphs, which emmbedded by lncRNAs based on ceRNA theory. (1) This package provides the `getExampleData` to return example data and environment variables. (see the section 2). (2) This package provides the `getInteUMGraph` function to reconstruct KEGG signal pathways by embedding lncRNAs into undirect KEGG signal pathway graphs.(see the section 3). (3) This package provides the `getLocSubGraph` function to locate lncRNAs competitively regulated signal subpathways by topologically analyzing the "lenient distance" of lncRNAs and genes, based on reconstructed pathways.(see the section 4). (4) This package provides the `identifyGraphW` function to identify the significantly enriched signal 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) This package provides the `updateOrgEnvir` function to updata the organism-specific environment variables.(see the section **??**).

2 get candidate lncRNA-mRNA interaction

We can use function `getExampleData` to return example data and environment variables, such as the data of candidate lncRNA-mRNA interaction, the data of undirect KEGG metabolic pathway graphs with genes as nodes.

```
> #obtain the data for candidate lncRNA-mRNA interaction.
> interaction<-GetExampleData(exampleData="pp")
> # view first six rows of data
> interaction[1:6,]

          LncEns         GeneEns
1 ENSG00000189149 ENSG00000142192
2 ENSG00000214870 ENSG00000142192
3 ENSG00000237697 ENSG00000142192
4 ENSG00000212694 ENSG00000142192
5 ENSG00000244625 ENSG00000142192
6 ENSG00000247796 ENSG00000142192

> #obtain the data for undirect KEGG metabolic pathway graphs with genes as nodes
>  g2<-GetExampleData(exampleData="g2")
> #obtain example data of mathed mRNA-lncRNA expression profiles
>  #GeneExp<-GetExampleData(exampleData="GeneExp")
>  #LncExp<-GetExampleData(exampleData="LncExp")
```

3 Reconstruct KEGG signal pathways

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

3.1 Get the co-express lncRNA-mRNA interactions

The function `getLncGenePairs` can calculated co-expression coefficient for any pair of relations in the candidate LncRNA-mRNA interaction based on matched LncRNA and mRNA expression profiles, those relations had reached a significant positive threshold were retained.

```
> #obtain example data of mathed mRNA-lncRNA expression profiles
>  GeneExp<-GetExampleData(exampleData="GeneExp")
>  LncExp<-GetExampleData(exampleData="LncExp")
> #calculated co-expression coefficient,the significant positive threshold is 0.025
>  LncGenePairs<-getLncGenePairs(GeneExp,LncExp,a=0.025)
> #obtain the data for undirect KEGG metabolic pathway graphs with genes as nodes
>  g2<-GetExampleData(exampleData="g2")
> # get reconstructed undirect pathway graph list
>  #interUMGraph<-getInteGraphList(g2,LncGenePairs)
>
```

3.2 Embed competitively regulated lncRNAs to undirect KEGG signal pathway graphs

The function `getInteGraphList` can competitively regulated lncRNAs into undirect KEGG signal pathway graphs with genes as nodes. With integrated graph list, we can offer the additional interested genes sets to identify the condition-specific pathways competitively regulated by lncRNAs.

```
> #obtain the data for undirect KEGG metabolic pathway graphs with genes as nodes
>   g2<-GetExampleData(exampleData="g2")
>  #obtain example data of mathed mRNA-lncRNA expression profiles
```

```
>  #GeneExp<-GetExampleData(exampleData="GeneExp")
>  #LncExp<-GetExampleData(exampleData="LncExp")
> #calculated co-expression coefficient,the significant positive threshold is 0.025
>  #LncGenePairs<-getLncGenePairs(GeneExp,LncExp,a=0.025)
> # get reconstructed undirect pathway graph list
> # To improve efficiency, a fraction of signal pathway as case
>   LncGenePairs<-GetExampleData(exampleData="LncGenePairs")
>   interUMGraph<-getInteGraphList(g2[42:45],LncGenePairs)
> ### Integrate lncRNAs of competitive regulation into KEGG pathway graphs ###
>   ##LncGenePairs<-GetExampleData(exampleData="LncGenePairs")
>   ##inteUMGraph<-getInteUMGraph(LncGenePairs)
```

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

```
> # visualize the reconstructed undirect pathway
>   #LncGenePairs<-GetExampleData(exampleData="LncGenePairs")
>   #inteUMGraph<-getInteUMGraph(LncGenePairs)
>   plotGraphL(interUMGraph[[1]],vertex.label=getNodeLabel)
```

Figure 1 shows the reconstructed undirect p53 signaling pathway.

4 Locate KEGG metabolic subpathways

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

```
> ### Integrate lncRNAs of competitive regulation into KEGG pathway graphs ###
>   LncGenePairs<-GetExampleData(exampleData="LncGenePairs")
>   #inteUMGraph<-getInteUMGraph(LncGenePairs)
>   # To improve efficiency, a fraction of signal pathway as case
>   LncGenePairs<-GetExampleData(exampleData="LncGenePairs")
>   interUMGraph<-getInteGraphList(g2[42:45],LncGenePairs)
> ### get user-interested lncRNAs and genes sets.
> ##geneLnc<-c(getBackground(type="gene")[1:3000],unique(LncGenePairs[1,]))
>   geneLnc<-GetExampleData(exampleData="geneLnc")
> # get locate subpathways.
>   sub<-getLocSubGraphLnc(geneLnc,interUMGraph,type="gene_lncRNA",n=1,s=8)
>
```

5 Identify the significantly enriched subpathways

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

```
> ### Integrate lncRNAs of competitive regulation into KEGG pathway graphs ###
>   #LncGenePairs<-GetExampleData(exampleData="LncGenePairs")
>   #inteUMGraph<-getInteUMGraph(LncGenePairs)
> ### get user-interested lncRNAs and genes sets.
> ##geneLnc<-c(getBackground(type="gene")[1:3000],unique(LncGenePairs[1,]))
>   geneLnc<-GetExampleData(exampleData="geneLnc")
```

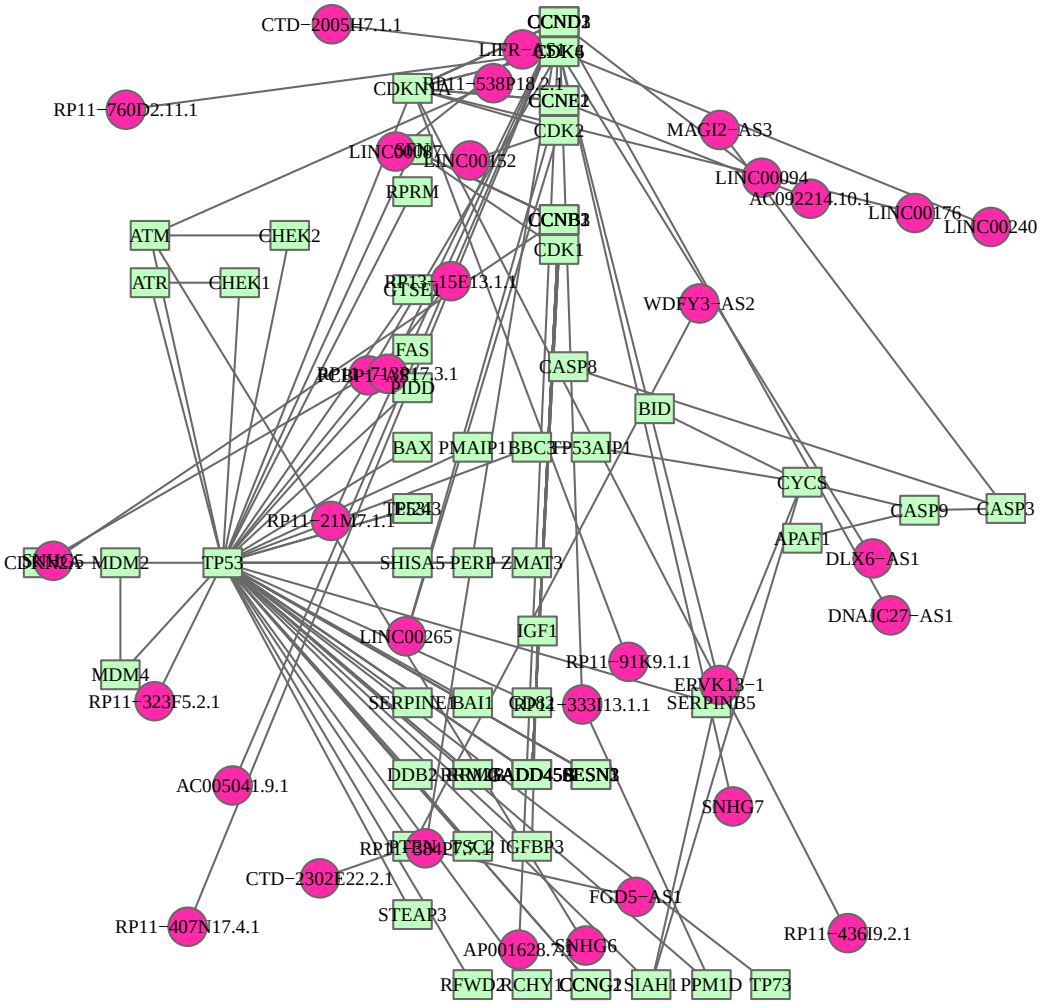

Figure 1: The visualization of reconstructed undirect p53 signaling pathway.

```
> # get locate subpathways.
>   #sub<-getLocSubGraphLnc(geneLnc,interUMGraph,type="gene_lncRNA",n=1,s=8)
>   sub<-GetExampleData(exampleData="sub")
>   # To improve efficiency, a fraction of signal subpathway as case
>   SubcodeLncResult<-identifyLncGraphW(geneLnc,sub[50:55],type="gene_lncRNA",bet=1)
>   #SubcodeLncResult<-identifyLncGraphW(geneLnc,sub,type="gene_lncRNA",bet=1)
>   #resultT<-printGraphW(SubcodeLncResult,detail=TRUE)
>   #write.table(resultT,file="result.txt",sep="\t",row.names=F,quote=F)
```

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

```
>   plotAnnGraph("path:04916_1",sub,SubcodeLncResult,gotoKEGG=FALSE,vertex.label=getNodeLabel)

IGRAPH 7b3e582 UN-B 22 35 -- path:hsa04916
+ attr: name (g/c), number (g/c), org (g/c), title (g/c), image (g/c),
| link (g/c), name (v/c), id (v/c), names (v/c), type (v/c), reaction
| (v/c), link (v/c), graphics_name (v/c), graphics_fgcolor (v/c),
| graphics_bgcolor (v/c), graphics_type (v/c), graphics_x (v/c),
| graphics_y (v/c), graphics_width (v/c), graphics_height (v/c),
| graphics_coords (v/c), id (e/c), names (e/c), type (e/c), reaction
| (e/c), graphics_name (e/c), directed (e/l)
+ edges from 7b3e582 (vertex names):
 [1] 1856  --2775   1856  --HOTAIR 815   --163688 815   --810    818   --163688
 [6] 818   --810    818   --MEG3   163688--23236  810   --23236  2775  --11211
+ ... omitted several edges
```

Figure 2 shows the reconstructed undirect Calcium signaling pathway.

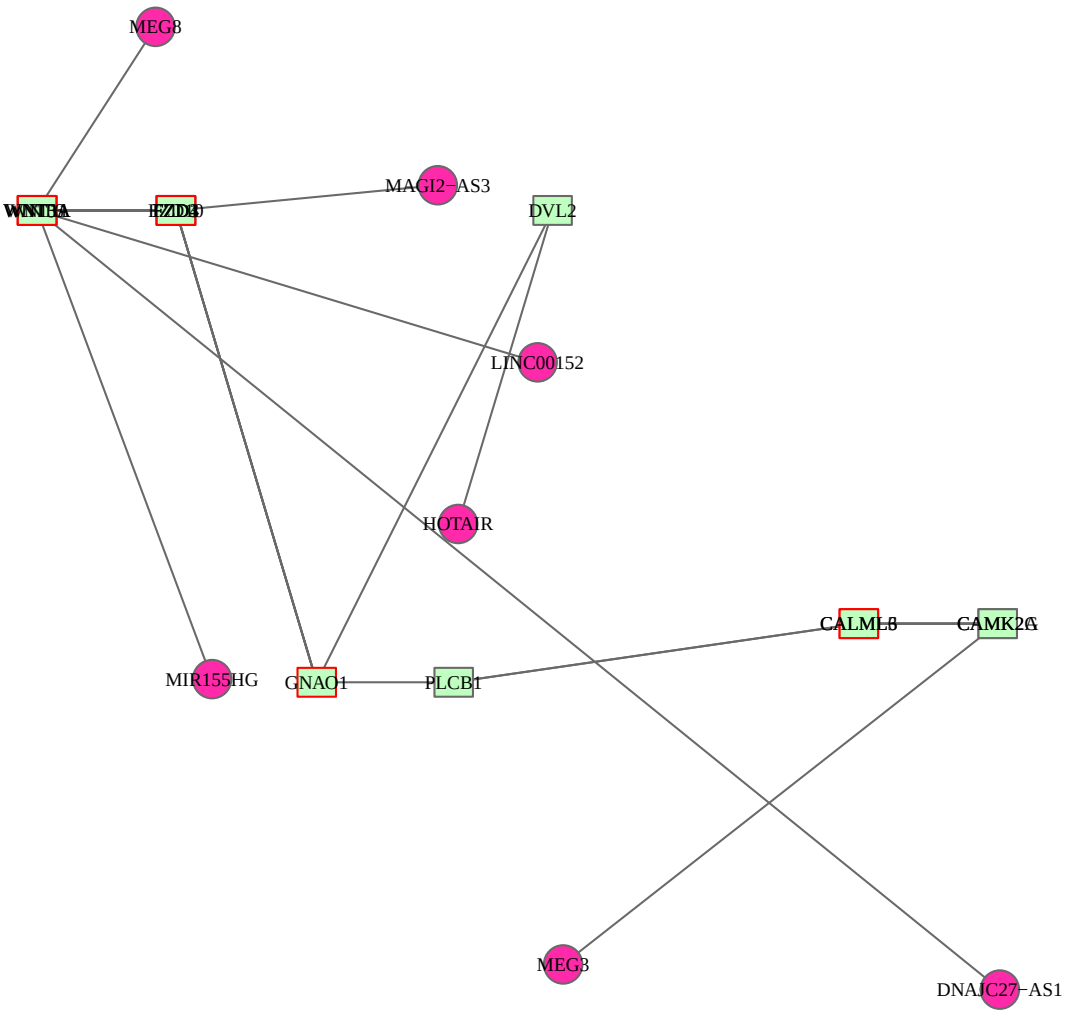

Figure 2: The visualization of reconstructed undirect Calcium signaling pathway.

6 Session Info

The script runs within the following session:

```
R version 4.4.2 (2024-10-31)
Platform: x86_64-pc-linux-gnu
Running under: Ubuntu 24.04.1 LTS

Matrix products: default
BLAS:   /usr/lib/x86_64-linux-gnu/openblas-pthread/libblas.so.3
LAPACK: /usr/lib/x86_64-linux-gnu/openblas-pthread/libopenblasp-r0.3.26.so;  LAPACK version 3.12.0

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

time zone: Etc/UTC
tzcode source: system (glibc)

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

other attached packages:
[1] SubpathwayLNCE_1.0  BiasedUrn_2.0.12    RBGL_1.83.0
[4] graph_1.85.0        BiocGenerics_0.53.3 generics_0.1.3
[7] igraph_2.1.1

loaded via a namespace (and not attached):
 [1] xfun_0.49        magrittr_2.0.3   maketools_1.3.1  glue_1.8.0
 [5] knitr_1.49       pkgconfig_2.0.3  buildtools_1.0.0 stats4_4.4.2
 [9] lifecycle_1.0.4  cli_3.6.3        compiler_4.4.2   sys_3.4.3
[13] tools_4.4.2      evaluate_1.0.1   rlang_1.1.4
```

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.