

How To Use LncPath

Junwei Han, Zeguo Sun

August 20, 2024

Contents

1 Overview	1
2 Finding the differentially expressed genes from a expression profile	1
3 Identifying pathways coordinately regulated by user interested lncRNAs	2
4 Gain insight into the details of each pathway	4
4.1 Plot the running enrichment score of a pathway	4
4.2 Check the detail of genes of each pathway.	4
4.3 Draw the heat map of genes in a certain pathway.	4
5 Session Info	8

1 Overview

This vignette illustrates how to easily use the LncPath package. The package can prioritize pathways coordinately regulated by lncRNAs based on a network diffusion strategy. We firstly constructed a lncRNA-mRNA relationship network by integrating the co-expression pairs between lncRNAs and mRNAs with the protein-protein interaction pairs. The lncRNAs user inputted will be mapped into the lncRNA-mRNA relationship network to evaluate the extent of each gene influenced by the lncRNAs based on a network diffusion strategy. We then built a ranked gene list based on the extent of influence. Finally, we mapped the genes of each pathway into the rank gene list and calculated the pathway enrichment score(ES) using a weighted Kolmogorov-Smirnov statistic. The permutation analysis was performed to selecting significant pathways.

2 Finding the differentially expressed genes from a expression profile

This section introduces how to find significantly differentially expressed genes from an expression profile. We provided two ways to find differentially expressed genes from an expression profile, student's t-test and fold change. For each strategy, a threshold is defined for selecting significant differentially expressed genes.

```
> #obtain the expression profile data
> Profile <- getExampleData("Profile")
> Profile[1:10, 1:10]
```

	A10U	A14Y	A14Y.1	A16R	A14Y.2
MIR143HG	4.817622e-05	2.714833e-05	3.938226e-05	1.583580e-05	4.759946e-06
NPY6R	2.776750e-05	4.436461e-05	1.471648e-05	7.396511e-06	7.028282e-07
AL078621.4	4.745406e-05	5.929525e-05	4.890773e-05	6.076707e-05	1.558605e-05
RP4-791K14.2	1.298476e-05	4.597001e-06	8.998136e-06	4.714152e-06	5.665172e-07
RP11-399019.5	3.404506e-05	1.921294e-05	3.255811e-05	1.910943e-05	3.607389e-06
RP11-357C3.3	4.899702e-05	4.459881e-05	5.840064e-05	2.664671e-05	1.246263e-05
AL589743.1	1.281410e-07	2.004831e-07	8.481405e-08	5.464789e-08	1.655219e-08
RP11-175K6.1	1.720569e-06	1.892652e-06	1.807399e-06	2.654562e-06	4.858749e-07
PGM5-AS1	1.953075e-04	2.153569e-04	1.682733e-04	3.074801e-04	7.435103e-05
CTC-228N24.3	8.655112e-06	1.044819e-05	1.108406e-05	9.040964e-06	3.394420e-06
	A14Y.3	A10U.1	A10U.2	A10U.3	A13Y
MIR143HG	3.587065e-06	4.103880e-05	3.444775e-05	2.057273e-06	1.200087e-06
NPY6R	2.470614e-06	2.180325e-05	5.335282e-06	1.468536e-07	0.000000e+00
AL078621.4	2.598227e-05	2.196803e-05	1.136785e-04	1.174064e-07	1.741630e-07
RP4-791K14.2	1.222805e-06	7.377760e-06	3.551054e-06	1.007499e-07	1.760444e-07
RP11-399019.5	6.122954e-06	3.500004e-05	2.828171e-05	1.354029e-06	1.137123e-06
RP11-357C3.3	1.738906e-05	5.062899e-05	3.225693e-05	1.828165e-05	1.404827e-05
AL589743.1	2.047051e-07	2.418385e-08	4.843106e-07	2.112712e-05	6.100757e-06
RP11-175K6.1	8.300081e-07	6.021785e-07	2.264556e-06	2.618080e-07	2.890107e-07
PGM5-AS1	1.005791e-04	8.785996e-05	4.693726e-04	2.389673e-07	0.000000e+00
CTC-228N24.3	5.648797e-06	7.110228e-06	7.912642e-06	4.127547e-06	5.448670e-06

```

> #obtain the labels of the samples of the expression profile, the label vector is a vector of 0/1s,
> # 0 represents the case sample and 1 represents the control sample
> Labels <- getExampleData("Labels")
> Labels[1:10]

```

```
[1] 0 0 0 0 0 0 0 0 1 1
```

```

> ##find differentially expressed genes, using t-Test defaultly
> options(stringsAsFactors = FALSE)
> SigGenes <- findSigGenes(Profile, Labels, Method = "tTest", FdrCut = 0.01)
> head(SigGenes)

```

```
$Up
```

```
[1] "RP11-166D19.1" "LINC00476" "LRFN5" "MORF4L2"
[5] "MEIS1"
```

```
$Low
```

```
[1] "SFT2D1" "SRSF9" "MRGBP" "MTHFD1L" "TICRR" "ATP13A1"
```

3 Identifying pathways coordinately regulated by user interested lncRNAs

This section introduces how to identify pathways coordinately regulated by user interested lncRNAs. A vector of lncRNAs should be inputted, and they will be mapped into the lncRNA-mRNA relationship network as seed nodes to perform a network diffusion strategy. Here we constructed a huge lncRNA-mRNA network constructed by ingegrating a lncRNA-mRNA co-expression network and the protein-protein interaction network. Considering the huge network may be time consuming, we provided a litte example network for little trial. A weighted Kolmogorov-Smirnov statistic is used to prioritize

the pathways regulated by the user inputted lncRNAs. Now three pathway databases are supported: KEGG, Reactome and Biocarta. The pathways with the number of genes between the user defined limit will be kept for further analysis to avoid potential bias. The permutation analysis was performed to filter significant pathways and the times of permutations can be set by the user.

```

> #get lncRNA-mRNA interaction network
> NetLncPath <- getNet();
> dim(NetLncPath);

[1] 295698      2

> print(head(NetLncPath), row.names = FALSE)

      V1      V2
PLEKHF2 SCYL3
PLEKHF2 STT3B
PLEKHF2 RAD21
PLEKHF2 GOLPH3
PLEKHF2 UBE2D3
PLEKHF2 NME3

> #get example lncRNA sets
> SigLncs <- getExampleData("SigLncs")
> print(head(SigLncs), row.names = FALSE)

[1] "ENSG00000262117" "ENSG00000236824" "ENSG00000240498" "ENSG00000235123"
[5] "ENSG00000234741" "ENSG00000130600"

> #get the example lncRNA-mRNA interaction network
> ExampleNet <- getExampleData("ExampleNet")
> print(head(ExampleNet), row.names = FALSE)

      V1      V2
SPRY2 C6orf48
GNB2L1 RPS18
GNB2L1 TPT1
GNB2L1 IMPDH2
GNB2L1 RPS19
GNB2L1 POLR1D

> #evaluate the rate of pathways regulated by lncRNA sets
> Result <- lncPath(SigLncs, ExampleNet, Weighted = TRUE, PathwayDataSet = "KEGG", nperm = 100,
+ minPathSize = 0, maxPathSize = 500)

Now start the random walking...
[1] "ENSG00000234741" "ENSG00000251562" "ENSG00000214548" "ENSG00000236824"
[5] "ENSG00000229807" "ENSG00000130600" "ENSG00000228630" "ENSG00000281560"
[9] "ENSG00000235123"
9 of 14 were mapped to the huge net with 267 Nodes.
Now, calculating running scores of each pathway...Now, do the purtabationsions...

> ## Generate a table of the summary of each pathway
> PathwaySummaryTable <- lncPath2Table(Result)
> print(head(PathwaySummaryTable), row.names = FALSE)

```

Gene Set Name	Gene Set Size
KEGG_PURINE_METABOLISM	3
KEGG_PYRIMIDINE_METABOLISM	2
KEGG_RIBOSOME	11
KEGG_ALANINE_ASPARTATE_AND_Glutamate_METABOLISM	1
KEGG_Glutathione_METABOLISM	1
KEGG_NITROGEN_METABOLISM	1

Enrichment Scores	Normalized Enrichment Scores	P Value	False Discovery Rate
-0.62222	-0.838196386522472	0.96	0.96
-0.65217	-0.878542215612421	0.9	0.96
0.57517	1	0.46	0.96
-0.70213	-0.945843638695355	<NA>	<NA>
-0.91489	-1.23245394244085	<NA>	<NA>
-0.70213	-0.945843638695355	<NA>	<NA>

4 Gain insight into the details of each pathway

4.1 Plot the running enrichment score of a pathway

The function `plotRunningES` can plot global cumulative running enrichment scores of each gene of a certain pathway.

```
> #get an example result data
> Result <- getExampleData("Result")

> #plot the running score of the KEGG RIBOSOME pathway
> plotRunningES(Result, Name = "KEGG_RIBOSOME")
```

Figure 1 shows the running scores of each gene in the KEGG RIBOSOME pathway.

4.2 Check the detail of genes of each pathway.

The function `geneSetDetail` can show the detail of each gene in a certain pathway.

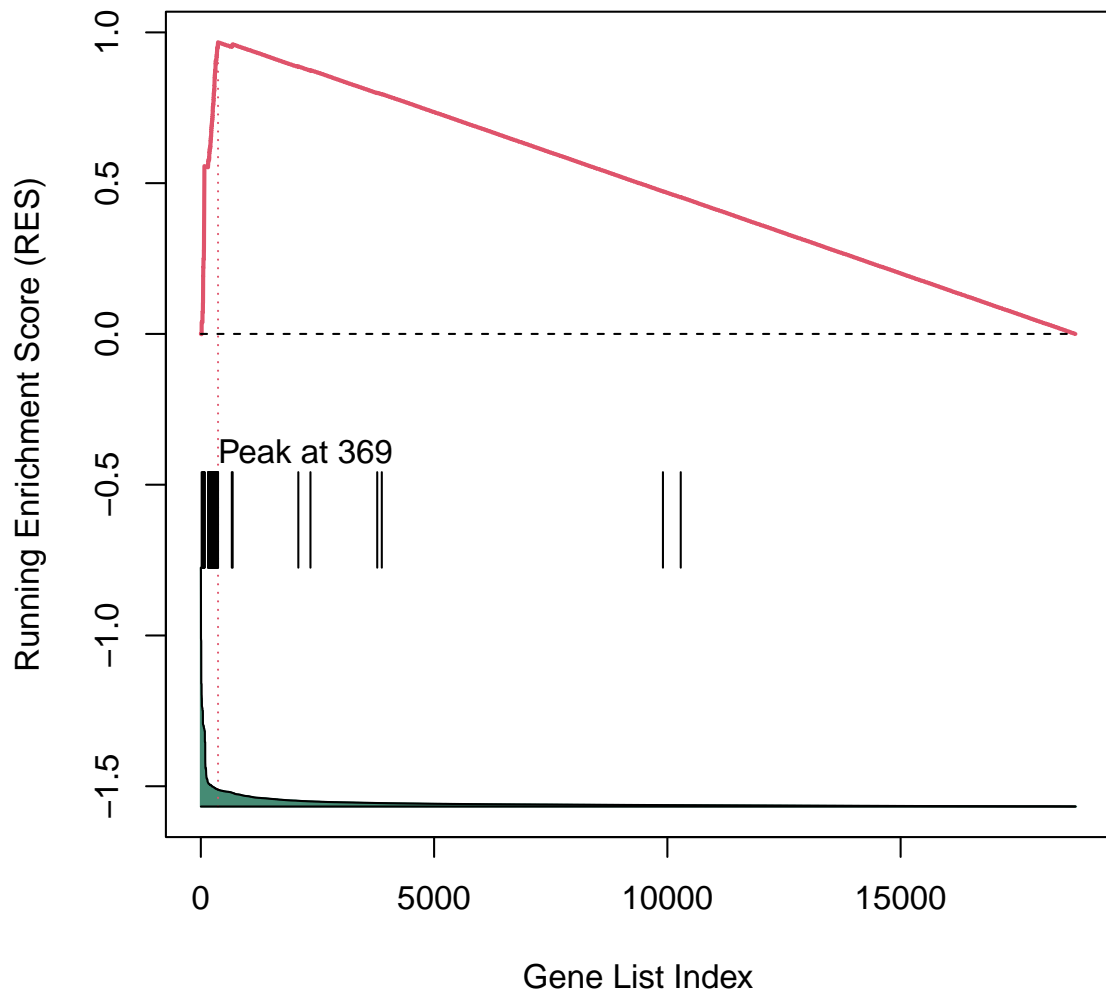
```
> #get an example result data
> Result <- getExampleData("Result")
> #get the details of genes in the KEGG_RIBOSOME pathway
> Detail <- geneSetDetail(Result, Name = "KEGG_RIBOSOME")
> head(Detail)
```

#	GENE	LIST	LOC	S2N	RES	CORE_ENRICHMENT
1	1	RPS19	18	0.0401	0.0392	YES
2	2	RPL37	36	0.0349	0.0732	YES
3	3	RPS17	45	0.0298	0.103	YES
4	4	RPS11	48	0.0295	0.132	YES
5	5	RPL29	49	0.0295	0.161	YES
6	6	RPL7	50	0.0294	0.191	YES

4.3 Draw the heat map of genes in a certain pathway.

The function `drawAHeatMap` can draw a heatmap of the genes in a certain pathway based on the expression profile user specified.

KEGG_RIBOSOME



Number of genes: 18745 (in list), KEGG_RIBOSOME (in gene set)

Figure 1: The running scores of each gene in the KEGG RIBOSOME pathway.

```
> #get an example result data
> Result <- getExampleData("Result")
> #get example data
> Profile <- getExampleData("Profile")
> Labels <- getExampleData("Labels")

> #Draw the heatmap of genes in KEGG_RIBOSOME pathway
> drawAHeatMap(Result, Name = "KEGG_RIBOSOME", PCExpr = Profile, Labels = Labels)

NULL
```

Figure 2 shows the heatmap of genes in KEGG RIBOSOME pathway based on the example expression profile.

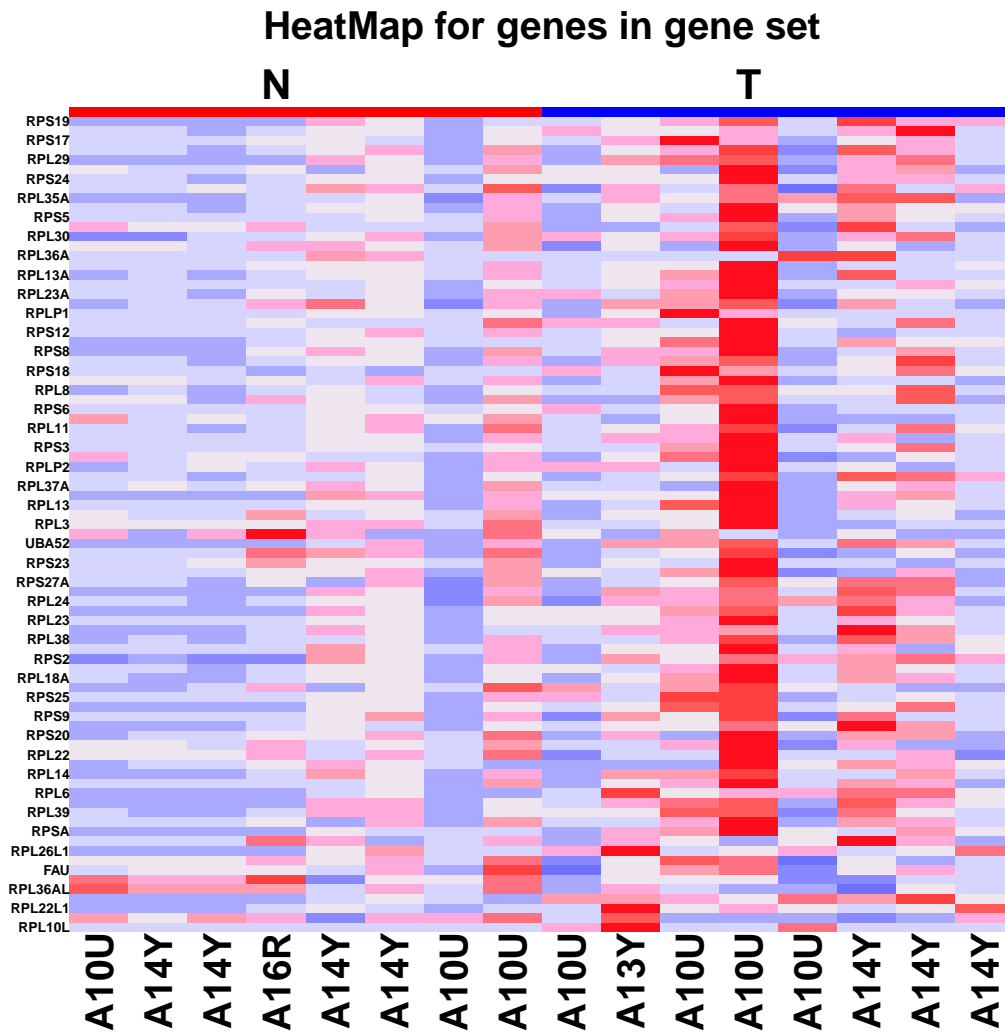


Figure 2: The heatmap of genes in KEGG RIBOSOME pathway based on the example expression profile.

5 Session Info

The script runs within the following session:

R version 4.4.1 (2024-06-14)

Platform: x86_64-pc-linux-gnu

Running under: Ubuntu 24.04 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/libopenblas-p0.3.26.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] LncPath_1.1  igraph_2.0.3
```

loaded via a namespace (and not attached):

```
[1] xfun_0.47      Matrix_1.7-0    lattice_0.22-6  magrittr_2.0.3
[5] maketools_1.3.0 glue_1.7.0      knitr_1.48      pkgconfig_2.0.3
[9] buildtools_1.0.0 lifecycle_1.0.4 cli_3.6.3       grid_4.4.1
[13] compiler_4.4.1 sys_3.4.2       tools_4.4.1     rlang_1.1.4
```

References

[Subramanian *et al.*, 2005] Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. et al. (2005) Gene set enrichment analysis: a knowledgebased approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102, 15545-15550.

[Liao Q *et al.*, 2011] Liao Q, Liu C, Yuan X, Kang S, Miao R, Xiao H, Zhao G, Luo H, Bu D, Zhao H, et al: Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res* 2011, 39:3864-3878.

[Guo X *et al.*, 2013] Guo X, Gao L, Liao Q, Xiao H, Ma X, Yang X, Luo H, Zhao G, Bu D, Jiao F, et al: Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks. *Nucleic Acids Res* 2013, 41:e35.