mirIntegrator: Integrating miRNAs into signaling pathways

Diana Diaz and Sorin Draghici
Department of Computer Science, Wayne State University, Detroit MI 48201

June 9, 2015

Abstract

mirIntegrator is an R package for integrating microRNAs (miRNAs) into signaling pathways to perform pathway analysis using both mRNA and miRNA expressions. Typical pathway analysis methods help to investigate which pathways are relevant to a particular phenotype understudy. The input of these methods are the fold change of mRNA of two different phenotypes (e.g. control versus disease), and a set of signaling pathways. Researchers investigating miRNA cannot perform pathway analysis using traditional methods because current pathways datasets do not contain miRNA-gene interactions. mirIntegrator package aims to fill this gap by:

1. Integrating miRNAs into signaling pathways,
2. Generating a graphical representation of the augmented pathways, and
3. Facilitating the use of pathway analysis techniques when studying miRNA and mRNA expression levels.

1 Integration of miRNA into signaling pathways

The main functionality of the mirIntegrator package is the integration of miRNAs into signaling pathways. The input of this functionality are a set of signaling pathways like KEGG pathways [1] or Reactome [2], and a miRNA-target interaction database like mirTarBase [3] or TargetScan [4]. The output is a set of augmented signaling pathways. Each augmented pathway contains the original sets of genes and interactions plus the set of miRNAs involved in the pathways and their miRNA-target interactions. These interactions are the biological miRNA repression to their target genes and are represented in the model as negative interactions.

Here we show an example of the method functionality. Let us say that some researchers need to integrate KEGG [1] human signaling pathways with miRNA interactions from miRTarBase [3]. Researchers must first obtain the list of pathways as a list of `graph::graphNEL` objects. The nodes of each pathway represent the genes involved in the pathway and the edges represent the biological interactions among those genes (activation or repression). The second step is to obtain a miRNA-target interactions dataset as a `data.frame` with the columns `"miRNA"` and `"Target.ID"`. Notice that the symbols used to identify the `"Target.ID"` column on the miRNA-target interactions dataset must be the same symbols used on the nodes of the pathways. i.e. If the genes are identified by entrezID on the pathways' dataset, then the miRNA-targets dataset must identify the genes by entrezID as well. Once the researchers have these two datasets, they can use the function *integrate_mir*.

To demonstrate this functionality, mirIntegrator package includes the object `mirTarBase` which is a copy of the experimentally validated miRNA-target interactions database miRTarBase [3]. We downloaded the miRTarBase database from `http://mirtarbase.mbc.nctu.edu.tw/` on

April 1st, 2015. A complete script describing how this database was downloaded and formated is included in this package on '/inst/scripts/get_mirTarBase.R'.

Here an example of how researchers can generate the list of augmented pathways from five KEGG pathways and mirTarBase interactions using the function *integrate_mir*:

```
> require("mirIntegrator")
> data(kegg_pathways)
> data(mirTarBase)
> kegg_pathways <- kegg_pathways[18:20] #delete this for augmenting all pathways.
> augmented_pathways <- integrate_mir(kegg_pathways, mirTarBase)
> head(augmented_pathways)
```

```
$`path:hsa04122`
A graphNEL graph with directed edges
Number of Nodes = 20
Number of Edges = 19

$`path:hsa04130`
A graphNEL graph with directed edges
Number of Nodes = 77
Number of Edges = 113

$`path:hsa04140`
A graphNEL graph with directed edges
Number of Nodes = 85
Number of Edges = 82
```

The result is a list of pathways where each pathway is a `graph::graphNEL` object. When researchers need to see the details of a particular pathway, they can do so by simply using the KEGG pathway id of the pathway of interest. For example, the pathway "path:hsa04122" can be reach with the following instruction:

```
> augmented_pathways$"path:hsa04122"
```

```
A graphNEL graph with directed edges
Number of Nodes = 20
Number of Edges = 19
```

2 Graphical output

mirIntegrator incorporates a functionality to produce a graphical representation of the final pathways. This is useful when researchers need to visualize the nodes that were added to the pathway. For instance, if they need to see how the pathway of "Sulfur relay system" (path:hsa04122) has changed, they can plot the augmented pathway using the function `plot_augmented_pathway`. Here an example, Figure 1 is the output of these instructions:

```
>   data(names_pathways)
> plot_augmented_pathway(kegg_pathways$"path:hsa04122",
+                        augmented_pathways$"path:hsa04122",
+                        names_pathways["path:hsa04122"] )
```

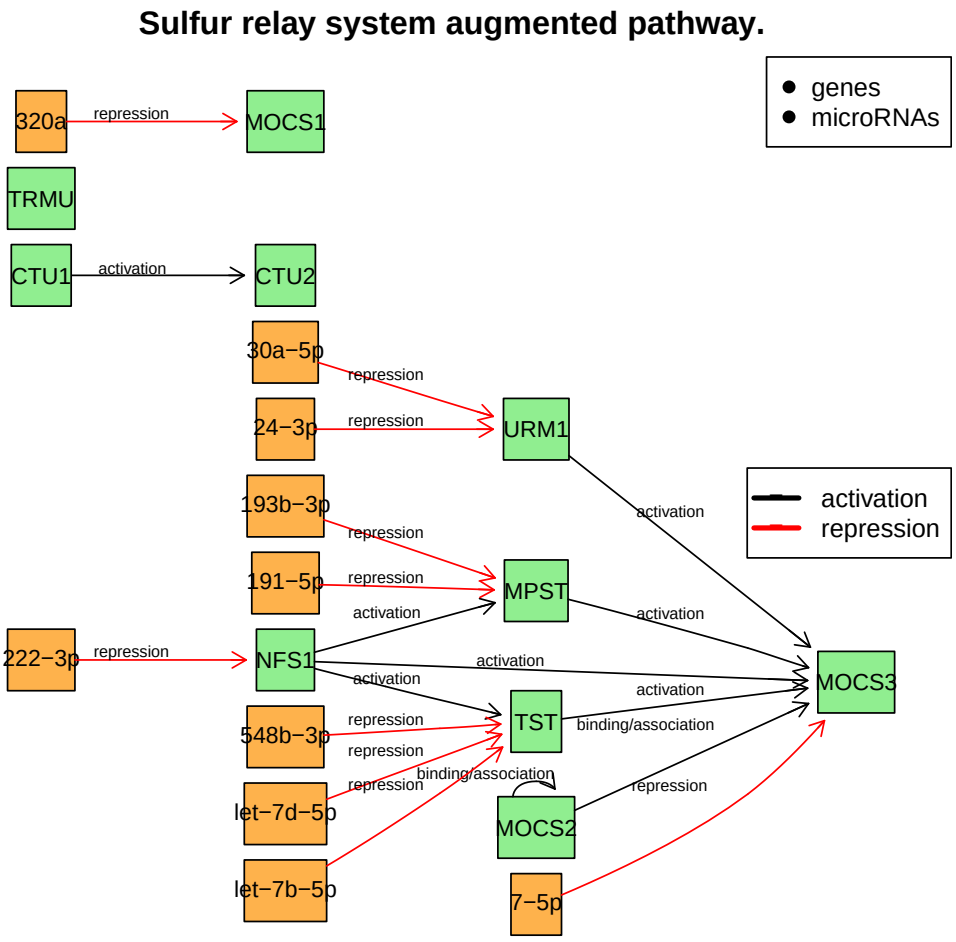

Figure 1: Visualization of the Sulfur relay system augmented pathway using the function `plot_augmented_pathway`.

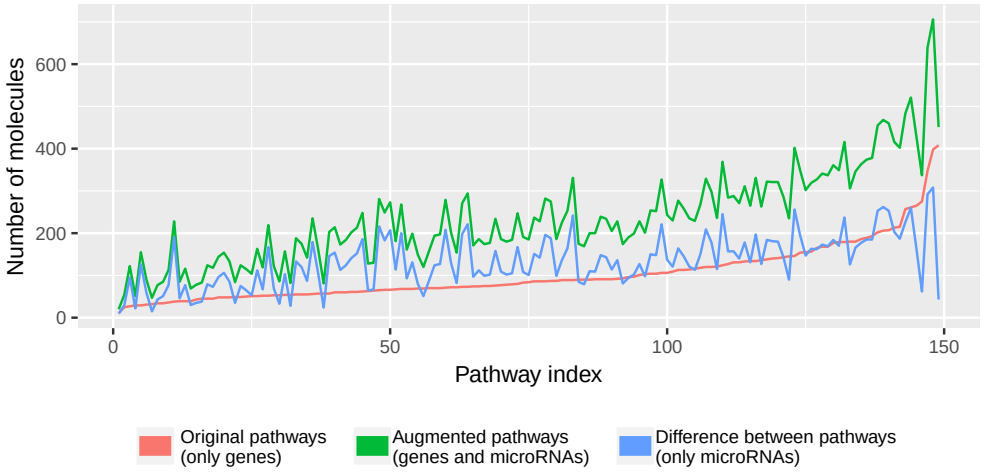

Figure 2: Plotting of the change of pathways' order using the function `plot_change`.

Another useful function is `plot_change` which can be used to see how much the order of the pathways have changed. To demonstrate this functionality, the mirIntegrator package includes a copy of KEGG human signaling pathways. We obtained these KEGG pathways using the RONTOTools package [5]. A complete script describing how this dataset was obtained is included in this package on '/inst/scripts/get_kegg_pathways.R'. Here an example of the use of the function `plot_change`. The result is shown on Figure 2:

```
>   data(augmented_pathways)
>   data(kegg_pathways)
>   data(names_pathways)
>   plot_change(kegg_pathways,augmented_pathways, names_pathways)
```

This package also includes a function to generate of a pdf file with the plots of the list of augmented pathways. Here an example of this functionality:

```
>   data(augmented_pathways)
>   data(kegg_pathways)
>   data(names_pathways)
>   pathways2pdf(kegg_pathways[18:20],augmented_pathways[18:20],
+                names_pathways[18:20], "three_pathways.pdf")
```

3 Pathway analysis of miRNA and mRNA: a case study

The main purpose of the pathways augmentation process is to analyze miRNA and mRNA expressions at the same time. For this reason, we show here how to analyze a multiple sclerosis dataset using the mirIntegrator package. The dataset that we analyzed was published by Jernas, M., et. al. [6] whom collected heparin-anticoagulated peripheral blood from 21 multiple sclerosis (MS) patients and nine healthy controls. Ten of the 21 samples were used to profiled mRNA expression, and the 11 remaining were used to profiled miRNA expression. These datasets are accessible at NCBI GEO database [7] with accession GSE43592. We preprocessed the datasets using the limma package [8]. For demonstration purposes, we included the preprocessed datasets on this package.

```
>   data(GSE43592_miRNA)
>   data(GSE43592_mRNA)
```

Once researchers have the data and the augmented pathways, they can run the pathway analysis method that they prefer. We suggest to use ROntoTools package [5] because it takes in account the topology of the pathways (the method implemented on ROntoTools is explained on [9]). We show here how to perform impact pathway analysis [9] using the ROntoTools package with our augmented pathways:

```
>   require(graph)
>   require(ROntoTools)
>   data(GSE43592_mRNA)
>   data(GSE43592_miRNA)
>   data(augmented_pathways)
>   data(names_pathways)
>   lfoldChangeMRNA <- GSE43592_mRNA$logFC
>   names(lfoldChangeMRNA) <- GSE43592_mRNA$entrez
>   lfoldChangeMiRNA <- GSE43592_miRNA$logFC
>   names(lfoldChangeMiRNA) <- GSE43592_miRNA$entrez
>   keggGenes <- unique(unlist( lapply(augmented_pathways,nodes) ) )
>   interGMi <- intersect(keggGenes, GSE43592_miRNA$entrez)
>   interGM <- intersect(keggGenes, GSE43592_mRNA$entrez)
>   ## For real-world analysis, nboot should be >= 2000
>   peRes <- pe(x= c(lfoldChangeMRNA, lfoldChangeMiRNA ),
+               graphs=augmented_pathways, nboot = 200, verbose = FALSE)
>   message(paste("There are ", length(unique(GSE43592_miRNA$entrez)),
+                 "miRNAs meassured and",length(interGMi),
+                 "of them were included in the analysis."))
>   message(paste("There are ", length(unique(GSE43592_mRNA$entrez)),
+                 "mRNAs meassured and", length(interGM),
+                 "of them were included in the analysis."))
>   summ <- Summary(peRes)
>   rankList <- data.frame(summ,path.id = row.names(summ))
>   tableKnames <- data.frame(path.id = names(names_pathways),names_pathways)
>   rankList <- merge(tableKnames, rankList, by.x = "path.id", by.y = "path.id")
>   rankList <- rankList[with(rankList, order(pAcc.fdr)), ]
>   head(rankList)
```

```
       path.id                    names_pathways   totalAcc  totalPert
1 path:hsa03008 Ribosome biogenesis in eukaryotes 0.03170446 0.06340892
2 path:hsa03013                     RNA transport 0.24112918 0.41898535
4 path:hsa03018                   RNA degradation 0.08603518 0.17207037
5 path:hsa03320           PPAR signaling pathway 0.03963057 0.07133503
6 path:hsa03460           Fanconi anemia pathway 0.08483085 0.11653531
7 path:hsa04010           MAPK signaling pathway 0.71264826 1.09004098
  totalAccNorm totalPertNorm     pPert      pAcc pPert.fdr  pAcc.fdr
1   -0.6196714    -0.6196714 0.3333333 0.3333333 0.7603648 0.7348064
2   -0.8474377    -0.8592665 0.3233831 0.3233831 0.7603648 0.7348064
```

```
4   -0.7813859    -0.7813859 0.3134328 0.3134328 0.7603648 0.7348064
5   -0.8017812    -0.8017812 0.3930348 0.3930348 0.7603648 0.7348064
6   -0.6329488    -0.6329488 0.3930348 0.3930348 0.7603648 0.7348064
7   -0.7629982    -0.7300710 0.4079602 0.4029851 0.7603648 0.7348064
```

4 Citing mirIntegrator

The algorithms and methods for integrating miRNA and mRNA included on this package are in publication process.

References

[1] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Research*, vol. 28, pp. 27–30, January 2000.

[2] D. Croft, A. F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M. R. Kamdar, B. Jassal, S. Jupe, L. Matthews, B. May, S. Palatnik, K. Rothfels, V. Shamovsky, H. Song, M. Williams, E. Birney, H. Hermjakob, L. Stein, and P. D'Eustachio, "The Reactome pathway knowledgebase," *Nucleic Acids Research*, vol. 42, no. D1, pp. D472–D477, 2014.

[3] S.-D. Hsu, Y.-T. Tseng, S. Shrestha, Y.-L. Lin, A. Khaleel, C.-H. Chou, C.-F. Chu, H.-Y. Huang, C.-M. Lin, S.-Y. Ho, T.-Y. Jian, F.-M. Lin, T.-H. Chang, S.-L. Weng, K.-W. Liao, I.-E. Liao, C.-C. Liu, and H.-D. Huang, "miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions," *Nucleic Acids Research*, vol. 42, pp. D78–D85, Jan. 2014.

[4] B. P. Lewis, C. B. Burge, and D. P. Bartel, "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets," *Cell*, vol. 120, pp. 15–20, Jan. 2005.

[5] C. Voichita and S. Draghici, *ROntoTools: R Onto-Tools suite*, 2014. R package version 1.2.0.

[6] M. Jernas, C. Malmestrom, M. Axelsson, I. Nookaew, H. Wadenvik, J. Lycke, and B. Olsson, "MicroRNA regulate immune pathways in t-cells in multiple sclerosis (MS)," *BMC immunology*, vol. 14, p. 32, 2013.

[7] R. Edgar, M. Domrachev, and A. E. Lash, "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository," *Nucleic Acids Research*, vol. 30, no. 1, pp. 207–210, 2002.

[8] G. K. Smyth, *Limma: linear models for microarray data*, pp. 397–420. New York: Springer, 2005.

[9] C. Voichita, M. Donato, and S. Draghici, "Incorporating gene significance in the impact analysis of signaling pathways," in *Machine Learning and Applications (ICMLA), 2012 11th International Conference on*, vol. 1, (Boca Raton, FL, USA), pp. 126–131, IEEE, 12-15 Dec. 2012.