German Medical Data Sciences: A Learning Healthcare System
U. Hübner et al. (Eds.)

doi:10.3233/978-1-61499-896-9-183

Visualization and Analysis of miRNAs Implicated in Amyotrophic Lateral Sclerosis Within Gene Regulatory Pathways

Hamid HAMZEIY[a,†], Rabia SULUYAYLA[b,c,†], Christoph BRINKROLF[d], Sebastian Jan JANOWSKI[d], Ralf HOFESTÄDT[d], Jens ALLMER[b,e,1]
[a]*NDAL, Department of Molecular Biology and Genetics, Faculty of Science, Bogazici University, Istanbul, Turkey*
[b]*Department of Molecular Biology and Genetics, Faculty of Science, Izmir Institute of Technology, Izmir, Turkey*
[c]*Institute of Biochemistry and Molecular Cell Biology, University Hospital Aachen*
[d]*Department of Bioinformatics and Medical Informatics, Faculty of Technology, Bielefeld University, Bielefeld, Germany*
[e]*Applied Bioinformatics, Bioscience, Wageningen University and Research, Wageningen, the Netherlands*
[†]*Equal Contribution*

Abstract. MicroRNAs (miRNAs), approximately 22 nucleotides long, post-transcriptionally active gene expression regulators, play active roles in modulating cellular processes. Gene regulation and miRNA regulation are intertwined and the main aim of this study is to facilitate the analysis of miRNAs within gene regulatory pathways. VANESA enables the reconstruction of biological pathways and supports visualization and simulation. To support integrative miRNA and gene pathway analyses, a custom database of experimentally proven miRNAs, integrating data from miRBase, TarBase and miRTarBase, was added to DAWIS-M.D., which is the main data source for VANESA. Analysis of human KEGG pathways within DAWIS-M.D. showed that 661 miRNAs (~1/3 recorded human miRNAs) lead to 65,474 interactions. *hsa-miR-335-5p* targets most genes in our system (2,544); while the most targeted gene (with 71 miRNAs) is *NUFIP2* (Nuclear Fragile X Mental Retardation Protein Interacting Protein 2). Amyotrophic Lateral Sclerosis (ALS), a complex neurodegenerative disease, was chosen as a proof of concept model. Using our system, it was possible to reduce the initially several hundred genes and miRNAs associated with ALS to eight genes, 19 miRNAs and 31 interactions. This highlights the effectiveness of the implemented system to distill important information from otherwise hard to access, highly convoluted and vast regulatory networks.

Keywords. microRNAs, amyotrophic lateral sclerosis, metabolic networks and pathways, gene expression regulation, KEGG, miRBase, miRTarBase

[1] Corresponding Author, Jens Allmer, Applied Bioinformatics, Bioscience, Wageningen University and Research, Wageningen, The Netherlands; E-mail: jens@allmer.de.

1. Introduction

Amyotrophic Lateral Sclerosis (ALS) is the third most common neurodegenerative disease following Alzheimer's and Parkinson's disease. *ATXN2*, a major risk factor for ALS [1], has been shown to be important in miRNA function in flies. Furthermore, *TDP-43* and *FUS* take part in the processing of both coding and non-coding RNA and are suggested to influence miRNA biogenesis [2, 3]. Such findings suggest a significant role for miRNAs in the onset and progression of ALS and motivates further investigation.

Mature miRNAs (~22 nt) are derived from distinct hairpin structures (pre-miRNAs) present in primary RNA transcripts and are post-transcriptional regulators of gene expression. They function via binding to complementary sequences on their target messenger RNAs (mRNAs). MiRNAs can target many mRNAs with multiple target sites. Conversely, mRNAs can be targeted by several miRNAs leading to mRNA degradation or blocking of translation. Currently, miRBase (miRBase 21) is the primary database for miRNA sequences and annotations. TarBase and miRTarBase contain miRNA target information and interactions derived from experimental and computational techniques.

KEGG and similar databases (e.g.: Reactome) contain large bodies of information on gene regulatory pathways. The roles of miRNAs as important regulators of cellular processes, has led to an increasing number of studies utilizing biological pathway analysis. Most such studies lack miRNA visualization methods within large biological networks and focus on predicted miRNAs. DIANA miRPath v.3.0 [4] is an online tool which combines miRNAs with KEGG pathways but lacks the necessary tools for network editing and simulation. Another tool is miTALOS [5] which uses miRNA target prediction, KEGG and NCI PID data [6].

Network reconstruction is important for investigating the impact of miRNAs on biological networks. For example, miRGen [7], integrates various biological databases and includes miRNAs. Such networks should be analyzed in combination with other resources and VANESA [8] (https://agbi.techfak.uni-bielefeld.de/vanesa/), a systems biology software, provides a suitable platform for this, as it draws data from many databases within DAWIS-M.D. [9] and allows for construction, merging, visualization, and analysis of biological networks. It also provides free access to our custom miRNA database and was employed in this study to integrate multiple resources yet avoid huge interaction networks. The KEGG ALS pathway has 575 interactions including miRNAs (Figure 1A). Merging it with a miRNA enriched GeneMANIA [10] pathway (Figure 1B) resulted in 694 interactions (Figure 1D); but the intersection of the two networks, reduced the complexity to only 31 (Figure 1C). Thereby, we were able to show that this approach was able to help distill information from multiple larger regulatory networks.

2. Materials and Methods

In-house scripts were used to collect data from miRBase 21 (http://www.mirbase.org/ftp.shtml) containing both sequence and annotation information for the experimentally proven human miRNAs. TarBase V6.0 (http://diana.cslab.ece.ntua.gr/tarbase/tarbase_download.php) was used for obtaining information regarding the target sequences of experimentally proven human miRNAs. MiRTarBase 6.1 (http://mirtarbase.mbc.nctu.edu.tw/php/download.php) was downloaded and the data was modeled and imported into our database and DAWIS-M.D. MySQL Workbench 5.2 CE was used for creating our miRNA database.

3. Results

The addition of our combined miRNA-target database to DAWIS-M.D. enables the analysis and visualization of miRNAs within biological pathways. MiRNAs can be added to gene pathways if they are co-expressed with genes (e.g.: within an intron). MiRNAs that come from intergenic regions can also be added if they target a gene in the pathway. This extends the applications of VANESA and allows for a variety of different analyses on networks. Advantages of VANESA include techniques such as Petri net modeling and graph analysis as well as direct access to a variety of databases available within DAWIS-M.D. This allows drill down analysis of the function of miRNAs or families of miRNAs from a parent network to other related or targeted networks.

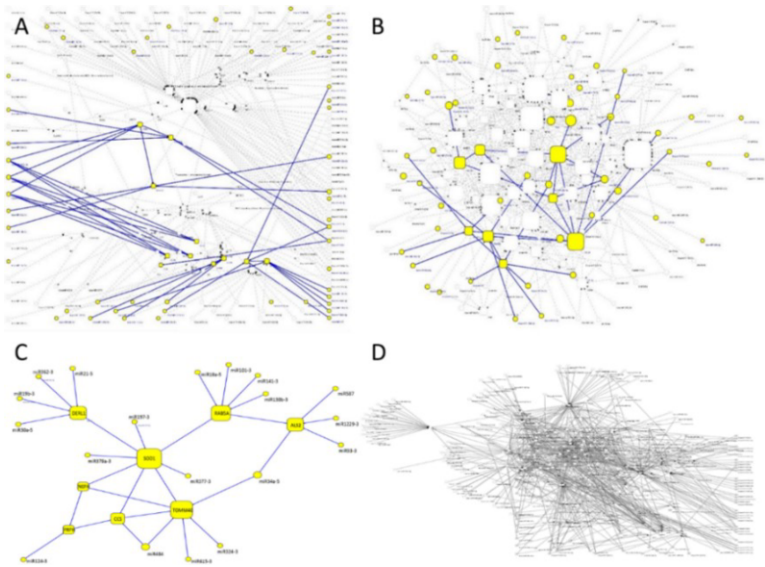

Figure 1. ALS KEGG pathway enriched with miRNAs (Panel A). Square nodes represent genes and circular nodes represent miRNAs. The colored edges in Panel B highlight the overlaps between the ALS network from KEGG and GeneMANIA (Panel B). Intersection of the two miRNA enriched ALS pathways from KEGG and GeneMANIA (Panel C). Merged KEGG and GeneMania ALS pathways enriched with miRNAs (Panel D).

2,588 miRNAs were extracted from miRBase along with 10,966 targets from TarBase and 34,777 targets from miRTarBase and integrated into our database, publicly accessible through VANESA. 3,022 genes and 346 miRNAs were found to be common in both TarBase and miRTarBase. In total, the 2,588 mature miRNA extracted from miRBase have 45,344 mRNA targets, targeting 12,347 distinct genes.

For human KEGG pathways, 99.15% showed miRNA mediated regulation. Within the 234 miRNA regulated KEGG pathways 65,473 miRNA gene interactions were found. Such high numbers show how versatile miRNAs are in regulating events within the cell. Some miRNAs such as *has-miR-335-5p* have higher potential activity by targeting 2,544 genes uniquely whereas the median for all miRNAs is 100 targets, but this may change due to pending miRNA target validations. This pattern is also seen at the gene level where some genes such as *Nfat5* (13) and *NUFIP2* (71) are targeted by multiple miRNAs.

When enriching a network built from the 30 known ALS implicated genes and *ATXN2*, the main risk factor for ALS, our system identified 119 miRNA interactions. Furthermore, two genes namely *ERBB4* and *CHMP2B* contained three and one miRNA

in their transcription unit, respectively. The most targeted ALS associated gene is *BLC2* (targeted by 47 miRNAs), while *DAO* is the only gene not having a known associated miRNA. *VCP*, *LMNB2*, *FUS*, *TARDBP*, *SOD1*, *LMNB1* and *TAF15* present the ALS associated genes within our network which are highly targeted by miRNAs (Figure 1B).

The ALS KEGG pathway was enriched with miRNAs using VANESA requiring that miRNAs are either co-expressed with any of the genes in the pathway or are directly targeting genes in the pathway. The enriched network consisted of 204 nodes, with 39 genes, 142 miRNAs and 242 interactions (Figure 1A). Regulatory networks tend to grow quickly with increasing number of genes involved and including miRNAs further adds to the already complex gene regulatory networks. The KEGG pathway for ALS is focused on the *SOD1* related pathogenesis mechanism. An alternative ALS-related GeneMANIA network was constructed, imported, and enriched with miRNAs in VANESA. All ALS associated genes along with *ATNX2*, as the main risk factor for ALS, were collected from literature and the resulting 31 genes were processed with GeneMANIA to obtain further interactions. The resulting network of 50 genes was downloaded and enriched with miRNAs under the same constraints as for KEGG resulting in a total of 260 nodes, consisting of genes and miRNAs. 575 interactions were found between 50 genes and 210 miRNAs (Figure 1B). The enriched KEGG and GeneMANIA networks were merged in VANESA resulting in a large complex interaction graph. This merging resulted in a total of 293 nodes (77 genes and 199 miRNAs) with 694 interactions (Figure 1D). Since such a complex network convolutes understanding and complicates experimental analysis, the intersection of the two networks was established in VANESA. The intersection graph contains only the shared nodes and edges, significantly reducing the complexity of the network to 27 nodes (8 genes and 19 miRNAs) with 31 interactions (Figure 1C). Out of the 8 genes, 4 (*ALS2*, *NEFH*, *PRPH*, and *SOD1;* collected from ALSoD) have previously been implicated in ALS. Among the 19 miRNAs, 2 (*hsa-miR-21-5p* [11] and *hsa-miR-124-3p* [12]; with 1351 targets) were previously shown to have a role in ALS.

4. Discussion

Currently, large amounts of biological data are available and the pool is growing at an increasing pace. This database and tool inflation has led to data fragmentation and thus data integration has gained increasing importance. DAWIS-M.D. facilitates data integration including resources such as KEGG and BRENDA as well as our custom miRNA database. Using VANESA which can query DAWIS-M.D., we were able to observe the significant impact of miRNAs on biological pathways, underlining the great importance of miRNAs for pathway analysis.

We demonstrate the ability of our system using ALS as an example. Previous studies have identified differentially expressed miRNAs both in disease models and patient samples such as *hsa-miR-143-3p*, a *TDP-43* binding miRNA, in the CSF and serum of sporadic ALS patients [13]. Here, we show that this miRNA along with *hsa-miR-365a-3p*, *has-miR-29a-3p* and *hsa-miR-21-5p* target *BCL2* in ALS KEGG (Figure 1A).

Intersecting miRNA enriched GeneMANIA and KEGG networks reduced the network size making it feasible for experimental validation efforts (Figure 1C). Among the 8 genes, 4 have been implicated in ALS before (*SOD1*, *ALS2*, *PRPH* and *NEFH*). Out of the 19 miRNAs, 2 (*miR-21-5p* and *miR-124-3p*) have been associated with ALS previously. These findings demonstrate how our system can help to elucidate some of

the possible miRNA-mediated pathology mechanisms in complex diseases such as ALS, but it is clear that experimental validation must be performed for confirmation.

5. Conclusion

Our findings demonstrate how our system and approach can help to further elucidate some of the possible miRNA-mediated pathological mechanisms in complex diseases such as ALS. The stability of miRNAs and their potential for use as biomarkers make these molecules strong candidates for further more in-depth studies especially in respect to the cross-talk among regulatory networks.

6. Conflicts of interest

The authors declare no conflicts of interest.

References

[1] A. C. Elden, H. J. Kim, M. P. Hart, A. S. Chen-Plotkin, B. S. Johnson, X. Fang, M. Armakola, F. Geser, R. Greene, et al., Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS, *Nature* **466**, (2010) 1069.

[2] Y. Kawahara and A. Mieda-Sato, TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes, *Proc. Natl. Acad. Sci.* **109**, (2012) 3347.

[3] M. Morlando, S. Dini Modigliani, G. Torrelli, A. Rosa, V. Di Carlo, E. Caffarelli and I. Bozzoni, FUS stimulates microRNA biogenesis by facilitating co-transcriptional Drosha recruitment, *EMBO J.* **31**, (2012) 4502.

[4] I. S. Vlachos, K. Zagganas, M. D. Paraskevopoulou, G. Georgakilas, D. Karagkouni, T. Vergoulis, T. Dalamagas and A. G. Hatzigeorgiou, DIANA-miRPath v3.0: Deciphering microRNA function with experimental support, *Nucleic Acids Res.* **43**, (2015) W460.

[5] M. Preusse, F. J. Theis and N. S. Mueller, miTALOS v2: Analyzing tissue specific microRNA function, *PLoS One* **11**, Public Library of Science, (2016) 1.

[6] M. D. Mailman, M. Feolo, Y. Jin, M. Kimura, K. Tryka, R. Bagoutdinov, L. Hao, A. Kiang, J. Paschall, et al., The NCBI dbGaP database of genotypes and phenotypes, *Nat. Genet.* **39**, (2007) 1181.

[7] M. Megraw, P. Sethupathy, B. Corda and A. G. Hatzigeorgiou, miRGen: A database for the study of animal microRNA genomic organization and function, *Nucleic Acids Res.* **35**, (2007) D149.

[8] C. Brinkrolf, S. J. Janowski, B. Kormeier, M. Lewinski, K. Hippe, D. Borck and R. Hofestädt, VANESA - a software application for the visualization and analysis of networks in system biology applications., *J. Integr. Bioinform.* **11**, (2014) 239.

[9] K. Hippe, B. Kormeier, S. J. Janowski, T. Töpel and R. Hofestädt, DAWIS-M.D. 2.0 - A Data Warehouse Information System for Metabolic Data, *Inform. 2010 Serv. Sci. - Neue Perspekt. für die Inform. Beiträge der 40. Jahrestagung der Gesellschaft für Inform. e.V.*, (2010) 720.

[10] D. Warde-Farley, S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, et al., The GeneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function, *Nucleic Acids Res.* **38**, (2010) W214.

[11] O. Butovsky, S. Siddiqui, G. Gabriely, A. J. Lanser, B. Dake, G. Murugaiyan, C. E. Doykan, P. M. Wu, R. R. Gali, et al., Modulating inflammatory monocytes with a unique microRNA gene signature ameliorates murine ALS, *J. Clin. Invest.* **122**, American Society for Clinical Investigation, (2012) 3063.

[12] A. Emde, C. Eitan, L.-L. L.-L. Liou, R. T. Libby, N. Rivkin, I. Magen, I. Reichenstein, H. Oppenheim, R. Eilam, et al., Dysregulated miRNA biogenesis downstream of cellular stress and ALS-causing mutations: a new mechanism for ALS, *EMBO J.* **34**, (2015) 2633.

[13] A. Freischmidt, K. Müller, A. C. Ludolph and J. H. Weishaupt, Systemic dysregulation of TDP-43 binding microRNAs in amyotrophic lateral sclerosis, *Acta Neuropathol. Commun.* **1**, (2013) 42.