Interactive Visual Analysis of miRNA Target Prediction Results

Daekyoung Jung[1], Sehi L'Yi[1], Bohyoung Kim[2], Jinwook Seo[1]
[1]Human-Computer Interaction Laboratory, Seoul National University, Seoul, Korea
[2]Division of Biomedical Engineering, Hankuk University of Foreign Studies, Yongin-si, Gyeonggi-do, Korea
{rati, shlyi}@hcil.snu.ac.kr, bkim@hufs.ac.kr, jseo@snu.ac.kr

***Abstract*—Integrated analysis of mRNA and miRNAs is essential to comprehend regulation of gene expression. In this paper, we present a case study with miRTarVis, a recently introduced visual analytics tool that supports effective visualizations of the miRNA target prediction results. In the case study, we evaluate the feasibility of miRTarVis by applying it to analyzing miRNA-mRNA expression profiles from TCGA (The Cancer Genome Atlas) breast cancer dataset. Our study results show that miRTarVis can be effective in confirming the previously reported miRNA-mRNA interactions, and it has potential to help researchers generate new hypotheses when it is applied to new dataset.**

Keywords—miRNA Target Prediction; miRNA-mRNA Interaction; Graph Visualizations.

I. INTRODUCTION

Recent advancement of next generation sequencing (NGS) enabled researchers to acquire more accurate sequencing information with higher throughput. However, using NGS involves some challenges due to its large data size. There are billions of reads for whole-genome sequencing and ten millions of reads for transcriptomes and methylomes. According to Zhang et al. [1], NGS experiment generally generates terabytes of raw data. As a result, the importance of analysis tools for bioinformatics is increasing, since such tools are essential to achieve full benefit in large NGS data [2] [3].

NGS technologies enables researchers to measure important genomic factors in gene regulation. MicroRNA (miRNA) is one of the important genomic factors since it is an important gene expression regulator. In addition, regulatory RNAs contributes to epigenetic processes that control differentiation and development [4]. Therefore, integrated analysis of mRNA and miRNA (i.e., one of the well studied regulatory RNAs) is essential to comprehend the regulation of the gene expression.

Some analysis tools, such as MMIA [5], miRConnX [6], and MAGIA [7], were introduced to analyze miRNA-mRNA expression profile data using miRNA prediction algorithms. They integrated various sequence-based and miRNA-mRNA expression profile based prediction algorithms for more accurate miRNA target prediction. However, there still is a demand for more accurate miRNA target prediction with miRNA-mRNA expression data.

In addition, such tools help users to easily analyze miRNA-mRNA expression data by supporting user-friendly interfaces or visualizations for the miRNA-mRNA interactions. For example, mirConnX [8] and MAGIA2 [9] used a node-link diagram to visualize a miRNA-mRNA interaction network. However, the visualizations in such tools are limited in terms of a scalability and user interaction. For example, miRConnX and MAGIA are not proper to visualize large miRNA-mRNA interactions. In addition, it is difficult to explore the miRNA-mRNA interaction network in the current tools.

According to Keim [10], visual data exploration should follow Visual Information Seeking Mantra [11] for the scalable exploration: Overview first, zoom and filter, and details on demand. However, existing visualization tools cannot fully support the Visual Information Seeking Mantra due to the lack of user interactions.

One of the recent tools that is designed based on the Visual Information Seeking Mantra and support comprehensive target prediction algorithms is miRTarVis [12]. In this paper, we briefly introduce this tool and present the result of an additional case study using TCGA (The Cancer Genome Atlas) breast cancer dataset to further evaluate the feasibility of miRTarVis.

II. SYSTEM OVERVIEW

The visual analysis procedure in miRTarVis consists of four steps (Fig. 1): (1) load, (2) filter, (3) predict, and (4) visualize. Users first load miRNA-mRNA expression profiles in miRTarVis. Then, they filter the profile data to focus on the subset of miRNAs and mRNAs for the next steps. By using various miRNA target prediction algorithms, users can predict miRNA-mRNA interactions. Finally, users can see visualizations for the miRNA-mRNA interaction network to let the users grasp the structure of the network.

In miRTarVis, two novel visual representations for the miRNA-mRNA interaction networks are supported: enhanced node-link diagram and bipartite treemap. The enhanced node-link diagram (Fig. 1) intuitively shows overall structure of a miRNA-mRNA interaction network. To enhance the visual exploration of the network, users can select one of four graph layout algorithms (i.e., ISOM layout [13], KK layout [14], force-directed layout, and circular layout).

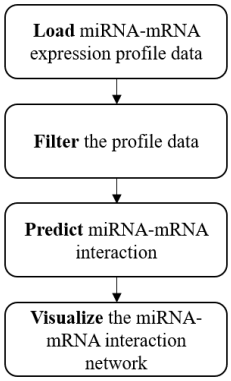

Figure 1. A visual analysis procedure in miRTarVis consists of four steps: load, filter, predict, and visualize.

Since the node-link diagram has occlusion problem, miRTarVis also provides another visualization technique (i.e., bipartite treemap, Fig. 2) that expands a Treemap visualization [15]. In the bipartite treemap, the size of mRNA nodes represents the significance of the prediction that is calculated by the prediction step. In addition, the color of node represents the fold change values.

III. Case Study

Using miRTarVis, we analyzed miRNA-mRNA expression profile data from TCGA (The Cancer Genome Atlas) breast cancer dataset. We downloaded miRNA-mRNA expression profile data from 60 cell lines. Among them, 50 samples are cancer cell lines while 10 samples are normal cell lines. We selected the miRNASeq and MiRNASeqV2 types for TCGA download data parameter.

In the load step of miRTarVis, we set the data type as unpaired two-sample data type because the number of cancer cells and normal cells are different. We set the t-test type as unequal variance since the variances of miRNAs and mRNAs are expected to be different, and we select the t-test mode as two-tailed t-test. When a user load miRNA-mRNA expression profile data, histograms appears to show the distribution of fold change of miRNAs and mRNAs. In filter step, a user can remain only the significant miRNAs and mRNAs.

In prediction step, we use correlation analysis (i.e., Pearson coefficient) with only negative correlation, mutual information, MINE, GenMiR++, TargetScan, and miRanda. In visualization step, the predicted miRNA-mRNA regulatory network is visualized by enhanced node-link diagram (Fig. 2) and bipartite treemap (Fig. 3).

As shown in the bipartite treemap (Fig. 3), the miRNA that has the most predicted target mRNA was hsa-miR-29b-3p. According to previous studies [16] [17], miR-29 plays an important role in development of cancer. In addition, Fig. 2 shows that CCND2 is an important mRNA that is predicted to be regulated by multiple miRNAs (i.e., hsa-let-7a-3p, hsa-let-7g-5p, has-miR-141-3p, and has-miR-29b-3p). We could verify that regulation of CCND2 gene by let-7a miRNA plays an important role in cancer development from a literature [18].

Using miRTarVis, we could effectively analyze the miRNA-mRNA expression profiles of TCGA breast cancer data. The visualizations generated by miRTarVis helped us to confirm previously reported miRNA-mRNA interactions from the literatures. This shows the possibility that miRTarVis mgiht also enable users to find new insights from miRNA-mRNA interactions that can be further verified by a biological experiment.

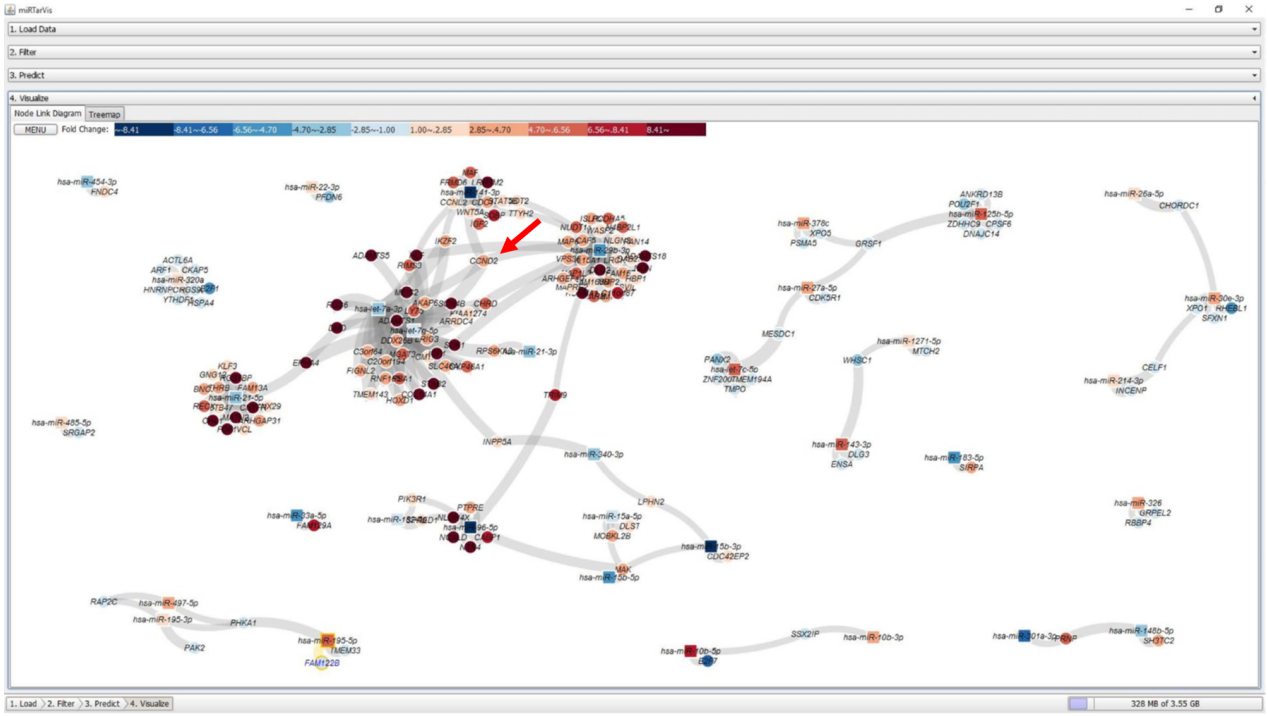

Figure 2. An enhanced node-link diagram shows the miRNA-mRNA expression profile data from TCGA breast cancer dataset. The thickness of links represents how significant the prediction is. Red and blue colors represent up- and down-regulated fold changes, respectively. In this visualization, CCND2 is predicted to be regulated by multiple miRNAs (i.e., hsa-let-7a-3p, hsa-let-7a-5p, hsa-miR-29b-3b, and hsa-miR-141-3p).

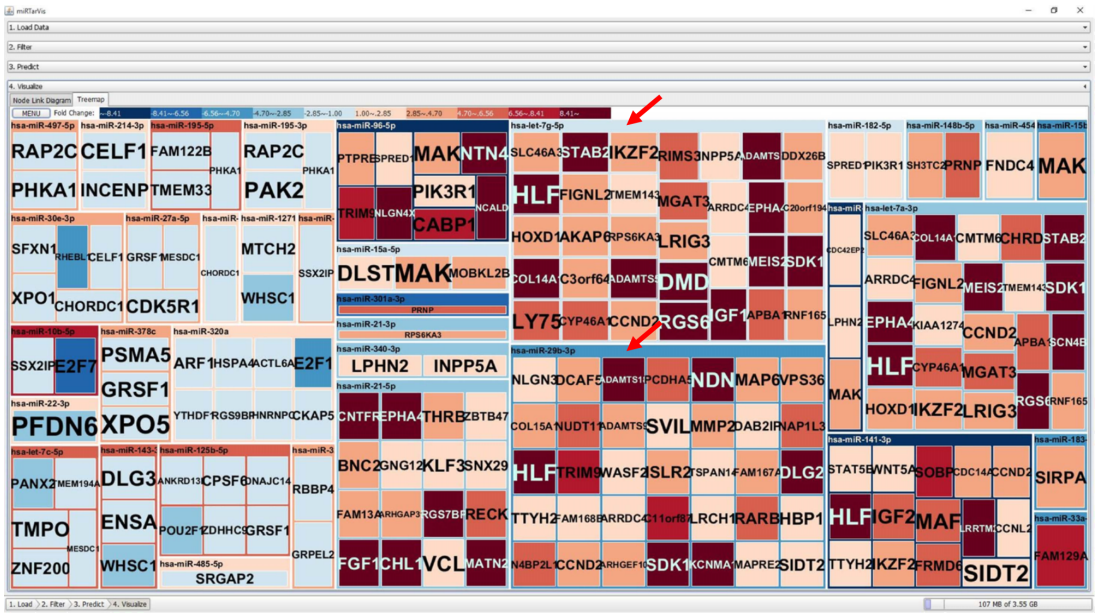

Figure 3. The bipartite treemap shows the predicted miRNA-mRNA regulatory network using TCGA breast cancer data. Red and blue colors represent up- and down-regulated fold changes. Two miRNAs (i.e., hsa-let-7g-5p and hsa-miR-29b-3p) are predicted to have many number of target mRNAs. We were able to expect that these two miRNAs play an important role.

IV. Conclusion

In this paper, we demonstrate miRTarVis, a recently introduced visual analytics tool that supports various miRNA target prediction algorithms with effective visualizations. We conducted a case study to further evaluate the feasibility of miRTarVis using TCGA breast cancer dataset. The case study results demonstrated that miRTarVis can be effective in confirming the previously reported miRNA-mRNA interactions, and it has potential to help researchers generate new hypotheses when it is applied to new datasets by exploring two interactive visualizations (i.e., enhanced node-link diagram and bipartite treemap).

Acknowledgment

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government(MEST) (No. NRF-2011-0030813).

References

[1] S. Zhang, Q. Li, J. Liu and X. J. Zhou, "A novel computational framework for simultaneous integration of multiple types of genomic data to identify microRNA-gene regulatory modules," Bioinformatics, vol. 27, no. 13, pp. i401--i409, 2011.

[2] J. Zhang, R. Chiodini, A. Badr and G. Zhang, "The impact of next-generation sequencing on genomics," Journal of genetics and genomics, vol. 38, no. 3, pp. 95-109, 2011.

[3] A. Teufel, M. Krupp, A. Weinmann and P. R. Galle, "Current bioinformatics tools in genomic biomedical research (Review)," International journal of molecular medicine, vol. 17, no. 6, pp. 967-973, 2006.

[4] K. V. Morris and J. S. Mattick, "The rise of regulatory RNA," Nature reviews. Genetics, vol. 15, no. 6, p. 423, 2014.

[5] S. Nam, M. Li, K. Choi, C. Balch, S. Kim and K. P. Nephew, "MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression," Nucleic acids research, p. gkp294, 2009.

[6] G. T. Huang, C. Athanassiou and P. V. Benos, "mirConnX: condition-specific mRNA-microRNA network integrator," Nucleic acids research, p. gkr276, 2011.

[7] G. Sales, A. Coppe, A. Bisognin, M. Biasiolo, S. Bortoluzzi and C. Romualdi, "MAGIA, a web-based tool for miRNA and Genes Integrated Analysis," Nucleic acids research, p. gkq423, 2010.

[8] G. T. Huang, C. Athanassiou and P. V. Benos, "mirConnX: condition-specific mRNA-microRNA network integrator," Nucleic acids research, p. gkr276, 2011.

[9] A. Bisognin, G. Sales, A. Coppe, S. Bortoluzzi and C. Romualdi, "MAGIA2: from miRNA and genes expression data integrative analysis to microRNA--transcription factor mixed regulatory circuits (2012 update)," Nucleic acids research, p. gks460, 2012.

[10] D. A. Keim, "Information visualization and visual data mining," Visualization and Computer Graphics, IEEE Transactions on, vol. 8, no. 1, pp. 1-8, 2002.

[11] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in Visual Languages, 1996. Proceedings., IEEE Symposium on, 1996.

[12] D. Jung, B. Kim, R. J. Freishtat, M. Giri, E. Hoffman, and J. Seo, “miRTarVis: an interactive visual analysis tool for microRNA-mRNA expression profile data”. In BMC proceedings, vol. 9, no. 6, pp. 1, 2015.

[13] B. Meyer, "Self-organizing graphs—a neural network perspective of graph layout," in Graph Drawing, 1998.

[14] T. Kamada and S. Kawai, "An algorithm for drawing general undirected graphs," Information processing letters, vol. 31, no. 1, pp. 7-15, 1989.

[15] M. Bruls, K. Huizing and J. J. Van Wijk, Squarified treemaps, Springer, 2000.

[16] Y. Pekarsky, U. Santanam, A. Cimmino, A. Palamarchuk, A. Efanov, V. Maximov, S. Volinia, H. Alder, C.-G. Liu, L. Rassenti and others, "Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181," Cancer research, vol. 66, no. 24, pp. 11590-11593, 2006.

[17] J. L. Mott, S. Kobayashi, S. F. Bronk and G. J. Gores, "mir-29 regulates Mcl-1 protein expression and apoptosis," Oncogene, vol. 26, no. 42, pp. 6133-6140, 2007.

[18] Q. Dong, P. Meng, T. Wang, W. Qin, W. Qin, F. Wang, J. Yuan, Z. Chen, A. Yang and H. Wang, "MicroRNA let-7a inhibits proliferation of human prostate cancer cells in vitro and in vivo by targeting E2F2 and CCND2," PloS one, vol. 5, no. 4, p. e10147, 2010