Prediction of Disease related miRNA through Transcription Factor(TF) Network

Jihwan Ha*, Yunku Yeu
Department of Computer Science,
Yonsei University
Seoul, Republic of Korea
jh0499@cs.yonsei.ac.kr

Sanghyun Park
Department of Computer Science,
Yonsei University
Seoul, Republic of Korea
sanghyun@cs.yonsei.ac.kr

***Abstract* — Since MicroRNA(miRNA) has been proved to be related to diseases, there has been continuous researches on miRNA. However, functions of miRNA and their oncogenic mechanisms are not confirmed evidently. In this paper, we suggest a method of finding disease related miRNA through prioritization using integrated miRNA dataset, and transcription factor(TF). It is anticipated that transcription factor could span the knowledge of scope that how miRNA regulates genes to occur specific disease. Based on miRNA network, we applied RNA-Seq data. RNA-Seq is a new way of approach that has been recently developed. It has an advantage of precise measurement and deep-sequencing technology. Many studies using this method have already proved that it produces more accurate result than using microarray data. In a network, based on differentially expressed nodes, we focused on DAG (directed acyclic graph). Together these findings might provide new insights on miRNA regulating the occurrence of diseases.**

Keywords— miRNA; Transcription Factor; network; DAG

I. INTRODUCTION

miRNAs are small non coding RNAs consists of 19-25 nucleotides which play important role in regulating genes.[1] According to previous studies, the major role of miRNAs was just delivering genetic codes. However further studies revealed that miRNAs are not only genetic code messenger but also plays role in almost every activities of life including cell proliferations.[2]

Although there have been numerous attempts on finding relationship between diseases and miRNAs, lack of data sets and functions led to insufficient results. To date, numerous analysis approaches have been proposed(Qinghua Jiang , 2010; Sheng-Da Hsu1, 2014).[3][4]

In this paper we suggest idea that by using integrated various data sets(miRecords, miRTarbase, miR2Disease) and miRNA network which consists of main network and sub network. Main network consists of miRNA nodes and edges. Edges are connected between two nodes based on their common target genes. Sub network consists of Transcription Factor(TF) which is usually known as promoter. Transcription Factor(TF) is a protein that binds to specific DNA sequence, regulates genetic codes. Applying Transcription factor to the sub network we might figure out how they influence miRNA. It is crucial to understand whether miRNA and TF work on same reactions.

And also, RNA-Seq data was used to see how every single node is differentially expressed on specific disease. RNA-Seq is a new way of approach that has several advantages over microarray technology, better estimate of absolute expression levels and ability of detecting new transcripts.

In this paper, we focused on DAG(directed acyclic graph) In a network with nodes that are differentially expressed on RNA-Seq. We might find a potential ability of miRNA which regulates genes to cause the diseases .

The remainder of this paper is organized as follows. In section 2, we describe how miRNA network is composed and how disease related miRNA can be detected. In section 3, we conclude by summarizing our work and discuss our future work.

II. METHOD AND MATERIALS

Compared to other data sets, the number of miRNA data set is relatively small. So integration of data sets(miRecords, miRTarbase, miR2Disease) is also needed .[5][6][7][8]

We constructed a network based on common target genes between two nodes as shown in the Figure 1. Nodes present miRNA and edges are connected based on common target genes. Because feature of miRNA is that if the miRNAs' function is similar, they might share similar target genes.

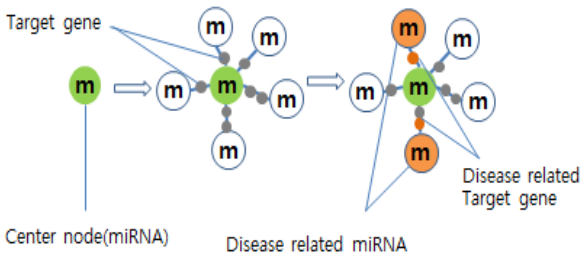

Fig.1 Making a miRNA network based on target genes

Every node has sub network which consists of Transcription Factor(TF), affecting the target gene of the miRNA.

ISSN 1975-4736 ©MITA 2014

Due to the limitation of microarray methods, RNA-seq uses recently developed deep-sequencing techniques. They use sequence-based approach to distinguish cDNA sequences.[9] We applied RNA seq to miRNA network.

We hypothesize that when nodes are differentially expressed with RNA-Seq, select two nodes with most highly expressed and highest degrees. Nodes have to form DAG(directed acyclic graph). Our assumption is that other nodes in DAG might be related to disease because of their connectivity to disease related miRNA as shown in the Figure 2. When Correlation Coefficient on each node is above 0.5 we assume that the node is related to the specific disease.[10]

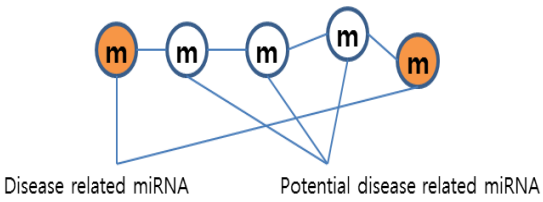

Fig. 2 Part of network that forms DAG

Based on scoring method every node in DAG(directed acyclic graph) is scored by their similarity distance between disease related nodes. Weight on each value of node is determined by following equation. Prioritization is performed based on scores of each nodes.

$$Weight = \frac{number\ of\ disease\ related\ nodes\ in\ DAG}{number\ of\ nodes\ in\ DAG} \quad (1)$$

Comparing Pearson Correlation Coefficient(PCC) on each node there might be a clue that whether transcription factor affects miRNA as a role of activator or inhibitor. Positive Pearson Correlation Coefficient means TF performs as activator, and negative Pearson Correlation Coefficient means it might performs as inhibitor.

III. CONCLUDION & FUTURE DIRECTIONS

Our approach provides a computational platform to integrate knowledge from different sources, such as miRNA and Transcription Factor(TF), and revealing potentially disease related miRNA.

Our expecting contribution is finding out disease related miRNA and revealing functions of transcription factor that could be affected by the miRNA.

However, due to the lack of data set and unveiled functions of miRNA there is an obstacle to extract the evidence of whether miRNAs are related to disease or not. We expect that we can solve these problems by using gene ontology(GO term) and various other investigations from our further researches.

REFERENCES

[1] V. Narry Kim, "MicroRNA biogenesis: coordinated cropping and dicing," Nature Reviews Molecular Cell Biology, vol.6, 2005

[2] David P.Barte, "MicroRNAs: genomics, biogenesis, mechanism and function," Cell, vol.6, 2005.

[3] Qinghua Jiang, Yangyang Hao, Guohua Wang, Liran Juan, Tianjiao Zhang, Mingxiang Teng, Yunlong Liu,Yadong Wang ,"Prioritization of disease microRNAs through a human phenome-microRNAome network", BMC Systems Biology, 2010

[4] Sheng-Da Hsu1, Yu-Ting Tseng, Sirjana Shrestha, Yu-Ling Lin, Anas Khaleel, Chih-Hung Chou1, Chao-Fang Chu1, Hsi-Yuan Huang, Ching-Min Lin, Shu-Yi Ho, "miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions", Nucleic Acids Research, vol.42, D78–D85, 2014

[5] Feifei Xiao, Zhixiang Zuo, Guoshuai Cai, Shuli Kang, Xiaolian Gao and Tongbin Li, "miRecords: an integrated resource for microRNA-target interactions", Nucleic Acids Research, vol.37, D105–.D110, 2009

[6] miRecords [Internet], http://mirecords.biolead.org/

[7] miR2Disease [Internet] http://www.mir2disease.org

[8] mirTarbase[Internet], http://mirtarbase.mbc.nctu.edu.tw/php/download.php

[9] Zhong Wang, Mark Getstein and Michael Snyder, "RNA-Seq: a revolutionary tool for transcriptomics", Nature Reviews Genetics, Vol 10, 2009

[10] Yang Li, Chengxiang Qiu, Jian Tu, Bin Geng, Jichun Yang, Tianzi Jiang,Qinghua Cui, "HMDD v2.0: a database for experimentally supported human microRNA and disease associations",Nucleic Acids Research,Vol. 42,D1070–.D1074, 2014