Discovery of Functional miRNA-mRNA Regulatory Modules with Computational Methods

Bing Liu[1], Jiuyong Li[1], Anna Tsykin[2]
[1]School of Computer and Information Science
University of South Australia
[2]Hanson Institute and Division of Human Immunology
Institute of Medical and Veterinary Science, Australia

Abstract

The identification of miRNAs and their target mRNAs and the construction of their regulatory networks may give new insights to biological procedures. This study proposes a computational method to discover the functional miRNA-mRNA regulatory modules (FMRMs), that is, groups of miRNAs and their target mRNAs that are believed to participate cooperatively in post-transcriptional gene regulation under specific conditions. The proposed method identifies negatively regulated patterns of miRNAs and mRNAs which associate with cancer and normal conditions respectively in a prostate cancer data set. GO and the literature also suggest that they may relate with prostate cancer. It can potentially identify the biologically relevant chains of 'miRNA → target gene → condition'.

1 Introduction

MicroRNAs (miRNAs) are a group of single-stranded, non-coding RNAs with ˜21-23 nucleotides in length. The mature miRNAs are cleaved from 70-110 nucleotide 'hairpin' like precursors with a double-stranded region containing one or more single-stranded loops [1]. miRNAs target protein coding mRNAs through complementary base-pairing for cleavage, repressing translation and causing protein degradation [2, 3]. Since the first miRNA, *lin-4*, was discovered in 1993, hundreds of miRNAs have been identified and sequenced in plants, animals, and viruses [4, 5]. As a group, miRNAs may directly regulate at least 30% of the genes in the human genome [6].

Increasing number of evidences suggest that miRNAs play important roles in cell differentiation, proliferation, growth, mobility, and apoptosis [7, 8, 9]. miRNAs negatively regulate their target mRNAs [10], and act as rheostats to make fine-scale adjustments to protein output [11]. Consequently, misregulation

of miRNA function may lead to human diseases, including cancers [12]. Recent studies have reported differentially regulated miRNAs in diverse cancer types such as breast cancer [13], lung cancer [14], prostate cancer [15], colon cancer [16] and ovarian cancer [17]. Thus, identifying miRNAs and their target mRNAs, and further building their regulatory networks may give new insights to biological procedures.

Several computational methods have been proposed for miRNA studies. Previous works largely focus on the genome-wide discovery of miRNAs [5] and prediction of putative target mRNAs [18]. For comprehensive reviews of these methods, we should like to refer readers to works of Zhang et al.[19] and Yoon et al. [20]. These methods have identified a large number of miRNAs and their target genes. For example, miRBase has deposited 5071 miRNAs and their target genes from 58 species [21] up to date.

An open question is how miRNAs associate with different conditions by regulating their *bona fide* target mRNAs. An answer to this question may also solve several problems in miRNA research. First, most target prediction methods search different parts of the miRNA-target space heuristically with different criteria [22]. They usually have very different results as Praveen et al. showed in their work [22]. However, only a small number of predicted miRNA-target interactions have been validated by interventional experiments in the laboratory [22]. Second, the computational algorithms used to predicate miRNA targets have limited accuracy. The complementarity between miRNAs and the binding sites of their target mRNAs is usually not long enough to be statistically significant. Again, there are relatively few miRNA targets which have experimental supports [22]. miRNAs and the predicted target mRNAs simply relying on the sequence complementarity may not be biologically relevant. Therefore, empirical methods are in need to find out true miRNA targets and reduce the false discovery rate. Third, according to the biological observation, multiple miRNAs regulate one message and one miRNA may have several target genes conversely [23]. It reflects that multiple miRNAs control the translation cooperatively, while single miRNA may involve several regulatory networks. This multiplicity of targets and cooperative signal integration on target genes are key features of the control of translation by miRNAs [24]. Thus, the discovery of the functional regulation networks which associate miRNAs and their target mRNAs with conditions is crucial for understanding the regulatory mechanism of miRNAs in complex cellular systems. It also will help depict the regulatory pathways of 'miRNA $\rightarrow$ target gene $\rightarrow$ condition'.

Some related works have been done previously at different levels. Yoon et al. [25] proposed a prediction method for miRNA regulatory modules (MRMs) at the sequence level. Their method is based on an observation that the binding strength of miRNAs and target mRNAs is modest and similar when multiple binding sites exist on a target. Predictions based on sequence only may lead to high false discovery rates. More precise predictions need more information like expression profiles.

Huang et al. [26] and Joung et al. [27] integrate both sequence information and expression profiles of miRNAs and mRNAs to identify the relevant miRNA-

mRNA pairs. The integrated approaches potentially reduce false discovery rate and facilitate the interventional experiments to validate the *bona fide* targets of miRNAs. Huang et al. [26] adopted Bayesian networks to model the regulatory mechanism of miRNAs. They found the miRNA target pairs through multiple tissues with relatively low false discovery rate. Joung et al. [27] utilized a bi-clustering approach to discover MRMs. The binding strength between miRNAs and mRNAs based on complementary base pairing as well as their expression profiles across various conditions are taken into account. They demonstrated that utilizing diverse resources including sequence information and expression profiles of miRNAs and mRNAs can achieve better prediction.

Unlike previous works as discussed above, this work identifies miRNAs and mRNAs regulatory modules within differentiated conditions. Specifically, this study proposes a computational method to discover the functional miRNA-mRNA regulation modules (FMRMs), that is, groups of miRNAs and their target mRNAs which are believed to participate cooperatively in post-transcriptional gene regulation under specific conditions, prostate cancer for example. The concept of FMRMs is an extension of MRMs introduced by Yoon et al. [25] but distinct from their works by associating MRMs with conditions. Many miRNAs have unique tissue-specific or developmental expression patterns such that each human tissue characterized by a specific set of miRNAs that may form a defining characteristic of that tissue [28]. Also, a large number of researches [29, 30] have suggested that the expression patterns of miRNAs are highly distinct in different tissues and cancers. We intend to discover the group of miRNA-mRNA target pairs which associate with certain condition, like prostate cancer and normal condition demonstrated in our study. It involves how each miRNA-mRNA pair expresses in certain modules besides the simple target. It is not general target prediction like many other researches (Huang et al., 2006 [26]). By associating the miRNA-mRNA modules with conditions in the model may help identify the functional groups which are involved in conditions directly, thus to reduce the false discovery rate of target prediction as well as to build the miRNA-mRNA regulatory networks.

2 Method

The goal of this study is to discover biologically relevant targets of miRNAs, and further to identify FMRMs for conditions with computational methods. It makes use of computationally predicted miRNA target information and corresponding expression profiles. The integration of expression information of miRNAs and mRNAs with their target information may discover the biologically relevant miRNA-mRNA duplexes. It can further identify functional miRNA-mRNA regulatory networks. The target information derived from sequence complementary base-pairing between miRNAs and mRNAs defines their putative networks. According to the biological observation that miRNAs negatively regulate their target mRNAs, their expression profile then will be applied to discover the condition related miRNA-mRNA modules within the putative net-

works. Thus, this problem can be divided into two sub-problems: i) to discover all the putative networks given the target information of miRNAs and mRNAs; ii) to derive FMRMs on expression data given the putative networks.

2.1 Modeling of the problem

Given the target information of miRNA, the relation between miRNAs and their targeted genes can be defined by a bipartite graph. Let S be a set of miRNAs and T a set of target mRNAs. The targeting relation is a bipartite graph $G = (S, T, E)$ with the edge set $E = \{(s, t)|$ miRNA $s \in S$ binds target $t \in T\}$ [25]. We further define the putative networks with at least m miRNAs and n mRNAs by a set of maximal bicliques $G_{m,n}$.

Definition 1: Putative networks $G_{m,n} = \{U, V, E'\}$, where $U \subseteq S$ and $|U| \geq m$, $V \subseteq T$ and $|V| \geq n$, and $E' \subseteq E$ are a set of maximal bicliques in G. They are subgraphs derived from G where miRNAs are fully connected with mRNAs. The numbers of miRNAs and mRNAs are no less than m and n respectively.

Given a bipartite graph derived from the target information between miRNAs and mRNAs, the solution for the first problem is to enumerate the maximal bicliques in which the numbers of the vertices from two sets are no less than m and n. Each maximal biclique consists of two sets of vertices U and V, denoting miRNAs and mRNAs respectively. They exhibit a full connectivity between vertex groups.

According to the biological observation that miRNAs negatively regulate their target genes, the second problem is to discover the co-regulated miRNA and mRNA groups on their expression data within each putative network.

Let P^S and P^T be the expression profiles of miRNAs and target mRNAs respectively. Both are profiled from L samples under K conditions $C = \{c_i|c_i \in (c_1, c_2, \cdots c_K)\}$. If the data sets are about two states of samples, then $K = 2$. For example c_1 refers to normal, and c_2 to cancer tissues. P^S can be denoted as a matrix in the size of $L \times |S|$ where the rows are samples and the columns are miRNAs. Correspondingly, P^T is denoted as a $L \times |T|$ matrix where the rows are samples and columns are mRNAs. Now we can define FMRM on expression profiles given the putative networks.

Definition 2: Functional miRNA-mRNA regulatory module (FMRM) G_F^i is subgraph of $G_{m,n}$ associating with condition $c_i, c_i \in (c_1, c_2, \cdots c_K)$ where the expressions of mRNAs and miRNAs are negatively regulated. That is, $G_F^i = \{U', V', E''\}$, where $U' \subseteq U$, $V' \subseteq V$, $E'' \subseteq E'$, and U' negatively regulates V' associating with condition c_i.

From the view of association rule mining [31], the second problem can be related with discovering the associations between conditions and miRNA-mRNA duplexes on their expression profiles. Association rule mining is a data mining technique to discover the frequent items that occur together frequently in a database. Given putative networks, the frequent items denoted by miRNAs and mRNAs with opposite patterns on expression profiles could be biologically relevant miRNA-mRNA modules. They potentially participate same biological process leading to the condition, cancer for example.

Example 1: Given G is a targeting relation between miRNAs $\{s_1, s_2, s_3, s_4\}$ and mRNAs $\{g_1, g_2, g_3, g_4, g_5, g_6\}$ (Figure 1 (a)), its putative networks with $m \geq 2$ and $n \geq 2$ are maximal bicliques denoted by miRNAs-mRNAs pairs $\{(s_1, s_2), (g_1, g_3, g_5)\}$, $\{(s_2, s_4), (g_2, g_3, g_6)\}$, $\{(s_1, s_2, s_3), (g_3, g_5)\}$, and $\{(s_3, s_4), (g_3, g_4)\}$ (Figure 1 (b)). Each maximal biclique defines a sub-data set from expression profiles of miRNA and mRNA. We further categorize their expression values to up-regulation and down-regulation, denoted by '↑' and '↓' respectively. It allows frequent itemset mining to discover the association between miRNA-mRNA duplexes and conditions on the reduced expression data sets. We are only interested in the association that miRNAs and mRNAs have opposite expression patterns. Suppose that from the sub-data set defined by the first maximal biclique $\{(s_1, s_2), (g_1, g_3, g_5)\}$ in Figure 1 (b), if frequent itemset mining discovers that the frequent itemset $\{s_1 \downarrow, s_2 \downarrow, g_1 \uparrow, g_5 \uparrow\}$ associates with cancer, we can construct relation $\{(s_1, s_2) \downarrow, (g_1, g_5) \uparrow \Rightarrow \textit{cancer}\}$ (Figure 2 (a) top panel). It suggests that miRNA s_1 and s_2 target mRNA g_1 and g_5. The down-regulation of s_1 and s_2 potentially leads to the up-regulation of their targets g_1 and g_5. As a group, this regulation associates with cancer. Similarly, suppose associations discovered on expression profiles for other three maximal bicliques are $\{s_2 \downarrow, g_6 \uparrow \Rightarrow \textit{cancer}\}$, $\{s_1 \downarrow, (g_3, g_5) \uparrow \Rightarrow \textit{cancer and } s_3 \downarrow, g_3 \uparrow \Rightarrow \textit{cancer}\}$, and $\{s_3 \downarrow, g_4 \uparrow \Rightarrow \textit{cancer}\}$. Finally, the individual patterns discovered from each putative network can be merged in terms of the common patterns of miRNAs and mRNAs, that is, the similar regulation of miRNAs and mRNAs. Thus, the functional regulatory module of miRNAs and their target mRNAs associating with cancer can be discovered (Figure 2 (b)).

3 Results

This method makes use of both miRNA target information and expression profiles of miRNAs and mRNAs. miRNA target information gives the basic relation between miRNAs and mRNAs. FMRMs are discovered on expression profiles given the putative networks which are derived from the target information. We tested this method with the public data including the miRNA target information and expression profiles.

3.1 Preparation of the data set

Several algorithms can predict miRNAs target mRNAs [22]. To demonstrate this method, we used the miRBase Targets Version 5.0 [21].

Expression profiles of miRNAs and mRNAs are extracted from an experiment conducted by Lu et al. [29]. We used 12 prostate samples where 6 samples were from cancer tissues and the other 6 samples were from normal tissues. We are interested in the functional miRNA-mRNA regulatory modules which are associated with conditions, cancer and normal in this case. Differentially expressed miRNAs and mRNAs are identified first for FMRM discovery. Among 217 miRNAs and 16063 mRNAs from the 12 samples, 159 miRNAs and 780

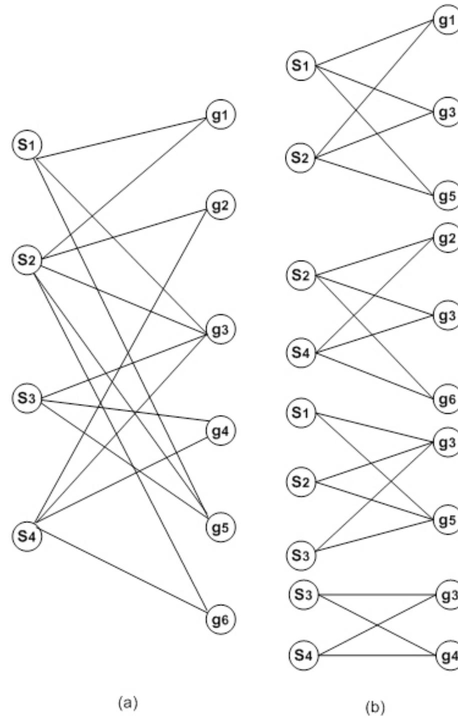

Figure 1: An Example of the bipartite graph and all its possible maximal bicliques with the number of nodes of both sides no less than 2. (a) A bipartite graph showing the putative targeting constructed by complementary base-pairing of miRNAs and mRNAs; (b) maximal bicliques with $m \geq 2$ and $n \geq 2$ of the putative networks in (a).

mRNAs are identified as differentially expressed with p-value$\leq$ 0.05 (see the additional file 1). Of these, 94 miRNAs and 152 mRNAs are linked together in miRBase. They constitute 623 binding pairs which were used in our experiments (see the additional file 2).

In order to integrate the expression profiles of miRNAs and mRNAs, we further categorized their expression values into up-regulation and down-regulation by discretization. The mean of miRNA(mRNA) expression values across the samples is the cutoff for each miRNA(mRNA). Values greater than the mean are categorized to up-regulation, otherwise down-regulation. It allows to combine expression profiles of miRNAs and mRNAs together and discover their regulatory pattern with association rule mining methods.

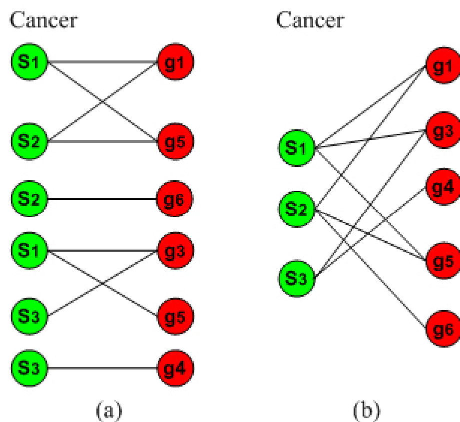

Figure 2: Example of FMRMs. (a) Association patterns discovered on exprssion profiles for each putative network. (b) FMRM associated with cancer can be identified by merging the redundant patterns. The down-regulation of s_1, s_2, and s_3 (colored in green) potentially leads to the up-regulation of their targets g_1, g_3, g_4, g_5, and g_6 (colored in red). As a group, this regulation associates with cancer.

Module	#miRNA-mRNA pairs	#miRNAs	#mRNAs
Cancer module	158	69	39
Normal module	166	60	38

Table 1: Numbers of miRNAs, mRNAs, and miRNA-mRNA pairs in Discovered FMRMs. General down-regulation of miRNAs and up-regulation of mRNAs are showed in cancer module, while the opposite regulation patters are discovered in normal module. Details in the additional file 3.

3.2 Discovery of FMRMs

In this test, 158 putative networks are identified. Association are further discovered on expression profiles within each putative network. After pruning the redundant relations, there are 158 pairs of miRNAs and mRNAs associating with cancer, and 166 pairs associating with normal (Table 1). A large amount of miRNA-mRNA pairs from cancer and normal modules are common but with opposite regulation patterns corresponding to cancer and normal respectively. Of these, miRNAs show down-regulation while their target mRNAs are up-regulated in the prostate cancer modules. On the contrary, miRNAs are up-regulated while their target mRNAs are down-regulated in normal modules in general.

miRNAs	Targeted mRNAs	p-Value
hsa-miR-107	*MRPS16, C12orf11, BCAM, ABHD12, FAM82B, PTX3, APOBEC3B*	1.80e-04
hsa-miR-24	*SLC12A6, MST1R, ABHD12, CHKB, PROC, DDX56, PVRL2*	1.90e-04

Table 2: miRNA targets in the cancer module queried from g:Profiler. The targeted genes discovered are highlighted.

3.3 Demonstration of negative correlation

We expect the miRNA-mRNA pairs identified in FMRMS are negatively correlated, thus we calculated the Pearson's correlation coefficients between miRNAs and mRNAs identified in the modules. It is supposed that the miRNAs negatively correlate with their target mRNAs on the expression profiles, while no correlation if there is no regulation between them. To compare with them, we also calculated the correlation of random miRNA-mRNA pairs. As displayed in Figure 3, the correlation coefficient of random pairs distribute around 0 randomly, while the identified pairs in cancer and normal modules are largely different from it. The t-test shows the difference with p-values at 6.64×10^{-45} for cancer module and 6.91×10^{-51} for normal module. Furthermore, we estimated the distribution of mean of correlation of random modules by sampling miRNA-mRNA pairs randomly for 10000 times. The mean of the correlation is largely around 0 to 0.05, while those of cancer and normal modules are less than -0.5 (Figure 4). It clearly shows that the pairs identified in the FMRMs are largely negatively correlated in both modules.

3.4 Validation with Gene Ontology

To validate the FMRMs discovered by this method, we queried the Gene Ontology the identified genes of FMRMs by g:Profiler [32] and GO::TermFinder [33] for annotation. They are tools to find common high-level knowledge including pathways, biological processes, molecular functions, and subcellular localizations accompanied by statistical significance to the list of given genes. They derive the annotations from several databases including Gene Ontology [34].

From the cancer module, it suggests that *hsa-miR-107* targets *MRPS16, C12orf11, BCAM, ABHD12*, and *FAM82B*. *hsa-miR-24* targets *SLC12A6, MST1R, ABHD12*, and *CHKB*. They are largely consistent with the query results of g:Profiler (Table 2). Similarly, the target pairs of normal modules are also consistent with the query results at significant levels (Table 3).

In addition, we queried the genes in the modules with GO::TermFinder to find significantly over-expressed GO terms. From both cancer and normal modules, it consistently shows that *NAB2* and *LSM7* participate in mRNA processing and mRNA metabolic process (Table 4) while they display opposite regulation patterns in different modules (up-regulation in cancer and down-regulation

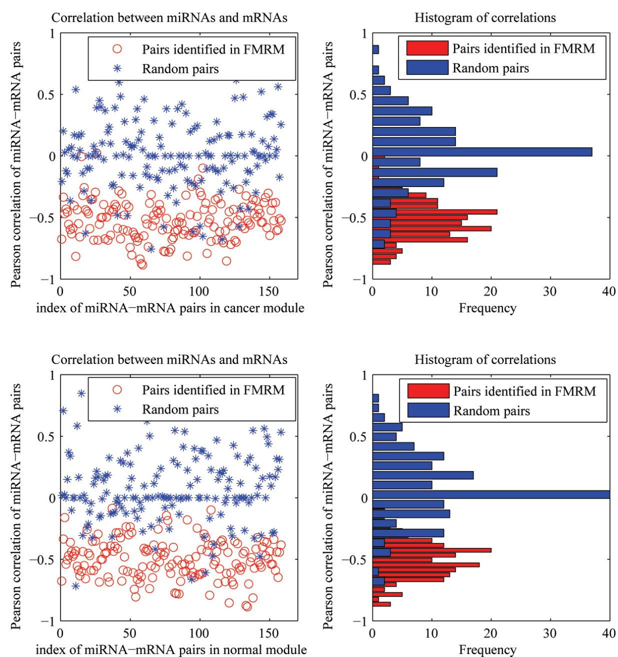

Figure 3: Comparison of correlation of miRNA-mRNA pairs. Correlation of miRNAs-mRNAs identified in cancer module (top-left) and normal module (bottom-left) are colored in red, while that of random pairs are colored in blue. The corresponding distributions are given at the right sides. It shows the miRNA-mRNA paris identified in both modules are largerly negatively correlated. They are differentiated from the random pairs with p-value at $6.64e-45$ for cancer module and $6.91e-51$ for normal module.

in normal). It suggests that *NAB2* and *LSM7* targeted by miRNAs identified in FMRMs may participate in the mRNA metabolism which is very important to tumor development.

3.5 Supports from the literature

Further supports from the literature indicate that *NAB2* functions together with another gene to modulate mRNA stability and strengthen a model where nuclear events are coupled to the control of mRNA turnover in the cytoplasm [35]. In addition, *LSM7* has been reported a role in mRNA degradation [36]. They are targeted by a group of miRNAs in our FMRMs. Given the regulation roles of miRNAs, *NAB2* and *LSM7* may be two of the genes triggered by miRNAs and function in mRNA degradation.

Several evidences suggest that many miRNAs in the identified FMRMs may be related with cancers. Among the miRNAs which target *NAB2* and *LSM7*

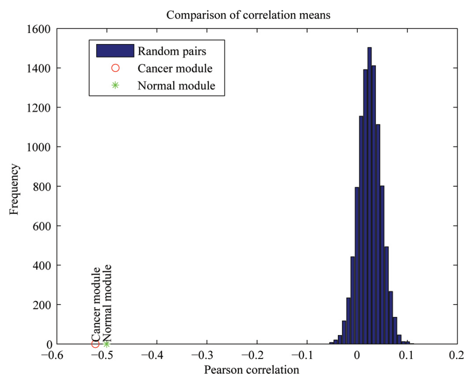

Figure 4: Distribution of the mean of the correlation of randomly chosen miRNA-mRNA modules. It was estimated by 10000 random sampling of modules including 160 miRNA-mRNA pairs.

miRNAs	**Targeted mRNAs**	**p-Value**
hsa-miR-24	*MST1R, CYB561D2, ABHD12, PROC, DDX56, CHKB, PVRL2, SLC12A6*	2.36e-05
hsa-miR-10a	*FAM82B, JUNB, BYSL, C19orf10, PROC, ARMC9, PVRL2*	1.78e-04
hsa-miR-107	*FAM82B, ABHD12, MRPS16, BCAM, C12orf10, APOBEC3B, PTX3*	1.80e-04

Table 3: miRNA targets in the normal module queried from g:Profiler. The targeted genes discovered are highlighted.

in FMRMs, *hsa-miR-15a* and *hsa-miR-16* have been strongly suggested the involvement of human cancers, including prostate cancer. They are frequently deleted or down-regulated in prostate cancer as well as other cancers [37, 38]. Our FMRMs display that *hsa-miR-15a* and *hsa-miR-16* have the consistent down-regulation patterns with these studies. In addition, it suggests that *let-7* family members may act as tumor suppressors [39]. In the discovered FMRMs, *let-7a, let-7c, let-7d, let-7e, let-7f, let-7g,* and *let-7i* show general down-regulation in cancer module, but up-regulation in normal module, suggesting that they play important roles in prostate cancer as well.

Several genes targeted by miRNAs in FMRMs may related with cancers. Oncogene *PIM3*, one of targets of *hsa-miR-15a* identified in FMRMs, belongs to a family of protooncogenes. It has been reported that *PIM3* was detected to be expressed in human pancreatic cancer tissue and pancreatic cancer cell lines, but not in normal tissue [40]. It is mapped to chromosome 22q where recently received great interest for prostate cancer. It has been shown that this region is related with prostate cancer [41]. Our modules suggest that *hsa-*

Annotated Genes	Item	Value
NAB2 *LSM7*	GO ID	GO:0006397
	Term	mRNA processing
	p-value	0.01573
	Genome frequency of use	151 out of 7150 genes
	GO ID	GO:0016071
	Term	mRNA metabolism
	p-value	0.03118
	Genome frequency of use	213 out of 7150 genes

Table 4: Enriched GO terms. Significantly over-expressed GO terms are mRNA processing and mRNA metabolism. Given the regulation roles of miRNAs, NAB2 and LSM7 may be two of the genes triggered by miRNAs and function in mRNA degradation.

miR-15a targets *PIM3* and their dysregulation may associate with prostate cancer. Similarly, *JUNB*, one of the genes identified in the FMRMs shows up-regulation in cancer but down-regulation in normal. It has been identified as a key transcriptional regulator of myelopoiesis and a potential tumor suppressor gene [42]. In addition, *FAT*, another identified gene in the FMRMs, encodes a tumor suppressor gene. Its recessive mutations lead to hyperplastic overgrowth of the imaginal discs, indicating that contact-dependent cell interactions may play an important role in regulating growth [43]. These genes are identified in FMRMs by our method, suggesting they may associate with prostate cancer.

4 Discussion

miRNAs comprise one of abundant classes of gene regulatory molecules in multicellular organisms. They potentially influence the output of many protein-coding genes by binding to and inhibiting mRNAs. Increasing number of evidences suggest that miRNAs have critical roles in diseases including cancers. Computational methods modeling miRNA-mRNA regulatory modules will help understand the complex biological procedures.

Many models have been proposed to predict miRNA-mRNA regulatory modules. Yoon et al. [25], Huang et al. [26], and Joung et al. [27] proposed models at different levels for prediction. Based on different biological observation, their models are either on sequence level or integration of sequence and expression profiles of miRNAs and mRNAs. Among them, Huang et al. [26] and Joung et al. [27] demonstrated that utilizing diverse resources including sequence and expression profiles of miRNAs and mRNAs can achieve better prediction.

The method proposed in this study goes further by associating the conditions with miRNA-mRNA regulatory modules. The FMRMs defined in this work consist of miRNAs, target mRNAs, and associated conditions, cancer and normal in the demonstrated case. The miRNAs and mRNAs identified in FM-

RMs show many meaningful discoveries supported by GO and the literature. By associating the miRNA-mRNA pairs with conditions, the method may eventually identify the regulatory networks of 'miRNA → target gene → condition'. It will bring new chance to the treatment for genetic diseases, such as cancer.

This method is independent of sequence target information and expression profiles. It is possible to refine the FMRMs with the continued update of target information. It also can be applied to different expression profiles when they are available.

5 Conclusions

We proposed a computational method for identifying functional regulatory miRNA-mRNA modules using predicted miRNA targets as well as expression profiles of miRNAs and mRNAs. It incorporates heterogeneous information to discover the biologically relevant miRNA-mRNA groups. The FMRMs identified in this study include the negatively correlated miRNA-mRNA pairs which associate with prostate cancer and normal condition. They display many meaningful discoveries supported by GO and the literature. By associating miRNA-mRNA pairs with conditions, it potentially can identify the biologically relevant targets of miRNAs and chains of 'miRNA → target gene → condition'. It will give new insight into the biological procedures at the molecular level.

6 Implementation

Differentially expressed miRNAs and mRNAs were identified using the empirical Bayes approach which ranks genes on a combination of magnitude and consistency of differential expression [44]. miRNAs and mRNAs with p-values less than 0.05 were identified as differentially expressed. It was analyzed in the freely available statistical programming and graphics environment R (http://cran.r-project.org). MICA algorithm [45] was used to enumerate maximal bicliques which were used to define the putative networks in this method. Cut-off values for the number of miRNA and mRNA were set at 2. Association between miRNA-mRNA and conditions were mined using the Apriori algorithm [46]. The minimum support was set at 0.3. Only class association rules were mined instead of general association rules. It was implemented on Matlab 7.1 with weka API [47].

7 Authors contributions

BL conceived of the study and carried out the computational experiment. BL, JYL and AT drafted the manuscript. All authors read and approved the final manuscript.

8 Acknowledgements

The authors thank Associate Professor Cristin Print at School of Medical Sciences, University of Auckland for constructive discussions.

This research has been supported by ARC DP0559090.

References

[1] Mariana Lagos-Quintana, Reinhard Rauhut, Winfried Lendeckel, and Thomas Tuschl. Identification of Novel Genes Coding for Small Expressed RNAs. *Science*, 294(5543):853–858, 2001.

[2] David P. Bartel. Micrornas: genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281–197, January 2004.

[3] Lin He and Gregory J. Hannon. Micrornas: small rnas with a big role in gene regulation. *Nature Reviews Genetics*, 5:522–531, July 2004.

[4] Eugene Berezikov, Edwin Cuppen, and Ronald H. A. Plasterk. Approaches to microrna discovery. *Nature Genetics*, 38(Suppl.):S2–7, June 2006.

[5] Isaac Bentwich, Amir Avniel, Yael Karov, Ranit Aharonov, Shlomit Gilad, Omer Barad, Adi Barzilai, Paz Einat, Uri Einav, Eti Meiri, Eilon Sharon, Yael Spector, and Zvi Bentwich. Identification of hundreds of conserved and nonconserved human micrornas. *Nature Genetics*, 37:766–770, June 2005.

[6] Benjamin P. Lewis, Christopher B. Burge, and David P. Bartel. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microrna targets. *Cell*, 120:15–20, January 2005.

[7] Victor Ambros. The functions of animal micrornas. *Nature*, 431:350–355, September 2004.

[8] Tingting Du and Phillip D. Zamore. Beginning to understand microrna function. *Cell Research*, 17:661–663, August 2007.

[9] Natascha Bushati and Stephen M. Cohen. microrna functions. *The Annual Review of Cell and Developmental Biology*, 23:175–205, May 2007.

[10] Lee P. Lim, Nelson C. Lau, Philip Garrett-Engele, Andrew Grimson, Janell M. Schelter, John Castle, David P. Bartel, Peter S. Linsley, and Jason M. Johnson. Microarray analysis shows that some micrornas downregulate large numbers of target mrnas. *Nature*, 433:769–773, February 2005.

[11] Daehyun Baek, Judit Villn, Chanseok Shin, Fernando D. Camargo, Steven P. Gygi, and David P. Bartel. The impact of micrornas on protein output. *Nature*, 445:64–71, September 2008.

[12] Chunxiang Zhang. MicroRNomics: a newly emerging approach for disease biology. *Physiol. Genomics*, 33(2):139–147, 2008.

[13] Marilena V. Iorio, Manuela Ferracin, Chang-Gong Liu, Angelo Veronese, Riccardo Spizzo, Silvia Sabbioni, Eros Magri, Massimo Pedriali, Muller Fabbri, Manuela Campiglio, Sylvie Menard, Juan P. Palazzo, Anne Rosenberg, Piero Musiani, Stefano Volinia, Italo Nenci, George A. Calin, Patrizia Querzoli, Massimo Negrini, and Carlo M. Croce. MicroRNA Gene Expression Deregulation in Human Breast Cancer. *Cancer Res*, 65(16):7065–7070, 2005.

[14] Nozomu Yanaihara, Natasha Caplen, Elise Bowman, Masahiro Seike, Kensuke Kumamoto, Ming Yi, Robert M. Stephens, Aikou Okamoto, Jun Yokota, Tadao Tanaka, George Adrian Calin, Chang-Gong Liu, Carlo M. Croce, and Curtis C. Harris. Unique microrna molecular profiles in lung cancer diagnosis and prognosis. *Cancer Cell*, 9:189–198, March 2006.

[15] Kati P. Porkka, Minja J. Pfeiffer, Kati K. Waltering, Robert L. Vessella, Teuvo L.J. Tammela, and Tapio Visakorpi. MicroRNA Expression Profiling in Prostate Cancer. *Cancer Res*, 67(13):6130–6135, 2007.

[16] Yukihiro Akao, Yoshihito Nakagawa, and Tomoki Naoe. Microrna-143 and -145 in colon cancer. *DNA and Cell Biology*, 26(5):311–320, 2007.

[17] Hua Yang, William Kong, Lili He, Jian-Jun Zhao, Joshua D. O'Donnell, Jiawang Wang, Robert M. Wenham, Domenico Coppola, Patricia A. Kruk, Santo V. Nicosia, and Jin Q. Cheng. MicroRNA Expression Profiling in Human Ovarian Cancer: miR-214 Induces Cell Survival and Cisplatin Resistance by Targeting PTEN. *Cancer Res*, 68(2):425–433, 2008.

[18] Azra Krek, Dominic Grun, Matthew N Poy, Rachel Wolf, Lauren Rosenberg, Eric J Epstein, Philip MacMenamin, Isabelle da Piedade, Kristin C Gunsalus, Markus Stoffel, and Nikolaus Rajewsky. Combinatorial microrna target predictions. *Nature Genetics*, 37(5):495–500, May 2005.

[19] Baohong Zhang, Xiaoping Pan, Qinglian Wang, George P. Cobb, and Todd A. Anderson. Computational identification of micrornas and their targets. *Computational Biology and Chemistry*, 30(6):395–407, 2006.

[20] Sungroh Yoon and Giovanni De Micheli. Computational identification of micrornas and their targets. *Birth Defects Res C Embryo Today*, 78(2):118–128, June 2006.

[21] Sam Griffiths-Jones, Harpreet Kaur Saini, Stijn van Dongen, and Anton J. Enright. miRBase: tools for microRNA genomics. *Nucl. Acids Res.*, 36(suppl_1):D154–158, 2008.

[22] Praveen Sethupathy, Molly Megraw, and Artemis G Hatzigeorgiou. A guide through present computational approaches for the identification of mammalian microrna targets. *Nature Methods*, 3:881–886, November 2006.

[23] Anton J Enright, Bino John, Ulrike Gaul, Thomas Tuschl, Chris Sander, and Debora S Marks. Microrna targets in drosophila. *Genome Biology*, 5(1):R1, December 2003.

[24] Bino John, Anton J. Enright, Alexei Aravin, Thomas Tuschl, Chris Sander, and Debora S. Marks. Human microrna targets. *PLoS Biol*, 2(11):e363, October 2004.

[25] Sungroh Yoon and Giovanni De Micheli. Prediction of regulatory modules comprising microRNAs and target genes. *Bioinformatics*, 21(suppl_2):ii93–100, 2005.

[26] Jim C. Huang, Quaid D. Morris, and Brendan J. Frey. Detecting microrna targets by linking sequence, microrna and gene expression data. *Research in Computational Molecular Biology*, 3909/2006:114–129, August 2006.

[27] Je-Gun Joung, Kyu-Baek Hwang, Jin-Wu Nam, Soo-Jin Kim, and Byoung-Tak Zhang. Discovery of microRNA mRNA modules via population-based probabilistic learning. *Bioinformatics*, 23(9):1141–1147, 2007.

[28] David P. Bartel. Micrornas: Genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281–297, January 2004.

[29] Jun Lu, Gad Getz, Eric A. Miska, Ezequiel Alvarez-Saavedra, Justin Lamb, David Peck, Alejandro Sweet-Cordero, Benjamin L. Ebert, Raymond H. Mak, Adolfo A. Ferrando, James R. Downing, Tyler Jacks, Robert R. Horvitz, and Todd R. Golub. Microrna expression profiles classify human cancers. *Nature*, 435(7043):834–838, June 2005.

[30] Stefano Volinia, George A. Calin, Chang-Gong Liu, Stefan Ambs, Amelia Cimmino, Fabio Petrocca, Rosa Visone, Marilena Iorio, Claudia Roldo, Manuela Ferracin, Robyn L. Prueitt, Nozumu Yanaihara, Giovanni Lanza, Aldo Scarpa, Andrea Vecchione, Massimo Negrini, Curtis C. Harris, and Carlo M. Croce. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2257–2261, 2006.

[31] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. In *SIGMOD '93: Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, pages 207–216, New York, NY, USA, 1993. ACM.

[32] Juri Reimand, Meelis Kull, Hedi Peterson, Jaanus Hansen, and Jaak Vilo. g:Profiler–a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucl. Acids Res.*, 35(suppl_2):W193–200, 2007.

[33] Elizabeth I. Boyle, Shuai Weng, Jeremy Gollub, Heng Jin, David Botstein, J. Michael Cherry, and Gavin Sherlock. GO::TermFinder–open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, 20(18):3710–3715, 2004.

[34] The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nature Genetics*, 25:25–29, 2000.

[35] Luciano H. Apponi, Seth M. Kelly, Michelle T. Harreman, Alexander N. Lehner, Anita H. Corbett, and Sandro R. Valentini. An Interaction between Two RNA Binding Proteins, Nab2 and Pub1, Links mRNA Processing/Export and mRNA Stability . *Mol. Cell. Biol.*, 27(18):6569–6579, 2007.

[36] Emmanuelle Bouveret, Guillaume Rigaut, Anna Shevchenko, Matthias Wilm, and Bertrand Seraphin. A sm-like protein complex that participates in mrna degradation. *The EMBO Journal*, 19:1661C1671, February 2000.

[37] Jin-Tang Dong, James C. Boyd, and Henry F. Frierson Jr. Loss of heterozygosity at 13q14 and 13q21 in high grade, high stage prostate cancer. *The Prostate*, 49(3):166–171, 2001.

[38] George Adrian Calin, Calin Dan Dumitru, Masayoshi Shimizu, Roberta Bichi, Simona Zupo, Evan Noch, Hansjuerg Aldler, Sashi Rattan, Michael Keating, Kanti Rai, Laura Rassenti, Thomas Kipps, Massimo Negrini, Florencia Bullrich, and Carlo M. Croce. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24):15524–15529, 2002.

[39] George Adrian Calin, Cinzia Sevignani, Calin Dan Dumitru, Terry Hyslop, Evan Noch, Sai Yendamuri, Masayoshi Shimizu, Sashi Rattan, Florencia Bullrich, Massimo Negrini, and Carlo M. Croce. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(9):2999–3004, 2004.

[40] Ying-Yi Li, Boryana K. Popivanova, Yuichiro Nagai, Hiroshi Ishikura, Chifumi Fujii, and Naofumi Mukaida. Pim-3, a Proto-Oncogene with Serine/Threonine Kinase Activity, Is Aberrantly Expressed in Human Pancreatic Cancer and Phosphorylates Bad to Block Bad-Mediated Apoptosis in Human Pancreatic Cancer Cell Lines. *Cancer Res*, 66(13):6741–6747, 2006.

[41] Nicola J. Camp, James M. Farnham, and Lisa A. Cannon-Albright. Localization of a Prostate Cancer Predisposition Gene to an 880-kb Region on Chromosome 22q12.3 in Utah High-Risk Pedigrees. *Cancer Res*, 66(20):10205–10212, 2006.

[42] Emmanuelle Passegue, Wolfram Jochum, Marina Schorpp-Kistner, Uta Mohle-Steinlein, and Erwin F. Wagner. Chronic myeloid leukemia with increased granulocyte progenitors in mice lacking junb expression in the myeloid lineage. *Cell*, 104(1):21–32, January 2001.

[43] Peter J. Bryant, Beth Huettner, Lewis I. Held Jr., Jan Ryerse, and Janos Szidonya. Mutations at the fat locus interfere with cell proliferation control and epithelial morphogenesis in drosophila. *Developmental Biology*, 129(2):541–554, October 1998.

[44] Gordon K. Smyth. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Artical 3, February 2004.

[45] Gabriela Alexe, Sorin Alexe, Yves Crama, Stephan Foldes, Peter L. Hammer, and Bruno Simeone. Consensus algorithms for the generation of all maximal bicliques. *Discrete Appl. Math.*, 145(1):11–21, 2004.

[46] Keyun Hu, Yuchang Lu, Lizhu Zhou, and Chunyi Shi. Integrating classification and association rule mining: A concept lattice framework. In *RSFDGrC '99: Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing*, pages 443–447, London, UK, 1999. Springer-Verlag.

[47] Ian H. Witten and Eibe Frank. *Data Mining: Practical machine learning tools and techniques.* Morgan Kaufmann, San Francisco, 2005.

Additional Files

Additional file 1 — Differentially expressed miRNAs and mRNAs

Additional file 2 — Binding pairs of miRNAs and mRNAs linked by miRBase in this study

Additional file 3 — Discovered functional miRNA-mRNA regulatory modules

Additional files are available online at http://sim.cis.unisa.edu.au/datamining.