A Novel Method to Detect Functional microRNA Regulatory Modules by Bicliques Merging

Cheng Liang, Yue Li, and Jiawei Luo

Abstract—MicroRNAs (miRNAs) are post-transcriptional regulators that repress the expression of their targets. They are known to work cooperatively with genes and play important roles in numerous cellular processes. Identification of miRNA regulatory modules (MRMs) would aid deciphering the combinatorial effects derived from the many-to-many regulatory relationships in complex cellular systems. Here, we develop an effective method called BiCliques Merging (BCM) to predict MRMs based on bicliques merging. By integrating the miRNA/mRNA expression profiles from The Cancer Genome Atlas (TCGA) with the computational target predictions, we construct a weighted miRNA regulatory network for module discovery. The maximal bicliques detected in the network are statistically evaluated and filtered accordingly. We then employed a greedy-based strategy to iteratively merge the remaining bicliques according to their overlaps together with edge weights and the gene-gene interactions. Comparing with existing methods on two cancer datasets from TCGA, we showed that the modules identified by our method are more densely connected and functionally enriched. Moreover, our predicted modules are more enriched for miRNA families and the miRNA-mRNA pairs within the modules are more negatively correlated. Finally, several potential prognostic modules are revealed by Kaplan-Meier survival analysis and breast cancer subtype analysis. **Availability:** BCM is implemented in Java and available for download in the supplementary materials, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/ TCBB.2015.2462370.

Index Terms—MicroRNA regulatory module; maximal bicliques, bicliques merging, survival analysis, breast cancer subtype analysis

1 Introduction

MicroRNAs (miRNAs) are a family of small, non-coding RNAs that regulate gene expression in a sequence-specific manner [1]. Two main mechanisms for such effects are degradation of the target mRNA, and inhibition of its translation [2]. They are predicted to control the activity of more than 60 percent of all protein-coding genes in mammals and participate in the regulation of numerous cellular processes [3]. The link between dysregulated miRNA expression and diverse cancer types has been well established [4], such as breast cancer (BRCA) [5], lung cancer [6], ovarian cancer [7] and etc. Therefore, understanding the functional roles of miRNAs is critical to elucidating the development and inhibition of the tumorigenesis, and further facilitating the discovery of the potential drug targets.

Great efforts have been made to study the functions of miRNAs in cellular processes using experimental and computational approaches. Early methods mainly focused on the prediction of miRNA-mRNA relationships using sequence information or expression profiles [8]. For instance, based on the sequence complementarity and interspecies conversation, John et al. successfully predicted the fragile X mental retardation protein (FMRP)-associated miRNAs [9]. Moreover, Lu et al. devised a bead-based hybridization profiling method for miRNAs from human cancer samples and accurately classified poorly differentiated tumours using the generated miRNAs expression profiles [10]. Importantly, studies have indicated that miRNAs regulate mRNAs synergistically since a limited number of miRNAs are able to control the expression of a much larger set of genes [11], [12], [13]. Indeed, despite their limited number, miRNAs may be in charge of more evolutionarily robust and potent regulatory effects through coordinated collective actions [14]. Thus, it is important to detect the functional modules involved in complex regulations between miRNAs and their targets to elucidate the collaborative regulatory mechanism of miRNAs.

Identification of miRNA regulatory modules (MRMs) has greatly advanced our understanding of miRNA regulatory networks. MRMs are considered as groups of miRNAs and target genes that participate coopratively in post-transcriptional gene regulation [15]. Several approaches have been proposed to detect biologically meaningful MRMs [11], [14], [15], [16], [17], [18], [19], [20], [21], [22]. Briefly, Yoon and De Micheli developed an algorithm to predict MRMs each having similar targets using only sequence information without considering their expression profiles [15]. Joung et al. developed a probabilities search method to identify MRMs with both the predicted miRNA targets information and their expression data sets. They adopted a co-evolutionary learning procedure that evolves cooperatively for two populations of miRNAs and mRNAs respectively [16]. However, their algorithm needs prior setting of several parameters of the fitness function, which would substantially affect the final outcomes. Peng et al. searched for MRMs by first enumerating all maximal bicliques in the miRNA-mRNA regulatory network as candidates and then filtering statistically insignificant modules according to statistical tests [11]. However, the approach tends to generate modules with only one miRNA and only operates on unweighted networks. Liu et al. employed a probabilistic graphical model called correspondence latent Dirichlet allocation (Corr-LDA) to separately sample from the miRNA and mRNA pools assuming that miRNA and mRNA within the same MRM participate in the same latent functions [19]. Corr-LDA allows detecting the MRMs with or without the miRNA-mRNA target information. Nonetheless, their approach requires a predefined number of modules, which might be difficult to decide beforehand. Zhang et al. proposed a sparse network-regularized multiple non-negative matrix factorization (SNMNMF) framework which could integrate not only the predicted miRNA-gene regulation and expression information but also the protein-protein interaction (PPI) and DNA-protein interaction networks to construct the MRMs [21]. SNMNMF also requires a predefined number of modules, and it cannot guarantee that all the identified modules contain both miRNA and mRNA. Pio et al. proposed a novel biclustering algorithm that is able to automatically determine the optimal number of biclusters. The algorithm starts with an initial set of bicliques and then it iteratively defines the hierarchical organization of biclusters according to a bottom-up strategy [23]. However, it does not take into account the miRNA and mRNA expression profiles. Recently, our research group developed a new method called Mirsynergy based on a greedy-based algorithm ClusterONE to detect the MRMs [14]. It operates in two clustering stages, where at the first stage they only consider maximizing the miRNA-miRNA synergy and then add/remove genes to/from each miRNA cluster at the second stage to further maximize the synergy score based on miRNA-gene and gene-gene interactions.

In this work, we propose a novel computational model called BiCliques Merging (BCM) to systematically detect functional MRMs based on bicliques merging. Our method exploits the miRNA/mRNA expression profiles, target site predictions as well as the gene-gene interactions. Furthermore, the module number is

- *C. Liang is with the College of Information Science and Electronic Enginerring, Hunan University, Changsha 410082, Hunan, China. E-mail: alcs417@hnu.edu.cn.*
- *Y. Li is with the Computer Science & Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139. E-mail: liyue@mit.edu.*
- *J.W. Luo is with the College of Information Science and Electronic Enginerring, Hunan University, Changsha 410082, Hunan, China. E-mail: luojiawei@hnu.edu.cn.*

Manuscript received 22 Dec. 2014; revised 25 May 2015; accepted 22 July 2015. Date of publication 31 July 2015; date of current version 1 June 2016.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2015.2462370

automatically determined during the merging process. We apply our method to breast cancer and thyroid cancer (THCA) datasets downloaded from TCGA. Comparing with alternative formalisms, we show that the modules identified by our method are more densely connected and functionally enriched. Moreover, BCM-based modules are more enriched for miRNA families and exhibit more negative expression correlations between miRNAs and mRNAs. Through a literature survey we find miRNAs and genes involved in the detected modules are strongly implicated in cancer. Several MRMs are predicted as potential prognostic biomarkers in BRCA and THCA using Kaplan-Meier (KM) survival analysis. The breast cancer subtypes analysis further demonstrates some interesting insights of the identified modules.

2 Method

2.1 Description of the Algorithm

To facilitate the analysis of cancer miRNA regulatory network, we propose a novel algorithm called BCM to effectively detect MRMs consisting of three steps. Prior to the three steps, we first construct the weighted miRNA-mRNA regulatory network with miRNA/mRNA expression data and predicted miRNA-gene target information as described in Section 3.2. At the first step, we enumerate all the maximal bicliques without considering the edge weights, which results in a large number of completely connected bipartite miRNA-mRNA subgraphs. At the second step, we perform a random shuffling test to assess the statistical significance for each module and only keep those significant modules as candidates. At the third step, we iteratively merge the candidate modules based on their overlaps as well as the gene-gene interactions. The program terminates when there are no modules could be merged. Details of each step are described as follows.

2.2 Bicliques Enumeration

Let W denote the $N \times M$ weighted miRNA regulatory network obtained from LASSO, G denote the $N \times N$ binary gene-gene interaction network, where N is the number of genes and M is the number of miRNAs. Here we define a miRNA-mRNA regulatory module u as a bipartite subgraph in which the miRNAs and mRNAs are densely connected and highly correlated. Suppose u_i is a miRNA and u^j is a mRNA in module u, $u^j{}_i$ is the edge weight between them. Similarly, $u_{\cdot}$ is the set of all miRNAs and $u^{\cdot}$ is the set of all mRNAs in module u. To assess the density of the modules for subsequent analysis, we herein define a density function $D(u)$:

$$D(u) = \frac{\sum_{i=1}^{|u_{\cdot}|} \sum_{j=1}^{|u^{\cdot}|} W(u_i^j)}{|u_{\cdot}| * |u^{\cdot}|}.$$

The density score reflects the number of regulations within this module as well as the strengths of the correlations between the miRNA set and mRNA set.

We search for all the maximal bicliques using the algorithm MICA [24] in the unweighted alternative of W, where the non-zero edge weights are transformed to 1. A biclique is defined as a complete bipartite subgraph [24]. Since the maximal bicliques containing a single miRNA or mRNA can be directly derived from the network, we only focus on maximal bicliques with at least two miRNAs and mRNAs. We then calculate the density score for all the identified maximal biclique and stored them in a list S as the input for the statistical test.

2.3 Statistical Significance of Bicliques

To evaluate the statistical significance of the density scores of each biclique in S, we perform a random edge weight shuffling test. We start with the observed network and repeatedly swapping the edge weights of randomly chosen pairs of interactions while keeping the miRNA-mRNA regulatory relationship unchanged. This ensures that we keep the consistency of the maximal bicliques identified across all random networks. Switching is prohibited if the source/target nodes of the chosen interactions are the same. In each randomization step, switching operation is performed a predefined number of times. To sufficiently shuffle the given network, we set the number of switching operations as the square of the given network size. We then repeat the randomization process for 1,000 times. For each biclique, we recalculate the density score from the randomly generated edge weights and compare it with the observed density to derive an empirical p-value. The p-value is the percentage of times that the expected density score is equal to or greater than the observed one. Statistically significant bicliques (p-value < 0.005, by default) are then considered as candidate modules for the merging process.

2.4 Bicliques Merging

Once the candidate modules are generated, we carry out an iterative merging process to associate modules that are closely related. To measure the closeness between two modules, we herein introduce a scoring function *overlapping_score(u,v)* defined as:

$$overlapping\_score(u,v) = \frac{\sum_{i=1}^{|u_{\cdot}|}\left(\sum_{j=1}^{|u^{\cdot}|} W(u_i^j)\right) + \sum_{i=1}^{|v_{\cdot}|}\left(\sum_{j=1}^{|v^{\cdot}|} W(v_i^j) + \sum_{j=1}^{|u^{\cdot}|} W(v_i^j)\right)}{|\cup(u_{\cdot}, v_{\cdot})| * |\cup(u^{\cdot}, v^{\cdot})|}.$$

Meanwhile, to facilitate the merging process where two modules are densely connected by the gene-gene interactions between them, we propose another scoring function which is defined as:

$$ggi\_enrichment\_score(u,v) = \frac{2 * \sum_{i=1}^{|u^{\cdot}|} \sum_{j=1}^{|v^{\cdot}|} G(u_i^j)}{|\cup(u^{\cdot}, v^{\cdot})| * (|\cup(u^{\cdot}, v^{\cdot})| - 1)}.$$

Based on the two functions, an overall scoring function *score(u,v)* is obtained:

$$score(u,v) = overlapping\_score(u,v) + ggi\_enrichment\_score(u,v).$$

The rationale behind the *overlapping_score* function is to quantify the overlap between module u and v as well as the corresponding interaction intensity imposed on the overlap. More specifically, the function reflects the relative edge strength gained from merging the two modules. Therefore, the higher the overlapping score is, the more closely connected the two modules are. For the *ggi_enrichment_score* function, it measures the number of gene-gene interactions in the merged mRNA set. In other words, it takes advantage of gene-gene interactions and acts as an informative reference to guide the merging process. For each merging operation, we obtain a sum score from the overall scoring functions. If the sum score obtained by merging two modules is larger than the density of both modules, we will then merge them. In summary, the density function is used as the standard to evaluate the density of the modules, while the *score* function evaluates the closeness between two modules. They coo-peratively drive the merging process until it reaches an optimal set of modules. In general, it is infeasible to get solutions for merging a list of subgraphs due to the high computational complexity caused by combinatorial explosion [25]. Thus, we adopt a greedy-based approach to iteratively merge the modules in the candidate list. The process starts with ranking the candidate modules in decreasing order according to their densities. Then for each candidate, it scans the rest of the modules in the list and records the one that could generate the largest score through merging. The process subsequently generates a new candidate by merging these two modules and recalculates its density. If the current candidate cannot be merged with any other modules, the program

Algorithm 1 Bicliques merging procedure

$M_{cand} \leftarrow \{m_1, m_2, ..., m_k\}; M_{new} \leftarrow \varnothing; M_{final} \leftarrow \varnothing$
while($|M_{cand}| > 0$) *do*
  $M_{cand} \leftarrow sort(M_{cand})$
  for each $m_i \in M_{cand}$ $(i = 1,...,k)$ *do*
    $m_{best} \leftarrow null$
    for each $m_j \in M_{cand}$ $(j = 1,...,k, j \neq i)$ *do*
      if $((score(m_i, m_j) > density(m_i)) \,\&\&\, (score(m_i, m_j) > density(m_j)))$
        $m_{new} \leftarrow merge(m_i, m_j)$
        if $(density(m_{new}) > density(m_{best}))$ $m_{best} \leftarrow m_{new}$
      end if
    end for
    if $(m_{best} != null)$ $M_{best} \leftarrow M_{best} \cup m_{best}$; *else* $M_{dist} \leftarrow M_{dist} \cup m_i$
  end for
  $M_{cand} \leftarrow \varnothing$
  if $(|M_{best}| > 0)$
    $M_{cand} \leftarrow M_{best} \cup M_{final}$
    $filter(M_{cand})$
    $M_{best} \leftarrow \varnothing; M_{final} \leftarrow \varnothing$
  end if
end while
return M_{final}

will still keep it in the candidate list for the following iterations. After each iteration, the program removes the modules which are contained by other modules in the candidate list to eliminate repetitive modules. It repeats these steps until no new candidates are generated. The pseudocode for the merging process is outlined in Algorithm 1.

3 Data Sets

3.1 Data Sources

The breast cancer and thyroid cancer datasets used in this study were obtained from TCGA project. The miRNA and mRNA expression profiles for 331/551 BRCA/THCA samples were downloaded from TCGA data portal (http://cancergenome.nih.gov/), each measuring 710 miRNAs and 13306 mRNAs, respectively. The data were further log2-transformed and mean-centered. The clinical data for the samples were also downloaded from the TCGA data portal. The miRNA-gene target site information were collected from TargetScanHuman 6.2 database [26]. We removed regulations involving miRNAs or genes that were not present in the expression data. Gene-gene interaction network was constructed by combining the transcription factor binding site (TFBS) and protein-protein interaction data processed from TRANSFAC [27] and BioGrid [28] databases, respectively. Only interactions with both genes presented in the weighted miRNA regulatory network were kept.

3.2 Preprocessing

In our previous study, we have shown that LASSO achieved the best performance among four methods according to the number of validated interactions identified based on miRTarBase [14]. Therefore, we followed the same pipeline to construct miRNA regulatory network with edge weights for downstream analysis. Here, we considered that a gene is regulated by a miRNA, only if the miRNA is predicted to target this gene and their expressions are negatively correlated [29]. We used expression of miRNAs that have at least one target mRNA to model the corresponding mRNA expression across samples. The fitted linear coefficients calculated for miRNA/mRNA pairs were considered as edge weights, where only negative coefficients were retained. We further filtered out the miRNAs (genes) having no correlations with any genes (miRNAs) in the network. The resulting weighted miRNA-mRNA regulatory network for BRCA/THCA contains 688/696 miRNAs, 9174/9759 mRNAs and 47336/64286 putative miRNA regulations. Interestingly, the BRCA and THCA networks contain a very similar number of miRNAs and mRNAs, but a considerable difference in the number of regulations. This is likely due to the differences in the expression patterns of miRNAs between the two cancer types [30], [31].

3.3 Evulation of Functional Significance of MRMs

Function and pathway analysis of the identified MRMs is conducted by looking for enrichment in gene ontology (GO) terms in biological process and KEGG pathways. The annotations of GO (Biological Process) terms were obtained programmatically from Ensembl-BiomaRt API (Ensembl version 67) and the canonical pathways (CPs) were downloaded from MsigDB (c2.cp.v4.0.symbols.gmt) [32]. GO terms with fewer than five genes or with evidence codes equal to electronic annotation (IEA), non-traceable author statement (NAS) or no biological data available (ND) were discarded, resulting in 2007 GO-BP terms and 10,315 unique genes. The statistical significance of the genes involved in each MRM with respect to each GO term or pathway is calculated by hypergeometric test and further corrected for multiple testing.

We also evaluate for MRMs the enrichment in miRNA families. The miRNA family information was obtained from the miRBase family annotation (v20, June 2013) [33], where only homo sapiens miRNAs were kept. The oncomirs were obtained from [34] (2009) and [35] (2011), while the oncogenes were obtained from COSMIC (v71) [36]. The oncomirs and oncogenes are used to identify the MRMs that are related to cancer.

4 Experiments

To compare the performance of BCM, we applied two leading methods: SNMNMF [21] and Mirsynergy [14] on the same datasets. Both methods have been previousely compared on the same data sets as well [14]. The source codes of both methods were downloaded from their websites (http://zhoulab.usc.edu/SNMNMF/ and www.cs.utoronto.ca/~yueli/Mirsynergy.html). We set the number of required modules to 50 for SNMNMF as suggested and kept other parameters as default settings throughout the analysis [14], [21].

4.1 Comparison of Module Sizes and Densities

We tested BCM, Mirsynergy and SNMNMF on BRCA and THCA obtained from TCGA data portal as described. The BCM first detected 70,384 and 206,961 maximal bicliques for each dataset that meet our requirement. We then carried out the statistical test to filter out insignificant bicliques. The distribution of the observed density for bicliques is significantly shifted to the left of the expected one (p-value $<$ 2.2e-16 for both datasets, Wilcoxon signed rank test), which indicates that most bicliques have very low regulatory strength (supplementary file, Fig. S1, available online). After the randomization test, 209/1033 bicliques were left as significant candidates for each dataset (p-value $<$ 0.005). The merging process eventually generated 43/41 MRMs for BRCA/THCA dataset, which has 5.65/8.07 miRNAs and 28.37/14.29 mRNAs per module on average (Table 1). For Mirsynergy and SNMNMF, they each identified 53/50 and 39/39 modules in BRCA/THCA dataset, and the respective averaged number of miRNAs (mRNAs) per module are 5.77/7.6 (29.92/32.86) and 2.62/2.23 (71.56, 74.82) (Table 1). To compare the density of the modules predicted by three methods, we calculated for each Mirsynergy-MRMs and SNMNMF-MRMs their density according to Eq. (1). The averaged density value for BCM, Mirsynergy and SNMNMF are 0.279, 0.119 and 0.001 for BRCA and 0.165, 0.110 and 0.002 for THCA, respectively (Table 1). Moreover, modules identified by our method are significantly more densely connected comparing with Mirsynergy-MRMs and SNMNMF-MRMs in both datasets (p-value $<$ 1.92e-06 and

TABLE 1
Performance Summary of BCM, Mirsynergy, and SNMNMF

Cancer	Method	M#	Ave (mi)	Ave (m)	Density	miRFam#	MiMEC	Time (hrs)
BRCA	BCM	43	5.65	28.37	0.279	15	-0.149	1
	Mir[1]	53	5.77	29.92	0.119	10	-0.077	1.5
	SNM[2]	39	2.62	71.56	0.001	2	-0.045	24+
THCA	BCM	41	8.07	14.29	0.165	13	-0.112	1.5
	Mir	50	7.6	39.86	0.110	7	-0.082	2
	SNM	39	2.23	74.82	0.002	7	-0.008	24+

M# is the number of modules; Ave(mi) and Ave(m) are the average number of miRNAs and mRNAs per module; miR_Fam is the number of modules enriched in at least one miRNA family; MiMEC is the averaged miRNA-mRNA expression correlation across all modules; Time is the number of hours algorithms took to run.
Note: *1. Mir is the abbreviation of Mirsynergy; 2 SNM is the abbreviation of SNMNMF.*

2.60e-15 for BRCA, p-value $<$ 1.84e-04 and 8.16e-15 for THCA, Wilcoxon rank sum test). Therefore, our method is in better agreement with the concept of miRNA modules, where a set of miRNAs closely regulate a specified cohort of mRNAs. A full list of the identified modules in both datasets is provided in Table S1 and several representative modules are illustrated in supplementary file Fig. S2, available online.

4.2 miRNA Family Enrichment Analysis

Previous studies have shown that miRNAs from the same family tend to be involved in the same disease or function synergistically [12]. We herein conduct a miRNA family enrichment analysis for the modules identified by BCM, Mirsynergy and SNMNMF. We counted for each method the number of modules that are significantly enriched in at least one miRNA family using hypergeometric test (q-value $<$ 0.05 after multiple testing correction). Encouragingly, for BRCA and THCA datasets, 15 and 13 of the BCM-MRMs are significantly enriched in at least one miRNA family, while only 10, 7 are found for Mirsynergy and 2, 7 for SNMNMF (Table 1, supplementary file Fig. S3, available online), indicating that MRMs predicted by our method are in better agreement with the evidences of synergistic regulation of miRNAs from the same family. We further investigated the details of the BCM-MRMs enriched in miRNA families for both datasets. For example, module 42 in BRCA contains fourteen miRNAs, seven of which (hsa-let-7b, hsa-let-7c, hsa-let-7d, hsa-let-7e, hsa-let-7g, hsa-let-7i, hsa-miR-98) belong to the miRNA family MIPF0000002. Moreover, all the seven miRNAs have been reported to play important roles in various cancers in a synergistic manner [37], [38], [39]. To take another example, two of the three miRNAs (hsa-miR-449a, hsa-miR-449b) in module 13 in THCA are from the same miRNA family MIPF0000133, where the two miRNAs have been shown to cooperatively function as a tumor suppressor or apoptosis promoter [40], [41]. A list of the miRNA family enrichment details for BCM-MRMs for BRCA and THCA datasets is provided in Table S2.

4.3 Functional Enrichment Analysis

To compare the biological significance of the MRMs identified by BCM, Mirsynergy and SNMNMF, we computed their enrichment in GO biological process (GO-BP) terms or canonical pathways accordingly. To adjust the results for different module sizes, we calculated for each module the number of significant GO-BP terms or CPs with respect to the size of the corresponding gene set as the normalized enrichment score (NES) [32], [42]. As illustrated in Fig. 1, for both GO-BP terms and CPs, BCM produced modules with significantly higher enrichment score comparing with Mirsynergy and SNMNMF in both BRCA and THCA datasets (for BRCA, the p-value is 0.052/0.087 comparing with Mirsynergy, and 4.72e-05/0.0014 for SNMNMF; for THCA, the p-value is 7.53e-05/0.016 comparing with Mirsynergy, and 2.44e-06/0.002 for SNMNMF, exact Wilcoxon rank sum test), which demonstrated that BCM-MRMs are more functionally enriched.

We further explored the correlations between the module density and GO-BP terms/CPs NES for modules identified by three methods. According to the definition of MRMs, it is reasonable to expect a positive correlation between the module density and function enrichment score. To make fair comparisons, we only kept modules with non-zero density or NES values. Intriguingly, for both datasets, we observed significant positive Pearsoncorrelations for BCM-MRMs, whereas negative correlations were obtained for SNMNMF-MRMs and inconsistent correlations for Mirsynergy-MRMs (Figs. 2 and 3). These observations showed that the denser the BCM-MRMs are, the more functionally enriched they tend to be, which further supported the effectiveness of our method.

We then investigated the details of enriched GO-BP terms or CPs for BCM-MRMs in both datasets. Interestingly, we found the identified modules are involved in a variety of cancer-related processes or pathways. For instance, both module 9 in BRCA and module 7 in THCA containing the oncogene PDGFRA are enriched in *negative regulation of platelet activation, platelet-derived growth factor receptor signaling pathway*, which is in accordance with the functional roles of PDGFRA reported in [43]; module 1 in THCA contains four miRNA members hsa-let-7c, hsa-let-7d, hsa-miR-29c and hsa-miR-98, all of which are oncomirs. It is enriched in various

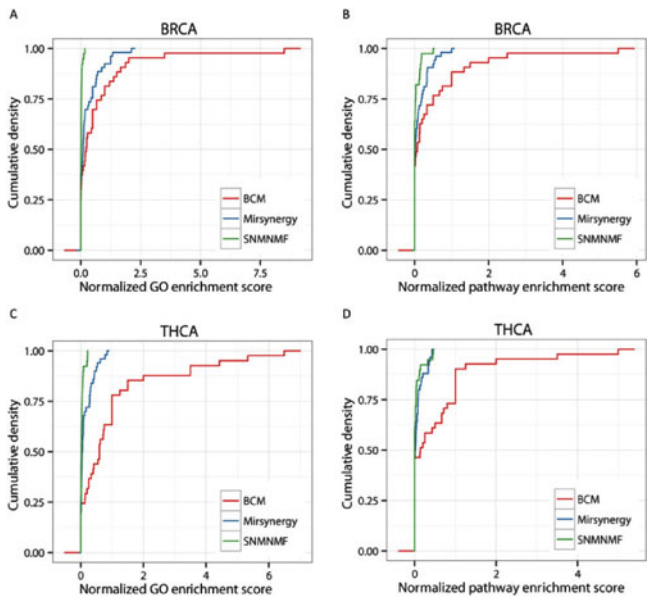

Fig. 1. The cumulative density functions (CDF) for the normalized functional enrichment score of modules identified by BCM, Mirsynergy, and SNMNMF in two datasets.

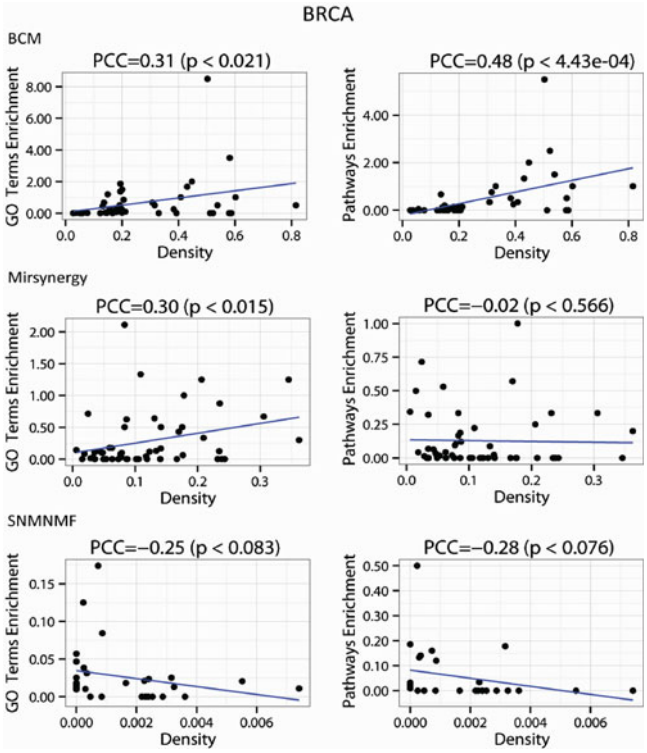

Fig. 2. The Pearson correlation coefficients (PCC) between the module density and the normalized functional enrichment score in BRCA.

cancer-related GO processes and pathways such as *positive regulation of epithelial to mesenchymal transition*, *transforming growth factor beta receptor signaling pathway* and so on. Moreover, one of its enriched pathways *KEGG_ECM_RECEPTOR_INTERACTION* is

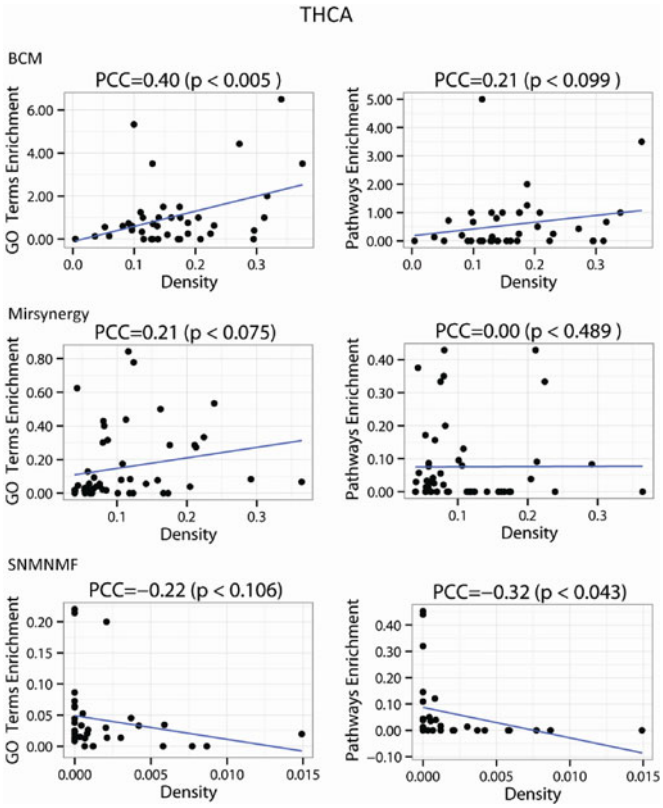

Fig. 3. The same as Fig. 2 except for THCA.

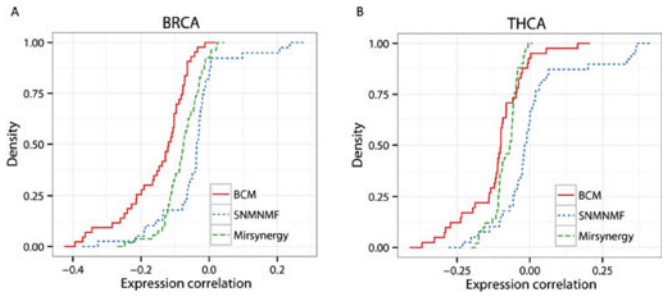

Fig. 4. The cumulative density functions (CDF) of the averaged expression correlation within each module identified by BCM, Mirsynergy and SNMNMF in (A) BRCA and (B) THCA.

recently suggested to play an essential role in cancer biology [44]. Another module 18 in BRCA is involved in *activation of signaling protein activity involved in unfolded protein response, ER-associated protein catabolic process and endoplasmic reticulum unfolded protein response*, while the four oncogenes in this module: CDX2, CEBPA, IL2 and PLAG1 have all been shown to be related with breast cancers[45], [46], [47], [48], leading to the possibility that the miRNA regulators hsa-miR-181c and hsa-miR-4262 within this module might play a role in breast cancer. A list of the enriched GO-BP terms and CPs in BCM-MRMs for both datasets is provided in the supplementary materials (Table S3), available online.

4.4 Co-Expression Correlation within MRMs

We next examined the miRNA-mRNA expression correlation (MiMEC) within the MRMs predicted by BCM, Mirsynergy and SNMNMF for both datasets. Since miRNAs are deemed as the repressors of their target mRNAs, we expected a strong negativity of the miRNA-mRNA expression correlation involved in the predicted modules.

We used the BRCA and THCA expression data to calculate for each module the averaged expression correlation of all miRNA-mRNA pairs. The cumulative distribution functions for the averaged expression correlation across modules from each method were then taken for comparison (Fig. 4). As expected, for both datasets, all of the identified BCM-MRMs exhibit modest negative MiMEC with an average of –0.149 and –0.112 (Table 1), which are almost 2, 1.5 folds more negative than those of Mirsynergy (–0.077, –0.082), and 4, 15 folds as more negative as those of SNMNMF (–0.045, –0.008). Moreover, miRNA-mRNA pairs within MRMs predicted by our method are significantly more negatively correlated than those identified by Mirsynergy and SNMNMF for both datasets (p-value<3.68e-03 & 3.03e-08 for BRCA, p-value < 0.064 & 1.45e-05 for THCA. two-sample Kolmogorov-Smirnov test). However, we noticed that the correlation coefficients are generally weak across all identified modules (supplementary file, Table S5, Figs. S4, S5 and S6) , available online. To find out possible reasons for this observation, we further calculated the mean Pearson correlation coefficients for all miRNA-mRNA pairs in the BRCA and THCA expression data, respectively. For BRCA/THCA, we found that 504/548 out of 710 miRNAs had average coefficients greater than 0 with their targets. Moreover, none of these miRNAs had average coefficients greater than 0.06/0.09 or less than -0.07/-0.11. Taken together, it indicated that a majority of the pearson correlation coefficients of miRNA-mRNA pairs were weak in both datasets, which also accounted for the weak correlation performance across modules.

4.5 Survival Analysis

To examine the potential diagnostic power of our identified modules, we performed the Kaplan-Meier survival analysis for the samples used in this study based on TCGA clinical data. The

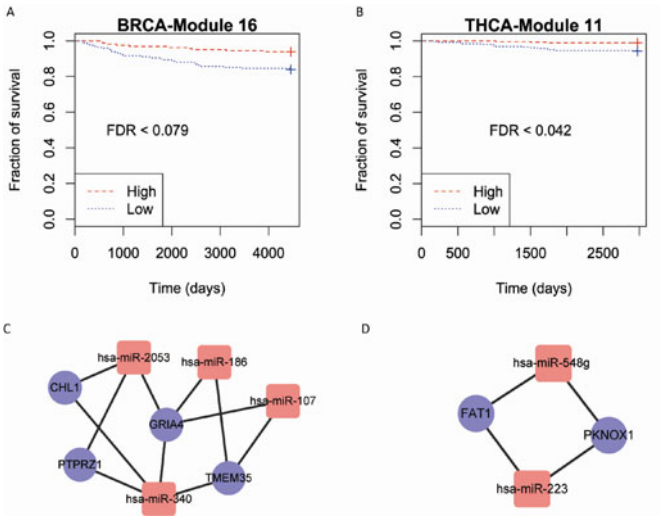

Fig. 5. (A) and (B) are two BCM-MRMs revealed by the Kaplan-Meier survival analysis in BRCA and THCA, respectively. As observed, the patients with higher miRNA expression level in both modules are at lower risk level. (C) and (D) are the visualization of the two clusters.

summary of clinical information for both datasets can be found in supplementary file Table S6, available online. We characterized the level of associations of patient survival time from their initial diagnosis to death with the averaged miRNA expression levels of BCM-MRMs. The miRNA expression data were first normalized to the range [0, 1] and then the patients were divided into two groups according to the higher or lower module-averaged miRNA expression levels than the sample means. The significance of each module was determined by the log-rank test and further adjusted by multiple test correction using Benjamini-Hochberg method over all modules. We found three and two modules resulting in significantly different survival rates (FDR < 0.1) for BRCA and THCA datasets, respectively (Fig. 5, supplementary file Figs. S7 and S8), available online. For BRCA, oncogene PDGFRA is involved in the potential prognostic module 9, and oncomirs let-7e and miR-29c are involved in module 40. Interestingly, we found that miR-340 was invovled in all three modules. A recent study has shown that induction of miR-340 expression was able to suppress tumor cell migration and invasion, which is in agreement with the survival analysis that higher module expression levels resulted in higher fraction of survival, suggesting that miR-340 might play an important role in the pathogenesis of breast cancer[49]. For THCA, miR-223 invovled in module 11 was known to be related to thyroid cancer [50]. To validate the statistical power gained from the modules, we further tested using the expression profiles of individual miRNA or miRNA-mRNA pair alone. For BRCA, none of the miRNAs nor miRNA-mRNA pairs survived the multiple testing correction (FDR < 0.1). For THCA, we identified the oncomir hsa-miR-223 in module 11 was able to significantly differentiate the two patient groups (FDR $= 0.055$). The involvement of the other miRNA hsa-miR-548g in this module further enhanced the significance of survival analysis (FDR $= 0.041$). Fig. 5 illustrates the KM survival curves for BRCA-module 16 and THCA-module 11 with the corresponding visualization of the two clusters.

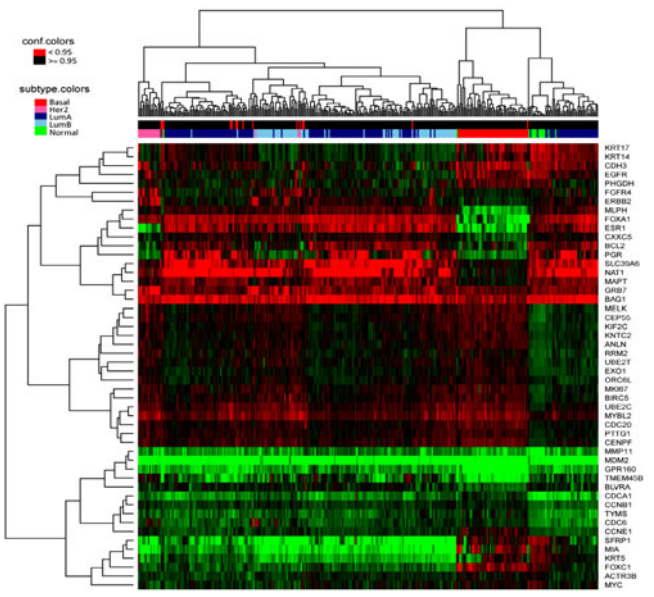

Fig. 6. Subtype classification heatmap for the 331 BRCA samples using PAM50 classifier.

TABLE 2
Modules with Top 10 Classification Accuracies

Index	Accuracy (%)	miRNA#	mRNA#
41	86.40	17	67
40	85.20	15	81
42	83.69	14	326
43	83.38	28	95
38	82.78	8	74
34	81.27	2	158
28	80.66	2	44
37	80.66	4	52
21	79.15	2	59
39	77.95	17	27

Cost $= 4$, *Gamma* $= 0.03125$.

4.6 Breast Cancer Subtype Analysis

We also tested the BRCA subtype classification performance for the identified modules. To this end, we downloaded the PAM50 classifier to predict for each sample its subtype [51]. Of the 331 BRCA samples analyzed in our test, 52 were Basal, 19 were Her2, 170 were Luminal A, 76 were Luminal B and 14 were Normal-like (Fig. 6). After the sample subtypes were properly assigned, support vector machine (SVM) was adopted to perform the classification analysis for each identified module [52]. All the miRNA/mRNA expression values were first scaled to the range [-1, 1]; then for each module, we used module members as features for classification. Radial basis function (RBF) was chosen as the kernel function, and the best values of the two paramters cost (C) and gamma (γ) in the kernel function were obtained by a grid-search approach using cross-validation; finally, 10-fold cross-validation was used to evaluate the classification accuracy. As can be observed in Table 2, module-41 has the best classification performance, and eight modules have classification accuracies higher than 80 percent (supplementary material, Table S4, available online). Moreover, we noticed that module-40 was capable of predicting not only the survival rates but also subtypes of patients (see supplementary file Fig. S7, available online). Our analysis showed that the identified modules could be used for breast cancer subtype classifications to some extent, especially that miRNA could also be considered as useful features in the classifications.

5 Conclusion

miRNAs have been established as key metastasis regulators in diverse cancer types. The ability of these small non-coding RNAs to regulate gene expression has generated much interests in exploiting them as potential therapeutic target modules in human diseases [4]. Although great efforts have been made to

elucidate the precise regulatory functions of miRNAs based on miRNA regulatory modules, it remains a challenge due to the complexity of the combinatorial regulations and cooperative mechanisms between miRNAs and genes.

In this article, we developed a flexible and effective approach BCM to identify regulatory modules using miRNA/mRNA expression profiles, miRNA-mRNA target site information and gene-gene interactions. A previous method proposed by Peng et al. [11] only considered the statistically significant maximal bicliques as regulatory modules, which produced a large amount of MRMs in spite of the relatively small size of the network and were difficult to tackle for further analysis. It is therefore unrealistic to apply such method to detect regulatory modules in large scale networks as the one used in this study. BCM extends from their method by iteratively merging the bicliques with the guidance of the gene-gene interactions as well as their overlaps. To quantify the closeness for modules, we proposed a scoring function to facilitate the merging process. The greedy-based merging strategy enables our method to fast and effectively finish the iterations.

To demonstrate the utility of our method, we compared BCM with Mirsynergy and SNMNMF on BRCA and THCA datasets from TCGA. We showed that the modules identified by BCM were more densely connected and functionally enriched. Specifically, we observed a positive correlation between the module density and the functional enrichment score for BCM-MRMs, corroborating thegeneral concept of network modules. Moreover, the miRNA-mRNA pairs in BCM-MRMs were significantly more negatively correlated than those in Mirsyerngy-MRMs and SNMNMF-MRMs, which is in better agreement with the assumption that miRNAs inhibit gene expression. Last, the comparison of run-times for all three methods on both datasets illustrated the efficiency of our algorithm. Based on the Kaplan-Meier survival analysis, we found several prognostically promising MRMs in both datasets. The breast cancer subtype analysis further revealed that the identified modules could be considered as useful features in subtype prediction. Despite the superiority of our method, we noticed that BCM is sensitive to the density of the input weight matrix, which sometimes affect the statistical significance of the candidate modules.

Together, the presented work provides a framework to efficiently analyze disease-specific miRNA regulatory networks by detecting biologically meaningful miRNA regulatory modules. It may be utilized to identify the functional roles of miRNAs in various physiologic and pathologic processes. With more miRNA/mRNA expression data measured in large patient cohort becoming increasingly abundant, our method described in this article can serve as a powerful tool to discover miRNA regulatory patterns in complex diseases.

ACKNOWLEDGMENTS

This work was supported by Natural Sciences and Engineering Research Council (NSERC) Canada Graduate Scholarship, National Natural Science Foundation of China (Grant NO. 61240046), Hunan Provincial Natural Science Foundation of China (Grant NO.13JJ2017) and Collaboration and Innovation Center for Digital Chinese Medicine of 2011 Project of Colleges and Universities in Hunan Province. J.W. Luo is the corresponding author of this paper.

REFERENCES

[1] L. He and G. J. Hannon, "MicroRNAs: Small RNAs with a big role in gene regulation," *Nature Rev. Genetics*, vol. 5, no. 7, pp. 522–531, Jul. 2004.

[2] R. Shalgi, D. Lieber, M. Oren, and Y. Pilpel, "Global and local architecture of the mammalian microRNA-transcription factor regulatory network," *PLoS Comput. Biol.*, vol. 3, no. 7, pp. 1291–1304, Jul. 2007.

[3] M. R. Fabian, N. Sonenberg, and W. Filipowicz, "Regulation of mRNA translation and stability by microRNAs," *Ann. Rev. Biochem.*, vol. 79, pp. 351–379, 2010.

[4] N. Pencheva and S. F. Tavazoie, "Control of metastatic progression by microRNA regulatory networks," *Nature Cell Biol.*, vol. 15, no. 6, pp. 546–554, Jun. 2013.

[5] H. Dvinge, A. Git, S. Graf, M. Salmon-Divon, C. Curtis, A. Sottoriva, Y. Zhao, M. Hirst, J. Armisen, E. A. Miska, S. F. Chin, E. Provenzano, G. Turashvili, A. Green, I. Ellis, S. Aparicio, and C. Caldas, "The shaping and functional consequences of the microRNA landscape in breast cancer," *Nature*, vol. 497, no. 7449, pp. 378–382, May 16, 2013.

[6] N. Yanaihara, N. Caplen, E. Bowman, M. Seike, K. Kumamoto, M. Yi, R. M. Stephens, A. Okamoto, J. Yokota, T. Tanaka, G. A. Calin, C. G. Liu, C. M. Croce, and C. C. Harris, "Unique microRNA molecular profiles in lung cancer diagnosis and prognosis," *Cancer Cell*, vol. 9, no. 3, pp. 189–198, Mar. 2006.

[7] M. V. Iorio, R. Visone, G. Di Leva, V. Donati, F. Petrocca, P. Casalini, C. Taccioli, S. Volinia, C. G. Liu, H. Alder, G. A. Calin, S. Menard, and C. M. Croce, "MicroRNA signatures in human ovarian cancer," *Cancer Res.*, vol. 67, no. 18, pp. 8699–8707, Sep. 15, 2007.

[8] P. Maziere and A. J. Enright, "Prediction of microRNA targets," *Drug Discovery Today*, vol. 12, nos. 11/12, pp. 452–458, Jun. 2007.

[9] B. John, A. J. Enright, A. Aravin, T. Tuschl, C. Sander, and D. S. Marks, "Human MicroRNA targets," *PLoS Biol.*, vol. 2, no. 11, p. e363, Nov. 2004.

[10] J. Lu, G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz, and T. R. Golub, "MicroRNA expression profiles classify human cancers," *Nature*, vol. 435, no. 7043, pp. 834–838, Jun. 9, 2005.

[11] X. Peng, Y. Li, K. A. Walters, E. R. Rosenzweig, S. L. Lederer, L. D. Aicher, S. Proll, and M. G. Katze, "Computational identification of hepatitis C virus associated microRNA-mRNA regulatory modules in human livers," *BMC Genomics*, vol. 10, p. 373, 2009.

[12] J. A. Xu, C. X. Li, Y. S. Li, J. Y. Lv, Y. Ma, T. T. Shao, L. D. Xu, Y. Y. Wang, L. Du, Y. P. Zhang, W. Jiang, C. Q. Li, Y. Xiao, and X. Li, "MiRNA-miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features," *Nucleic Acids Res.*, vol. 39, no. 3, pp. 825–836, Feb. 2011.

[13] E. C. Lai, "Predicting and validating microRNA targets," *Genome Biol.*, vol. 5, no. 9, p. 115, 2004.

[14] Y. Li, C. Liang, K. C. Wong, J. Luo, and Z. Zhang, "Mirsynergy: Detecting synergistic miRNA regulatory modules by overlapping neighbourhood expansion," *Bioinformatics*, vol. 30, no. 18, pp. 2627–2635, Sep. 15, 2014.

[15] S. Yoon and G. De Micheli, "Prediction of regulatory modules comprising microRNAs and target genes," *Bioinformatics*, vol. 21, no. Suppl 2, pp. ii93–ii100, Sep. 1, 2005.

[16] J. G. Joung, K. B. Hwang, J. W. Nam, S. J. Kim, and B. T. Zhang, "Discovery of microRNA-mRNA modules via population-based probabilistic learning," *Bioinformatics*, vol. 23, no. 9, pp. 1141–1147, May 1, 2007.

[17] D. H. Tran, K. Satou, and T. B. Ho, "Finding microRNA regulatory modules in human genome using rule induction," *BMC Bioinformatics*, vol. 9, no. suppl. 12, p. S5, 2008.

[18] B. Liu, J. Li, and A. Tsykin, "Discovery of functional miRNA-mRNA regulatory modules with computational methods," *J. Biomed. Informat.*, vol. 42, no. 4, pp. 685–691, Aug. 2009.

[19] B. Liu, L. Liu, A. Tsykin, G. J. Goodall, J. E. Green, M. Zhu, C. H. Kim, and J. Li, "Identifying functional miRNA-mRNA regulatory modules with correspondence latent dirichlet allocation," *Bioinformatics*, vol. 26, no. 24, pp. 3105–3111, Dec. 15, 2010.

[20] V. Jayaswal, M. Lutherborrow, D. D. F. Ma, and Y. H. Yang, "Identification of microRNA-mRNA modules using microarray data," *BMC Genomics*, vol. 12, p. 138, Mar. 6, 2011.

[21] 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, Jul. 1, 2011.

[22] T. D. Le, L. Liu, A. Tsykin, G. J. Goodall, B. Liu, B. Y. Sun, and J. Y. Li, "Inferring microRNA-mRNA causal regulatory relationships from expression data," *Bioinformatics*, vol. 29, no. 6, pp. 765–771, Mar. 15, 2013.

[23] G. Pio, M. Ceci, D. D'Elia, C. Loglisci, and D. Malerba, "A novel biclustering algorithm for the discovery of meaningful biological correlations between microRNAs and their target genes," *BMC Bioinformatics*, vol. 14, no. Suppl 7, p. S8, 2013.

[24] G. Alexe, S. Alexe, Y. Crama, S. Foldes, P. L. Hammer, and B. Simeone, "Consensus algorithms for the generation of all maximal bicliques," *Discrete Appl. Math.*, vol. 145, no. 1, pp. 11–21, Dec. 30, 2004.

[25] K. Fenner, J. F. Gao, S. Kramer, L. Ellis, and L. Wackett, "Data-driven extraction of relative reasoning rules to limit combinatorial explosion in biodegradation pathway prediction," *Bioinformatics*, vol. 24, no. 18, pp. 2079–2085, Sep. 15, 2008.

[26] R. C. Friedman, K. K. H. Farh, C. B. Burge, and D. P. Bartel, "Most mammalian mRNAs are conserved targets of microRNAs," *Genome Res.*, vol. 19, no. 1, pp. 92–105, Jan. 2009.

[27] E. Wingender, X. Chen, R. Hehl, H. Karas, I. Liebich, V. Matys, T. Meinhardt, M. Pruss, I. Reuter, and F. Schacherer, "TRANSFAC: An integrated system for gene expression regulation," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 316–319, Jan. 1, 2000.

[28] C. Stark, B. J. Breitkreutz, A. Chatr-Aryamontri, L. Boucher, R. Oughtred, M. S. Livstone, J. Nixon, K. Van Auken, X. Wang, X. Shi, T. Reguly, J. M. Rust, A. Winter, K. Dolinski, and M. Tyers, "The BioGRID Interaction Database: 2011 update," *Nucleic Acids Res.*, vol. 39, no. Database issue, pp. D698–D704, Jan. 2011.

[29] Y. Li, J. Xu, H. Chen, J. Bai, S. Li, Z. Zhao, T. Shao, T. Jiang, H. Ren, C. Kang, and X. Li, "Comprehensive analysis of the functional microRNA-mRNA regulatory network identifies miRNA signatures associated with glioma malignant progression," *Nucleic Acids Res.*, vol. 41, no. 22, p. e203, Dec. 2013.

[30] G. A. Calin and C. M. Croce, "MicroRNA signatures in human cancers," *Nature Rev. Cancer*, vol. 6, no. 11, pp. 857–866, Nov. 2006.

[31] N. Rosenfeld, R. Aharonov, E. Meiri, S. Rosenwald, Y. Spector, M. Zepeniuk, H. Benjamin, N. Shabes, S. Tabak, A. Levy, D. Lebanony, Y. Goren, E. Silberschein, N. Targan, A. Ben-Ari, S. Gilad, N. Sion-Vardy, A. Tobar, M. Feinmesser, O. Kharenko, O. Nativ, D. Nass, M. Perelman, A. Yosepovich, B. Shalmon, S. Polak-Charcon, E. Fridman, A. Avniel, I. Bentwich, Z. Bentwich, D. Cohen, A. Chajut, and I. Barshack, "MicroRNAs accurately identify cancer tissue origin," *Nature Biotechnol.*, vol. 26, no. 4, pp. 462–469, Apr. 2008.

[32] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov, "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 43, pp. 15545–15550, Oct. 25, 2005.

[33] A. Kozomara and S. Griffiths-Jones, "miRBase: Annotating high confidence microRNAs using deep sequencing data," *Nucleic Acids Res.*, vol. 42, no. Database issue, pp. D68–D73, Jan. 2014.

[34] R. Spizzo, M. S. Nicoloso, C. M. Croce, and G. A. Calin, "SnapShot: MicroRNAs in cancer," *Cell*, vol. 137, no. 3, p. 586, May 1, 2009.

[35] I. Koturbash, F. J. Zemp, I. Pogribny, and O. Kovalchuk, "Small molecules with big effects: The role of the microRNAome in cancer and carcinogenesis," *Mutation Res.-Genetic Toxicol. Environ. Mutagenesis*, vol. 722, no. 2, pp. 94–105, Jun. 17, 2011.

[36] S. A. Forbes, D. Beare, P. Gunasekaran, K. Leung, N. Bindal, H. Boutselakis, M. Ding, S. Bamford, C. Cole, S. Ward, C. Y. Kok, M. Jia, T. De, J. W. Teague, M. R. Stratton, U. McDermott, and P. J. Campbell, "COSMIC: Exploring the world's knowledge of somatic mutations in human cancer," *Nucleic Acids Res.*, vol. 43, no. Database issue, pp. D805–D811, Jan. 2015.

[37] S. Wang, Y. Tang, H. Cui, X. Zhao, X. Luo, W. Pan, X. Huang, and N. Shen, "Let-7/miR-98 regulate Fas and Fas-mediated apoptosis," *Genes Immunity*, vol. 12, no. 2, pp. 149–154, Mar. 2011.

[38] A. Wendler, D. Keller, C. Albrecht, J. J. Peluso, and M. Wehling, "Involvement of let-7/miR-98 microRNAs in the regulation of progesterone receptor membrane component 1 expression in ovarian cancer cells," *Oncol. Rep.*, vol. 25, no. 1, pp. 273–279, Jan. 2011.

[39] Y. Yang, T. Ago, P. Zhai, M. Abdellatif, and J. Sadoshima, "Thioredoxin 1 negatively regulates angiotensin II-induced cardiac hypertrophy through upregulation of miR-98/let-7," *Circulation Res.*, vol. 108, no. 3, pp. 305–313, Feb. 4, 2011.

[40] M. Lize, S. Pilarski, and M. Dobbelstein, "E2F1-inducible microRNA 449a/b suppresses cell proliferation and promotes apoptosis," *Cell Death Differentiation*, vol. 17, no. 3, pp. 452–458, Mar. 2010.

[41] A. Martin, A. Jones, P. J. Bryar, M. Mets, J. Weinstein, G. Zhang, and N. A. Laurie, "MicroRNAs-449a and -449b exhibit tumor suppressive effects in retinoblastoma," *Biochem. Biophys. Res. Commun.*, vol. 440, no. 4, pp. 599–603, Nov. 1, 2013.

[42] P. Jia, L. Wang, A. H. Fanous, X. Chen, K. S. Kendler, and Z. Zhao, "A bias-reducing pathway enrichment analysis of genome-wide association data confirmed association of the MHC region with schizophrenia," *J. Med. Genetics*, vol. 49, no. 2, pp. 96–103, Feb. 2012.

[43] T. Wei, L. N. Zhang, Y. Lv, X. Y. Ma, L. Zhi, C. Liu, F. Ma, and X. F. Zhang, "Overexpression of platelet-derived growth factor receptor alpha promotes tumor progression and indicates poor prognosis in hepatocellular carcinoma," *Oncotarget*, vol. 5, no. 21, pp. 10307–10317, Oct. 15, 2014.

[44] M. Krupp, T. Maass, J. U. Marquardt, F. Staib, T. Bauer, R. Konig, S. Biesterfeld, P. R. Galle, A. Tresch, and A. Teufel, "The functional cancer map: A systems-level synopsis of genetic deregulation in cancer," *BMC Med. Genomics*, vol. 4, p. 53, Jun. 30, 2011.

[45] Z. C. Zhou, J. Wang, Z. H. Cai, Q. H. Zhang, Z. X. Cai, and J. H. Wu, "Association between vitamin D receptor gene Cdx2 polymorphism and breast cancer susceptibility," *Tumor Biol.*, vol. 34, no. 6, pp. 3437–3441, Dec. 2013.

[46] S. Gery, S. Tanosaki, S. Bose, N. Bose, J. Vadgama, and H. P. Koeffler, "Down-regulation and growth inhibitory role of C/EBPalpha in breast cancer," *Clin. Cancer Res.*, vol. 11, no. 9, pp. 3184–3190, May 1, 2005.

[47] I. Garcia-Tunon, M. Ricote, A. Ruiz, B. Fraile, R. Paniagua, and M. Royuela, "Interleukin-2 and its receptor complex (alpha, beta and gamma chains) in in situ and infiltrative human breast cancer: an immunohistochemical comparative study," *Breast Cancer Res.*, vol. 6, no. 1, pp. R1–R7, 2004.

[48] D. Bu, G. Tomlinson, C. M. Lewis, C. Zhang, E. Kildebeck, and D. M. Euhus, "An intronic polymorphism associated with increased XRCC1 expression, reduced apoptosis and familial breast cancer," *Breast Cancer Res. Treatment*, vol. 99, no. 3, pp. 257–265, Oct. 2006.

[49] Z. S. Wu, Q. Wu, C. Q. Wang, X. N. Wang, J. Huang, J. J. Zhao, S. S. Mao, G. H. Zhang, X. C. Xu, and N. Zhang, "miR-340 inhibition of breast cancer cell migration and invasion through targeting of oncoprotein c-Met," *Cancer*, vol. 117, no. 13, pp. 2842–2852, Jul. 1, 2011.

[50] M. T. Tetzlaff, A. Liu, X. W. Xu, S. R. Master, D. A. Baldwin, J. W. Tobias, V. A. Livolsi, and Z. W. Baloch, "Differential expression of miRNAs in papillary thyroid carcinoma compared to multinodular goiter using formalin fixed paraffin embedded tissues," *Endocrine Pathol.*, vol. 18, no. 3, pp. 163–173, Fall 2007.

[51] J. S. Parker, M. Mullins, M. C. Cheang, S. Leung, D. Voduc, T. Vickery, S. Davies, C. Fauron, X. He, Z. Hu, J. F. Quackenbush, I. J. Stijleman, J. Palazzo, J. S. Marron, A. B. Nobel, E. Mardis, T. O. Nielsen, M. J. Ellis, C. M. Perou, and P. S. Bernard, "Supervised risk predictor of breast cancer based on intrinsic subtypes," *J. Clin. Oncol.*, vol. 27, no. 8, pp. 1160–1167, Mar. 10, 2009.

[52] C. C. Chang, and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, p. 23, 2011.