RNA BIOLOGY
2018, VOL. 15, NO. 8, 1093–1105
https://doi.org/10.1080/15476286.2018.1502590

RESEARCH PAPER

MiRNA-BD: an evidence-based bioinformatics model and software tool for microRNA biomarker discovery

Yuxin Lin [a], Wentao Wu [a], Zhandong Sun [a], Li Shen [a,b], and Bairong Shen [a,c,d]

[a]Center for Systems Biology, Soochow University, Suzhou, Jiangsu, China; [b]Department of Genetics & Systems Biology Institute, Yale University School of Medicine, West Haven, CT USA; [c]Center for Translational Biomedical Informatics, Guizhou University School of Medicine, Guiyang, China; [d]Institute for Systems Genetics, West China Hospital, Sichuan University, Chengdu, China

ABSTRACT

MicroRNAs (miRNAs) are small non-coding RNAs with the potential as biomarkers for disease diagnosis, prognosis and therapy. In the era of big data and biomedical informatics, computer-aided biomarker discovery has become the current frontier. However, most of the computational models are highly dependent on specific prior knowledge and training-testing procedures, very few are mechanism-guided or evidence-based. To the best of our knowledge, untill now no general rules have been uncovered and applied to miRNA biomarker screening. In this study, we manually collected literature-reported cancer miRNA biomarkers and analyzed their regulatory patterns, including the regulatory modes, biological functions and evolutionary characteristics of their targets in the human miRNA-mRNA network. Two evidences were statistically detected and used to distinguish biomarker miRNAs from others. Based on these observations, we developed a novel bioinformatics model and software tool for miRNA biomarker discovery (http://sysbio.suda.edu.cn/MiRNA-BD/). In contrast to routine methods that focus on miRNA synergic functions, our method searches for vulnerable sites in the miRNA-mRNA network and considers the independent regulatory power of miRNAs, i.e., single-line regulations between miRNAs and mRNAs. The performance comparison demonstrates the generality and precision of our model, which identifies miRNA biomarkers for cancers as well as other complex diseases without training or specific prior knowledge.

ARTICLE HISTORY
Received 4 March 2018
Revised 8 July 2018
Accepted 11 July 2018

KEYWORDS
Evidence-based bioinformatics model; miRNA biomarker discovery; miRNA-mRNA network analysis; single-line regulation mode

Introduction

MicroRNAs (miRNAs) are a class of small non-coding RNAs approximately comprising 22–24 nucleotides. They regulate the expression of more than 60% of human protein-coding genes by base-pairing with their target messenger RNAs (mRNAs) [1,2]. Accumulating studies indicate that miRNAs play critical roles in various biological processes [3], and that the aberrant expression of these small molecules often leads to the occurrence and development of complex human diseases, including cancers [4,5], diabetes [6], and cardiovascular diseases [7]. Due to their high sensitivity and specificity in patient samples, extensive efforts have been made to infer miRNAs as biomarkers for disease management and clinical decision making [8–10].

Most of the recent studies for miRNA biomarker discovery have been experimental, where differentially expressed (DE) miRNAs were first extracted from large-scale expression profiles and low-throughput experiments were then conducted to validate the results and understand disease pathogenesis [11–13]. However, detecting miRNA biomarkers using only experimental techniques is often time consuming and costly due to the complexity of biological systems. Thus, computational methods and data-driven models should preferably be developed [14]. For example, Madden et al. [15] proposed a model that combines correspondence and co-inertia analysis to screen for disease-associated miRNAs by analyzing gene expression data and miRNA-gene interactions. By considering the synergic functions of miRNAs, Xu et al. [16] constructed a miRNA target-dysregulated network and defined four important features for prioritizing miRNAs associated with human prostate cancer. Cun et al. [17] developed an R-package called netClass to integrate biological data and network analysis for biomarker signature discovery. Zhao et al. [18] developed a framework that infers cancer-related miRNAs based on gene expression profiles by the combined analysis of miRNA-gene relationships and dysfunctional pathway clustering. These approaches are scientifically valuable but the models obtained are often dependent on the training-and-testing procedure employed, and thus the precision of the predictions is strongly related to the quality of the training data. In addition, no general rules or mechanistic reasoning can be extracted from these models. It is widely acknowledged that the detection of disease-associated miRNAs is not the same as biomarker miRNA discovery, where the latter must be an indicator of a systematic change in state from health to disease, not only associated with the disease [19].

Until now, most of the studies have focused on the functional synergism among miRNAs as well as the multiple-to-multiple relationships between miRNAs and mRNAs. However, few have considered the sub-structure, especially the vulnerable regulatory sites in the miRNA-mRNA regulatory network. Similar to the protein-protein interaction network (PPIN) [20], the degree distribution of the human

CONTACT Bairong Shen bairong.shen@suda.edu.cn Center for Systems Biology, Soochow University, Suzhou, Jiangsu, China
Supplemental data can be accessed here.

© 2018 Informa UK Limited, trading as Taylor & Francis Group

miRNA-mRNA regulatory network also follows a power law, which implies that there are few miRNAs with stronger regulatory power [21]. In addition, some genes are regulated independently by an individual miRNA. In this study, we reconstructed the human miRNA-mRNA regulatory network by focusing on the vulnerable regulatory interactions, i.e., the single-line regulatory relationships between miRNAs and mRNAs. We discovered two features as evidences for characterizing biomarker miRNAs. First, they have a greater capacity for regulating genes independently compared with others and they tend to uniquely target genes with high degrees, but not hubs, in the human PPIN according to our cross-scale investigation. Second, statistical analysis of their functions showed that biomarker miRNAs are likely to target more transcription factor (TF) genes. Based on these two features, we developed an evidence-based bioinformatics model and software tool called MiRNA-BD for MicroRNA Biomarker Discovery. The executable file and related documents are available for downloading at: http://sysbio.suda.edu.cn/MiRNA-BD/. Our performance comparison indicated that the model detected miRNA biomarkers precisely without any model training. Translational applications to complex diseases, e.g., prostate cancer [21,22], colorectal cancer (CRC) [23], pediatric acute myeloid leukemia (AML) [24], sepsis [25], and acute coronary syndrome [26], demonstrated the good generality of the model, while further validations based on *in vitro* q-PCR experiments and pathway analyses convinced its predictive power [21–23].

Results

Biomarker miRNA characterization based on miRNA-mRNA network analyses

As shown in Figure 1, we classified the miRNA-mRNA relationship into four types based on their regulatory mechanisms, and considered their regulatory modes and the biological functions of their targets. Two measurements were defined as follows to quantify the regulatory power of single miRNAs in the network.

Measurement 1: The number of single-line regulation (NSR) is the number of genes that are independently or uniquely regulated by a single miRNA. Compared with single-line regulation, genes targeted by more than one miRNA, i.e., multiple-line regulation, would be more robust because the alteration of one of the regulatory links could be compensated for by others. We considered single-line regulation as a vulnerable structure in the network and alterations of these regulatory sites are important for the stability of the biological system.

Measurement 2: TF gene percentage (TFP) is the percentage of TF genes targeted by a single miRNA. TF genes are functionally important in the biological system, so this measurement can be used to quantify the functional importance of a given miRNA.

For example, the NSR values of the four miRNAs in Figure 1 are 2, 0, 1, and 3, and their TFP values are 0, 2/5, 1/2, and 3/5, respectively. We calculated NSR and TFP values for the 618 miRNAs in the human miRNA-mRNA regulatory network and two features were summarized as evidences for characterizing biomarker miRNAs as follows.

Evidence 1: Biomarker miRNAs tended to have high NSR values.

The distribution of the NSR values for miRNAs in the human miRNA-mRNA network followed a power law, as shown in Figure 2(a). Clearly, biomarker miRNAs had higher NSR values than others (p-value <2.2E-16, Wilcoxon signed-rank test), as illustrated in Figure 2(b), thereby indicating that the biomarker miRNAs had greater single-line regulatory power than other miRNAs. We categorized miRNAs in the human miRNA-mRNA network based on their NSR values. As shown in Figure 2(c), miRNAs in Group I had limited single-line regulatory power. For the miRNAs with NSR > 0, we further divided them into three groups (Group II-IV) where the cutoffs of 4 and 8 were the median and the third quartile of the NSR distribution, respectively. The results indicated that the biomarker miRNAs were distributed statistically significantly among those with higher NSR values. The percentages of miRNA biomarkers in the last three groups (Group II-IV) were significantly higher than that in the first group (p-value = 1.38E-11, Pearson's Chi-square test). Among the three groups with NSR > 0, the percentage of biomarker miRNAs in Group IV was significantly higher than that in Groups II and III with p-values of 4.17E-8 and 2.44E-4, respectively (Pearson's Chi-square test).

Evidence 2: Biomarker miRNAs tended to have high TFP values.

We compared the differences in the TFP values for the biomarker miRNAs and others. As illustrated in Figure 2(d), the biomarker miRNAs tended to have significantly higher TFP values (p-value = 1.38E-4, Wilcoxon signed-rank test). Together with the single-line regulatory power discussed above, we then checked the miRNAs with high NSR values (NSR ≥ 4, Groups III and IV in Figure 2(c)) in the network and found that 47.1% (106/225) of them had been previously reported as biomarkers (see Table S1). As shown in Figure 2

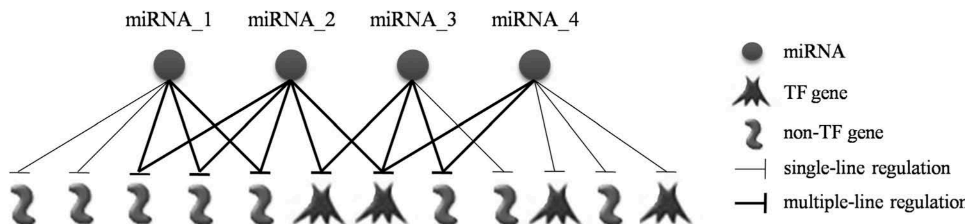


Figure 1. Four regulatory types of miRNAs in the human miRNA-mRNA network, i.e., TF/non-TF genes regulated by single/multiple miRNAs.

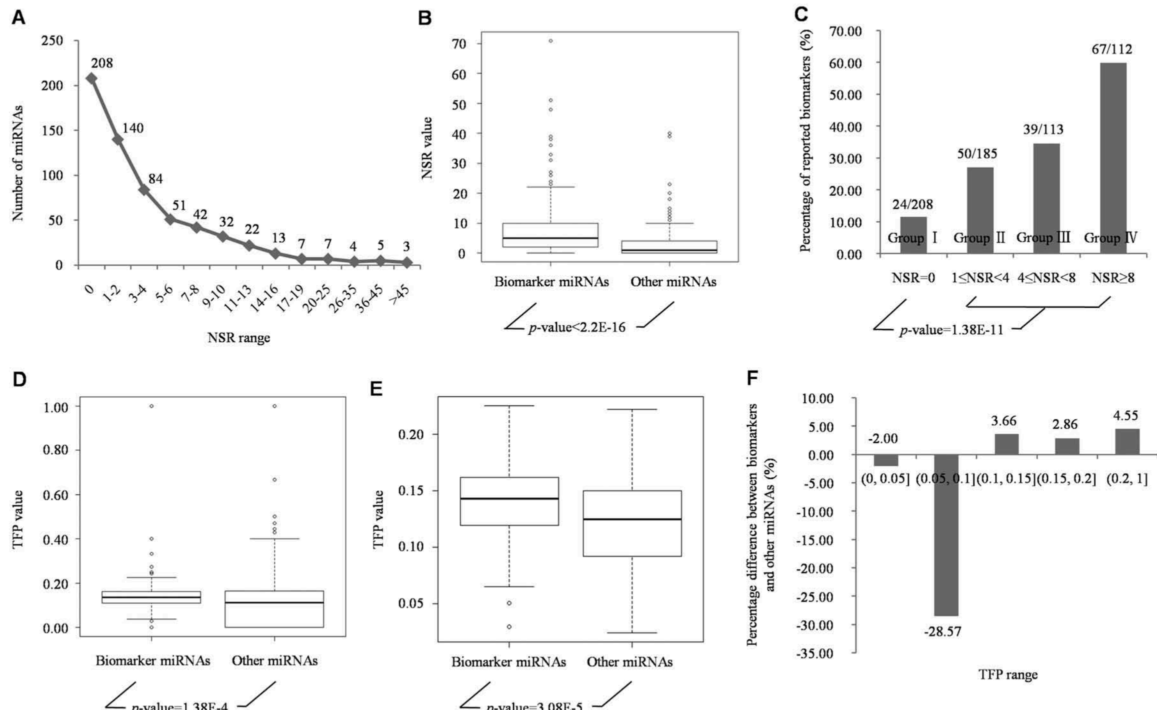


Figure 2. Characterization of miRNA biomarkers. (a) NSR distribution of miRNAs in the reconstructed human miRNA-mRNA network. (b) NSR distribution of biomarkers and other miRNAs. (c) Percentage of miRNA biomarkers according to their NSR values, i.e., 24/208 indicates that there are 208 miRNAs in the network with NSR = 0 and 24 of them are reported as biomarkers. (d) TFP distribution of biomarkers and other miRNAs. (e) TFP distribution of biomarkers and other miRNAs with high NSR values (NSR ≥ 4). (f) Percentage difference between biomarkers and other miRNAs with high NSR values (NSR ≥ 4).

(e), the biomarker miRNAs were likely to regulate more TF genes (p-value = 3.08E-5, Wilcoxon signed-rank test). According to Figure 2(f), the percentage of biomarker miRNAs was higher than that of other miRNAs when TFP > 0.1.

Functional and evolutionary analysis on targets of biomarker miRNAs

We mapped the genes independently regulated by biomarker miRNAs onto human PPIN and characterized their structural and functional traits in the network. As shown in Figure 3(a) and 3(b), the human PPIN had an approximately scale-free topological feature with a slope of −1.50 and the biomarker miRNAs tended to independently regulated genes with high degrees, but not the highest degree hubs (HDHs) in the network. After checking the annotations of HDH genes in both the NCBI Gene database [27] and Database of Essential Genes (DEG, version 11.0) [28], as shown in Table 1, we found that most of the HDH genes played crucial roles in biological processes such as cell cycle, proliferation, apoptosis, and signal transduction, and 75% (15/20) were known essential genes. In addition, HDHs in the PPIN were often highly enriched with lethal genes that are functionally important in cellular processes, whereas disease genes preferred the sites secondary to network hubs [29]. Thus, it is reasonable to suggest that the biomarker miRNAs refrain from regulating HDH genes independently because the dysregulation of these genes could be lethal for organisms.

According to previous studies, old genes with essential functions are often located in the centre of gene interaction networks, and new genes undergo long evolutionary processes before gradually integrating into the network from the peripheral sites to the hubs [20,30]. From the perspective of miRNA regulation, as illustrated in Figure 3(c), the statistical p-value of 0.48 (Wilcoxon signed-rank test) indicated that both the biomarkers and other miRNAs tended to regulate old genes (Branch 0 in the vertebrate phylogenetic tree; see the Materials and Methods). We checked the ages of the miRNA-targeted TFs and found that more than 64% (1,174/1,834) of these TF genes belonged to Branch 0 (i.e., the oldest genes), thereby indicating that the TF genes were usually the oldest in gene family. We further partitioned the old genes into TF or non-TF groups, and found that the biomarker miRNAs could regulate more TF genes (p-value = 7.60E-5, Wilcoxon signed-rank test, see Figure 3(d)) than others. It is suggested that the miRNAs identified as biomarkers tended to selectively target old genes with special functions during evolution.

Model implementation

Based on the evidences described above, we developed a novel bioinformatics model and a software tool called MiRNA-BD for miRNA biomarker discovery. To ensure the specificity of the miRNAs for certain disease conditions, as shown in Figure 4, the model first identifies a condition-specific miRNA-mRNA network from the reference, i.e., human miRNA-mRNA regulatory network (see the Materials and

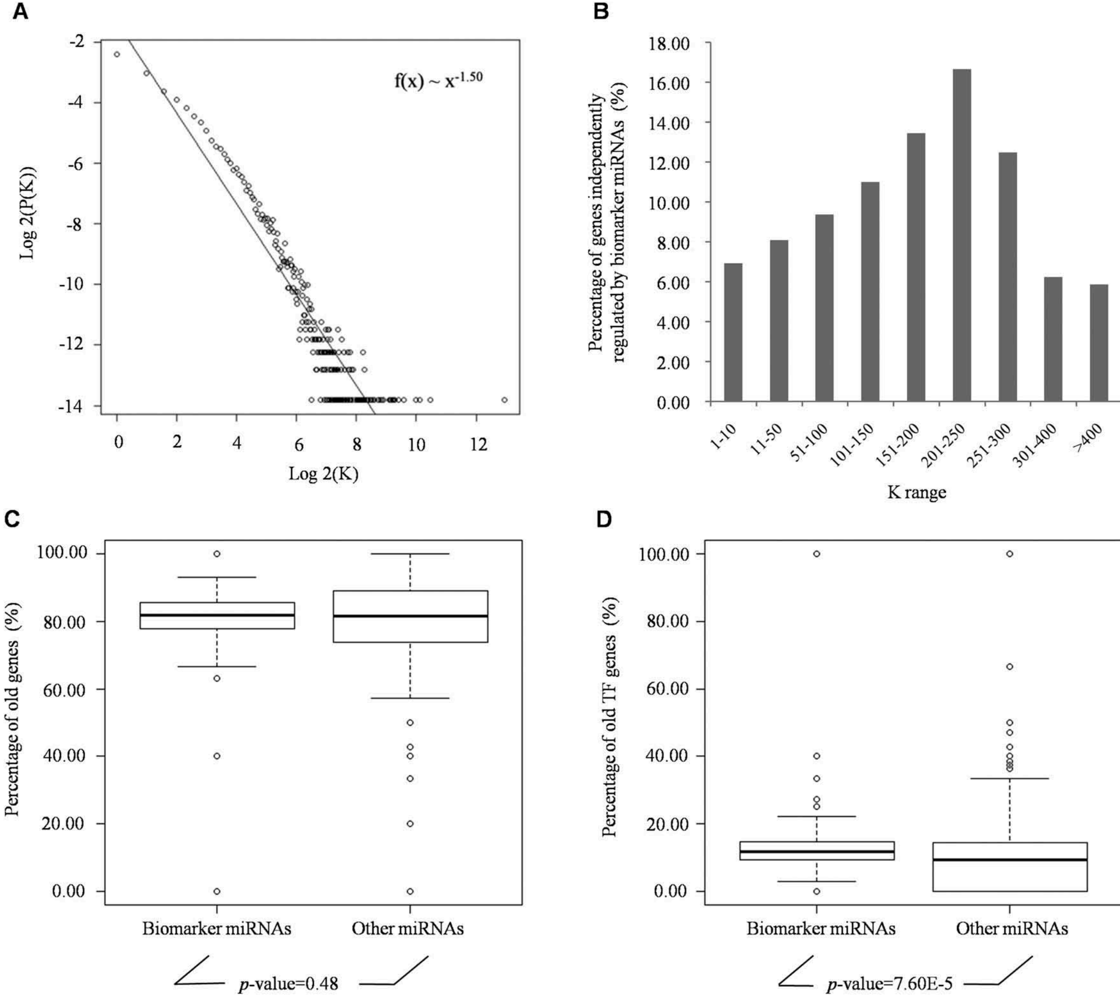


Figure 3. Results of functional and evolutionary analyses. (a) Degree distribution of genes in the human PPIN, where K represents the degree of genes and P(K) is the fraction of genes in the network with degree K to others. Both K and P(K) were 2-based log transformed. (b) Percentage of genes independently regulated by biomarker miRNAs, where K represents the degree of genes. (c) Percentages of old genes regulated by biomarker and other miRNAs. (D) Percentages of old TF genes regulated by biomarkers and other miRNAs.

Methods), according to the input data and it then calculates the two features, i.e., NSR and TFP values of miRNAs, based on the condition-specific network. Finally, miRNAs with significantly high NSR and TFP values (default threshold: p-value <0.05, Wilcoxon signed-rank test) are selected as candidate biomarkers for the disease considered.

The bioinformatics model was implemented as a Java program (MiRNA-BD, see Figure S1). The inputs comprised lists of disease-associated DE-miRNAs/mRNAs, or paired/unpaired miRNAs-mRNAs, although paired data are preferred. Here DE-miRNAs/mRNAs could be extracted from high-throughput expression profiles or low-throughput experiments by statistical analysis. Alternatively, the reference network and TF gene data set could be provided by the user, and the thresholds of the NSR and TFP values may also be adjusted based on the given p-value cutoffs. To better investigate miRNA biomarkers, the program integrated human miRNA sequence data from miRBase (release 21) [31], including 2,588 mature miRNAs and 1,881 precursors, and we added the secondary structure annotations for miRNA precursors predicted by RNAfold [32]. The Java executable file, user guide, and example data can be downloaded from the webpage at: http://sysbio.suda.edu.cn/MiRNA-BD/.

Application and performance evaluation

The NSR measure was previously applied to detect miRNA biomarkers for diagnosing and sub-typing prostate cancer [21,22]. In particular, Zhang et al. [21] identified 39 miRNAs as key players for prostate cancer diagnosis based on paired miRNA-mRNA expression profiles, where miR-648 was considered to be a novel molecular biomarker validated by *in vitro* experiments. Zhu et al. [22] inferred 11 miRNAs as potential biomarkers of castration-resistant prostate cancer based solely on mRNA expression data, where both experimental verification and functional analyses confirmed the predictive accuracy of the model. Peng et al. [23] screened nine miRNAs as putative biomarkers for predicting the response of CRC to preoperative chemoradiotherapy, and seven were confirmed by real-time q-PCR and western blot analysis. To improve the model, we integrated the TFP measure and applied it to the discovery of diagnostic miRNA

Table 1. Top 20 ranked genes with high degrees in the human PPIN.

Gene symbol	Degree in PPIN	Location	Description	Essential gene
UBC	7839	Chromosome 12, NC_000012.12	Ubiquitin C. The gene is associated with DNA repair, cell cycle, kinase modification, protein degradation.	Yes
ELAVL1	1415	Chromosome 19, NC_000019.10	ELAV like RNA binding protein 1. The gene plays roles in stabilizing ARE-containing mRNAs as well as in many diseases, including cancers.	Yes
SUMO2	1112	Chromosome 17, NC_000017.11	Small ubiquitin-like modifier 2. The gene is related to transcriptional regulation, nuclear transport and protein stability.	Unknown
CUL3	1009	Chromosome 2, NC_000002.12	Cullin 3. The gene is related to subsequent degradation of specific protein substrates.	Yes
KIAA0101	763	Chromosome 15, NC_000015.10	The gene is a predictive marker for hepatica cancer and is associated with poor survival of patients with esophageal cancer.	Unknown
COPS5	678	Chromosome 8, NC_000008.11	COP9 signalosome subunit 5. The gene is an important regulator in multiple signaling pathways.	Yes
SIRT7	621	Chromosome 17, NC_000017.11	Sirtuin 7. The gene may regulate epigenetic gene silencing as well as suppress recombination of rDNA.	Unknown
SUMO1	604	Chromosome 2, NC_000002.12	Small ubiquitin-like modifier 1. The gene is related to transcriptional regulation, nuclear transport and protein stability.	Yes
YWHAZ	594	Chromosome 8, NC_000008.11	The gene is able to mediate signal transduction.	Unknown
CAND1	572	Chromosome 12, NC_000012.12	Cullin-associated and neddylation-dissociated 1. The gene is related to prostate cell growth and is often disrupted in prostate cancer.	Unknown
GRB2	566	Chromosome 17, NC_000017.11	Growth factor receptor bound protein 2. The gene is involved in the signal transduction pathway.	Yes
TP53	564	Chromosome 17, NC_000017.11	Tumor protein p53. It is a tumor suppressor gene and mutations in this gene are related to multiple human cancers.	Yes
CUL1	549	Chromosome 7, NC_000007.14	Cullin 1. The gene promotes cell proliferation through cell cycle.	Yes
MYC	478	Chromosome 8, NC_000008.11	V-myc avian myelocytomatosis viral oncogene homolog. It is a transcription factor gene which plays roles in cell cycle, apoptosis and cellular transformation.	Yes
TRAF6	459	Chromosome 11, NC_000011.10	TNF receptor associated factor 6. The gene is able to mediate signal transduction.	Yes
HDAC1	442	Chromosome 1, NC_000001.11	Histone deacetylase 1. The gene is able to regulate eukaryotic gene expression.	Yes
EP300	422	Chromosome 22, NC_000022.11	E1A binding protein p300. The gene is important in cell proliferation and differentiation.	Yes
SRC	390	Chromosome 20, NC_000020.11	SRC proto-oncogene, non-receptor tyrosine kinase. The gene is able to regulate embryonic development and cell growth.	Yes
VHL	387	Chromosome 3, NC_000003.12	Von Hippel-Lindau tumor suppressor. The gene is associated with the development of cancers.	Yes
EWSR1	386	Chromosome 22, NC_000022.11	EWS RNA binding protein 1. The gene is important and functional in gene expression, cell signaling, RNA processing and transport.	Yes

Note: The symbol 'SIRT7' in bold means the gene is independently regulated by a biomarker miRNA.

biomarkers for pediatric AML [24]. *In vitro* q-PCR experiments showed that the predicted miR-155 (i.e., miR-155-5p) and miR-196b (i.e., miR-196b-5p) were significantly overexpressed in children with AML and in subgroups M4-M5, respectively. In addition to cancer studies, the model is suitable for miRNA biomarker discovery in other complex diseases such as sepsis [25] and acute coronary syndrome [26]. The details are summarized in Table 2.

In order to validate the robustness and efficiency of the reconstructed reference network, DE-miRNAs and DE-mRNAs identified in our previous study [24] were used to update the AML-specific miRNA-mRNA network. As illustrated in Table 3 and Figure 5, three miRNAs with significantly high NSR and TFP values were screened as pediatric AML biomarkers according to the MiRNA-BD model. In particular, two (miR-155-5p, miR-196b-5p) were consistent with our previous results and they were confirmed by biological experiments [24]. Interestingly, the novel miR-221-3p, which was not predicted in our previous study, was also associated with AML [33] and its primary miRNA pri-221 was reported as a molecular marker and putative oncogene in the progression of AML [34]. Furthermore, we utilized the (Database for Annotation, Visualization and Integrated Discovery (DAVID) [35] and Ingenuity Pathway Analysis (IPA) [36] to map the targets of the identified pediatric AML miRNA biomarkers from the AML-specific miRNA-mRNA network (see Figure S2) onto Kyoto Encyclopedia of Genes and Genomes (KEGG) [37] and IPA pathways, respectively. As shown in Figure 6, four of the top 10 significantly enriched pathways were common in both KEGG and IPA, including MAPK (Mitogen-Activated Protein Kinase) Signaling Pathway (ERK/MAPK Signaling), Pathways in Cancer (Molecular Mechanisms of Cancer), Colorectal Cancer (Colorectal Cancer Metastasis Signaling), and T Cell Receptor Signaling Pathway (T Cell Receptor Signaling). After searching for citations in PubMed, we found that the enriched signaling pathways directly or indirectly participated in the genesis and development of AML. For instance, the ERK/MSK MAPK signaling pathway is involved in the activation of Sp1/c-Myc and the regulation the expression of survivin to modulate drug resistance in leukemia stem cells [38]. The combination of drugs comprising azacitidine and cetuximab has good therapeutic effects in patients with AML following oxaliplatin for metastatic CRC, thereby indicating the interrelationship between the pathogenesis and treatment of AML and CRC metastasis [39]. Majeti et al. [40] first showed that

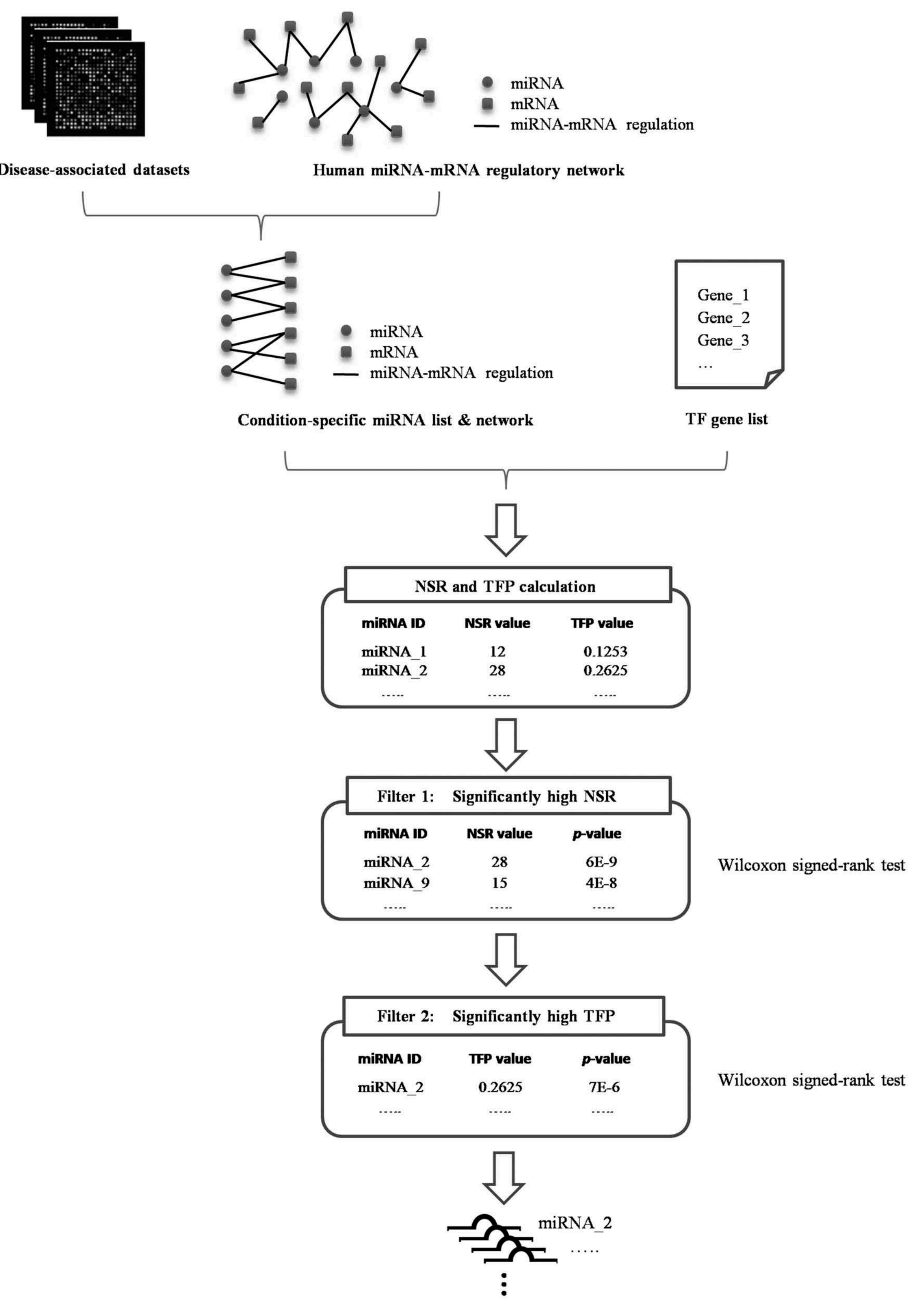


Figure 4. Schematic pipeline of the MiRNA-BD model. Here disease-associated data sets can be lists of DE-miRNAs/mRNAs, or paired/unpaired miRNAs-mRNAs, and miRNAs with significantly high NSR and TFP values are screened as candidate biomarkers (default threshold: p-value <0.05, Wilcoxon signed-rank test).

the T cell receptor signaling pathway is potentially related to the functional regulation of AML stem cells. In addition, pathways such as Cell Cycle [41], Wnt Signaling Pathway [42], Chronic Myeloid Leukemia [43], and Glucocorticoid Receptor Signaling [44] also have critical roles during AML initiation and evolution.

We also compared the performance of our model with those of three state-of-the-art models, as described in the

Table 2. Translational applications of NSR and TFP measurements for miRNA biomarker discovery.

Cancer type	Sample	Identified biomarker number	Validation: Literature report	*In vitro* q-PCR	Western blot	*In silico* analysis	PMID
Prostate cancer	**miRNA mRNA**	39	Yes	Yes	No	Yes	24,618,011
Castration-resistant prostate cancer	mRNA	11	Yes	Yes	Yes	Yes	26,540,468
Colorectal cancer	miRNA	9	Yes	Yes	Yes	Yes	27,903,980
Pediatric acute myeloid leukemia	miRNA mRNA	3	Yes	Yes	No	Yes	26,317,787
Sepsis	miRNA	10	Yes	No	No	Yes	24,809,055
Acute Coronary Syndrome	miRNA	AMI: 27 UA: 26	Yes	No	No	Yes	28,044,128

Note: *Sample* in bold denotes paired miRNA-mRNA expression data. Values in the *Identified biomarker number* column represent the number of miRNA biomarkers identified by our model. Abbreviations: AMI: acute myocardial infarction, UA: unstable angina, PMID: PubMed ID.

Table 3. Diagnostic miRNA biomarkers identified for pediatric acute myeloid leukemia based on the model with the reconstructed miRNA-mRNA reference network.

No.	miRNA symbol	NSR (p-value)	TFP (p-value)
1	miR-155-5p	37(9.69E-8)	0.1852(0.008)
2	miR-196b-5p	28(2.65E-5)	0.1733(0.027)
3	miR-221-3p	17(0.016)	0.1951(0.004)

Note: The miRNAs are listed in descending order of their NSR values.

Material and Methods section. As shown in Figure 7, our model achieved comparable and the overall best performance for biomarker discovery relative to the others. However, the miR_Path model was based on the guidance from dysfunctional pathway clustering, and thus it was highly dependent on the pathway knowledge related closely to certain disease types [18]. The cancer miRNA prioritization (CMP) model prioritizes human cancer miRNAs by calculating the functional consistency between the miRNA targets and cancer-related genes, so specific cancer gene sets are needed as prior knowledge [45]. Jiang's model employs both the disease phenotype similarities and miRNA-disease relationship as knowledge for disease-associated miRNA prioritization [46]. None of the other three models integrate the sub-structural information from the miRNA-mRNA network and the regulatory mechanism of miRNAs. Thus, to the best of our knowledge, our model is the first to treat miRNA regulatory mechanisms as important features for biomarker discovery. In conclusion, our model detects key miRNAs based on two statistical evidences (i.e., NSR and TFP), neither specific prior knowledge nor a training procedure is required during biomarker screening. Therefore, the model had good generality in translational applications. Finally, different types of RNA samples, i.e., miRNA or mRNA only, paired/unpaired miRNAs-mRNAs, can be selected as the model input according to specific research needs.

Discussion

It is widely recognized that the aberrant expression of miRNAs is associated with numerous human malignancies [47–49]. Many studies have focused on detecting miRNA biomarkers in order to provide precision clinical diagnosis and treatment [50–52]. Most of these studies identified miRNA biomarkers based on biological experiments, and bioinformatics approaches focusing on miRNA synergism are also well performed. The miRNA-mRNA networks were constructed sequentially, but few studies explored the hidden structures in the network and no general rules were extracted and applied to miRNA biomarker discovery.

In this study, we discovered two evidences where biomarker miRNAs have high single-line regulatory power and they tend to regulate more TF genes. The former was identified based on sub-structural analysis of the human miRNA-mRNA regulatory network. Compared with multiple-line regulation, it is reasonable to assume that unique or single-line regulatory interactions are more vulnerable, and they may be important for systematic state changes. The second evidence was identified by investigating both the biological functions and gene evolutionary patterns. It is important to consider TF genes because TFs are among the key factors that influence the functions of gene networks, where they regulate downstream genes during various biological processes, and the abnormal expression of TF genes is closely related to human carcinogenesis. Clearly, if more TFs are altered, gene expression and the biological system will be greatly affected, and thus miRNAs with high TFP values are more likely to contribute to pathogenesis and the change in state from health to disease, which is an essential feature of a biomarker.

Based on the characterization of biomarker miRNAs in the miRNA-mRNA network and the functional analyses of their targets, we developed a novel computational model

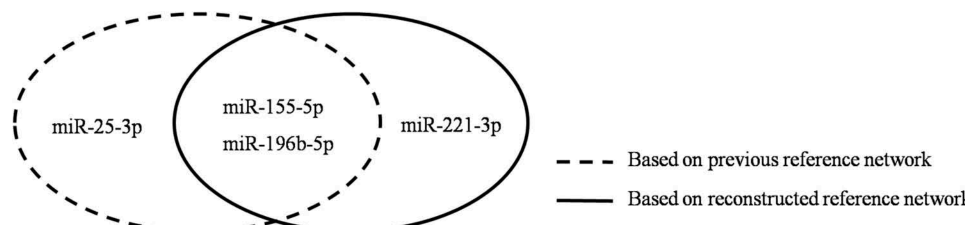


Figure 5. Comparison of results obtained from different reference networks. The miRNA symbols in dotted and solid ellipses are screened biomarkers based on previous and reconstructed human miRNA-mRNA regulatory networks, respectively.

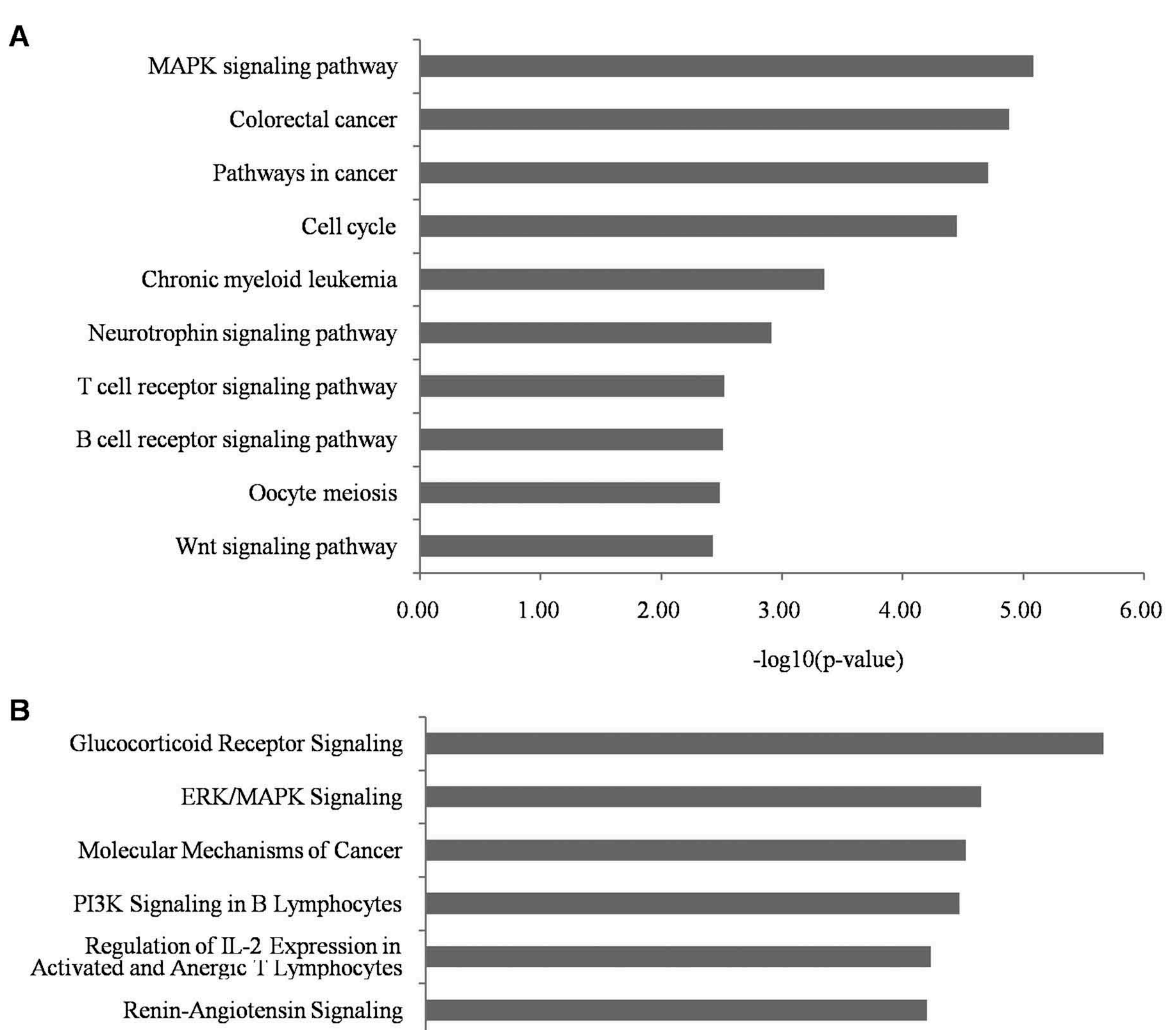
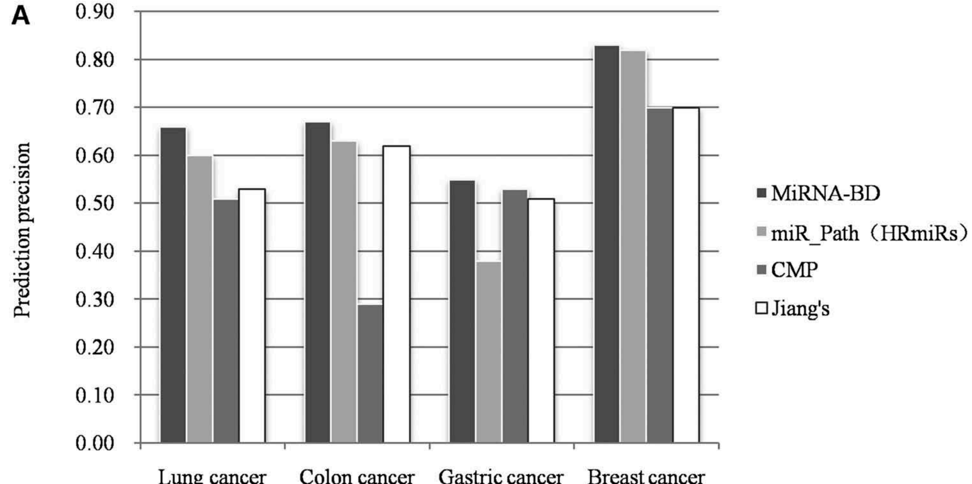


Figure 6. Pathway enrichment analysis for targets of identified pediatric AML miRNA biomarkers. The statistical significance levels (*p*-value) were negative 10-based log transformed and the top 10 significantly enriched pathways are listed. (a) Results of KEGG pathway enrichment analyses. (b) Results of IPA pathway enrichment analyses.

and software tool for miRNA biomarker discovery. Since the single-line regulation and the regulation of TF genes are important for the stability of network systems, we employed these two criteria as the principle and general rule for miRNA biomarker measurement. Compared with routine methods that rely greatly on training data or specific prior knowledge, experimental validations conducted in our previous studies showed that the proposed model can detect miRNA biomarkers with good generality and accuracy. However, some limitations still exist and we will improve the model as follows. First, only two features were discovered and considered in the present study. Although the feasibility and effectiveness of the model was validated in our application-based studies, the NSR and TFP are not the only attributes suitable for miRNA characterization. True biomarkers ought to be accurate, reproducible, and highly specific to a disease, and thus it is possible that a miRNA that does not target TF genes or that only binds influential mRNAs could also serve as a biomarker. Due to the complexity of disease development, more disease-specific signals should be rigorously weighted, e.g., the percentage of disease-associated genes, and the sensitivity and specificity of identified molecules, etc. At present, we have only investigated the number of independently regulated



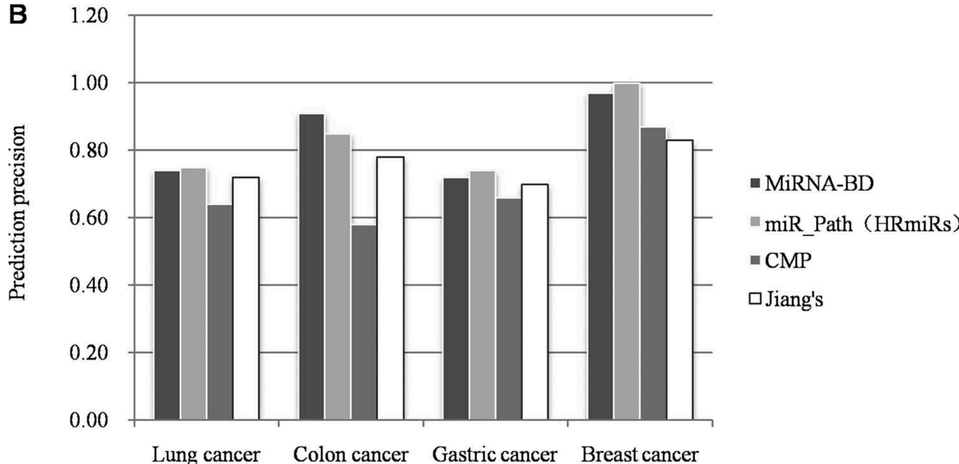


Figure 7. Comparison of the prediction precision. The prediction precision was defined as the percentage of reported cancer biomarker/associated miRNAs in the whole predicted set. (a) Precision of cancer biomarker miRNA prediction. (b) Precision of cancer associated miRNA prediction.

genes and TF genes, but the roles of genes are often not the same in different biological backgrounds, so more conditional annotations need to be measured. In addition, we are exploring the importance of other types of genes in biological systems, e.g., essential genes, house-keeping genes, tumor suppressor genes, and oncogenes, and the NSR requirement for a biomarker should also include information about the types of genes it is regulating. Second, the current model does not consider the fact that most RNA-based biomarkers are probably involved in an expression signature that includes a number of miRNAs. This is important because many miRNAs are pleiotropic and functional in various biological processes or disease states, and thus they would not be sufficiently specific to predict disease on their own. Considering the dynamic nature and heterogeneity of complex diseases, our next step is identifying network biomarkers that integrate dysfunctional miRNAs/mRNAs with their regulatory patterns for precision medicine and healthcare. Third, the miRNAs involved in disease pathogenesis do not always match with miRNAs that serve as biomarkers or that indicates disease status, and thus it might be more accurate to term the miRNAs identified by our model as 'key players' or 'candidate/putative biomarkers'. To address this issue, it is necessary to perform wet-lab verification using cell lines, model animals, or human samples in order to obtain better evaluations of the identified miRNAs as biomarkers. Finally, the software will be upgraded to a cross-platform version with a user-friendly interface, which is more convenient to use. In addition, only 618 miRNAs were included in our reconstructed network but more than 2,000 human mature miRNAs are recorded in miRBase (release 21) [31], so the network can be expanded when more miRNA-mRNA pairs are reported and validated, which will further increase our knowledge of miRNA biomarkers.

Materials and methods

Human miRNA-mRNA regulatory network reconstruction

The human miRNA-mRNA regulatory network was used as the reference for constructing the condition-specific miRNA-mRNA network. In this study, we reconstructed the network based on both experimentally validated and computationally

predicted miRNA–mRNA regulatory data. The experimental data were extracted from miRTarBase (version 4.5) [53], TarBase (version 6.0) [54], miRecords (version 4.0) [55], and miR2Disease [56]. The computationally predicted data included information from HOCTAR (version 2.0) [57], ExprTargetDB [58], and starBase (version 2.0) [59].

To reduce the number of false positives, miRNA-mRNA pairs validated based on low-throughput experiments (e.g., real-time PCR, etc.) and those present in at least two of the three computationally predicted databases were mainly selected for network reconstruction. We note that the nomenclature for miRNAs has changed since miRBase updated to the 17th version and several databases (e.g., miRTarBase, TarBase, and starBase) or studies refer to miRNAs based on the new rule (e.g., hsa-miR-21-5p and hsa-miR-21-3p), whereas some others (e.g., miRecords, miR2Disease, HOCTAR, and ExprTargetDB) still use the previous symbols (e.g., hsa-miR-21 and hsa-miR-21*). To reduce inconsistencies, we cleaned the data and the final reconstructed network comprised 48,868 regulatory pairs among 618 miRNAs and 9,526 target genes.

Cancer miRNA biomarkers and transcription factor gene data set

The literature-reported cancer miRNA biomarkers were manually extracted from citations in PubMed using the search term: '(cancer[tiab] OR carcinoma[tiab]) AND (microRNA*[tiab] OR miRNA*[tiab]) AND (biomarker*[tiab] OR marker*[tiab] OR indicator*[tiab] OR predict*[tiab])'. The biomarker potential of miRNAs was rechecked till June 2017 based on three criteria: 1) the miRNAs were identified in human samples, including tissues, blood, or cell lines, using high-throughput (e.g., microarrays, next-generation sequencing, etc.) or low-throughput methods (e.g., real-time PCR, etc.); 2) the miRNAs were highlighted as biomarkers/markers/indicators/predictors or candidate/potential/putative/latent biomarkers throughout the research article; and 3) the identified biomarkers were validated by either wet-lab experiments (e.g., real-time PCR, etc.) or bioinformatics analyses (e.g., receiver operating characteristic curve, clustering, functional enrichment, etc.). To avoid duplication, only one item was fully considered if a miRNA was reported as a biomarker in multiple studies. In particular, we standardized all of the miRNA symbols according to the records in miRBase (release 21) [31]. Finally, 180 miRNA biomarkers associated with more than 20 types of cancers were collected for further analysis and their details are shown in Table S1.

The TF gene data set was collected from the review by Vaquerizas et al. [60], which included 1,834 human TF genes. In addition, we manually added the aliases of each TF gene based on the records in the National Center for Biotechnology Information (NCBI) Gene database [27].

Human protein-protein interaction network

The human PPIN was obtained from Protein Interaction Network Analysis (PINA) version 2.0 [61]. This platform unifies PPIs from six publicly available databases comprising BioGRID [62], DIP [63], HPRD [64], IntAct [65], MINT [66], and MIPS/Mpact [67]. Duplicated items or those with ambiguous gene symbols in PPIs were filtered out. The final PPIN contained 14,441 genes and 107,802 interactive pairs. Herein, proteins were considered to be equivalent to their protein-coding genes.

Gene age or evolutionary information

To characterize gene evolution, human gene age information was retrieved from a study by Zhang et al. [68]. In brief, protein-coding genes were divided into 13 branches (labeled in order from Branch 0 to 12) by inferring the presence and absence of orthologs throughout the vertebrate phylogenetic tree [20], where genes in branches with higher numbers were considered relatively younger. Here old genes were strictly defined as those belonging to Branch 0. All of the gene symbols were unified using the ID conversion tool bioDBnet [69].

Pathway enrichment analyses

The targets of the predicted miRNA biomarkers were retrieved from the AML-specific miRNA-mRNA network and mapped onto KEGG [37] and IPA pathways using the DAVID [35] and IPA [36] programs, respectively. Here the top 10 pathways enriched with statistically significant p-values <0.01 were analyzed to determine their associations with AML development by mining the NCBI PubMed citations.

Performance comparison

To evaluate the predictive power of our model, we compared its performance with three state-of-the-art models: 1) the miR_Path model infers cancer-related miRNAs based on gene expression data and dysfunctional pathway clustering [18]; 2) the CMP model prioritizes cancer miRNAs by measuring the functional consistency of known cancer-related genes and miRNA target genes [45]; and 3) Jiang's model identifies disease-related miRNAs by constructing a human phenome-microRNAome network [46]. To ensure an unbiased comparison, data sets for four types of human cancers (i.e., lung cancer, colon cancer, gastric cancer, and breast cancer; see Table S2) tested by miR_Path were downloaded directly from the Gene Expression Omnibus (GEO) and used in the evaluation. The Student's t-test was employed to extract DE-genes from each data set (p-value <0.05). The common miRNAs obtained from two data sets for the same type of cancer were recognized as the final result and they were compared with the HRmiRs from miR_Path, which achieved the best performance [18]. The other two methods do not use gene expression data, so the same numbers of top ranked miRNAs from their predictions as those found in our model were selected for the comparison. Besides the records in the Human microRNA Disease Database (HMDD, version 2.0) as the gold standard [70], we searched for the identified miRNAs in PubMed citations (the search criterion was similar to that described above) to evaluate the prediction precision as the percentage

of reported cancer biomarker/associated miRNAs in the whole predicted set.

Acknowledgments

The authors gratefully thank the academic editor and the anonymous reviewers for their constructive suggestions, which helped to improve this manuscript. They also thank Prof. Luonan Chen at Key Laboratory of Systems Biology, Shanghai Institute of Biological Sciences, Chinese Academy of Sciences for providing the IPA analysis.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by the National Natural Science Foundation of China (Grant Nos. 31670851, 31470821, and 91530320) and National Key R&D programs of China (Grant No. 2016YFC1306605).

ORCID

Yuxin Lin http://orcid.org/0000-0001-6543-0317
Wentao Wu http://orcid.org/0000-0002-8119-5537
Zhandong Sun http://orcid.org/0000-0002-5805-0284
Li Shen http://orcid.org/0000-0002-9789-9360
Bairong Shen http://orcid.org/0000-0003-2899-1531

References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004 Jan 23;116(2):281–297. PubMed PMID: 14744438.
2. Cui Q, Yu Z, Purisima EO, et al. Principles of microRNA regulation of a human cellular signaling network. Mol Syst Biol. 2006;2:46. PubMed PMID: 16969338; PubMed Central PMCID: PMC1681519.
3. Ambros V. The functions of animal microRNAs. Nature. 2004 Sep 16;431(7006):350–355. . PubMed PMID: 15372042.
4. Esteller M. Non-coding RNAs in human disease. Nat Rev Genet. 2011 Dec;12(12):861–874. . PubMed PMID: 22094949.
5. Dvinge H, Git A, Graf S, et al. The shaping and functional consequences of the microRNA landscape in breast cancer. Nature. 2013 May 16;497(7449):378–382. PubMed PMID: 23644459.
6. Al-Kafaji G, Al-Mahroos G, Alsayed NA, et al. Peripheral blood microRNA-15a is a potential biomarker for type 2 diabetes mellitus and pre-diabetes. Mol Med Rep. 2015 Nov;12(5):7485–7490. PubMed PMID: WOS:000364889100037; English.
7. McManus DD, Freedman JE. MicroRNAs in platelet function and cardiovascular disease. Nat Rev Cardiol. 2015 Dec;12(12):711–717. . PubMed PMID: WOS:000365560300003; English.
8. Dang TM, Wong WC, Ong SM, et al. MicroRNA expression profiling of human blood monocyte subsets highlights functional differences. Immunology. 2015 Jul;145(3):404–416. PubMed PMID: 25707426; PubMed Central PMCID: PMC4479539.
9. Chen X, Ba Y, Ma L, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res. 2008 Oct;18(10):997–1006. PubMed PMID: 18766170.
10. Mitchell PS, Parkin RK, Kroh EM, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci USA. 2008 Jul 29;105(30):10513–10518. PubMed PMID: 18663219; PubMed Central PMCID: PMC2492472.
11. Fan C, Lin Y, Mao Y, et al. MicroRNA-543 suppresses colorectal cancer growth and metastasis by targeting KRAS, MTA1 and HMGA2. Oncotarget. 2016 Mar 8 PubMed PMID: 26968810. 10.18632/oncotarget.7989.
12. Qiu X, Zhang J, Shi W, et al. Circulating microRNA-26a in plasma and its potential diagnostic value in gastric cancer. PloS One. 2016;11(3):e0151345. PubMed PMID: 27010210; PubMed Central PMCID: PMC4806920.
13. Yau TO, Wu CW, Tang CM, et al. microRNA-20a in human faeces as a non-invasive biomarker for colorectal cancer. Oncotarget. 2016 Jan 12;7(2):1559–1568. PubMed PMID: 26621842.
14. Lin Y, Qian F, Shen L, et al. Computer-aided biomarker discovery for precision medicine: data resources, models and applications. Brief Bioinform. 2017 Nov 29 PubMed PMID: 29194464; eng. 10.1093/bib/bbx158.
15. Madden SF, Carpenter SB, Jeffery IB, et al. Detecting microRNA activity from gene expression data. BMC Bioinf. 2010;11:257. PubMed PMID: 20482775; PubMed Central PMCID: PMC2885376.
16. Xu J, Li CX, Lv JY, et al. Prioritizing candidate disease miRNAs by topological features in the miRNA target-dysregulated network: case study of prostate cancer. Mol Cancer Ther. 2011 Oct;10(10):1857–1866. 10.1158/1535-7163.MCT-11-0055. PubMed PMID: 21768329.
17. Cun Y, Frohlich H. netClass: an R-package for network based, integrative biomarker signature discovery. Bioinformatics. 2014 May 1;30(9):1325–1326. . PubMed PMID: 24443376.
18. Zhao XM, Liu KQ, Zhu G, et al. Identifying cancer-related microRNAs based on gene expression data. Bioinformatics. 2015 Apr 15;31(8):1226–1234. PubMed PMID: 25505085.
19. Chen J, Sun M, Shen B. Deciphering oncogenic drivers: from single genes to integrated pathways. Brief Bioinform. 2015 May;16(3):413–428. . PubMed PMID: 25378434.
20. Zhang W, Landback P, Gschwend AR, et al. New genes drive the evolution of gene interaction networks in the human and mouse genomes. Genome Biol. 2015 Oct 1;16:202. 10.1186/s13059-015-0772-4. PubMed PMID: 26424194; PubMed Central PMCID: PMCPMC4590697. eng.
21. Zhang W, Zang J, Jing X, et al. Identification of candidate miRNA biomarkers from miRNA regulatory network with application to prostate cancer. J Transl Med. 2014 Mar 11;12:66. PubMed PMID: 24618011; PubMed Central PMCID: PMCPMC4007708. eng.
22. Zhu J, Wang SG, Zhang WY, et al. Screening key microRNAs for castration-resistant prostate cancer based on miRNA/mRNA functional synergistic network. Oncotarget. 2015 Dec 22;6(41):43819–43830. PubMed PMID: WOS:000369908400049; English.
23. Zhu Y, Peng Q, Lin Y, et al. Identification of biomarker microRNAs for predicting the response of colorectal cancer to neoadjuvant chemoradiotherapy based on microRNA regulatory network. Oncotarget. 2017 Jan 10;8(2):2233–2248. PubMed PMID: 27903980; PubMed Central PMCID: PMC5356795.
24. Yan WY, Xu LH, Sun ZD, et al. MicroRNA biomarker identification for pediatric acute myeloid leukemia based on a novel bioinformatics model. Oncotarget. 2015 Sep 22;6(28):26424–26436. PubMed PMID: WOS:000363160100141; English.
25. Huang J, Sun Z, Yan W, et al. Identification of microRNA as sepsis biomarker based on miRNAs regulatory network analysis. Biomed Res Int. 2014;2014:594350. PubMed PMID: 24809055; PubMed Central PMCID: PMC3997997.
26. Zhu Y, Lin Y, Yan W, et al. Novel biomarker microRNAs for subtyping of acute coronary syndrome: a bioinformatics approach. Biomed Res Int. 2016;2016:4618323. PubMed PMID: 28044128; PubMed Central PMCID: PMC5156791.
27. Brown GR, Hem V, Katz KS, et al. Gene: a gene-centered information resource at NCBI. Nucleic Acids Res. (Database issue) 2015 Jan;43:D36–D42. PubMed PMID: 25355515; PubMed Central PMCID: PMC4383897.
28. Zhang R, Ou HY, Zhang CT. DEG: a database of essential genes. Nucleic Acids Res. 2004 Jan 1;32:D271–D272. (Database issue). PubMed PMID: 14681410; PubMed Central PMCID: PMC308758.
29. Han HW, Ohn JH, Moon J, et al. Yin and Yang of disease genes and death genes between reciprocally scale-free biological networks.

Nucleic Acids Res. 2013 Nov;41(20):9209–9217. PubMed PMID: WOS:000326746400010; English.
30. Warnefors M, Eyre-Walker A. The accumulation of gene regulation through time. Genome Biol Evol. 2011;3:667–673. PubMed PMID: 21398425; PubMed Central PMCID: PMC3157833.
31. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014 Jan;42((Database issue)):D68–D73. PubMed PMID: 24275495; PubMed Central PMCID: PMC3965103
32. Lorenz R, Bernhart SH, Honer Zu Siederdissen C, et al. ViennaRNA package 2.0. Algorithms Mol Biol. 2011;6:26. PubMed PMID: 22115189; PubMed Central PMCID: PMC3319429.
33. Cammarata G, Augugliaro L, Salemi D, et al. Differential expression of specific microRNA and their targets in acute myeloid leukemia. Am J Hematol. 2010 May;85(5):331–339. PubMed PMID: 20425795.
34. Rommer A, Steinleitner K, Hackl H, et al. Overexpression of primary microRNA 221/222 in acute myeloid leukemia. BMC Cancer. 2013;13:364. PubMed PMID: 23895238; PubMed Central PMCID: PMC3733744.
35. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. PubMed PMID: 19131956 Nat Protoc. 2009;41:44–57.
36. Kramer A, Green J, Pollard J Jr., et al. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014 Feb 15;30 (4):523–530. . PubMed PMID: 24336805; PubMed Central PMCID: PMC3928520.
37. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30. PubMed PMID: 10592173; PubMed Central PMCID: PMC102409.
38. Zhang Y, Chen HX, Zhou SY, et al. Sp1 and c-Myc modulate drug resistance of leukemia stem cells by regulating survivin expression through the ERK-MSK MAPK signaling pathway. Molecular Cancer. 2015;14:56, 10.1186/s12943-015-0326-0. PubMed PMID: 25890196; PubMed Central PMCID: PMC4357193
39. Hashimoto A, Takada K, Horiguchi H, et al. Combination chemotherapy of azacitidine and cetuximab for therapy-related acute myeloid leukemia following oxaliplatin for metastatic colorectal cancer. Case Rep Oncol. 2014 May;7(2):316–322. PubMed PMID: 24932174; PubMed Central PMCID: PMC4049016.
40. Majeti R, Becker MW, Tian Q, et al. Dysregulated gene expression networks in human acute myelogenous leukemia stem cells. Proc Natl Acad Sci USA. 2009 Mar 3 106(9):3396–3401. PubMed PMID: 19218430; PubMed Central PMCID: PMC2642659
41. Cao L, Xiao PF, Tao YF, et al. Microarray profiling of bone marrow long non-coding RNA expression in Chinese pediatric acute myeloid leukemia patients. Oncol Rep. 2016 Feb;35(2):757–770. PubMed PMID: 26573779.
42. Valencia A, Roman-Gomez J, Cervera J, et al. Wnt signaling pathway is epigenetically regulated by methylation of Wnt antagonists in acute myeloid leukemia. Leukemia. 2009 Sep;23(9):1658–1666. PubMed PMID: 19387464.
43. De La Puente P, Weisberg E, Muz B, et al. Identification of ILK as a novel therapeutic target for acute and chronic myeloid leukemia. Leuk Res. 2015 Sep 9 PubMed PMID: 26413753. DOI:10.1016/j.leukres.2015.09.005.
44. Skoog L, Ost A, Biberfeld P, et al. Prognostic significance of terminal transferase activity and glucocorticoid receptor levels in acute myeloid leukemia. Br J Cancer. 1984 Oct;50(4):443–449. PubMed PMID: 6593093; PubMed Central PMCID: PMC1976893.
45. Li X, Wang Q, Zheng Y, et al. Prioritizing human cancer microRNAs based on genes' functional consistency between microRNA and cancer. Nucleic Acids Res. 2011 Dec;39(22):e153. PubMed PMID: 21976726; PubMed Central PMCID: PMC3239203. eng.
46. Jiang Q, Hao Y, Wang G, et al. Prioritization of disease microRNAs through a human phenome-microRNAome network. BMC Syst Biol. 2010;4 Suppl 1:S2. PubMed PMID: 20522252; PubMed Central PMCID: PMC2880408. eng.
47. Allison DB, Cui X, Page GP, et al. Microarray data analysis: from disarray to consolidation and consensus. Nat Rev Genet. 2006 Jan;7(1):55–65. PubMed PMID: 16369572.
48. Calin GA, Croce CM. MicroRNA signatures in human cancers. Nature Rev Cancer. 2006 Nov;6(11):857–866. . PubMed PMID: 17060945.
49. Niu Y, Jin Y, Deng SC, et al. MiRNA-646-mediated reciprocal repression between HIF-1alpha and MIIP contributes to tumorigenesis of pancreatic cancer. Oncogene. 2018 Jan 18 PubMed PMID: 29343850. DOI:10.1038/s41388-017-0082-2.
50. Gasparini P, Cascione L, Fassan M, et al. microRNA expression profiling identifies a four microRNA signature as a novel diagnostic and prognostic biomarker in triple negative breast cancers. Oncotarget. 2014 Mar 15;5(5):1174–1184. PubMed PMID: 24632568; PubMed Central PMCID: PMC4012726.
51. Marcucci G, Maharry KS, Metzeler KH, et al. Clinical role of microRNAs in cytogenetically normal acute myeloid leukemia: miR-155 upregulation independently identifies high-risk patients. J Clin Oncol Off J Am Soc Clin Oncol. 2013 Jun 10;31(17):2086–2093. PubMed PMID: 23650424; PubMed Central PMCID: PMC3731981.
52. Bhagirath D, Yang TL, Bucay N, et al. MicroRNA-1246 is an exosomal biomarker for aggressive prostate cancer. Cancer Res. 2018 Feb 1 PubMed PMID: 29437039. DOI:10.1158/0008-5472.CAN-17-2069.
53. Hsu SD, Tseng YT, Shrestha S, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucleic Acids Res. (Database issue) 2014 Jan;42:D78–85. PubMed PMID: 24304892; PubMed Central PMCID: PMC3965058.
54. Vergoulis T, Vlachos IS, Alexiou P, et al. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucleic Acids Res. (Database issue) 2012 Jan;40:D222–D229. PubMed PMID: 22135297; PubMed Central PMCID: PMC3245116.
55. Xiao F, Zuo Z, Cai G, et al. miRecords: an integrated resource for microRNA-target interactions. Nucleic Acids Res. (Database issue) 2009 Jan;37:D105–D110. PubMed PMID: 18996891; PubMed Central PMCID: PMC2686554.
56. Jiang Q, Wang Y, Hao Y, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. (Database issue) 2009 Jan;37:D98–D104. PubMed PMID: 18927107; PubMed Central PMCID: PMC2686559.
57. Gennarino VA, Sardiello M, Mutarelli M, et al. HOCTAR database: a unique resource for microRNA target prediction. Gene. 2011 Jul 1;480(1–2):51–58. PubMed PMID: 21435384; PubMed Central PMCID: PMC3113163.
58. Gamazon ER, Im HK, Duan S, et al. Exprtarget: an integrative approach to predicting human microRNA targets. PloS One. 2010;5(10):e13534. PubMed PMID: 20975837; PubMed Central PMCID: PMC2958831.
59. Li JH, Liu S, Zhou H, et al. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. Nucleic Acids Res. (Database issue) 2014 Jan;42:D92–D97. PubMed PMID: 24297251; PubMed Central PMCID: PMC3964941.
60. Vaquerizas JM, Kummerfeld SK, Teichmann SA, et al. A census of human transcription factors: function, expression and evolution. Nat Rev Genet. 2009 Apr;10(4):252–263. PubMed PMID: 19274049.
61. Wu JM, Vallenius T, Ovaska K, et al. Integrated network analysis platform for protein-protein interactions. Nat Methods. 2009 Jan 6;1:75–77. 10.1038/Nmeth.1282. PubMed PMID: WOS:000262370200028; English.
62. Stark C, Breitkreutz BJ, Chatr-Aryamontri A, et al. The bioGRID interaction database: 2011 update. Nucleic Acids Res. 2011 Jan;39:D698–D704. PubMed PMID: WOS:000285831700111; English.
63. Salwinski L, Miller CS, Smith AJ, et al. The database of interacting proteins: 2004 update. Nucleic Acids Res. 2004 Jan 1;32:D449–D451. PubMed PMID: WOS:000188079000106; English.

64. Prasad TSK, Goel R, Kandasamy K, et al. Human protein reference database-2009 update. Nucleic Acids Res. 2009 Jan;37:D767–D772. PubMed PMID: WOS:000261906200134; English.
65. Aranda B, Achuthan P, Alam-Faruque Y, et al. The IntAct molecular interaction database in 2010. Nucleic Acids Res. 2010 Jan;38: D525–D531. PubMed PMID: WOS:000276399100085; English.
66. Ceol A, Aryamontri AC, Licata L, et al. MINT, the molecular interaction database: 2009 update. Nucleic Acids Res. 2010 Jan;38: D532–D539. PubMed PMID: WOS:000276399100086; English.
67. Guldener U, Munsterkotter M, Oesterheld M, et al. MPact: the MIPS protein interaction resource on yeast. Nucleic Acids Res. 2006 Jan 1;34:D436–D441. PubMed PMID: WOS:000239307700096; English.
68. Zhang YE, Vibranovski MD, Landback P, et al. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. PLoS Biol. 2010 Oct 5;8(10):e1000494. PubMed PMID: 20957185; PubMed Central PMCID: PMC2950125.
69. Mudunuri U, Che A, Yi M, et al. bioDBnet: the biological database network. Bioinformatics. 2009 Feb 15;25(4):555–556. PubMed PMID: 19129209; PubMed Central PMCID: PMC2642638.
70. Li Y, Qiu C, Tu J, et al. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. Nucleic Acids Res. (Database issue) 2014 Jan;42:D1070–D1074. PubMed PMID: 24194601; PubMed Central PMCID: PMC3964961.