An Information Entropy-based Method to Detect microRNA Regulatory Module

YI YANG^{1,†1,a)} YAN SONG^{1,†1,b)} BUWEN CAO^{2,c)}

Received: September 27, 2018, Accepted: November 12, 2018

Abstract: The detection of miRNA regulatory modules (MRMs) can facilitate the analysis of miRNA combination effects in aberrant transcriptional regulatory networks. Existing methods suffer from stochastic or require a predetermined number of regulatory modules. We here develop a parameter-free computational framework named DeMine to predict MRMs. Briefly, DeMine is an information entropy-based method implemented by three steps. It first transforms miRNA regulatory network into a miRNA-miRNA synergistic network, and then detects miRNA clusters by maximizing the referred cluster entropy density. After that, the co-regulated mRNAs are added into corresponding clusters to form final MRMs. Compared with existing methods Mirsynergy and BCM on synthetic dataset and three real cancer datasets: ovarian cancer (OVCA), breast cancer (BRCA) and thyroid cancer (THCA) datasets, the proposed method can not only exhibit higher accuracy but also extract larger cohesiveness-preserved MRMs. More importantly, DeMine can find more miRNAs as tumor markers for the diagnosis of tumors.

Keywords: microRNA regulatory module, information entropy, ovarian cancer

1. Introduction

MicroRNAs (miRNAs) are a class of non-coding RNA molecules of 20–22 nucleotides [1]. Abnormal miRNA is involved in tumorigenesis by causing target mRNA degradation or post-translational repression [2], [3], [4], [5], [6]. Studies have shown that multiple miRNAs are likely to synergistically regulate one or more biological processes [7], [8], [9]. Thus, the identification of miRNA regulatory modules (MRMs) can help decipher miRNA synergism in cancer.

A number of relevant approaches have been proposed to study miRNA synergism. Several early approaches were primarily based on modularity Q-value maximization, such as: a contraction-dilation method based on greedy techniques [10], a method based on simulated annealing [11], a method based on spectral optimization [12]. Although these approaches are statistically sound, they suffer from either one-sidedness or time-consuming. Briefly, these algorithms focus on the network topology but ignore the biological information implied in the network. On the other hand, many algorithms take a lot of time to train parameters to obtain an optimal solution.

Some approaches beyond topological characteristics have been developed in the last decade. These approaches consider not only the topological characteristics of miRNA-mRNA networks but also the respective functional information of miRNAs and their target genes to discover miRNA regulatory mod-

ules. Typical cases such as SNMNMF (Sparse Network-regularized Multiple Non-negative Matrix Factorization) [13], PIMiM (Protein Interaction-based MicroRNA Modules) [14] and Mirsynergy [15], which all were developed to detect miRNA synergism. An overarching strategy of these approaches is to enhance the confidence of MRMs identification by integrating multiple types of genetic data. However, the first two methods require a predefined number of MRMs, which may be data-dependent, while the performance of the last method is sensitive to the weights of miRNA-miRNA and miRNA-mRNA network. In addition, these methods rely on multiple parameters to constrain the size or density of the MRMs.

More recently, Liang et al. [16] described a bi-cliques merging (BCM) framework, and claimed that BCM can identify densely connected and functionally enriched modules. To restrict the module assignments, however, BCM requires two predefined density thresholds, which may be data-dependent and difficult to determine beforehand. Similarly, a path-based model called MFSP developed by Ding et al. [17] also requires manual predefinition of multiple parameters.

In this paper, we develop a novel computational framework named DeMine (detecting MRMs based on information entropy) to predict MRMs. The method transforms miRNA regulatory network into a miRNA-miRNA synergistic network, and detects miRNA clusters by maximizing the entropy density of clusters in the synergistic network. The clustering process of DeMine can be operated without any required parameters. We applied DeMine to some synthetic networks as well as three real cancer datasets, and identified a number of miRNA regulatory modules for each dataset. In simulation, we found that the MRMs identified by DeMine exhibited higher accuracy than other two alternatives Mirsynergy and BCM. Tested on ovarian, breast and thyroid

¹ Department of Information Technology, Hunan Women's University

² College of Information and Electronic Engineering, Hunan City University

^{†1} Presently with Hunan Women's University

a) snryou@126.com

b) song0305@126.com

c) cbwchj@126.com

Fig. 1 The detail processes of transforming the miRNA regulatory network into a miRNA-miRNA synergistic network. This is a small-scale simulated miRNA regulatory network containing seven miRNAs (A – G) and six mRNAs (a – f). A total of twelve unit events are occurred in this network. Taking miRNA A and B as an example, the two miRNAs co-regulate only one gene a, thus the synergistic score between them is 1/12.

cancer datasets, we discovered that the MRMs detected by DeMine present more significantly functional enrichment, coherent expression and thus are likely to be more biologically meaningful than the counterparts. Through Kaplan-Meier survival analysis, we detected several MRMs with significant prognostic associations for the cancers.

2. Theories and Methods

We here develop a novel computational framework for identifying miRNA regulatory modules based on the principle of information entropy. The process of module identification can be divided into a three-step operation. We first transform the miRNA regulatory network into a miRNA-miRNA synergistic network. Then we put the most similar miRNAs into a cluster by maximizing the “cluster entropy density”, which is defined as the scoring function of miRNA clustering. Finally, we add corresponding target mRNAs to each miRNA cluster to complete miRNA regulatory module identification.

2.1 Network Transform

A miRNA regulatory network can be described as a bipartite network $G = (U, V, E)$, where U and V represent miRNA set and mRNA set, respectively, and $E = \{<u_i, v_j> | u_i \in U, v_j \in V, i = 1, 2, \dots, M, j = 1, 2, \dots, N\}$ represents the interactions between miRNAs and mRNAs. Notably, there is no connection between the same type of nodes. Applying information entropy to the miRNA regulatory network is to treat two miRNAs together regulating a mRNA as an event. Taking each miRNA pair as a unit, the probability of each unit event occurring is the ratio of the number of events involved in the two miRNAs to all events in the network. Based on this hypothesis, DeMine first converts

the miRNA regulatory network into a miRNA-miRNA synergistic network, where the corresponding probability value represents the synergistic score of each miRNA pair. **Figure 1** exhibits the detail processes of transformation.

The synergistic score of miR_a and miR_b is described as Eq. (1):

$$p_{miR_a} = \frac{\sum_{j=1}^N (miR_a \rightarrow miR_j \cap miR_b \rightarrow miR_j)}{\sum_{i,k=1}^M \sum_{j=1}^N (miR_i \rightarrow miR_j \cap miR_k \rightarrow miR_j)}, \quad (1)$$

where M and N represent the number of miRNAs and mRNAs in a miRNA-miRNA synergistic network, respectively. Moreover, we define “individual probability” as Eq. (2) for ranking miRNAs:

$$p(miR_a, miR_b) = \sum_{i=1}^{|C|} p(miR_a, miR_i), \quad (2)$$

where miR_a is a miRNA that belongs to a network or is connected to a cluster, $|C|$ is the number of miRNAs in a network ($|C| = M$) or a cluster ($|C| = 1, \dots, M$), and miR_i is the i -th miRNA in a cluster.

2.2 Systematic Identification of miRNA Clusters

To evaluate the synergy between miRNAs within a module, we use information entropy to define a function called “cluster entropy” [18]. The function is as follows:

$$H = \sum_{i=1}^{|C|} \sum_{j=1}^{|C|} H(miR_i, miR_j), \quad (3)$$

where $H(miR_i, miR_j)$ represents the entropy of synergy between miR_i and miR_j , which is defined as:

$$H(miR_i, miR_j) = -p(miR_i, miR_j) \log_2 p(miR_i, miR_j), \quad (4)$$

Table 1 The change of entropy in DeMine clustering process.

miRNA	PmiR($ C = 7$)	PmiR($ C = CCL $)	dH	mH	dH $\geq$ mH	CCL
A	2/12					
B	3/12					
C	4/12	1/12	0.30	0.44	N	
D	5/12	0.00	0.00	0.00	Y	{D}
E	4/12	2/12	0.43	0.22	Y	{D, E}
F	3/12	2/12	0.60	0.34	Y	{D, E, F}
G	3/12	3/12	0.90	0.48	Y	{D, E, F, G}

PmiR: the probability of a miRNA with the current cluster ($|C| = |CCL|$) or network ($|C| = 7$); dH/mH: the entropy increment and entropy density of the cluster; CCL: the current cluster.

Entropy increment $dH = H_2 - H_1$ indicates the change of entropy after adding a miRNA to a cluster. H_1 and H_2 are the entropies of the cluster before and after the addition, respectively.

In order to determine the miRNA assignments, we define a clustering scoring function called “cluster entropy density” to reflect the information entropy density of a cluster. The cluster entropy density is defined as:

$$mH = \frac{H}{|C|}, \quad (5)$$

which denotes the averaged co-regulate strength between miRNAs in a cluster. In the clustering process of DeMine algorithm, miRNAs are continuously added to a miRNA cluster based on the entropy density increment, making the cluster entropy density reaches the maximum. The cluster regulation density also reaches the maximum with the cluster entropy density. Therefore, final modules obtained by maximizing entropy density are more cohesiveness-preserved. In addition, whether miRNA is added to a cluster or not only depends on the inequation $dH \geq mH$. The process of cluster entropy density maximization needs no parameters. Hence, it does not like simulated annealing algorithm which needs to repeatedly adjust the parameters initial temperature, cooling factor, etc. [19].

With the above definition, we formulate the discovery of miRNA clusters as an iterative clustering problem. In the clustering process, we first take a candidate miRNA with the largest individual probability from the miRNA set U as the cluster center and then grow the cluster by iteratively adding neighbors until the cluster entropy density no longer increases. When a cluster is completed, we remove the miRNAs from the vector U . We iteratively repeat the above clustering procedure. The entire cluster process is terminated when there is no candidate miRNA in U .

Table 1 describes the clustering miRNA process in the synergistic network in Fig. 1 D. By the table, D with the largest individual probability (5/12) in the network is sought out as the first cluster center and the initial cluster D . The individual probabilities of nodes linking with the current cluster are 2/12 for E , 1/12 for C , F and G . So E is selected based on the principle of maximum individual probability. D and E are merged to form a new cluster D, E within the limitation $dH \geq mH$. Similarly, F and G are added to the cluster, but C fails to meet $dH \geq mH$. At this point, the first cluster is completed. After that, C is found as another cluster center and nodes C, B, A form another cluster.

Fig. 2 Adding mRNAs to miRNA clusters. (A) We gain two clusters in the network described in Fig. 1. Five candidate target genes are selected under the limitation of co-regulation by at least two miRNAs. (B) We add the corresponding candidate mRNAs to the two clusters and get the final MRMs. It is worth noting that mRNA c is successively added into the two miRNA clusters because of its importance as a bottleneck node.

2.3 Adding mRNAs to miRNA Clusters

Once the miRNA clusters are determined, we add the corresponding candidate target genes to the clusters to complete MRMs. The candidate target genes are the mRNAs that are co-regulated by at least two miRNAs in the cluster. Therefore, the end results are clustering assignments of MRMs. **Figure 2** illustrates the detail processes of adding mRNAs to miRNA clusters.

3. Description of Proposed Algorithm

Algorithm 3.1 describes our proposed DeMine algorithm. It is a three-step method that does not require predefined module Number or other threshold parameters. The algorithm only needs to input miRNA regulatory network data to get the final miRNA regulatory modules. It first gets the probability matrix of the network. Then it selects the node with the largest individual probability as a cluster center. After that, it puts the most similar miRNAs into a cluster by maximizing the cluster entropy density. Finally, DeMine adds corresponding target mRNAs to each miRNA cluster to get final MRMs.

4. Experiment Result and Analysis

In order to test the performance of the proposed algorithm, we carried out experiments on synthetic data and real network data. We compared detection accuracy of the proposed algorithm with two existing algorithms Mirsynergy and BCM on some simulated networks. Moreover, we compared the biological significance of the modules detected by three methods on three real cancer datasets.

4.1 Experiment on Synthetic Data Set

We first run three algorithms on 100 computer-generated random miRNA regulatory networks to compare their detection accuracy. A random miRNA network example is illustrated in **Fig. 3**. More random networks are provided in supplementary material Fig. S1. These networks consist of 64 miRNAs and 64 mRNAs. The average degree of each miRNA or mRNA in the networks is 9. Each of the networks can be divided into four modules and each module in them has 16 miRNAs and 16 mRNAs.

Algorithm 3.1 DeMine Algorithm.

Input: miRNA-mRNA regulatory network $G = (U, V, E)$
Output: final MRM set MC

```

1:  $C \leftarrow \emptyset; MC \leftarrow \emptyset$ 
2: ###Step 1: network transformation
3:  $p_{matrix} \leftarrow$  compute the synergistic score of each edge using Eq. (1)
4:  $Mn \leftarrow$  compute the individual probability of each node using Eq. (2)
5: ###Step 2: formation of miRNA clusters
6: while (TRUE) do
7:    $maxP \leftarrow$  select the maximum PmiR from  $Mn$ 
8:    $mc \leftarrow maxP; Mn \leftarrow Mn - \{maxP\}$ 
9:   while (TRUE) do
10:    ### $N_{maxP}$  are adjacent nodes of  $maxP$ 
11:     $Mn' \leftarrow$  compute probability of  $N_{maxP}$  to  $MC$ 
12:     $maxE \leftarrow$  select the maximum PmiR from  $Mn'$ 
13:    if ( $dH \geq mH$ ) then
14:       $mc \leftarrow mc \cup \{maxE\}$ 
15:       $Mn' \leftarrow Mn' - \{maxE\}$ 
16:    end if
17:  end while
18:   $C \leftarrow C \cup \{mc\}$ 
19: end while
20: ###Step 3: adding mRNAs to miRNA clusters
21: for ( $mc \in C$ ) do
22:   for ( $mr \in V$  and  $mr$  regulated by two miRNAs) do
23:     $mc \leftarrow merge(mr, mc)$ 
24:   end for
25:   $MC \leftarrow MC \cup \{mc\}$ 
26: end for
27: return  $MC$ ;
    
```

Fig. 3 A sample of synthetic miRNA regulatory networks.

There exist different amounts of connections between the modules, which leads to different difficulty of module identification. We used the networks as benchmark networks to compare clustering accuracy and stability of the three algorithms.

To describe the difficulty of module identification, we introduced two variables Z_{in} and Z_{out} . One is edge number between internal nodes while the other is edge number between internal and external nodes of a module. The larger Z_{out} values, the more difficult module identification. Based on different Z_{out} values,

Table 2 Mean μ and δ of clustering accuracy.

algorithm	μ	δ
DeMine	0.951	0.292
Mirsynergy	0.932	0.532
BCM	0.913	0.715

μ : mean value of clustering accuracy; δ : mean variance of clustering accuracy.

Fig. 4 Clustering performances of the three methods.

we compared the averaged detection accuracy of DeMine with Mirsynergy and BCM. Clustering accuracy is widely used for evaluate the clustering results [20], [21], which is defined as follows:

$$Acc = \frac{\sum_{i=1}^n \delta(map(c_i), l_i)}{n}, \quad (6)$$

where n is the total number of clusters, c_i the cluster label of x_i , and l_i denotes the true class label. The function δ yields one if $x = y$, zero otherwise. $map(c_i)$ is the mapping function that maps each cluster label c_i to the equivalent label from the benchmark networks. As shown in **Table 2**, DeMine gained the largest mean and the smallest variances of clustering accuracy among three methods.

It can also be clearly seen from **Fig. 4** that our proposed method is distinguished from the other two methods on detection accuracy. From Fig. 4 A, we can see DeMine has higher module detection accuracy than the counterparts at different Z_{out} values. On the other hand, the fluctuation of clustering accuracy of DeMine is significantly smaller than the other two methods (see Fig. 4 B). The result on the synthetic networks confirmed that DeMine is superior to Mirsynergy and BCM on the accuracy and stability of module detection.

4.2 Experiment on Three Cancer Datasets

We used three miRNA-mRNA regulatory networks from human cancers involving ovarian, breast and thyroid cancer datasets processed by the Mirsynergy authors (Li et al.) [15]. We worked with the datasets provided by these authors, in order to evaluate DeMine and fairly compare it with two existing MRM detection methods Mirsynergy and BCM that do the same. The obtained miRNA regulatory networks contain 559/688/696 miRNAs, 12456/9174/9759 mRNAs and 16365/47336/64286 miRNA-mRNA regulatory interactions, respectively. In the modules, the average numbers of miRNAs (mRNAs) are 5.88/2.97/2.97 (23.34/82.29/95.12), respectively.

4.2.1 Comparison of Overall Performance

Following the pipeline of DeMine, we obtained 32, 28, 58 MRMs in OVCA, BRCA and THCA datasets, respec-

Table 3 Overall module profiles of DeMine, BCM, and Mirsynergy.

Cancer	Method	Num	AmiR	AmR	Dens	MiFam	MiMEC
OVCA	DeMine	32	5.88	23.34	0.70	14	−0.40
	Mirsynergy	84	4.76	7.57	0.39	12	−0.37
	BCM	36	2.08	2.28	0.99	8	0.24
BRCA	DeMine	28	2.97	82.29	0.75	15	−0.16
	Mirsynergy	53	5.77	24.15	0.03	10	−0.08
	BCM	43	5.65	28.37	0.28	15	−0.15
THCA	DeMine	58	2.97	95.12	0.79	16	−0.14
	Mirsynergy	50	7.60	32.26	0.02	7	−0.08
	BCM	41	8.07	14.29	0.16	13	−0.11

Num: module number; AmiR and AmR: averaged miRNA and mRNA number per module; Dens: averaged density per module; MiFam: module number enriched in at least one miRNA family; MiMEC: averaged miRNA-mRNA expression correlation.

tively (Table 3, Supplementary Table S1). For Mirsynergy and BCM, they each detected 84/53/50 and 36/43/41 MRMs in OVCA/BRCA/THCA dataset, and the respective mean number of miRNAs (mRNAs) per module are 4.76/5.77/7.60 (7.57/24.15/32.26) and 2.08/5.65/8.07 (2.28/28.37/14.29), respectively. DeMine-MRMs are significantly more densely connected comparing with Mirsynergy-MRMs and BCM-MRMs on the datasets. BCM has higher average density than DeMine on OVCA, but BCM-modules contain very few miRNAs and mRNAs. A full list of the MRMs discovered by DeMine on the three datasets is provided in supplementary file Table S1.

4.2.2 Structure Analysis of MRMs

Averaged density values of three methods are 0.70/0.75/0.79, 0.39/0.03/0.02, and 0.99/0.28/0.16 on three datasets. Community characteristic analyses of MRMs can help us understand synergistic regulation of miRNAs to their target genes. We herein conduct comparative analysis on density of per MRM to present the connectivity between miRNAs and mRNAs in detail. As shown in the left side of Fig. 5, the density distributions of modules detected by the three methods are significantly different. Densities of modules from DeMine are clearly larger than Mirsynergy and BCM on BRCA and THCA, which indicates that DeMine-MRMs are significantly more densely connected than Mirsynergy-MRMs and BCM-MRMs. Although BCM got the largest averaged density on OVCA, only a small amount of miRNAs and mRNAs is present in BCM-MRMs (Table 3). This means that BCM is difficult to detect large-scale synergistic regulation among miRNAs on OVCA. We also noticed that DeMine-modules contain more mRNAs than Mirsynergy and BCM on the three datasets, but fewer miRNAs than the other two methods on BRCA and THCA datasets (Table 3). The cause of this phenomenon is that microRNAs have higher synergy values in denser networks and the rapidly maximization of cluster entropy density in the DeMine clustering process makes DeMine difficult to gain larger-scale miRNA clusters.

In addition, we also calculated MiMEC (MiRNA-mRNA Expression Correlation) [13] of the discovered modules to assess the regulation strength between miRNAs and mRNAs within the modules. As illustrated in the right side of Fig. 5, DeMine can identify more and smaller negative anti-correlation modules than

Fig. 5 The CDF diagram for the density and MiMEC of modules detected by three methods.

the other two methods on three datasets. The statistical result suggests that MiRNAs regulate mRNAs more strongly in the DeMine modules.

All of the above results demonstrate that DeMine-MRMs conform to the cooperative regulation between miRNAs and their genes. Detailed topological structure and expression correlation of each DeMine-MRM on OVCA are provided in the supplementary Fig. S2 and Fig. S3. Though DeMine needs to make further efforts to improve performance of identifying large-scale miRNA synergistic regulation in the dense networks, it has the ability to gain more critical miRNAs and mRNAs than the counterparts. The conclusion will be verified in the following sections.

4.2.3 MiRNA Synergy Analysis

Some works have shown that miRNAs from the same family are inclined to form a cluster to accomplish a specified regulatory function [22], [23], [24]. So we applied miRNA family enrichment analysis to assess miRNA synergy in DeMine-modules. Family classification of miRNA hairpin sequences file was downloaded from miRBase (<http://www.mirbase.org/>). With hypergeometric test (q -value < 0.05), we counted for each method the MRMs enriched in one or more miRNA families. Encouragingly, for OVCA, BRCA, and THCA, 14/15/16 of DeMine-MRMs are enriched in one or more miRNA families, while 12/10/7, 8/15/13 are found for Mirsynergy and BCM (Table 3). The MRMs detected by DeMine on three datasets are all enriched more miRNA families than their counterparts. Some DeMine-MRMs from OVCA enriched in miRNA families are presented in Table 4. We further explored the details of the DeMine-MRMs enriched in miRNA families on ovarian cancer dataset. For example, five of ten miRNAs in MRM 3 are from the same miRNA family MIPF0000020. Furthermore, the five miRNAs play significant roles in varieties of cancers in a coordinated man-

Table 4 Some DeMine-MRMs from OVCA enriched in miRNA families.

No.	MiRNA	MiRNA Family
3	miR-518f, miR-518b, miR-518c, miR-518e, miR-520e	MIPF0000020
4	miR-518f, miR-518b, miR-518c, miR-518e	MIPF0000020
5	miR-30a, miR-30d, miR-30e	MIPF0000005
24	miR-517b, miR-517a, miR-517c, miR-518f, miR-548y, miR-548j	MIPF0000020, MIPF0000317

No.: the index of MRM; MiRNA: miRNAs in a MRM; MiRNA Family: miRNA families enriched in a MRM.

Table 5 Onco-miRNA enrichment profile of three methods.

Method	OncoMiR	Recall	OncoMRM	AvgMiR
DeMine	102	100.00%	24 (75.00%)	4.25
Mirsynergy	56	54.90%	68 (80.95%)	0.67
BCM	25	24.51%	25 (69.44%)	1.00

ner [25], [26], [27]. Similarly, in MRM 5, three of eleven miRNAs are from the same miRNA family MIPF0000005, which inhibit or promote tumor growth in a synergistic manner [28], [29]. More information of DeMine-modules on OVCA enriched in miRNA families is provided in the supplementary file Table S2.

4.2.4 miRNA-disease Correlation Analysis

A great deal of evidence indicates that miRNAs can be used as tumor markers for the diagnosis of tumors [30], [31], [32]. Therefore we conduct a miRNA-disease correlation analysis for the MRMs detected by three methods on OVCA. We first downloaded several miRNA sets associated with ovarian cancer from the HDMM [33] and two review articles [34], [35] to construct an ovarian cancer related miRNA benchmark dataset composed of 102 miRNAs. Excitingly, all 102 miRNAs from the benchmark dataset are detected in the Demine-MRMs, while only 56 (54.90%) are found for Mirsynergy and 25 (24.51%) for BCM (Table 5). This ratio is extremely important and meaningful. In addition, although the proportions of onco-MRMs detected by three methods are all high, DeMine-modules contain more different onco-miRNAs. The average numbers of onco-miRNA (avg-MiR) in each module detected by three methods are 4.25, 0.67 and 1.00, respectively. Therefore, DeMine-MRMs containing more miRNA markers can help us diagnose tumors more accurately.

Figure 6 are the heatmaps of the average expression values of 32 DeMine-modules from OVCA and all the 11 miRNAs belonging to MRM 6. From both heatmaps, it is clear that the obtained DeMine-modules and miRNAs effectively discriminate cancer samples from normal samples.

The above analysis demonstrated that fewer DeMine-MRMs have more potential to be tumor markers as the diagnosis of cancer than Mirsynergy and BCM. More information of miRNA-disease correlation in modules detected on OVCA is provided in the supplementary file Table S3.

4.2.5 Survival Analysis

To verify the potential diagnostic ability of the detected MRMs, we divided the set of OVCA samples from TCGA clinical data into two degrees of association. Patients with higher miRNA expression level than the mean miRNA expression value were put

Fig. 6 Heatmaps of 32 DeMine-modules and 11 selected miRNAs within DeMine-MRM 6. (A) Similar expression patterns are found in cancer samples compared to normal samples based on the average expression values of miRNAs. (B) Samples are divided into cancer and normal based on 11 differentially expressed miRNAs in MRM 6.

Fig. 7 (A) and (B) are two Kaplan-Meier survival analyses using DeMine-MRM3 and DeMine-MRM4 from ovarian cancer, respectively. As seen in the figure, the samples with higher miRNA expression level in the DeMine-MRMs faced higher risk.

into high expression group (red curve), and the rest were classified as low expression group (green curve). We then conducted Kaplan-Meier method to compare the survival characteristics of the two groups. Kaplan-Meier survival analysis was performed on each DeMine-MRM.

As seen from Fig. 7, the risk groups of different miRNA expression levels are clearly separated from the rest. The patients with lower miRNA expression level in both modules were at lower risk. Oncogenes CXCL12, RBBP8 participated in potential prognostic DeMine-MRM 3, and onco-miRNAs miR-34, miR-450b-5p participated in DeMine-MRM 4 (see Fig.S2). Some studies on CXCL12 and RBBP8 suggest that they are associated with tumor cell migration and invasion [36], [37]. Some recent studies have shown that miR-34 plays an important role in the process of ovarian cancer [38], [39]. The experiment results confirmed that selected genes and miRNAs of MRMs are highly associated with patient survival. Kaplan-Meier survival analyses using each DeMine-MRM from OVCA are provided in the supplementary Fig. S4.

5. Conclusion

In this paper, we introduced a novel clustering algorithm named DeMine to solve MRM detection required parameter problems using information entropy density as a scoring function. We first transform miRNA regulatory network into a miRNA-miRNA synergistic network, by which the initial center selection strategy is designed to improve cluster accuracy. Then, we grow miRNA clusters by maximizing cluster entropy density to accelerate the convergence. Finally, we add the strongly correlated mRNAs into corresponding clusters to form final MRMs. The clustering process of DeMine is operated without any required parameters. The accuracy of DeMine was tested on some synthetic networks with known modular structures. The experiment results indicated that DeMine boasts much higher accuracy on module detection than the other two existing methods. We also tested DeMine on three authentic cancer networks and detected 32/28/58 MRMs. The biological comparative analysis of these MRMs detected by three methods demonstrated that DeMine is able to extract more cohesiveness-preserved and higher diagnostic value MRMs. To sum up, the performance of DeMine is good and cluster entropy density maximization can resolve the stochastic result or predetermined module number problem. In the future, the algorithm will be further enhanced to detect large-scale miRNA synergistic modules in dense networks.

Acknowledgments The research is funded by the “Twelfth Five-year” Education Science Plan Key Project of Hunan province of China (Grant No.XJK014AXX002), the Key Project of the Education Department of Hunan Province (Grant No.17A037) and the Natural Science Foundation of Hunan province, China (Grant No.2018JJ2024).

References

- [1] He, L. and Hannon, G.J.: MicroRNAs: Small RNAs with a big role in gene regulation, *J. Nat Rev Genet*, Vol.10, No.2, pp.522–531 (2004).
- [2] Spizzo, Riccardo, et al.: SnapShot: MicroRNAs in Cancer, *J. Cell*, Vol.137, No.3, pp.586–586 (2009).
- [3] Kozomara, A. and Griffiths-Jones, S.: miRBase: Annotating high confidence microRNAs using deep sequencing data, *Nucleic Acids Research*, Vol.42, Database issue, D68 (2014).
- [4] Friedman, R.C. et al.: Most mammalian mRNAs are conserved targets of microRNAs, *Genome Research*, Vol.19, No.1, pp.92–105 (2008).
- [5] Papadopoulos, G.L. et al.: DIANA-mirPath: Integrating human and mouse microRNAs in pathways, *Bioinformatics*, Vol.25, No.15, pp.1991–1993 (2009).
- [6] Tsang, J.S., Ebert, M.S. and Van, O.A.: Genome-wide dissection of microRNA functions and cotargeting networks using gene set signatures, *Molecular Cell*, Vol.38, No.1, pp.140–153 (2010).
- [7] Boross, G., Orosz, K. and Farkas, I.J.: Human microRNAs co-silence in well-separated groups and have different predicted essentialities, *Bioinformatics*, Vol.25, No.8, pp.1063–1069 (2009).
- [8] Krek, A. et al.: Combinatorial microRNA target predictions, *Nature Genetics*, Vol.37, No.5, pp.495–500 (2005).
- [9] Xu, J. et al.: MiRNACmiRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features, *Nucleic Acids Research*, Vol.39, No.3, pp.825–836 (2011).
- [10] Mei, J. et al.: Revealing network communities through modularity maximization by a contraction-dilation method, *New Journal of Physics*, Vol.11, No.4, 043025 (2009).
- [11] Guimera, R., Salespardo, M. and Amaral, L.A.N.: Module identification in bipartite networks with applications to directed networks, *Physical Review E*, Vol.76, No.3 (2007).
- [12] Richardson, T., Mucha, P.J. and Porter, M.A.: Spectral tripartitioning of networks, *Physical Review E Statistical Nonlinear & Soft Matter Physics*, Vol.80, No.3, p.036111 (2009).
- [13] Zhang, S. et al.: 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.1401–1409 (2011).
- [14] Le, H.S. and Barjoseph, Z.: Integrating sequence, expression and interaction data to determine condition-specific miRNA regulation, *Bioinformatics*, Vol.29, No.13, pp.89–97 (2013).
- [15] Li, Y. et al.: Mirsynergy: Detecting synergistic miRNA regulatory modules by overlapping neighbourhood expansion, *Bioinformatics*, Vol.30, No.18, p.2627 (2014).
- [16] Liang, C., Li, Y. and Luo, J.: A Novel Method to Detect Functional microRNA Regulatory Modules by Biclques Merging, *IEEE/ACM Trans. Computational Biology & Bioinformatics*, Vol.13, No.3, pp.549–556 (2016).
- [17] Ding, P. et al.: A path-based measurement for human miRNA functional similarities using miRNA-disease associations, *Scientific Reports*, Vol.6, p.32533 (2016).
- [18] Eagle, N., Macy, M. and Claxton, R.: Network Diversity and Economic Development, *Science*, Vol.328, No.5981, pp.1029–1031 (2010).
- [19] Fortunato, S.: Community detection in graphs, *Physics Reports*, Vol.486, No.3, pp.75–174 (2009).
- [20] Xu, W., Liu, X. and Gong, Y.: Document clustering based on non-negative matrix factorization, *DBLP*, pp.267–273 (2003).
- [21] Gu, Q., Ding, C. and Han, J.: On trivial solution and scale transfer problems in graph regularized NMF, *International Joint Conference on Artificial Intelligence, AAAI Press*, pp.1288–1293 (2011).
- [22] Johnson, S.M. et al.: RAS Is Regulated by the let-7 MicroRNA Family, *Cell*, Vol.120, No.5, pp.635–647 (2005).
- [23] Lai, E.C.: Predicting and validating microRNA targets, *Genome Biology*, Vol.5, No.9, p.115 (2004).
- [24] Xu, J., Li, C.-X., Li, Y.-S., et al.: MiRNA-miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features, *Nucleic Acids Research*, Vol.39, No.3, pp.825–836 (2011).
- [25] Cherni, I. et al.: Abstract 171: Pre-miR-518b and pre-miR-598, novel serum biomarkers of de novo chemoresistance in advanced or unresectable NSCLC, *Cancer Research*, Vol.71, No.8, p.171, Supplement (2011).
- [26] Kinouchi, M. et al.: Involvement of miR-518c-5p to Growth and Metastasis in Oral Cancer, *Plos One*, Vol.9, No.12, p.e115936 (2014).
- [27] Rubie, C. et al.: Chemokine receptor CCR6 expression is regulated by miR-518a-5p in colorectal cancer cells, *Journal of Translational Medicine*, Vol.12, No.1, p.48 (2014).
- [28] Zhao, Y. et al.: The combined use of miRNAs and mRNAs as biomarkers for the diagnosis of papillary thyroid carcinoma, *International Journal of Molecular Medicine*, Vol.36, No.4, p.1097 (2015).
- [29] Son, J.W. et al.: MicroRNA Expression Profiles in Korean Non-Small Cell Lung Cancer, *Tuberculosis & Respiratory Diseases*, Vol.67, No.5, p.413 (2009).
- [30] Ahmed, F.E. et al.: Diagnostic MicroRNA Markers for Screening Sporadic Human Colon Cancer and Active Ulcerative Colitis in Stool and Tissue, *Cancer Genomics Proteomics*, Vol.6, No.5, pp.281–295 (2009).

- [31] Patterson, E.E. et al.: MicroRNA Profiling of Adrenocortical Tumors Reveals miR-483 as a Marker of Malignancy, *Molecular Cell*, Vol.117, No.8, pp.1630–1639 (2011).
- [32] Liu, J. et al.: MicroRNA-100 is a potential molecular marker of non-small cell lung cancer and functions as a tumor suppressor by targeting polo-like kinase 1, *BMC Cancer*, Vol.12, No.1, p.519 (2012).
- [33] Maagh, P. et al.: Pulmonary Vein Isolation in 2012: Is It Necessary to Perform a Time Consuming Electrophysical Mapping or Should We Focus on Rapid and Safe Therapies? A Retrospective Analysis of Different Ablation Tools, *International Journal of Medical Sciences*, Vol.10, No.1, pp.24–33 (2013).
- [34] Katz, B. et al.: MicroRNAs in Ovarian Cancer, *Human Pathology*, Vol.46, No.9, p.1245 (2015).
- [35] Langhe, R.: microRNA and Ovarian Cancer. microRNA: Cancer, *Springer International Publishing*, pp.1161–1169 (2015).
- [36] Sha, W. et al.: Silencing of CXCL12 performs a protective effect on C5b-9-induced injury in podocytes, *International Urology & Nephrology*, pp.1–10 (2018).
- [37] Soriabretones, I. et al.: Prognostic value of CtIP/RBBP8 expression in breast cancer, *Cancer Medicine*, Vol.2, No.6, p.774 (2013).
- [38] Li, R. et al.: MiR-34a suppresses ovarian cancer proliferation and motility by targeting AXL, *Tumor Biology*, Vol.36, No.9, pp.7277–7283 (2015).
- [39] Schmid, G. et al.: Expression and promotor hypermethylation of miR-34a in the various histological subtypes of ovarian cancer, *Bmc Cancer*, Vol.16, No.1, p.102 (2016).

ing.

Yi Yang is an associate professor of Hunan Womens University. He received his M.S. degree from Wuhan University of Technology in 2006. He began researching computational biology in 2014 while he was a visiting scholar at Hunan University. His current research interests are biological network analysis and data min-

Yan Song was born in 1983. She received her M.S. degree from Hunan Norm University in 2006. Currently, she is a lecturer of Hunan Womens University. She has published some research papers in various journals and proceedings of conferences. Her research interest is community detection in complex networks.

Appendix

A.1 Availability and Implementation

A.1.1 Source Code

The source code files for DeMine is available at <https://github.com/snryou/DeMine>.

A.1.2 Supplementary file.doc

In the file Supplementary_file.doc, there are some figures about random miRNA regulatory networks (Fig. S1), topological structure of DeMine-MRMs (Fig. S2), the scatter plot of the DeMine-MRMs expression correlations (Fig. S3) and survival analysis using DeMine-MRMs (Fig. S4).

A.1.3 Table S1_modules.xls

In the file Table S1_modules.xls, there are three worksheets which include detail information of modules detected by DeMine on the three datasets.

A.1.4 Table S2_miRNA family enrichment analysis.xls

In the file Table S2_miRNA family enrichment analysis.xls, there are three worksheets which include miRNA family enrichment analysis data of DeMine modules.

A.1.5 Table S3_MRMs enriched onco miRNAs.xls

In the file Table S3_MRMs enriched onco miRNAs.xls, there are three worksheets which include onco miRNAs enriched in MRMs detected by the three methods on ovarian cancer dataset.

Buwen cao was born in 1978. He received his D.E degree from Hunan University in 2016. Currently, he is a lecturer of Hunan City University. He has published more than ten research papers in various journals and proceedings of conferences. His research interests are biological network analysis and data mining.

(Communicated by *Susumu Goto*)