

RESEARCH ARTICLE

Predicting microRNA-disease association based on microRNA structural and functional similarity network

Tao Ding^{1,2}, Jie Gao^{1,*}, Shanshan Zhu¹, Junhua Xu¹, Min Wu¹

¹ School of Science, Jiangnan University, Wuxi 214122, China

² School of Mathematics Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK

* Correspondence: gaojie@jiangnan.edu.cn

Received August 30, 2018; Accepted October 31, 2018

Background: Increasing evidences indicate that microRNAs (miRNAs) are functionally related to the development and progression of various human diseases. Inferring disease-related miRNAs can be helpful in promoting disease biomarker detection for the treatment, diagnosis, and prevention of complex diseases.

Methods: To improve the prediction accuracy of miRNA-disease association and capture more potential disease-related miRNAs, we constructed a precise miRNA global similarity network (MSFSN) via calculating the miRNA similarity based on secondary structures, families, and functions.

Results: We tested the network on the classical algorithms: WBSMDA and RWRMDA through the method of leave-one-out cross-validation. Eventually, AUCs of 0.8212 and 0.9657 are obtained, respectively. Also, the proposed MSFSN is applied to three cancers for breast neoplasms, hepatocellular carcinoma, and prostate neoplasms. Consequently, 82%, 76%, and 82% of the top 50 potential miRNAs for these diseases are respectively validated by the miRNA-disease associations database miR2Disease and oncomiRDB.

Conclusion: Therefore, MSFSN provides a novel miRNA similarity network combining precise function network with global structure network of miRNAs to predict the associations between miRNAs and diseases in various models.

Keywords: miRNAs; hairpin structure; miRNA families; functional similarity; disease semantic; leave-one-out cross-validation

Author summary: microRNAs (miRNAs) are functionally related to the development and progression of various human diseases. To improve the prediction accuracy of miRNA-disease association and capture more potential disease-related miRNAs, we constructed a precise miRNA global similarity network via calculating the miRNA similarity of secondary structures, families, and functions. The novel miRNA similarity network combining precise function network with global structure network of miRNAs showed the better performance compared with others similarity network and could be used to predict the associations between miRNAs and diseases in various models.

INTRODUCTION

MicroRNAs are small length (~22 nt) non-coding RNAs that regulate the expression of a target messenger RNAs (mRNAs) by base pairing with their 3'-UTRs, and triggering their translational repression or degradation [1,2]. Accumulated studies have shown that miRNAs play

critical roles in various fundamental biological processes, such as cell development, proliferation, differentiation, apoptosis and signal transduction[3,4]. Not surprisingly, the dysregulation of miRNAs is obviously associated with the development and progression of complex diseases, such as schizophrenia, cardiovascular diseases and cancers [5].

It is crucial to discover the mechanism through which the regulatory mechanism of miRNA exerts its functions. It is usually expensive and time-consuming to identify the associations between miRNAs and diseases via experimental methods. Considering that a large number of datasets (miRNA functional similarity, disease semantic similarity, disease phenotype similarity, and miRNA-disease associations) are available, computational models are efficient in predicting disease-related miRNAs in that they can select the most promising associated miRNAs for further experimental disease studies. So, more and more researchers made further efforts and developed efficient computational models (HDMP [6], RWRMDA [7], WBSMDA [8], HGIMDA [9], RLSMDA [10], MCMDA [11], *etc.*) to infer the potential miRNA-disease associations.

Most of the known models were developed under the basic assumption that functionally similar miRNAs are regarded to be involved in similar diseases and vice versa [12,13]. So, the prediction performance of these methods is strongly relying on the miRNA functional similarity network. When the biological functions of miRNAs are unidentified, the similarity of miRNAs would not be calculated by the methods so that they cannot be applied to different computational models to predict related diseases. More specifically, only the disease-related miRNAs are used in these models so that the models are too weak to capture the disease-related miRNAs that are newly detected or uncovered their functions. Other methods such as Net-CBI [14], MCMDA [11] can work for diseases without any known associated miRNAs, they have achieved unsatisfactory performances.

To overcome these limitations, we try to make full use of the existing miRNA information to construct a global miRNA similarity network. A striking feature of these miRNAs is that their loci are usually clustering in the genome [15,16]. More in-depth studies, including miRNA co-expression and primary transcript identification, suggest that the clusters of proximal miRNAs are typically expressed as polycistronic, coregulated units and shared common biological functions [16]. Meanwhile, miRNA gene families are highly conserved in nucleotide sequence and secondary structure among closely related species over evolutionary time [17,18]. Notably, miRNAs with similar sequences or highly conserved secondary structures tend to play roles in the similar biological process [2,19,20].

Taken together, we proposed a novel measurement to construct a precise miRNA global network by calculating the similarity of miRNA nucleotide sequence, secondary structures, families, and their functions. Note that calculating the similarity value between any two miRNAs that will be extremely powerful and convenient by constructing their structural similarity, especially for

new detected miRNAs or miRNAs whose functions are unknown. Considering that factors, we applied new miRNA similarity network (miRNA structural and functional similarity network, MSFSN) into the models of WBSMDA and RWRMDA. In WBSMDA, Chen *et al.* [8] integrated known miRNA-disease associations, miRNA functional similarity, disease semantic similarity, and Gaussian interaction profile kernel similarity for miRNAs and diseases so that they can infer the associations of disease-related miRNAs. WBSMDA makes a breakthrough in that it is successful in predicting related miRNAs for new diseases without known related miRNAs and new miRNAs without known related diseases. Also, Chen *et al.* [7] proposed a RWRMDA approach to predict new human miRNA-disease associations by adopting the method of random walk on miRNA functional similarity network. RWRMDA is motivated based on the investigation that global similarity measures are better in predicting the associations between miRNAs and diseases. Through employing the new network into these two models, the exciting results, including leave-one-out cross-validation and case studies, show that MSFSN makes a breakthrough in that it succeeds in introducing the miRNA structure into miRNA similarity network and getting a superior performance to miRNA functional similarity network.

RESULTS

Leave-one-out cross-validation

Leave-one-out cross-validation (LOOCV) on known miRNA-disease associations is applied to evaluate the prediction performance of MSFSN in various models. Briefly, for each given disease d , one known disease-related miRNA is omitted, whereas WBSMDA/RWRMDA model is trained on the remaining candidate miRNAs to rank them. The prediction probability is then calculated based on the model prediction for the left-out miRNAs. This is repeated until every miRNA is left out once and the generate prediction probability values are then used for receiver-operating characteristics (ROC) analysis.

By varying the threshold, the true positive rate (TPR, sensitivity) and the false positive rate (FPR, 1-specificity) are calculated to obtain the ROC curves. TPR refers to the percentage of the test miRNA-disease associations which are ranked higher than the given threshold. And FPR refers to the percentage of negative miRNA-disease pairs below the threshold. The area under ROC curve (AUC) can be calculated to evaluate the ability of MSFSN. If $AUC = 1$, it indicates that the MSFSN has a perfect performance in algorithms.

To avoid the impact of the different data sets on results,

we unify miRNAs and diseases sets and obtain 5222 miRNA–disease associations, including 485 miRNAs and 325 diseases from HMDD2. Figure 1 illustrates the ROC curves of two models after LOOCV.

The performance comparisons in the framework of LOOCV are shown in Figure 1. As a result, WBSMDA and RWRMDA gain AUCs of 0.8171 and 0.9414 in LOOCV on MFSN, respectively. And for MSFSN, AUCs are 0.8212 and 0.9657, respectively. The results show that the performance of MSFSN is superior to MFSN's in both WBSMDA and RWRMDA. It is greatly demonstrated the MSFSN has a more reliable and valid prediction performance on miRNA–disease associations than MFSN.

Based on the assumptions that miRNAs with similar sequences or highly conserved secondary structures tend to play roles in the similar biological process. Especially, the miRNAs belonged to a family share similar biological functions. As for any newly found miRNAs which their functions or disease-related are still uncovered, we can calculate the similarity value by miRNA nucleotide sequences, hairpin structures, and families to predict their related-diseases. In addition, combining with the accurate MFSN, new network can be used to improve the prediction accuracy of miRNA–disease associations and uncover more disease-related miRNAs by bioinformatics.

Case studies

From the results, MSFSN has a superior performance of RWRMDA to WBSMDA. Three cancers (breast neoplasms, hepatocellular carcinoma, and prostate neoplasms) are presented here to evaluate the prediction ability of MSFSN through RWRMDA. The predicted results are validated by another two major miRNA–disease association databases, miR2Disease (updated in 2009) [21] and oncomiRDB (updated in 2014) [22], these two validation data sets are entirely independent of data sets used for prediction.

We present a case study for breast neoplasms, which is one of the most commonly occurring cancers among women and accounts for 22% of all female cancers [23]. We rank all miRNAs related to breast neoplasms. Among the top 50 predicted breast neoplasms-related miRNAs in MSFSN, 41 miRNAs have been confirmed to be associated with breast neoplasms by miR2Disease or oncomiRDB. However, only 37 miRNAs are related to breast neoplasms in MFSN. We respectively list the top 50 miRNAs and evidence for their associations with breast neoplasms in Table 1 (MSFSN) and Table 2 (MFSN). Recent literatures showed that the unconfirmed potential miRNA, hsa-mir-139, will may represent a useful signature for the identification of high-risk breast

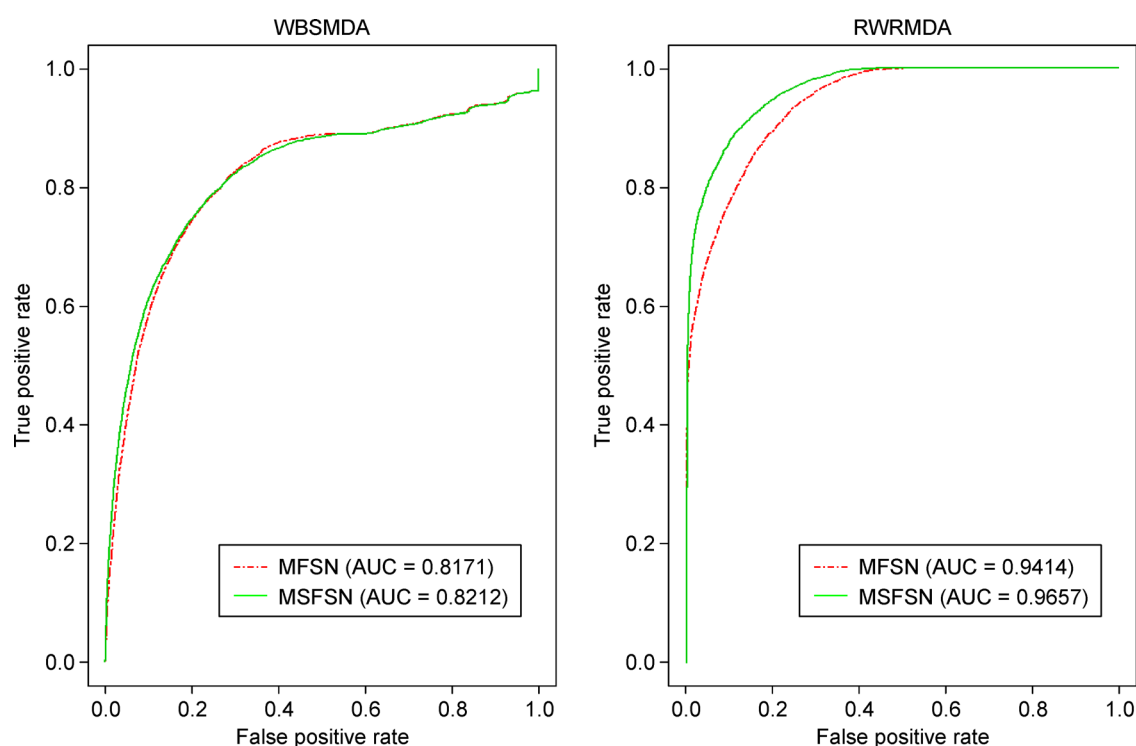


Figure 1. Performance comparisons between miRNA–disease association prediction models (WBSMDA and RWRMDA) concerning ROC curve and AUC based on MFSN and MSFSN for LOOCV, respectively.

Table 1 Prediction of the top 50 miRNAs which are associated with breast neoplasms based on MSFSN

miRNA	Evidence	miRNA	Evidence
hsa-mir-218	oncomiRDB	hsa-mir-141	miR2Disease; oncomiRDB
hsa-mir-200a	miR2Disease; oncomiRDB	hsa-let-7g	oncomiRDB
hsa-mir-10b	miR2Disease; oncomiRDB	hsa-mir-222	miR2Disease; oncomiRDB
hsa-mir-200c	miR2Disease; oncomiRDB	hsa-let-7a	miR2Disease; oncomiRDB
hsa-mir-135b	Unconfirmed	hsa-mir-497	miR2Disease; oncomiRDB
hsa-mir-200b	miR2Disease; oncomiRDB	hsa-mir-199b	oncomiRDB
hsa-mir-25	oncomiRDB	hsa-mir-151a	Unconfirmed
hsa-mir-302c	oncomiRDB	hsa-mir-34a	oncomiRDB
hsa-mir-93	oncomiRDB	hsa-mir-146b	miR2Disease; oncomiRDB
hsa-mir-205	miR2Disease; oncomiRDB	hsa-mir-34b	Unconfirmed
hsa-mir-135a	oncomiRDB	hsa-mir-148a	miR2Disease; oncomiRDB
hsa-mir-302b	Unconfirmed	hsa-mir-193b	miR2Disease; oncomiRDB
hsa-mir-107	Unconfirmed	hsa-mir-191	miR2Disease
hsa-mir-182	miR2Disease; oncomiRDB	hsa-mir-203	miR2Disease; oncomiRDB
hsa-mir-22	miR2Disease; oncomiRDB	hsa-mir-9	oncomiRDB
hsa-mir-101	oncomiRDB	hsa-mir-139	Unconfirmed
hsa-mir-27b	oncomiRDB	hsa-mir-30c	oncomiRDB
hsa-mir-224	oncomiRDB	hsa-mir-125b	miR2Disease; oncomiRDB
hsa-mir-125a	miR2Disease; oncomiRDB	hsa-mir-152	miR2Disease; oncomiRDB
hsa-mir-127	miR2Disease; oncomiRDB	hsa-mir-106b	oncomiRDB
hsa-mir-221	miR2Disease; oncomiRDB	hsa-mir-365b	Unconfirmed
hsa-mir-27a	miR2Disease; oncomiRDB	hsa-mir-195	miR2Disease; oncomiRDB
hsa-mir-124	oncomiRDB	hsa-mir-302d	Unconfirmed
hsa-let-7d	miR2Disease	hsa-mir-338	Unconfirmed
hsa-mir-183	oncomiRDB	hsa-mir-145	miR2Disease; oncomiRDB

The first column records top 1–25 related miRNAs. The second column records the top 26–50 related miRNAs.

cancer patients [24]. Moreover, miR-17 is associated with lymph node metastasis and receptor status of breast cancer patients [25]. All the datasets used in this paper are generated before the publication of this paper. Therefore, this successful independent literature validation gives a further strong support to the reliable performance demonstration of MSFSN. We further list the associations between other unconfirmed potential miRNAs. The newest miRNA-disease association databases deEMC2 [26] (updated in 2017) and HMDD V3.0 (updated in 2018) have been verified that all such miRNAs from new network are related to breast cancer.

Owing to the space limitations, we only present the predicted results for breast neoplasms, and implement two case studies for hepatocellular carcinoma and prostate neoplasms. For hepatocellular cancer, 38 miRNAs have been confirmed by various databases. Compared with MFSN, the percentage of disease-related miRNAs has increased from 72% to 76% in MSFSN (Supplementary

Table S1). And as for MSFSN, 41 out of the top 50 miRNAs that are related to prostate neoplasms are confirmed by the miR2Disease and oncomiRDB database, but MFSN only for miRNAs are verified (Supplementary Table S2). These favorable predicted examples sufficiently demonstrate that new miRNA similarity network would be of great significance to capture the potential relationships between miRNAs and diseases.

DISCUSSION

Increasing evidences indicate that miRNAs are closely related to the development and progression of various human diseases [5,27]. Under the basic assumption that functionally similar miRNAs are regarded to be involved in similar diseases and vice versa, researchers propose several computational models to uncover the new potential miRNA-disease associations by integrating

Table 2 Prediction of the top 50 miRNAs which are associated with breast neoplasms based on MFSN

miRNA	Evidence	miRNA	Evidence
hsa-mir-218	oncomiRDB	hsa-mir-224	oncomiRDB
hsa-mir-10b	miR2Disease; oncomiRDB	hsa-mir-107	Unconfirmed
hsa-mir-135a	oncomiRDB	hsa-mir-139	Unconfirmed
hsa-mir-200c	miR2Disease; oncomiRDB	hsa-let-7a	miR2Disease; oncomiRDB
hsa-mir-135b	Unconfirmed	hsa-mir-486	Unconfirmed
hsa-mir-183	oncomiRDB	hsa-mir-145	miR2Disease; oncomiRDB
hsa-mir-205	miR2Disease; oncomiRDB	hsa-mir-148a	miR2Disease; oncomiRDB
hsa-mir-27a	miR2Disease; oncomiRDB	hsa-mir-151a	Unconfirmed
hsa-mir-375	oncomiRDB	hsa-let-7d	miR2Disease
hsa-mir-25	oncomiRDB	hsa-mir-191	miR2Disease
hsa-mir-22	miR2Disease; oncomiRDB	hsa-mir-31	miR2Disease; oncomiRDB
hsa-mir-101	oncomiRDB	hsa-mir-296	Unconfirmed
hsa-mir-302b	Unconfirmed	hsa-mir-96	miR2Disease; oncomiRDB
hsa-mir-18b	oncomiRDB	hsa-mir-199b	oncomiRDB
hsa-mir-9	oncomiRDB	hsa-mir-193a	oncomiRDB
hsa-mir-127	miR2Disease; oncomiRDB	hsa-mir-143	miR2Disease; oncomiRDB
hsa-mir-200a	miR2Disease; oncomiRDB	hsa-mir-629	Unconfirmed
hsa-mir-182	miR2Disease; oncomiRDB	hsa-mir-100	Unconfirmed
hsa-mir-93	oncomiRDB	hsa-mir-193b	miR2Disease; oncomiRDB
hsa-mir-34b	Unconfirmed	hsa-mir-146b	miR2Disease; oncomiRDB
hsa-mir-200b	miR2Disease; oncomiRDB	hsa-mir-215	Unconfirmed
hsa-mir-124	oncomiRDB	hsa-mir-203	miR2Disease; oncomiRDB
hsa-mir-497	miR2Disease; oncomiRDB	hsa-mir-126	miR2Disease; oncomiRDB
hsa-mir-302c	oncomiRDB	hsa-mir-625	Unconfirmed
hsa-mir-34c	Unconfirmed	hsa-mir-125a	miR2Disease; oncomiRDB

The first column records top 1–25 related miRNAs. The second column records the top 26–50 related miRNAs.

miRNA functional similarity, disease similarity, and known miRNA-disease associations. For all of these models, the most critical step is to construct a miRNA similarity network. And with a precise similarity network, more disease-related miRNAs will be ranked among the candidate miRNAs. So, improving the prediction accuracy of miRNA-disease associations will be our primary concern. Besides, previous prediction models of miRNA-disease association cannot uncover the potential diseases associated with new miRNAs without any known related diseases.

So based on the hypothesis that (i) miRNA gene families are highly conserved in the sequence of seed region and hairpin structure among closely related species over evolutionary time; (ii) the majority of miRNA clusters is transcribed as a single unit and shared common biological functions; (iii) miRNAs with conserved

structures tend to play roles in the similar biological process; (iv) functionally similar miRNAs are regarded to be involved in similar diseases, we proposed a novel measurement to construct miRNA similarity global network according to miRNA sequences, structures, families, and functions. The MSFSN obtains reliable AUCs of 0.8212 and 0.9657 in the validation schema of LOOCV by WBSMDA and RWRMDA, respectively. Furthermore, we implemented simulated case studies for three important human diseases: breast neoplasms, hepatocellular carcinoma and prostate neoplasms on RWRMDA by using MSFSN. 82%, 76%, and 82% of top 50 potential miRNAs for these three important diseases are respectively confirmed miRNA-disease association databases by miR2Disease and oncomiRDB. It is anticipated that MSFSN would be a useful resource to improve the prediction accuracy of miRNA-disease

associations in various models and predict diseases associated with newly discovered miRNAs without any known related diseases.

The reasons why MSFSN achieves better performances are as follows. Firstly, we construct a global miRNA similarity network by using all of miRNAs nucleotide sequence, hairpin structures, families, and functions. The novel miRNA similarity network would offer lots of assistance in inferring more disease-related miRNAs. Besides, we did not abandon the miRNA functional similarity and constructed a more precise miRNA global similarity network based on structural and functional similarity of miRNA. Practical experiments show that the novel miRNA similarity network has a superior prediction performance on miRNA-disease associations. Further development of exploiting miRNA structure information to improve the precision of MSFSN may be needed for our future studies.

MATERIALS AND METHODS

Human miRNA-disease associations

The known miRNA-disease associations were downloaded from Human MicroRNA Disease Database (HMDD version 2.0 in September 2013) [28]. After integrating the name of miRNAs and diseases, we obtained 5222 miRNA-disease associations, including 485 miRNAs and 325 diseases. The adjacency matrix A is denoted as miRNA-disease associations, where the entity $A(m(i), d(j))$ is 1 if miRNA $m(i)$ is related to the disease $d(j)$, otherwise 0.

miRNA functional similarity network

In previous work, miRNA functional similarity score was calculated based on the assumption that functionally similar miRNAs tend to be associated with semantically similar diseases [12,13]. Therefore, Wang *et al.* [29] estimated functional similarity of two miRNAs by measuring the semantic similarity of their associated diseases. Based on the data sets from HMDD 2.0 and the method in Ref. [22], we recalculate the miRNA functional similarity score for each miRNA pair and retrieve miRNA functional similarity network $MFSN(m(i), m(j))$. The entity $MFSN(m(i), m(j))$ is the similarity score between miRNA $m(i)$ and $m(j)$.

miRNA structural and functional similarity network

This paper was aimed at constructing a global miRNA structural similarity network primarily. In addition, concerning the powerful miRNA functional similarity

network structured by former researchers, we further combine MFSN with miRNA structural similarity network and achieve a more precise global miRNA similarity network.

Nucleotide sequence and hairpin structure

A detailed analysis of miRNA gene expression showed that the coexpression of closely clustered miRNAs are generated as polycistronic primary transcripts [30]. And miRNAs are transcribed as long primary transcripts (pri-miRNAs) that are first trimmed into the hairpin intermediates (pre-miRNAs, with distinctive fold-back hairpin structures), and subsequently cleaved into mature miRNAs (containing identical seed sequences and interaction with target mRNAs) [17,30]. So, the sequences should be phylogenetically conserved in the precursor hairpin, particularly in the mature miRNA segment. In addition, conservation of hairpin structure in spite of sequence variation implies that the structure may be functionally important.

Considering the miRNA structural specificity (hairpin structure) and its functional seed region, we downloaded 485 human pre-miRNAs with their hairpin structures and nucleotide sequences from miRBase (version 21 in July 2014) [31] to construct their structure and sequence similarity. According to the previous works, for any pre-miRNAs, they can be represented as the brackets (“(“or”)”) and dots (“.”) [32]. The left bracket “(“ means that the paired nucleotide is located near the 5'-end. Similarly, the paired nucleotide near the 3'-end is expressed as right bracket “)”. And the unpaired nucleotide is replaced by dots “.”. Therefore, all the miRNA structures can be represented by “bracket and dot notation” chains. Then beginning from the 5'-terminal of miRNA chains, we scanned each base along the chain till its last base and substitute its upper case in bold italic for “(“ and the upper case for “)” in this base pair. Or else, the lower case is adopted. Figure 2 has showed the transferring processes from miRNA structures to linear strings.

Eventually, we obtained the linear strings of miRNA sequences and structures, and calculate its probability by following formula:

$$P_{a_i a_j} = \begin{cases} \frac{n_{a_i a_j}}{\sum_{k=1}^{12} n_{a_i a_k}}, & \text{if } \sum_{k=1}^{12} n_{a_i a_k} \neq 0 \\ 0, & \text{if } \sum_{k=1}^{12} n_{a_i a_k} = 0 \end{cases}, \quad i, j = 1, 2, \dots, 12, \quad (1)$$

where a_i represents the i -th element of $\{A, U, C, G, a, u, c,$

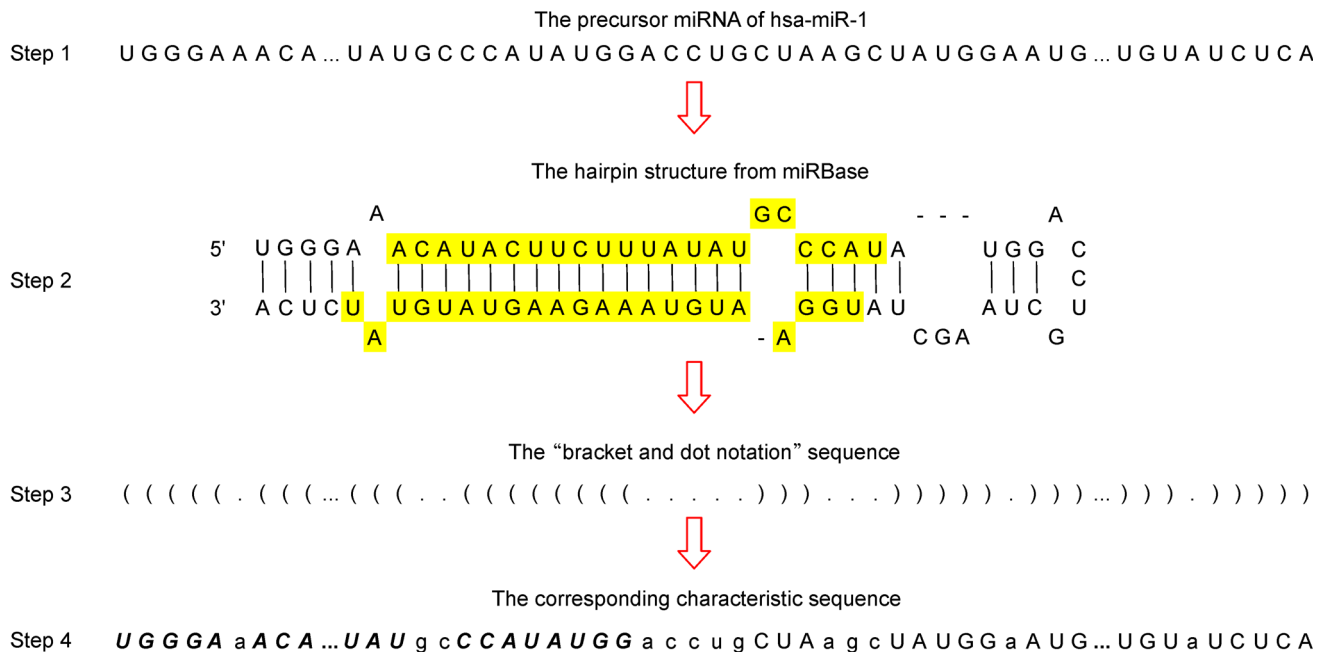


Figure 2. The basic idea of transforming complex nucleotide sequences and hairpin structures to linear strings. Because of the space limitations, here we have omitted some bases or brackets in Step 1, 3, and 4. The “...” is used to replace the missing bases or brackets. The nucleotide sequences in yellow shadow are the mature miRNAs (hsa-miR-1-5p and hsa-miR-1-3p).

g, A, U, C, G}; $n_{a_i a_j}$ represents the occurrence frequency of the event that base a_i is followed by base a_j and $P_{a_i a_j}$ is the occurrence frequency of the event that base a_i is followed by base a_j .

By arranging all these probabilities into a 144-dimensional vector TPV [33], the sequence and structure similarity score between miRNA $m(i)$ and $m(j)$ is defined as follow:

$$SS(m(i), m(j)) = \arccos \left(\frac{(TPV_i)(TPV_j)}{\|TPV_i\| \|TPV_j\|} \right). \quad (2)$$

miRNA families

It is reported that the members of miRNA family or clusters are more likely to associate with the similar diseases [29]. So, 310 miRNA families, containing 485

$$MSFSN = \begin{cases} \lambda MSSN + (1-\lambda)MFSN & \text{if } MFSN(m(i),m(j)) \neq 0, \\ MSSN & \text{if } MFSN(m(i),m(j)) = 0 \end{cases}, \quad (4)$$

where the entity $MSFSN(m(i), m(j))$ reflects the similarity value of miRNA $m(i)$ and $m(j)$ based on the method in [22], and λ is the weight parameter. We assigned the different weights in the Equation (4) for novel miRNA similarity network and calculate their AUCs by Leave-

miRNAs and derived from miRBase 21, will be considered into the novel miRNA similarity network. Adjacency matrix $SF(m(i), m(j))$ is defined to represent known miRNA-miRNA family relationships. If miRNA $m(i)$ and miRNA $m(j)$ belong to the same family, the entity $SF(m(i), m(j))$ is 1, otherwise 0.

After taking together, we obtain the miRNA structural similarity network $MSSN(m(i), m(j))$.

$$MSSN(m(i), m(j)) = \frac{SS(m(i), m(j)) + SF(m(i), m(j))}{2}. \quad (3)$$

Eventually, a novel miRNA similarity network based on miRNA structural and functional similarity is denoted as:

one-out cross-validation. In this paper, we chose $\lambda=0.5$ in the final network considering the outstanding performance of miRNA global similarity network. Figure 3 shows the main methodology of this paper proposed.

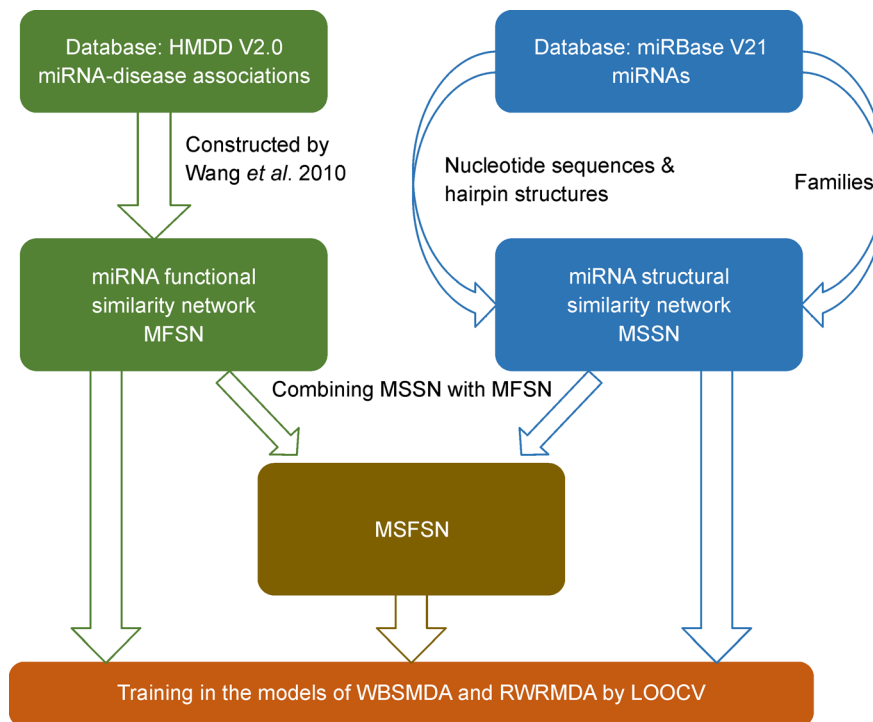


Figure 3. The overall structure of proposed system. This flowchart is a brief description of this paper's work. Constructing the miRNA structural similarity network and combining miRNA functional similarity network, we train the novel network in different models to show the prediction performance of miRNA-disease associations.

SUPPLEMENTARY MATERIALS

The supplementary materials can be found online with this article at <https://doi.org/10.1007/s40484-019-0170-0>.

ACKNOWLEDGEMENTS

This research is supported by Major Research Plan of National Natural Science Foundation of China (No. 91730301), Key Projects of National Natural Science Foundation of China (No.11831015) and the State Scholarship Fund of China (No. 201806790020).

COMPLIANCE WITH ETHICS GUIDELINES

The authors Tao Ding, Jie Gao, Shanshan Zhu, Junhua Xu and Min Wu declare that they have no conflict of interests.

This article does not contain any studies with human or animal subjects performed by any of the authors.

REFERENCES

- Bartel, D. P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116, 281–297
- Bartel, D. P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, 136, 215–233
- Karp, X. and Ambros, V. (2005) Encountering microRNAs in cell fate signaling. *Science*, 310, 1288–1289
- Miska, E. A. (2005) How microRNAs control cell division, differentiation and death. *Curr. Opin. Genet. Dev.*, 15, 563–568
- Mendell, J. T. and Olson, E. N. (2012) MicroRNAs in stress signaling and human disease. *Cell*, 148, 1172–1187
- Xuan, P., Han, K., Guo, M., Guo, Y., Li, J., Ding, J., Liu, Y., Dai, Q., Li, J., Teng, Z., *et al.* (2013) Prediction of microRNAs associated with human diseases based on weighted k most similar neighbors. *PLoS One*, 8, e70204
- Chen, X., Liu, M. X. and Yan, G. Y. (2012) RWRMDA: predicting novel human microRNA-disease associations. *Mol. Biosyst.*, 8, 2792–2798
- Chen, X., Yan, C. C., Zhang, X., You, Z. H., Deng, L., Liu, Y., Zhang, Y. and Dai, Q. (2016) WBSMDA: within and between score for miRNA-disease association prediction. *Sci. Rep.*, 6, 21106
- Chen, X., Yan, C. C., Zhang, X., You, Z. H., Huang, Y. A. and Yan, G. Y. (2016) HGIMDA: heterogeneous graph inference for miRNA-disease association prediction. *Oncotarget*, 7, 65257–65269
- Chen, X. and Yan, G. Y. (2014) Semi-supervised learning for potential human microRNA-disease associations inference. *Sci. Rep.*, 4, 5501
- Li, J. Q., Rong, Z. H., Chen, X., Yan, G. Y. and You, Z. H. (2017) MCMMA: matrix completion for miRNA-disease association prediction. *Oncotarget*, 8, 21187–21199
- Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W. and Cui,

- Q. (2008) An analysis of human microRNA and disease associations. *PLoS One*, 3, e3420
13. Bandyopadhyay, S., Mitra, R., Maulik, U. and Zhang, M. Q. (2010) Development of the human cancer microRNA network. *Silence*, 1, 6
 14. Chen, H. and Zhang, Z. (2013) Similarity-based methods for potential human microRNA-disease association prediction. *BMC Med. Genomics*, 6, 12
 15. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA*, 102, 15545–15550
 16. Baskerville, S. and Bartel, D. P. (2005) Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. *RNA*, 11, 241–247
 17. Kaczkowski, B., Torarinsson, E., Reiche, K., Havgaard, J. H., Stadler, P. F. and Gorodkin, J. (2009) Structural profiles of human miRNA families from pairwise clustering. *Bioinformatics*, 25, 291–294
 18. Ding, T., Gao, J. (2016) Prediction of human miRNA functions by gene cluster discriminant analysis. *Chinese Journal of Cell Biology* 12, 1467–1472
 19. Griffiths-Jones, S. (2010) miRBase: microRNA sequences and annotation. *Curr. Protoc. Bioinformatics*, Chapter 12, 12.9.1–10
 20. Washietl, S., Hofacker, I. L., Lukasser, M., Hüttenhofer, A. and Stadler, P. F. (2005) Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotechnol.*, 23, 1383–1390
 21. Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G. and Liu, Y. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, 37, D98–D104
 22. Khurana, R., Verma, V. K., Rawoof, A., Tiwari, S., Nair, R. A., Mahidhara, G., Idris, M. M., Clarke, A. R. and Kumar, L. D. (2014) OncomiRdbB: a comprehensive database of microRNAs and their targets in breast cancer. *BMC Bioinformatics*, 15, 15
 23. Zhang, W., Zeng, T., Liu, X. and Chen, L. (2015) Diagnosing phenotypes of single-sample individuals by edge biomarkers. *J. Mol. Cell Biol.*, 7, 231–241
 24. Rask, L., Balslev, E., Søkilde, R., Høgdaal, E., Flyger, H., Eriksen, J. and Litman, T. (2014) Differential expression of miR-139, miR-486 and miR-21 in breast cancer patients sub-classified according to lymph node status. *Cell Oncol. (Dordr.)*, 37, 215–227
 25. Stückerath, I., Rack, B., Janni, W., Jäger, B., Pantel, K. and Schwarzenbach, H. (2015) Aberrant plasma levels of circulating miR-16, miR-107, miR-130a and miR-146a are associated with lymph node metastasis and receptor status of breast cancer patients. *Oncotarget*, 6, 13387–13401
 26. Yang, Z., Wu, L., Wang, A., Tang, W., Zhao, Y., Zhao, H. and Teschendorff, A. E. (2017) dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res.*, 45, D812–D818
 27. Meola, N., Gennarino, V. A. and Banfi, S. (2009) microRNAs and genetic diseases. *PathoGenetics*, 2, 7
 28. Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T. and Cui, Q. (2014) HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.*, 42, D1070–D1074
 29. Wang, D., Wang, J., Lu, M., Song, F. and Cui, Q. (2010) Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics*, 26, 1644–1650
 30. Kim, V. N. (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.*, 6, 376–385
 31. Griffiths-Jones, S., Saini, H. K., van Dongen, S. and Enright, A. J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, 36, D154–D158
 32. Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, 125, 167–188
 33. Han, S. Q., Pei, Z. L., Wang, Q. F. and Shi, X. H. (2007). A characteristic sequences and normalized euclidean distance based method for RNA secondary structures comparison. In: *The International Conference on Bioinformatics and Biomedical Engineering*, 222–225