

MIPDH: A Novel Computational Model for Predicting microRNA–mRNA Interactions by DeepWalk on a Heterogeneous Network

Leon Wong, Zhu-Hong You,* Zhen-Hao Guo, Hai-Cheng Yi, Zhan-Heng Chen, and Mei-Yuan Cao



Cite This: *ACS Omega* 2020, 5, 17022–17032



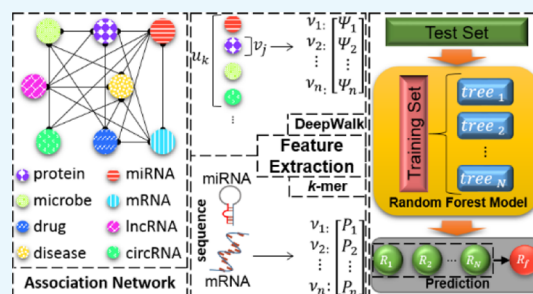
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: Analysis of miRNA–target mRNA interaction (MTI) is of crucial significance in discovering new target candidates for miRNAs. However, the biological experiments for identifying MTIs have a high false positive rate and are high-priced, time-consuming, and arduous. It is an urgent task to develop effective computational approaches to enhance the investigation of miRNA–target mRNA relationships. In this study, a novel method called MIPDH is developed for miRNA–mRNA interaction prediction by using DeepWalk on a heterogeneous network. More specifically, MIPDH extracts two kinds of features, in which a biological behavior feature is learned using a network embedding algorithm on a constructed heterogeneous network derived from 17 kinds of associations among drug, disease, and 6 kinds of biomolecules, and the attribute feature is learned using the k -mer method on sequences of miRNAs and target mRNAs. Then, a random forest classifier is trained on the features combined with the biological behavior feature and attribute feature. When implementing a 5-fold cross-validation experiment, MIPDH achieved an average accuracy, sensitivity, specificity and AUC of 75.85, 74.37, 77.33%, and 0.8044, respectively. To further evaluate the performance of MIPDH, other classifiers and feature descriptors are conducted for comparisons. MIPDH can achieve a better performance. Additionally, case studies on hsa-miR-106b-5p, hsa-let-7d-5p, and hsa-let-7e-5p are also implemented. As a result, 14, 9, and 9 out of the top 15 targets that interacted with these miRNAs were verified using the experimental literature or other databases. All these prediction results indicate that MIPDH is an effective method for predicting miRNA–target mRNA interactions.



1. INTRODUCTION

With the development of biology promoted by bioinformatics, the central dogma of molecular biology has been found to be hard to incorporate the advancement of the pattern of coding information in life activity for its complexity.^{1,2} However, the coding exons of protein-coding genes in the human genome account for 2% of the genome considering untranslated regions (UTRs). As one of the main classes of non-coding RNAs (ncRNA) proved to involve in disease as well as normal development and physiology, microRNA(miRNA), small ncRNA of ~22 nucleotides, are gaining prominence because increasing evidences suggest that there are not less than 60% of protein-coding genes regulated in translation.³ They can act as master regulators of a process and regulators for specific individual targets. The expression of genes can be regulated simultaneously by key miRNAs, and miRNA–targets can be regulated cooperatively by many kinds of miRNAs. Sequence complementarity among miRNAs and their mRNA targets is one of the most important activities, in which the RNA-induced silencing complex (RISC) guided to mRNAs by miRNAs can cause inhibition of protein translation and mRNA degradation.⁴

It is reported that there are some distinct targeting patterns of miRNAs in animals and plants. miRNA in plants has nearly

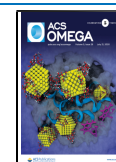
perfect sequence complementarity with target mRNA, forming the cleavage of double-stranded RNA.⁵ In animals, partial sequence complementarity among miRNAs and their target mRNAs makes the mechanisms far more sophisticated than in plants. The miRNA seed region (6–8 nt) plays an essential role in target regulation. More and more studies suggest that human miRNA target prediction is a complex task that resulted from a key mechanism that target mRNA is repressed for the interaction between its 3'UTR and the 5' end of corresponding miRNA.⁶

Biogenetical disorder in miRNA or misexpression regulated by miRNA can result in human diseases.⁷ The involvement of miRNA has been ascertained in nearly all sort of cellular pathways such as immunity, cardiovascular disease, carcinogenesis, normal development, diabetic nephropathy, atherosclerosis, and diabetes. It is reported that clinical application

Received: December 8, 2019

Accepted: March 6, 2020

Published: July 9, 2020



and therapeutic application with respect to miRNA has been investigated for its wide contribution in disease processes and gene regulation.⁸ More and more studies show that miRNAs are involved in many targets and found in many important biological processes and cellular pathways. Some specific miRNAs have been explored as therapeutic agents in clinical trials. For example, Nico et al. conducted the first in-human study on miR-16 mimic for the treatment of malignant pleural mesothelioma.⁹ Yan et al. investigated into the treatment for scleroderma by targeting miR-155 that functions as upregulation in skin tissues.¹⁰ Wang et al. proposed a novel therapeutic strategy in hepatocellular carcinoma, in which exosome miR-335 was identified to downregulate its target mRNAs after treatment.¹¹ Zhu et al. pointed out that a therapeutic strategy for modulating multidrug resistant in cancer cells lines A2780DX5 and KB-V1 is made by targeting miR-451 and miR-27a that are related to activating the expression of the P-glycoprotein.¹² In addition, many miRNAs have found to be the vital factors and therapeutic target in human diseases, such as miR-122 in liver biology,¹³ miR-29 in cardiac fibrosis,¹⁴ and miR-433 and miR-127 in gastric cancer.¹⁵ Such understanding on miRNA regulation in pathophysiology can usher in the new biomarker discovery and therapeutic advancement.

Many principles have been incorporated in miRNA target prediction algorithms to reach more effective performance, such as seed sequence complementarity, G-U wobble, free energy, target-site accessibility, target-site abundance, evolutionary conservation status, local AU flanking content, and pattern-based approach.^{16–23} Most of the prediction tools using seed sequence complementarity focus on 3'UTR of target genes, but it is not effective for studies on human genome for its poor characterized portions in 3'UTR. Therefore, it will suffer from the high false-negative rate. Free energy is the metric to assess the binding stability between miRNA and its target mRNA by computing the accessibility and hybridization with the mRNA secondary structure.²³ The predicted pair of miRNA and mRNA with free energy at a low level is regarded to interact with each other more likely. Massive computing performance is required to analyze various folding patterns of mRNA. G-U wobble is the controversial principle that a U nucleotide is allowed to pair a G nucleotide that originally pair with a C nucleotide.²² Wang suggest that such sequence-based imperfections to the target may be involved in the biological function.²⁴ Target-site accessibility is to assess how easy miRNA binds to its mRNA target. To a certain extent, the binding ability of miRNA will be weakened owing to target mRNA with a secondary structure after transcription.^{19,25} In miRNA-target prediction using such principle, the true-positive rate can be increased without considering the targets with lower conservation. Otherwise, taking the conservation regions that are not the substantial targets into account will result in a low true-positive rate.²⁶ However, the research suggests that less than 70% of the target sites confirmed by experiments are conserved. As for the conserved miRNA targets, about 30% of them have no effect on expression of their targets.²⁷

Before diverse miRNA-target prediction tools were expanded, identification of the miRNA-target was carried out using high-cost, time-consuming and labor-intensive biological approaches such as the immunoprecipitation of RISC components, RNA ligase mediated-5' rapid identification of cDNA ends (5' RLM-RACE), gene expression analysis, and luciferase reporter gene assay.²⁸

The increasing miRNAs have been found and their regulation functions on their targets have been also identified via manual bioexperiments. However, the number of the existing miRNA-target interactions identified by bioexperiments is still limited. It is an urgent work to identify such vital and valid interactions. An individual miRNA has many corresponding target genes, but it is a high costing task to validate the interactions between them. Owing to the rapid development of computer science, bioinformatics has been boosted using computer-aided technologies. The computational prediction approach becomes an important supplement to biological experimental approaches and accelerates the identification process because the potential candidates with a high confidence value can be yielded by efficient prediction algorithms with a lower cost so that it can reduce the false-positive rate in experiments.^{28–35}

There are many exiting miRNA-target prediction tools that are developed for researches, which allow them to choose the specific tools according to the corresponding research preferences. Kertesz et al. proposed a miRNA target prediction tool called probability of interaction by target accessibility (PITA) that assesses the binding site accessibility by computing the energy changes between free energy obtained from the forming miRNA-target complexes and energy cost of unpairing the target for miRNA accessibility.²⁵ PITA additionally combines target-site abundance, seed sequence complementarity, and G-U wobble. Still, a limitation of PITA should be noticed is that it mainly focuses on the 3'UTR. A pattern-based method named RNA22, not limited to the 3'UTR, can offer the most potential candidate list that has been not confirmed by experiments yet.²⁰ Other than these, there are many other prediction tools based on web techniques. miRDB is a miRNA target database and offers information of functional annotations as well as online prediction services that can return the prediction results based on their sequences submitted by users.³⁶ The prediction model called MirTarget is used to analyze the high-throughput data based on current studies on CLIP-RNA ligation sequencing.³⁷ In addition, functional annotation of miRNAs of miRDB is developed by combining many principles, including 3' compensatory pairing, seed sequence complementarity, target-site accessibility, local AU content, free energy, evolutionary conservation, and machine learning. 568 functional annotations in humans were discovered through the great effort on analyzing and arranging the literature. The STarMir webserver is a tool developed by using a logistic prediction model based on miRNA binding information obtained from CLIP researches.³⁸ Like miRDB, users can get the prediction results of binding sites by submitting miRNA sequences and target mRNA sequences. STarMir assesses the confidence of seed sites or nonseed sites in the coding sequence, the 3'UTR and 5'UTR of an mRNA, considering 3' compensatory pairing, seed sequence complementarity, local AU content, free energy, target-site accessibility, and G-U wobble.

Although many exiting miRNA-target prediction tools have been developed, more effort is still need to be made to boost the development of life science from all aspects that can be exerted for prediction. Recently, Huang et al. investigated the potential associations between lncRNA and miRNA using a graph-based method named EPLMI that used lncRNA–miRNA interactions and expression profile-based similarities of lncRNA and miRNA.³⁹ Qu et al. proposed to predict

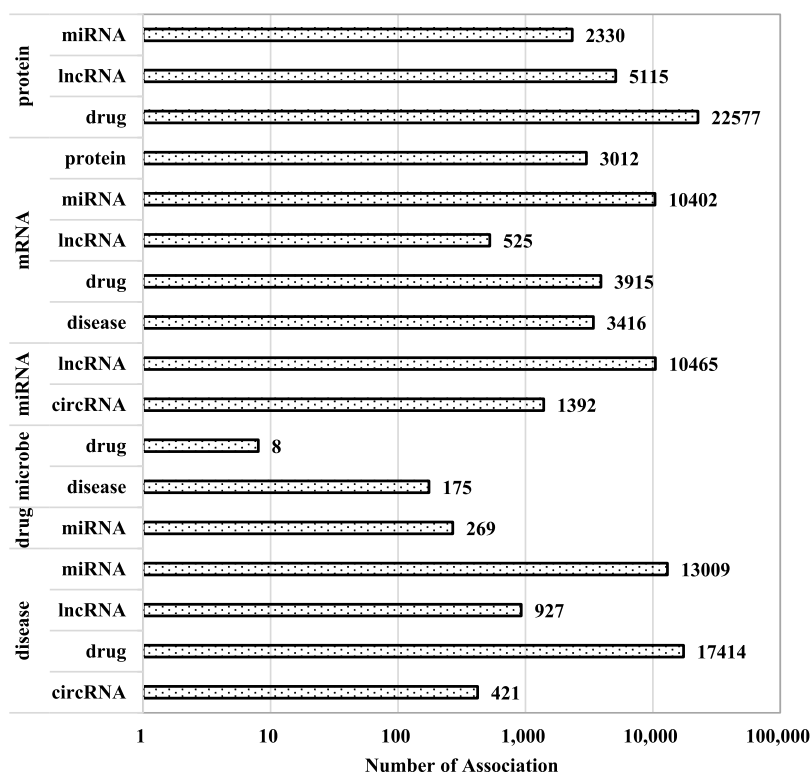


Figure 1. Quantity distribution of biological molecule associations.

associations between small molecules and miRNAs based on a triple layer heterogeneous network constructed by integrating similarities of small molecules, miRNAs, and disease.⁴⁰ Wang et al. employed a novel logistic model tree to detect the associations between miRNA and disease and multiresource bioinformation, including known association, disease semantics as well as similarities of the miRNA sequence and functional information, was obtained by feature fusion.³² Liang et al. introduced a novel model to identify miRNA-disease association based on a constructed heterogeneous network that consisted of known miRNA-disease associations and miRNA-target associations.⁴¹ Guo et al. proposed a novel method to construct a molecular association network among lncRNA, miRNA, disease, drug and protein, and predict the associations between any two nodes in the network.⁴²

In this study, we proposed a computational approach named MIPDH for miRNA-mRNA interaction prediction using DeepWalk on a heterogeneous network. In detail, to integrally learn the feature from 17 kinds of association, the DeepWalk method was introduced to represent latent features of each node in the constructed network using local information generated from truncated random walks.⁴³ the *k*-mer method was also employed for the final integrated feature that can help to yield a better performance. In addition, a random forest (RF) classifier was trained for predicting task. In order to obtain performance evaluation with the least error, 5-fold cross validation was implemented. We also compared our proposed method with different classifiers and feature descriptors. As results of 5-fold cross validation implemented on our proposed method, the best performance among the comparison experiments achieved average accuracy, sensitivity, precision, Matthews correlation coefficient (MCC), specificity, Matthews correlation coefficient (MCC) and AUC of 75.85 ± 0.63 , 74.37 ± 0.86 , 76.66 ± 1.00 , $77.33 \pm 1.34\%$, 0.6335 ± 0.0066 ,

and 0.8044 ± 0.0078 , respectively. Further, case studies performed on hsa-miR-106b-5p, hsa-let-7d-5p and hsa-let-7e-5p, and the top 15 predictions (14 of 15 in hsa-miR-106b-5p, 9 of 15 in hsa-let-7d-5p, and 9 of 15 in hsa-let-7e-5p) were verified from the PubMed literature and other databases. The prediction results yielded by MIPDH indicated that our proposed method can be applied to predict miRNA-target interactions on a large scale, which facilitates future miRNA biomarker discovery and the development of new therapy.

2. MATERIALS

In this work, the molecular association network is constructed by 17 kinds of association that are collected from 18 databases and the NCBI website, in which 2425 miRNA, 2730 lncRNA, 421 circRNA, 3024 mRNA, 4900 protein, 100 microbe, 5861 drug, and 1314 disease are involved. After data processing, a gold standard dataset comprising 90,171 associations and 20,775 nodes are obtained from 18 databases and the NCBI website (see Figure 1).^{44–61} In addition, sequence information of miRNA and mRNA is downloaded from NCBI using the Biopython package.

3. METHODOLOGY

3.1. DeepWalk for Network Embedding. The DeepWalk algorithm, as a widely-used network embedding algorithm, was employed on the whole molecular association network to vectorize the vertices (e.g., ncRNAs, diseases, proteins, and drugs). Given a network G with a set of vertices V and a set of edge E , that is $G = (V, E)$. The algorithm has two main points—RandomWalk and SkipGram. By performing RandomWalk on each vertex in V , a set of short walk-path is conducted to storage the local community information. By employing SkipGram, considering such a walk-path sequence as a word sequence is to update the representation of each

vertex in the network. The algorithm description is as Algorithm 1.

Denote a random walk-path sequence S_{v_i} rooted at vertex v_i using RandomWalk. Then each S_{v_i} is taken as input for SkipGram. By a configured parameter w as a windows size, the context vertices of v_i are denoted as $C(v_i) = (v_{i-w}, \dots, v_{i+w})/v_i$. The optimized objective function of the algorithm is defined as follow

$$\Pr(C(v_i)|\Psi(v_i)) = \prod_{\substack{j=i-k \\ j \neq i}}^{i+k} \Pr(v_j|\Psi(v_i)) \quad (1)$$

where $\Psi(v_i) \in \mathbb{R}^d$ is the representation of the vertex v_i . To solve feasibly, the Hierarchical Softmax algorithm is employed (see Algorithm 2). As the vertices are assigned to the leaves of a binary tree, the definition of the objective function is updated as following

$$\Pr(u_k|\Psi(v_i)) = \prod_{l=1}^{\log|V|} \Pr(b_l|\Psi(v_i)) \quad (2)$$

where u_k denotes a sequence of the tree node b_l rooted at b_0 . As a binary classifier is assigned to the parent of the node b_l , $\Pr(b_l|\Psi(v_i))$ can be defined as following

$$\Pr(b_l|\Psi(v_i)) = 1/(1 + e^{-\Psi(v_i) \cdot \Phi(b_l)}) \quad (3)$$

where $\Phi(b_l) \in \mathbb{R}^d$ is the representation of the vertex b_l 's parent.

Algorithm 1 DeepWalk(G, w, d, η, t)

Input: graph $G(V, E)$

windows size w

representation dimension d

epoch η

step size t

Output: representation $\Psi \in \mathbb{R}^{|V| \times d}$

1: Initialize Ψ

2: Build a binary Tree T

3: **for** $i = 0$ to η **do**

4: $V' = \text{Shuffle}(V)$

5: **for each** $v_i \in V'$ **do**

6: $S_{v_i} = \text{RandomWalk}(G, v_i, t)$

7: $\text{SkipGram}(\Psi, S_{v_i}, w)$

8: **end for**

9: **end for**

Algorithm 2 SkipGram(Ψ, S_{v_i}, w)

1: **for each** $v_j \in S_{v_i}$ **do**

2: **for each** $u_k \in S_{v_i}[j - w: j + w]$ **do**

3: $J(\Psi) = -\log \Pr(u_k|\Psi(v_j))$

4: $\Psi = \Psi - \alpha * \frac{\partial J}{\partial \Psi}$

5: **end for**

6: **end for**

3.2. Feature Extraction of RNA by Using k -mer. The k -Mer algorithm, one of the widely used feature extraction methods for the sequence, encodes a sequence information into a numeric feature vector that can be available as an input of the classifier.

Given a sequence with length L composed of 4 kinds of nucleotides of mRNA (i.e., A, T, C, and G), the possible subsequence can be enumerated using permutation and combination. Setting a specific value to the parameter k , there may be 4^k subsequences obtained from a sequence. Note

that the sliding window size is k . To the subsequence set $\{\alpha_1, \alpha_2, \dots, \alpha_4^k\}$, the number of each α_i , denoted as $n(\alpha_i)$, is computed by traversing the sequence to find a subsequence α_i with a sliding window of size k . After traversal, $n(\alpha_1, \alpha_2, \dots, \alpha_4^k)$ is obtained. Finally, the k -mer feature vector can be calculated as $n(\alpha_1, \alpha_2, \dots, \alpha_4^k)/(L - k + 1)$. For example, in the mRNA sequence "GGTGGTGCCTCAGCCATGGC", $p("G") = 8/20$, $p("CC") = 2/(20 - 2 + 1)$ and $p("GCC") = 2/(20 - 3 + 1)$.

To the miRNA sequence, nucleotides are different from mRNA (i.e., A, U, C, and G).

3.3. Constructing RF for Classification. The RF classifier, proposed by Breiman, is one of the classical classifiers that are widely used in machine learning and exiting studies on miRNA-target interaction prediction. To assemble a RF classifier, a specific number of decision trees that are trained, in which the individual decision trees is constructed with a subset of whole samples yielded using the bootstrap algorithm as well as the corresponding feature vectors selected randomly from all features. For classification, the final result is yielded by voting according to the decision tree results.

There are two import parameters in model training—the number of trees and dimension of the subset of the feature. In detail, to grow an unpruned decision tree, the training set is a two-dimensional matrix generated by randomly selecting about two-third of samples as well as user-defined dimension of feature that is far less than the whole set. There are many kinds of decision trees to ensemble a RF classifier, such as ID3, C4.5, and CART. To different types of decision trees, different methods are employed to split the node by one of the subset features that is with the smallest impurity. Additionally, out-of-bag data (OOB) that is the rest of the samples excluding the selected training samples, is used for generating OOB estimate of an error rate that is an internal unbiased estimate of the generalization error.

4. RESULTS

4.1. Performance Evaluation Using 5-Fold Cross Validation. For the purpose of the performance assessment, some widely used performance evaluation indicators that each of them measured different aspects of performance were made for all experiments. In measurement of the classification problem, a confusion matrix is constructed to illustrate how the test set is predicted correctly and incorrectly for each of the categories. Based on the confusion matrix, evaluation metrics introduced in our work include: accuracy, sensitivity, specificity, precision, and MCC. In detail, accuracy denotes the rate of overall correct prediction; sensitivity, also called the true positive rate (TPR), is the percentage of the positive actual samples predicted correctly; specificity, also called the true negative rate, is the percentage of the negative actual samples predicted correctly; precision, also called the positive predictive value, denotes the percentage of the positive predictive samples predicted correctly. MCC is the correlation coefficient between predictive and actual results in binary classification, which can roundly measure the performance especially for the unbalanced dataset. The MCC value lies between -1 and $+1$, where -1 indicates that the prediction results are contrary to the actual ones, and $+1$ indicates that it is a perfect prediction and 0 indicates that it is a random prediction.

Additionally, to better show the performance, the receiver operating characteristic (ROC) curves were plotted, and the

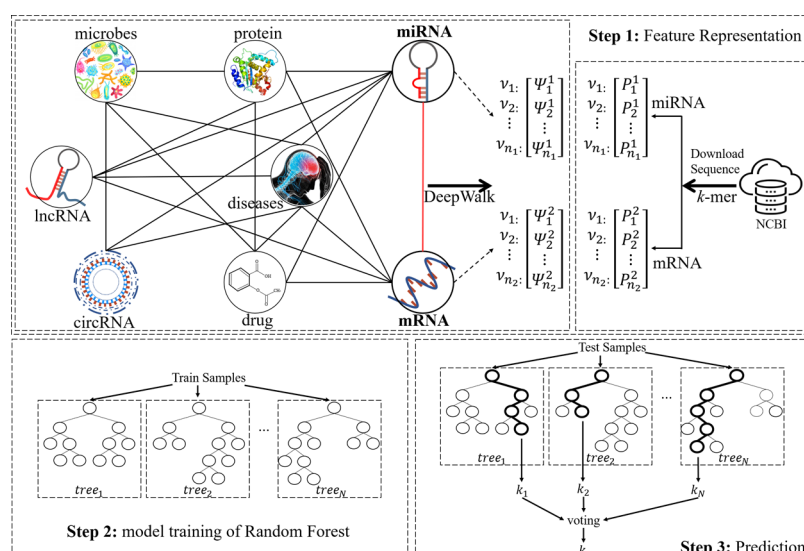


Figure 2. Flowchart of the computational process of MIPDH based on the biological behavior and attribute.

Table 1. 5-fold Cross Validation Results Performed by RF Classifier on Integrated Features of Attribute and Behavior

test set	accuracy (%)	sensitivity (%)	precision (%)	MCC	specificity (%)	AUC
1	75.22	74.74	75.46	0.6272	75.70	0.8014
2	75.70	74.83	76.15	0.6320	76.56	0.8012
3	75.98	72.81	77.74	0.6343	79.15	0.8100
4	75.36	74.16	75.98	0.6285	76.56	0.7933
5	77.00	75.31	77.94	0.6456	78.69	0.8159
average	75.85 ± 0.63	74.37 ± 0.86	76.66 ± 1.00	0.6335 ± 0.0066	77.33 ± 1.34	0.8044 ± 0.0078

corresponding values of the area under curve (AUC) were computed. The value of the AUC at 1 indicates a perfect prediction and the value at 0.5 denotes a random prediction. In order to derive a more accurate estimate, 5-fold cross validation was adopted.

In our study, all associations of our collected dataset were positive. To construct a balanced dataset, the negative samples with the same number of positive ones, were randomly generated from the adjacent matrix of the miRNA–mRNA interaction network where the values are 0. As TPR indicates the performance of model prediction positive interaction correctly, it is a significant indicator for measurement. To implement 5-fold cross validation, the positive samples and negative samples were respectively reordered in random and split into 5 independent subsets with the same number. In each turn of 5 experiments, each subset of the negative/positive set was used as the test set alternately, and the rest was as train set.

According to cross validation, the behaviors among molecule associations were constructed excluding the associations that were used as the test set. When employing the k -mer method, the parameter k is set as 3, which can achieve better effectiveness in the feature representation and model training. In each model training, grid search was employed to find the best parameters that can help to achieve the best performance.

In this work, we proposed a novel framework that miRNA–mRNA interactions can be predicted by the training RF model that utilized integrated features combining attributes generated by the k -mer algorithm with biological behaviors generated by the DeepWalk algorithm. For better understanding the framework, the flowchart was shown in Figure 2. From the results of 5-fold cross validation (see Table 1), the best performance obtained by the trained RF model achieved

77.00% accuracy, 75.31% sensitivity, 77.94% precision, 78.69% specificity, and 0.6456 MCC. In addition, ROCs were plotted in Figure 3, and the corresponding AUCs were calculated.

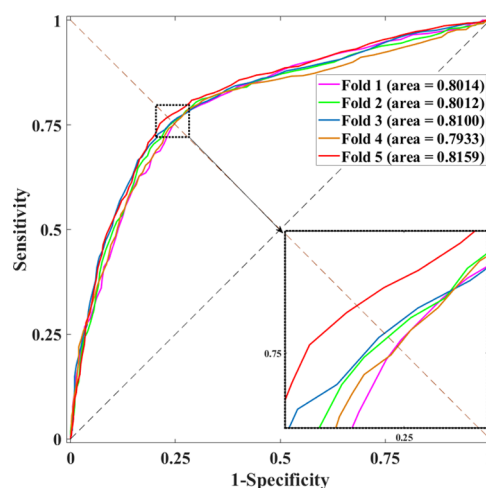


Figure 3. ROC curves performed by the RF classifier based on integrated features of attribute and behavior.

4.2. Comparison the Proposed Method with the Logistics Regression Model and Support Vector Machine. To illustrate the performance of our proposed method, the logistics regression (LR) model and support vector machine (SVM), two classical classifiers in machine learning, were employed for comparison. We implemented these model training with same train samples and test samples in each fold of cross validation. Specially, in order to yield

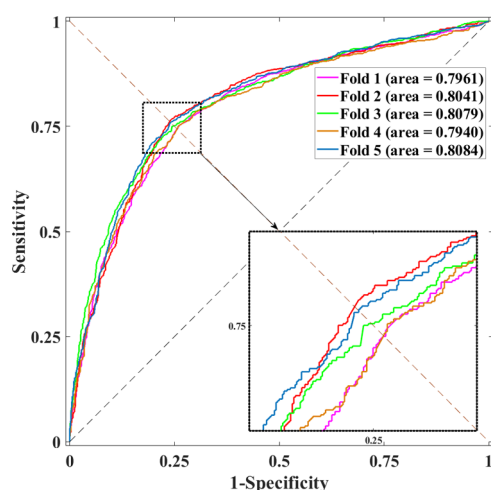
Table 2. 5-Fold Cross Validation Results Performed by the LR Classifier on Integrated Features of Attribute and Behavior

test set	accuracy (%)	sensitivity (%)	precision (%)	MCC	specificity (%)	AUC
1	68.30	69.45	67.89	0.5669	67.15	0.7315
2	67.77	70.12	66.97	0.5627	65.42	0.7408
3	69.07	70.22	68.64	0.5726	67.92	0.7443
4	67.29	67.72	67.14	0.5598	66.86	0.7293
5	68.76	68.18	68.98	0.5703	69.33	0.7433
average	68.23 \pm 0.72	69.13 \pm 1.14	67.92 \pm 0.89	0.5665 \pm 0.0053	67.33 \pm 1.44	0.7378 \pm 0.0070

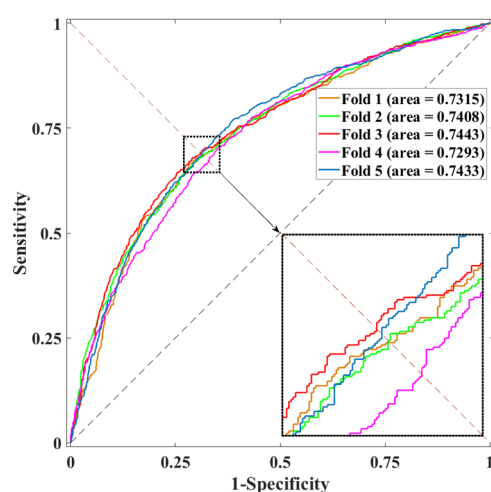
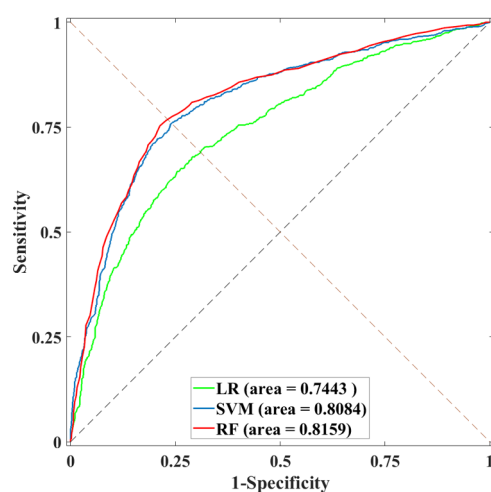
Table 3. 5-Fold Cross Validation Results Performed by SVM Classifier on Integrated Features of Attribute and Behavior

test set	accuracy (%)	sensitivity (%)	precision (%)	MCC	specificity (%)	AUC
1	74.50	74.74	74.38	0.6200	74.26	0.7961
2	76.03	75.60	76.26	0.6355	76.46	0.8041
3	74.78	73.78	75.29	0.6228	75.79	0.8079
4	74.21	73.58	74.51	0.6172	74.83	0.7940
5	75.60	74.83	76.00	0.6311	76.37	0.8084
average	75.02 \pm 0.77	74.51 \pm 0.83	75.29 \pm 0.85	0.6253 \pm 0.0077	75.54 \pm 0.97	0.8021 \pm 0.0067

better results, feature normalization was employed on biological behaviors, in which all elements lay between 0 and 1. As to achieve the best performance of the SVM, a grid search approach was used to get the best parameters (parameter c and parameter g) with the highest accuracy. From overall metrics in Tables 1–3, the best performance was achieved using the RF model. Specifically, the average sensitivity yielded by the SVM was higher than that using RF. Moreover, the corresponding ROCs were plotted and the respective ROCs with highest AUCs were yielded by three models (see Figures 4 and 5). The ROCs with highest AUCs yielded by three classifiers were plotted for comparison (see Figure 6).

**Figure 4.** ROC curves performed by the SVM classifier based on integrated features of attribute and behavior.

4.3. Comparison the Integrated Feature with Attribute Feature and Behavior Feature. To further validate the effectiveness of our proposed feature representation, we also implemented 5-fold cross validation on the RF model by only using behavior feature and attribute feature respectively. The 5-fold cross validation results yielded by respectively using attribute feature and behavior feature were listed in Tables 4 and 5, and the corresponding ROC curves were plotted (see Figures 7 and 8). In detail, the average accuracies of using the

**Figure 5.** ROC curves performed by the LR classifier based on integrated features of attribute and behavior.**Figure 6.** Performance comparison among RF, SVM, and LR models in terms of ROC curves and AUCs based on integrated features of attribute and behavior.

attribute feature and behavior feature were 74.47 ± 0.46 and $75.07 \pm 0.59\%$, sensitivities were 73.08 ± 0.66 and $73.99 \pm 1.00\%$, precisions were 75.17 ± 0.78 and $75.64 \pm 0.75\%$,

Table 4. 5-Fold Cross Validation Results Performed by the RF Classifier on Attribute Features

test set	accuracy (%)	sensitivity (%)	precision (%)	MCC	specificity (%)	AUC
1	73.78	73.29	74.01	0.6130	74.26	0.7921
2	74.45	73.58	74.88	0.6195	75.31	0.7953
3	74.40	71.95	75.66	0.6186	76.85	0.7968
4	74.69	73.49	75.30	0.6218	75.89	0.8021
5	75.02	73.10	76.03	0.6250	76.95	0.8097
average	74.47 \pm 0.46	73.08 \pm 0.66	75.17 \pm 0.78	0.6196 \pm 0.0044	75.85 \pm 1.12	0.7992 \pm 0.0069

Table 5. 5-Fold Cross Validation Results Performed by the RF Classifier on Behavior Features

test set	accuracy (%)	sensitivity (%)	precision (%)	MCC	specificity (%)	AUC
1	74.59	74.45	74.66	0.6209	74.74	0.7872
2	74.83	73.97	75.27	0.6233	75.70	0.7920
3	75.55	73.58	76.60	0.6303	77.52	0.7975
4	74.54	72.62	75.52	0.6202	76.46	0.7816
5	75.84	75.31	76.12	0.6336	76.37	0.8028
average	75.07 \pm 0.59	73.99 \pm 1.00	75.64 \pm 0.75	0.6257 \pm 0.0060	76.16 \pm 1.03	0.7922 \pm 0.0083

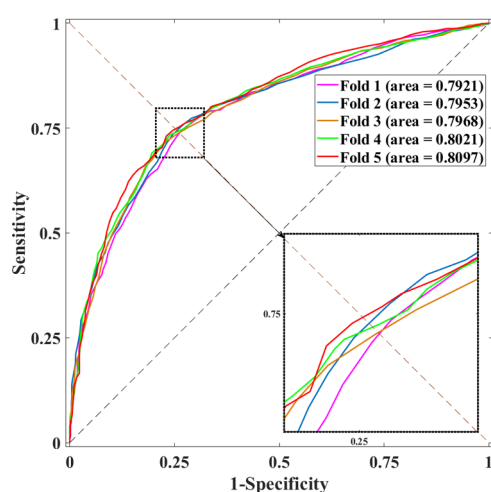


Figure 7. ROC curves performed by the RF classifier based on attribute features.

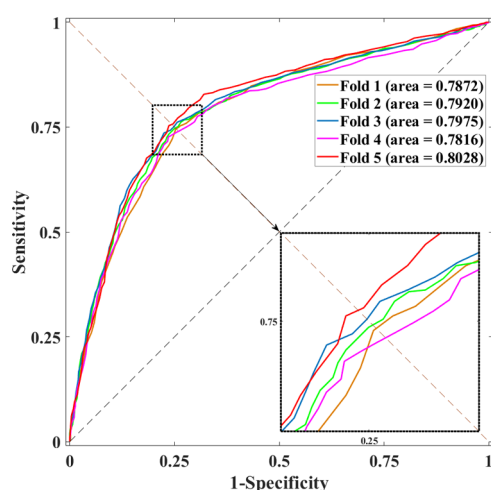


Figure 8. ROC curves performed by the RF classifier based on behavior features.

MCCs were 0.6196 ± 0.0044 and 0.6257 ± 0.0060 , specificities were 75.85 ± 1.12 and $76.16 \pm 1.03\%$, AUCs were 0.7992 ± 0.0069 and 0.7922 ± 0.0083 , respectively.

Compared with results yielded by using the integrated feature, using these two features separately cannot yield better results.

For better comparison between our proposed feature representation and other two feature representations, the ROCs of the best performance yielded using different kinds of feature representation were plotted in Figure 9. From all these results, it suggested that our proposed feature representation method could help to solve the miRNA–mRNA interaction prediction problem.

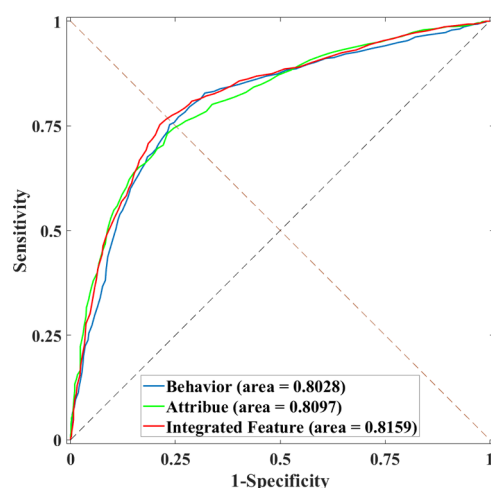


Figure 9. Performance comparison among behavior features, attribute features, and integrated features in terms of ROC curves and AUCs based on the RF classifier.

4.4. Case Studies. For the further performance evaluation of our proposed method, the case studies were implemented based on the RF model using the integrated feature generated using the *k*-mer method and DeepWalk method. In the experiment, the whole dataset of miRNA–mRNA interaction feature vectors, generated from the miRTarBase database, was set as a train set to train the RF model. Considering the known interaction from miRTarBase, the unknown interactions were set as the test set. After model training, the probability values of test samples were obtained and ranked. Then the test samples with higher probability values were verified in the miRWalk database that collects the up-to-date MTIs from the

PubMed literature, TargetScan and miRDB. Note that the negative samples of the train set were generated by random selection, and have no intersection with the miRWalk database. Here, we observed the top 15 potential target mRNAs for three miRNAs including hsa-let-7d-5p, hsa-let-7e-5p, and hsa-miR-106b-5p. Specifically, the results were shown in Tables 6–8. The potential target mRNAs found in the miRWalk database were marked with “PubMed”, “TargetScan”, and “miRDB”. If not, they were marked as “unconfirmed”.

Table 6. Top 15 mRNA Related to hsa-let-7d-5p Predicted by MIPDH

rank	mRNA	evidence	Rank	mRNA	evidence
1	TIMP3	unconfirmed	9	FBN1	PubMed TargetScan miRDB
2	CD44	unconfirmed	10	ITGB3	TargetScan miRDB
3	PTEN	unconfirmed	11	SMN1	PubMed TargetScan
4	NCAM1	unconfirmed	12	IL6R	PubMed TargetScan miRDB
5	AFTPH	unconfirmed	13	BACH1	TargetScan
6	ADAM9	unconfirmed	14	FAIM	TargetScan
7	BCL2L1	TargetScan	15	CCNE1	miRDB
8	MAP4K3	TargetScan miRDB			

Table 7. Top 15 mRNA Related to hsa-let-7e-5p Predicted by MIPDH

rank	mRNA	evidence	Rank	mRNA	evidence
1	CDK4	unconfirmed	9	TIMP3	PubMed
2	CALN1	TargetScan	10	TRIM71	PubMed TargetScan miRDB
3	ZBTB7A	unconfirmed	11	BCL2L1	TargetScan miRDB
4	VDR	unconfirmed	12	TGFBR3	PubMed TargetScan miRDB
5	IGFBP5	unconfirmed	13	MDM4	PubMed TargetScan miRDB
6	GRM3	unconfirmed	14	KLF9	TargetScan miRDB
7	ALDH5A1	unconfirmed	15	PAPPA	TargetScan miRDB
8	MYC	PubMed			

MiRNA hsa-let-7d-5p is one of the miRNAs that have significant expression levels in ovarian cancer ranked seventh in the most common cancer contracted by women. Although there are some treatments such as chemotherapy, radiation therapy, and hormone therapy, it is not optimistic about the five-year survival rate at around 30%. Predicting target mRNAs associated with hsa-let-7d-5p can help to find a novel biomarker. In our prediction for hsa-let-7d-5p, 9 of the top 15 target mRNA prediction were verified (see Table 6). MiRNA hsa-let-7e-5p have been found to be related to more than one type of disease such as rectal carcinoma, pathological cardiac hypertrophy, and type 2 diabetes. In the prediction for hsa-let-7e-5p, 9 of the top 15 target mRNA prediction were verified (see Table 7). It was reported that miRNA hsa-miR-106b-5p is involved in many diseases such as attention-deficit/

Table 8. Top 15 mRNA Related to hsa-miR-106b-5p Predicted by MIPDH

rank	mRNA	evidence	rank	mRNA	evidence
1	PPP2R5C	unconfirmed	9	NTRK2	miRDB
2	FXN	miRDB	10	ATAT1	PubMed
3	SLC6A4	PubMed	11	FLT1	TargetScan miRDB
4	FAS	PubMed miRDB	12	NLN	PubMed TargetScan miRDB
5	GPD2	TargetScan	13	PBX3	PubMed TargetScan miRDB
6	MCL1	PubMed TargetScan miRDB	14	PGR	PubMed TargetScan miRDB
7	EGLN1	TargetScan miRDB	15	RASA1	PubMed TargetScan
8	PAX6	miRDB			

hyperactivity disorder, hepatocellular, and Alzheimer's disease.^{62–64} In the case study of hsa-miR-106b-5p, 14 of the top 15 target mRNA prediction were validated (see Table 8).

5. DISCUSSION

In this study, we proposed a novel computational method named MIPDH for predicting miRNA-target interaction based on diversified interaction pairs. A special aspect of our proposed method is the feature representation method that make use of network embedding to learn the neighbor information of each. The feature generated by employing such a method called behavior defines the pathway of each biomolecule in the constructed biomolecular network. In order to improve the performance, the attribute feature generated using the *k*-mer method was introduced. In comparison experiments among different classifiers and feature representation, our proposed method yielded the best results. Additionally, case studies were implemented to validate the predicted candidates from other databases or literatures. These results released that an insight into miRNA–mRNA interaction can be provided using a novel method MIPDH.

There are many reasons for good performance. First, the DeepWalk method can learn the feature from the large network efficiently. It is the key to integrate the exiting knowledge among biomolecules, disease, and drug. Second, the RF classifier is a powerful model for classification. It is stable, robust to outliers, and less impacted by noise. Third, the *k*-mer method was introduced to improve the performance. The sequence-based feature can be the supplement for the network-based feature. Therefore, our proposed method can yield anticipated results.

Even though MIPDH can achieve reliable prediction results, some limitations should be worth notice. Imbalanced data amounts collected from the current database might cause prediction bias. Moreover, the dataset we used was generated as each miRNA could interact with as much mRNAs as possible and vice versa, but the available ratio of miRNA-target interaction was only 0.7741% ($10,402/(1985 \times 677) \times 100\%$), which revealed that the MTI problem has not been well-studied.

In further research, we will investigate how to learn the information from a massive biomolecular network more sufficiently and effectively, and achieve more remarkable performance by the reconstructing network, by using a more

efficient network embedding method and improving the process framework.

AUTHOR INFORMATION

Corresponding Author

Zhu-Hong You – The Xinjiang Technical Institute of Physics and Chemistry and Xinjiang Laboratory of Minority Speech and Language Information Processing, Chinese Academy of Sciences, Urumqi 830011, China; University of Chinese Academy of Sciences, Beijing 100049, China;
Email: zhuhongyou@ms.xjb.ac.cn

Authors

Leon Wong – The Xinjiang Technical Institute of Physics and Chemistry and Xinjiang Laboratory of Minority Speech and Language Information Processing, Chinese Academy of Sciences, Urumqi 830011, China; University of Chinese Academy of Sciences, Beijing 100049, China; orcid.org/0000-0002-6864-991X

Zhen-Hao Guo – The Xinjiang Technical Institute of Physics and Chemistry and Xinjiang Laboratory of Minority Speech and Language Information Processing, Chinese Academy of Sciences, Urumqi 830011, China; University of Chinese Academy of Sciences, Beijing 100049, China

Hai-Cheng Yi – The Xinjiang Technical Institute of Physics and Chemistry and Xinjiang Laboratory of Minority Speech and Language Information Processing, Chinese Academy of Sciences, Urumqi 830011, China; University of Chinese Academy of Sciences, Beijing 100049, China; orcid.org/0000-0001-8339-396X

Zhan-Heng Chen – The Xinjiang Technical Institute of Physics and Chemistry and Xinjiang Laboratory of Minority Speech and Language Information Processing, Chinese Academy of Sciences, Urumqi 830011, China; University of Chinese Academy of Sciences, Beijing 100049, China

Mei-Yuan Cao – Guang Dong Polytechnic College, Zhaoqing 526100, Guangdong, China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.9b04195>

Author Contributions

L.W. conceived the project, developed the prediction method, designed the experiments, analyzed the result and wrote the manuscript. L.W. implemented the experiments, analyzed the result and revised the manuscript. Z.-H.Y. and Z.-H.C. analyzed the result and revised the manuscript. Z.-H.G., H.-C.Y., and M.-Y.C. analyzed the result. All authors read and approved the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is supported by the NSFC Excellent Young Scholars Program, under Grants 61722212, in part by the National Science Foundation of China under Grants 61873212, 61572506. The authors would like to thank all the editors and anonymous reviewers for their constructive advices.

REFERENCES

(1) Széll, M.; Bata-Csörgő, Z.; Kemény, L. The enigmatic world of mRNA-like ncRNAs: their role in human evolution and in human diseases. *Seminars in Cancer Biology*; Elsevier, 2008; pp 141–148.

(2) Ambros, V. The functions of animal microRNAs. *Nature* **2004**, *431*, 350.

(3) Betel, D.; Wilson, M.; Gabow, A.; Marks, D. S.; Sander, C. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* **2008**, *36*, D149–D153.

(4) Guttman, M.; Rinn, J. L. Modular regulatory principles of large non-coding RNAs. *Nature* **2012**, *482*, 339.

(5) Dai, X.; Zhuang, Z.; Zhao, P. X. psRNATarget: a plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **2018**, *46*, W49–W54.

(6) Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **2009**, *136*, 215–233.

(7) Alvarez-Garcia, I.; Miska, E. A. MicroRNA functions in animal development and human disease. *Development* **2005**, *132*, 4653–4662.

(8) Ab Mutalib, N.-S.; Sulaiman, S. A.; Jamal, R. Computational Tools for microRNA Target Prediction. *Computational Epigenetics and Diseases*; Elsevier, 2019; pp 79–105.

(9) van Zandwijk, N.; Pavlakis, N.; Kao, S. C.; Linton, A.; Boyer, M. J.; Clarke, S.; Huynh, Y.; Chrzanowska, A.; Fulham, M. J.; Bailey, D. L.; Cooper, W. A.; Kritharides, L.; Ridley, L.; Pattison, S. T.; MacDiarmid, J.; Brahmbhatt, H.; Reid, G. Safety and activity of microRNA-loaded minicells in patients with recurrent malignant pleural mesothelioma: a first-in-man, phase 1, open-label, dose-escalation study. *Lancet Oncol.* **2017**, *18*, 1386–1396.

(10) Yan, Q.; Chen, J.; Li, W.; Bao, C.; Fu, Q. Targeting miR-155 to treat experimental scleroderma. *Sci. Rep.* **2016**, *6*, 20314.

(11) Wang, F.; Li, L.; Piontek, K.; Sakaguchi, M.; Selaru, F. M. Exosome miR-335 as a novel therapeutic strategy in hepatocellular carcinoma. *Hepatology* **2018**, *67*, 940–954.

(12) Zhu, H.; Wu, H.; Liu, X.; Evans, B. R.; Medina, D. J.; Liu, C.-G.; Yang, J.-M. Role of MicroRNA miR-27a and miR-451 in the regulation of MDR1/P-glycoprotein expression in human cancer cells. *Biochem. Pharmacol.* **2008**, *76*, 582–588.

(13) Coulouarn, C.; Factor, V. M.; Andersen, J. B.; Durkin, M. E.; Thorgerirsson, S. S. Loss of miR-122 expression in liver cancer correlates with suppression of the hepatic phenotype and gain of metastatic properties. *Oncogene* **2009**, *28*, 3526.

(14) Van Rooij, E.; Sutherland, L. B.; Thatcher, J. E.; DiMaio, J. M.; Naseem, R. H.; Marshall, W. S.; Hill, J. A.; Olson, E. N. Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13027–13032.

(15) Guo, L.-H.; Li, H.; Wang, F.; Yu, J.; He, J.-S. The tumor suppressor roles of miR-433 and miR-127 in gastric cancer. *Int. J. Mol. Sci.* **2013**, *14*, 14171–14184.

(16) Betel, D.; Koppal, A.; Agius, P.; Sander, C.; Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* **2010**, *11*, R90.

(17) Lewis, B. P.; Burge, C. B.; Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **2005**, *120*, 15–20.

(18) Marín, R. M.; Vaníček, J. Efficient use of accessibility in microRNA target prediction. *Nucleic Acids Res.* **2010**, *39*, 19–29.

(19) Mahen, E. M.; Watson, P. Y.; Cottrell, J. W.; Fedor, M. J. mRNA secondary structures fold sequentially but exchange rapidly in vivo. *PLoS Biol.* **2010**, *8*, No. e1000307.

(20) Miranda, K. C.; Huynh, T.; Tay, Y.; Ang, Y.-S.; Tam, W.-L.; Thomson, A. M.; Lim, B.; Rigoutsos, I. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell* **2006**, *126*, 1203–1217.

(21) Lewis, B. P.; Shih, I.-h.; Jones-Rhoades, M. W.; Bartel, D. P.; Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **2003**, *115*, 787–798.

(22) Doench, J. G.; Sharp, P. A. Specificity of microRNA target selection in translational repression. *Genes Dev.* **2004**, *18*, 504–511.

(23) Yue, D.; Liu, H.; Huang, Y. Survey of computational algorithms for microRNA target prediction. *Curr. Genomics* **2009**, *10*, 478–492.

(24) Wang, X. Composition of seed sequence is a major determinant of microRNA targeting patterns. *Bioinformatics* **2014**, *30*, 1377–1383.

- (25) Kertesz, M.; Iovino, N.; Unnerstall, U.; Gaul, U.; Segal, E. The role of site accessibility in microRNA target recognition. *Nat. Genet.* **2007**, *39*, 1278.
- (26) Sturm, M.; Hackenberg, M.; Langenberger, D.; Frishman, D. TargetSpy: a supervised machine learning approach for microRNA target prediction. *BMC Bioinf.* **2010**, *11*, 292.
- (27) Sethupathy, P.; Corda, B.; Hatzigeorgiou, A. G. TarBase: A comprehensive database of experimentally supported animal microRNA targets. *RNA* **2006**, *12*, 192.
- (28) Thomson, D. W.; Bracken, C. P.; Goodall, G. J. Experimental strategies for microRNA target identification. *Nucleic Acids Res.* **2011**, *39*, 6845–6853.
- (29) Wong, L.; Huang, Y.-A.; You, Z. H.; Chen, Z. H.; Cao, M. Y. LNRLMI: Linear neighbour representation for predicting lncRNA-miRNA interactions. *J. Cell. Mol. Med.* **2020**, *24*, 79–87.
- (30) Chen, Z.-H.; You, Z.-H.; Li, L.-P.; Wang, Y.-B.; Wong, L.; Yi, H.-C. Prediction of Self-Interacting Proteins from Protein Sequence Information Based on Random Projection Model and Fast Fourier Transform. *Int. J. Mol. Sci.* **2019**, *20*, 930.
- (31) Jiang, H.-J.; Huang, Y.-A.; You, Z.-H. Predicting Drug-Disease Associations via Using Gaussian Interaction Profile and Kernel-Based Autoencoder. *BioMed Res. Int.* **2019**, *2019*, 1–11.
- (32) Wang, L.; You, Z.-H.; Chen, X.; Li, Y.-M.; Dong, Y.-N.; Li, L.-P.; Zheng, K. LMTRDA: Using logistic model tree to predict MiRNA-disease associations by fusing multi-source information of sequences and similarities. *PLoS Comput. Biol.* **2019**, *15*, No. e1006865.
- (33) Yi, H.-C.; You, Z.-H.; Zhou, X.; Cheng, L.; Li, X.; Jiang, T.-H.; Chen, Z.-H. ACP-DL: A Deep Learning Long Short-Term Memory Model to Predict Anticancer Peptides Using High-Efficiency Feature Representation. *Mol. Ther. Nucleic Acids* **2019**, *17*, 1–9.
- (34) Zheng, K.; You, Z.-H.; Wang, L.; Zhou, Y.; Li, L.-P.; Li, Z.-W. MLMDA: a machine learning approach to predict and validate MicroRNA–disease associations by integrating of heterogeneous information sources. *J. Transl. Med.* **2019**, *17*, 260.
- (35) Zheng, K.; You, Z.-H.; Wang, L.; Li, Y.-R.; Wang, Y.-B.; Jiang, H.-J. MISSIM: Improved miRNA-Disease Association Prediction Model Based on Chaos Game Representation and Broad Learning System. *International Conference on Intelligent Computing*; Springer, 2019; pp 392–398.
- (36) Wong, N.; Wang, X. miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.* **2014**, *43*, D146–D152.
- (37) Wang, X. Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. *Bioinformatics* **2016**, *32*, 1316–1322.
- (38) Kanoria, S.; Rennie, W.; Liu, C.; Carmack, C. S.; Lu, J.; Ding, Y. STarMir Tools for Prediction of microRNA binding sites. *RNA Structure Determination*; Springer, 2016; pp 73–82.
- (39) Huang, Y.-A.; Chan, K. C. C.; You, Z.-H. Constructing prediction models from expression profiles for large scale lncRNA–miRNA interaction profiling. *Bioinformatics* **2017**, *34*, 812–819.
- (40) Qu, J.; Chen, X.; Sun, Y.-Z.; Li, J.-Q.; Ming, Z. Inferring potential small molecule–miRNA association based on triple layer heterogeneous network. *J. Cheminf.* **2018**, *10*, 30.
- (41) Ding, L.; Wang, M.; Sun, D.; Li, A. A novel method for identifying potential disease-related miRNAs via a disease–miRNA–target heterogeneous network. *Mol. BioSyst.* **2017**, *13*, 2328–2337.
- (42) Guo, Z.-H.; Yi, H.-C.; You, Z.-H. Construction and Comprehensive Analysis of a Molecular Association Network via lncRNA–miRNA–Disease–Drug–Protein Graph. *Cells* **2019**, *8*, 866.
- (43) Perozzi, B.; Al-Rfou, R.; Skiena, S. Deepwalk: Online learning of social representations. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM, 2014; pp 701–710.
- (44) Zhao, Z.; Wang, K.; Wu, F.; Wang, W.; Zhang, K.; Hu, H.; Liu, Y.; Jiang, T. circRNA disease: a manually curated database of experimentally supported circRNA-disease associations. *Cell Death Dis.* **2018**, *9*, 475.
- (45) Yuan, J.; Wu, W.; Xie, C.; Zhao, G.; Zhao, Y.; Chen, R. NPInter v2.0: an updated database of ncRNA interactions. *Nucleic Acids Res.* **2013**, *42*, D104–D108.
- (46) Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; Liu, Y.; Mandal, R.; Neveu, V.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082.
- (47) Tong, Z.; Cui, Q.; Wang, J.; Zhou, Y. TransmiR v2.0: an updated transcription factor-microRNA regulation database. *Nucleic Acids Res.* **2018**, *47*, D253–D258.
- (48) Rizkallah, M. R.; Gamal-Eldin, S.; Saad, R.; Aziz, R. K. The pharmacomicrobiomics portal: a database for drug-microbiome interactions. *Curr. Pharmacogenomics Pers. Med.* **2012**, *10*, 195–203.
- (49) Piñero, J.; Bravo, A.; Queralt-Rosinach, N.; Gutiérrez-Sacristán, A.; Deu-Pons, J.; Centeno, E.; García-García, J.; Sanz, F.; Furlong, L. I. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **2017**, *45*, D833.
- (50) Miao, Y.-R.; Liu, W.; Zhang, Q.; Guo, A.-Y. lncNASNP2: an updated database of functional SNPs and mutations in human and mouse lncRNAs. *Nucleic Acids Res.* **2017**, *46*, D276–D280.
- (51) Ma, W.; Zhang, L.; Zeng, P.; Huang, C.; Li, J.; Geng, B.; Yang, J.; Kong, W.; Zhou, X.; Cui, Q. An analysis of human microbe–disease associations. *Briefings Bioinf.* **2016**, *18*, 85–97.
- (52) Liu, X.; Wang, S.; Meng, F.; Wang, J.; Zhang, Y.; Dai, E.; Yu, X.; Li, X.; Jiang, W. SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics* **2012**, *29*, 409–411.
- (53) Li, Y.; Qiu, C.; Tu, J.; Geng, B.; Yang, J.; Jiang, T.; Cui, Q. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.* **2013**, *42*, D1070–D1074.
- (54) Hewett, M.; Oliver, D. E.; Rubin, D. L.; Easton, K. L.; Stuart, J. M.; Altman, R. B.; Klein, T. E. PharmGKB: the pharmacogenetics knowledge base. *Nucleic Acids Res.* **2002**, *30*, 163–165.
- (55) Fan, C.; Lei, X.; Fang, Z.; Jiang, Q.; Wu, F.-X. CircR2Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases. *Database* **2018**, *2018*, bay044.
- (56) Davis, A. P.; Grondin, C. J.; Johnson, R. J.; Sciaky, D.; McMorran, R.; Wieggers, J.; Wieggers, T. C.; Mattingly, C. J. The comparative toxicogenomics database: update 2019. *Nucleic Acids Res.* **2018**, *47*, D948–D954.
- (57) Chou, C.-H.; Shrestha, S.; Yang, C.-D.; Chang, N.-W.; Lin, Y.-L.; Liao, K.-W.; Huang, W.-C.; Sun, T.-H.; Tu, S.-J.; Lee, W.-H.; Chiew, M.-Y.; Tai, C.-S.; Wei, T.-Y.; Tsai, T.-R.; Huang, H.-T.; Wang, C.-Y.; Wu, H.-Y.; Ho, S.-Y.; Chen, P.-R.; Chuang, C.-H.; Hsieh, P.-J.; Wu, Y.-S.; Chen, W.-L.; Li, M.-J.; Wu, Y.-C.; Huang, X.-Y.; Ng, F. L.; Buddhakosai, W.; Huang, P.-C.; Lan, K.-C.; Huang, C.-Y.; Weng, S.-L.; Cheng, Y.-N.; Liang, C.; Hsu, W.-L.; Huang, H.-D. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **2017**, *46*, D296–D302.
- (58) Cheng, L.; Wang, P.; Tian, R.; Wang, S.; Guo, Q.; Luo, M.; Zhou, W.; Liu, G.; Jiang, H.; Jiang, Q. LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.* **2018**, *47*, D140–D144.
- (59) Chen, G.; Wang, Z.; Wang, D.; Qiu, C.; Liu, M.; Chen, X.; Zhang, Q.; Yan, G.; Cui, Q. LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.* **2012**, *41*, D983–D986.
- (60) Bhattacharya, A.; Cui, Y. SomamiR 2.0: a database of cancer somatic mutations altering microRNA–ceRNA interactions. *Nucleic Acids Res.* **2015**, *44*, D1005–D1010.
- (61) Bao, Z.; Yang, Z.; Huang, Z.; Zhou, Y.; Cui, Q.; Dong, D. LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.* **2018**, *47*, D1034–D1037.

(62) Yılmaz, Ş. G.; Erdal, M. E.; Özge, A. A.; Sungur, M. A. Can peripheral MicroRNA expression data serve as epigenomic (upstream) biomarkers of Alzheimer's disease? *OMICS: J. Integr. Biol.* **2016**, *20*, 456–461.

(63) Zadehbagheri, F.; Hosseini, E.; Bagheri-Hosseinabadi, Z.; Rekabdarkolaei, H. M.; Sadeghi, I. Profiling of miRNAs in serum of children with attention-deficit hyperactivity disorder shows significant alterations. *J. Psychiatr. Res.* **2019**, *109*, 185–192.

(64) Liu, X.; Wang, S.; Xu, J.; Kou, B.; Chen, D.; Wang, Y.; Zhu, X. Extract of *Stellerachamaejasme* L (ESC) inhibits growth and metastasis of human hepatocellular carcinoma via regulating micro-RNA expression. *BMC Complementary Altern. Med.* **2018**, *18*, 99.