

OPEN ACCESS

EDITED BY

Hongmin Cai,
South China University of
Technology, China

REVIEWED BY

Pu-Feng Du,
Tianjin University, China
Cheng Liang,
Shandong Normal University, China

*CORRESPONDENCE

Changsong Ding
dingcs1975@hnu.edu.cn
Guihua Duan
duangh@csu.edu.cn

SPECIALTY SECTION

This article was submitted to
Precision Medicine,
a section of the journal
Frontiers in Medicine

RECEIVED 09 August 2022

ACCEPTED 25 October 2022

PUBLISHED 17 November 2022

CITATION

Yan C, Ding C and Duan G (2022)
PMMS: Predicting essential miRNAs
based on multi-head self-attention
mechanism and sequences.
Front. Med. 9:1015278.
doi: 10.3389/fmed.2022.1015278

COPYRIGHT

© 2022 Yan, Ding and Duan. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

PMMS: Predicting essential miRNAs based on multi-head self-attention mechanism and sequences

Cheng Yan^{1,2}, Changsong Ding^{1*} and Guihua Duan^{3*}

¹School of Informatics, Hunan University of Chinese Medicine, Changsha, China, ²School of Computer and Information, Qiannan Normal University for Nationalities, Duyun, China, ³School of Computer Science and Engineering, Central South University, Changsha, China

Increasing evidence has proved that miRNA plays a significant role in biological progress. In order to understand the etiology and mechanisms of various diseases, it is necessary to identify the essential miRNAs. However, it is time-consuming and expensive to identify essential miRNAs by using traditional biological experiments. It is critical to develop computational methods to predict potential essential miRNAs. In this study, we provided a new computational method (called PMMS) to identify essential miRNAs by using multi-head self-attention and sequences. First, PMMS computes the statistic and structure features and extracts the static feature by concatenating them. Second, PMMS extracts the deep learning original feature (BiLSTM-based feature) by using bi-directional long short-term memory (BiLSTM) and pre-miRNA sequences. In addition, we further obtained the multi-head self-attention feature (MS-based feature) based on BiLSTM-based feature and multi-head self-attention mechanism. By considering the importance of the subsequence of pre-miRNA to the static feature of miRNA, we obtained the deep learning final feature (WA-based feature) based on the weighted attention mechanism. Finally, we concatenated WA-based feature and static feature as an input to the multilayer perceptron model to predict essential miRNAs. We conducted five-fold cross-validation to evaluate the prediction performance of PMMS. The areas under the ROC curves (AUC), the F1-score, and accuracy (ACC) are used as performance metrics. From the experimental results, PMMS obtained best prediction performances (AUC: 0.9556, F1-score: 0.9030, and ACC: 0.9097). It also outperformed other compared methods. The experimental results also illustrated that PMMS is an effective method to identify essential miRNA.

KEYWORDS

microRNA, essential miRNA, bi-directional long short-term memory, multi-head self-attention mechanism, deep learning

Introduction

MicroRNAs (miRNAs) are a class of typical non-coding RNAs (ncRNAs) of 22 nucleotide (nt) in length, express endogenously, and regulate gene expression on the posttranscriptional level. During miRNA biogenesis, Drosha and Dicer process the primary transcript (pri-miRNA) through a precursor hairpin (pre-miRNA) to the mature miRNA (1). In other words, RNA polymerase II transcribes the nuclear genes of miRNA to generate pri-miRNAs. The pre-miRNA is produced from pri-miRNAs based on the hairpin structure with enzymatic cleavage. miRNAs often interact with 3' untranslated region (3'UTR) of a target mRNA to mediate mRNA degradation and/or translational repression (2). Furthermore, some studies also demonstrated that they also interact with other regions, such as 5'UTR, gene promoters, and coding sequence (3).

The first two known miRNAs, namely *lin-4* and *let-7*, were derived from *Caenorhabditis elegans* and were discovered more than 20 years ago (4, 5). Until recently, thousands of currently annotated miRNAs have been identified in a variety of species from plants and animals to viruses (6, 7). miRNAs can be released into the extracellular environment and transported to the target cells by vesicles, including exosomes, or *via* binding to proteins. Once expressed, miRNAs are integrated into the RISC and guide the repression of a target mRNA by base complementarity within the “seed” sequence of the miRNA (8). This process resulted in either repression or degradation of the target mRNA and affected the differentiation and proliferation of cells. In addition, many pieces of evidence have proven that miRNAs play essential roles in some important biological processes, such as cell growth, proliferation (9), differentiation (10), and development (11).

Furthermore, many studies demonstrated that miRNAs are highly correlated with human complex diseases, and essential miRNA is crucial in animal development and human diseases. For example, miR-144-3p was lowly expressed in non-small cell lung cancer (NSCLC) and might function as a potential tumor biomarker in the prognosis prediction for NSCLC (12). miR-200c-141 and miR-200b-200a-429 were downregulated in human breast cancer stem cell (BCSC), normal human and murine mammary stem/progenitor cells, and embryonal carcinoma cells (13). miR-200c inhibited the clonal expansion of breast cancer cells and suppressed the growth of embryonal carcinoma cells *in vitro*. In colorectal cancer (CRC), miRNAs-21 is one of the most important miRNAs and is also emerging as a biomarker in CRC, with good potential as a diagnostic and therapeutic target (14). miR-125a-5p could be considered a regulator of glycolipid metabolism in type 2 diabetes mellitus (T2DM), which can inhibit hepatic lipogenesis and gluconeogenesis and elevate glycogen synthesis by targeting STAT3 (15). The expression of miR-145 is significantly downregulated in dedifferentiated vascular smooth muscle cells (VSMCs) and in balloon-injured

arteries, which can be considered a potential therapeutic target (16). miR-228 was also upregulated in osteoarthritis (OA), and can be considered a biomarker (17). In addition, after knocking out miR-15b, B cell lymphoproliferative disorders in mice have been observed (18). After knocking out miR-144, the incidence of spontaneous B lymphoma and acute myeloid leukemia in aged mice increased (19).

Due to the importance mentioned earlier and the necessity of miRNA-disease associations and essential miRNAs, a growing number of databases have been developed. miRbase was an online repository for nomenclature and annotation, and the latest release (v22) contains microRNA sequences from 271 organisms: 38,589 hairpin precursors and 48,860 mature microRNAs (20). miRgator was also a miRNA portal for deep sequencing, expression profiling, and mRNA targeting (21). In addition, many miRNA-disease association databases have also been established, which include miR2Disease (22), human microRNA disease database (HMDD) (23, 24), ExcellmiRDB (25), and miRCancer (26). miR2Disease was a manually curated database and represented an exhaustive resource of miRNA deregulation in different human diseases. HMDD was also an miRNA-disease association database and was developed in 2007, and the HMDD v3.2 gathered more than 35,547 experimentally confirmed entries of miRNA-disease association containing about 1,206 miRNA genes and 893 diseases from 19,280 papers. ExcellmiRDB was also a user-friendly and curated online database about miRNA-disease associations, which includes 1,108 extracellular miRNAs-biofluid relationships and 2,773 extracellular miRNA-disease derived from 108 papers selected from >600 PubMed abstracts. miRCancer collected 878 associations between 236 miRNAs and 79 human cancers through the processing of >26,000 published articles. Besides, Cui et al. (27) established an essential miRNA benchmark dataset to predict potential essential miRNAs.

Due to the importance and necessity of miRNA and the development of the miRNA-disease association database and essential miRNA benchmark dataset, a growing number of computational methods have been proposed for the prediction of miRNA-disease association and essential miRNA. For example, BNPMMA was a typical network-based method to predict potential miRNA-disease associations by using the known miRNA-disease association network, integrated miRNA similarity network, and integrated disease similarity network (28). DNRLMF-MDA was a miRNA-disease association prediction method based on dynamic neighborhood regularized logistic matrix factorization. The main feature of DNRLMF-MDA was that known miRNA-disease associations are assigned higher importance levels than unknown miRNA-disease associations (29). Chen et al. also proposed a method to predict new miRNA-disease association by completing the missing miRNA-disease association based on the known associations and the integrated miRNA similarity and disease similarity (30). Based on the matrix decomposition and heterogeneous

graph inference model, MDHGI was proposed to predict potential miRNA-disease associations (31). It improved the prediction accuracy (ACC) by taking full advantage of matrix decomposition before the construction of a heterogeneous network. Zhou et al. also proposed a method of neural inductive matrix completion with the graph convolutional network (NIMCGCN) for identifying miRNA-disease association (32). NIMCGCN first extracted the latent feature representations and disease from the miRNA and disease similarity networks. PDMDA was also an miRNA-disease association prediction method based on the graph neural network (33). It can predict not only miRNA-disease association but also predict the association type. miES was the first essential miRNA prediction method by using pre-miRNA and miRNA sequences and it also constructed the benchmark dataset of essential miRNA (27). PESM was also an essential miRNA prediction method and improved the prediction performance by adding new features and gradient boosting machines model (34). In addition, XGEM was also an essential miRNAs prediction method by applying the XGBoost framework with Classification and Regression Trees (CART) on various types of sequence-based features (35).

Although we have obtained some progress in predicting essential miRNA based on the development of computing technology and essential miRNA benchmark dataset, it is critical to propose new computational method to improve the prediction ACC. In this study, we provided a method (predicting essential miRNAs based on the multi-head self-attention and sequences, PMMS) to predict essential miRNA. PMMS first calculates the static feature based on the pre-miRNA and miRNA sequences by statistics and the Vienna RNA Package (34). The deep learning original feature (BiLSTM-based feature) is obtained by bi-directional Long Short-Term Memory (BiLSTM) and pre-miRNA sequences. The multi-head self-attention feature (MS-based feature) of miRNA is extracted based on multi-head self-attention mechanism and BiLSTM-based feature. The deep learning final feature (WA-based feature) of miRNA is obtained by using a weight attention mechanism with a static structure feature and MS-based feature of miRNA. Finally, we obtained the final feature of miRNA by concatenating the static feature and WA-based feature, and then take feature, and then taken it as input into Multilayer perceptron (MLP) to predict essential miRNA. We conducted five-fold cross-validation to evaluate the prediction performance of PMMS and compared it with other computational methods which includes PESM, miES, GaussianNaiveBayes (Gaus NB), and support vector machines (SVM) models. The the areas under the ROC curves (AUC), F1-score, and ACC are used as metrics. The experimental results showed that PMMS obtains the best prediction performance according to AUC value of 0.9556. The ACC (0.9037) and F1-score (0.9030) of PMMS in the five-fold cross-validation were higher than that of other methods, respectively, and it also proved that it can achieve better results.

Materials

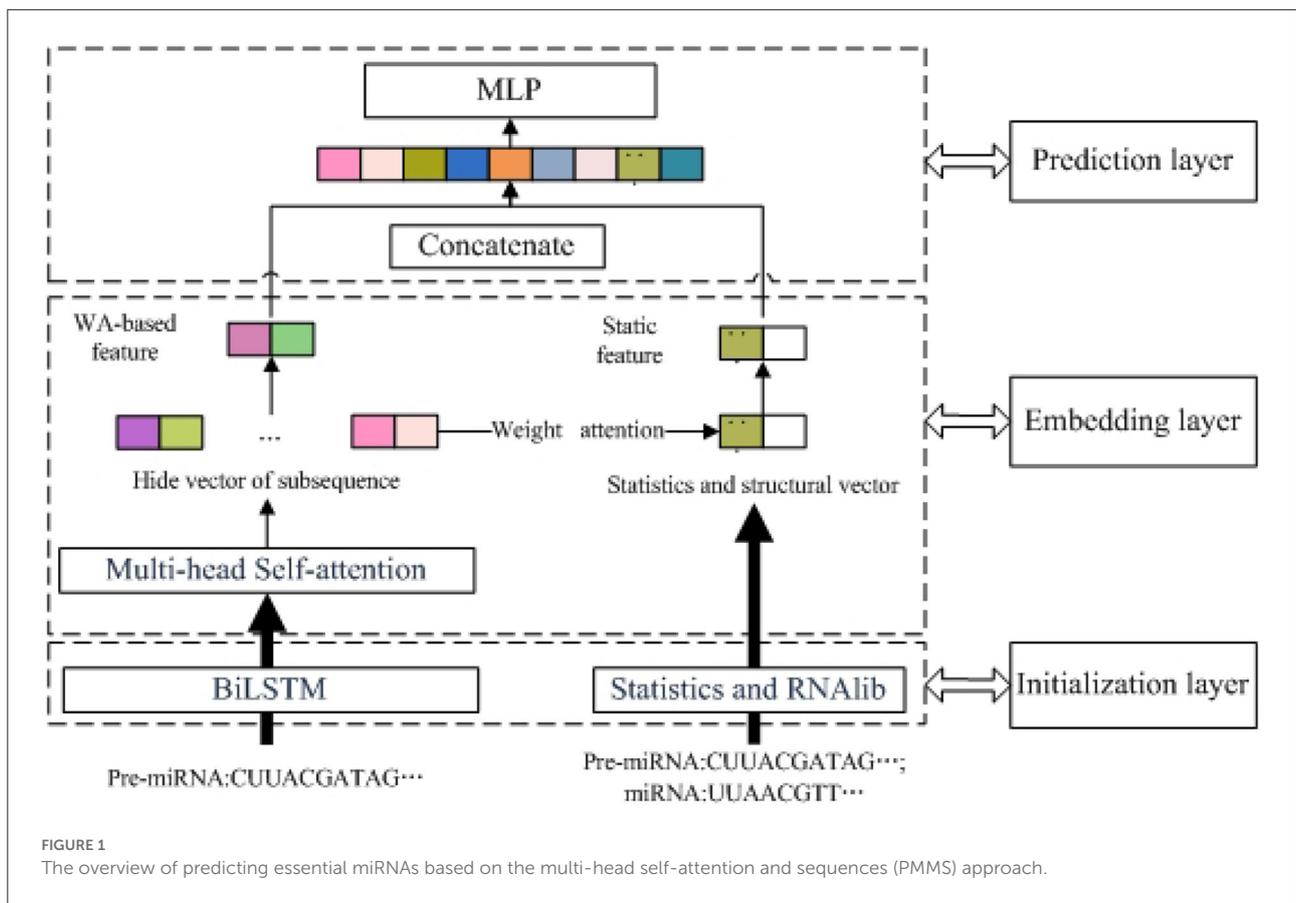
In this study, we also used the benchmark dataset of essential miRNA which was also used in PESM and miES. This benchmark dataset includes 77 known essential miRNAs that were confirmed by knocking out gene experiments. It also includes the same number of negative samples which are randomly selected from the unknown essential miRNAs. In addition, by considering the production process of mature miRNA and the hairpin structure of pre-miRNA, we also used the pre-miRNA and miRNA sequences which are downloaded from the miRbase database. miRbase provided the nomenclature and annotation and pre-miRNA sequences and mature-miRNA sequences of humans, rats, and mice. The latest release (v22) contains miRNA sequences from 271 organisms: 38,589 hairpin precursors and 48,860 mature miRNAs.

Methods

As shown in Figure 1, PMMS mainly contains three layers. The left section of the initialization layer extracted the BiLSTM-based feature by k-mer and the BiLSTM model. The right section of the initialization layer obtained the static feature (statistic and structure feature) by calculating the number of nucleotides and RNaLib package. In addition, we obtained the MS-based feature through the multi-head self-attention mechanism and BiLSTM-based feature. In the embedding layer, we also obtained the WA-based feature based on the MS-based feature and weight attention mechanism. Finally, the miRNA final feature is obtained by concatenating the WA-based feature and static feature and is input into the MLP model to predict essential miRNA. The identification of essential miRNA is also a typical binary-classification problem, and MLP has also been successfully applied to the data classification problem.

Statistic and structure feature

Based on the production process of mature miRNA, we calculated the statistic and structure feature of miRNA from pre-miRNA sequence and mature miRNA sequence. We all know that RNA polymerase II transcribes the nuclear genes of miRNA to generate pri-miRNAs which produced pre-miRNA by enzymatic cleavage on the hairpin structure. Then, the mature miRNA is produced through cleaving pre-miRNA. Since the mature miRNA sequence is the subsequence of the pre-miRNA sequence, we let non-mature miRNA denote the rest of the pre-miRNA sequence after cleaving miRNAs. First, we calculated the single nucleotide base content $S \in \{U, C, G\}$ in pre-miRNA, miRNA, and non-mature miRNA sequences. Three features with dimensionality 3 were obtained, respectively. The sequence lengths of pre-miRNA and miRNA sequences were



also calculated as the feature with dimensionality 1, respectively. In addition, we also calculated the number of dinucleotide pairs $S, Z \in \{U, C, G\}$ in pre-miRNA and miRNA sequences. Thus, we obtained features with dimensionality 9 from them, respectively. The cleavage site base class is another feature and is divided into three categories: (1) 1 represents all cleavage sites of mature-miRNAs from the same pre-miRNAs are *U*; (2) 0 represents not all cleavage sites are *U*; (3) -1 represents all are *non-U*. Based on the hairpin structure contained in all pre-miRNAs which can produce mature miRNA, we also calculated the structure feature by the Vienna RNA Package. The minimum free energy (MFE) and nMFE (minimum free energy and it is divided by its length) are the feature with dimensionality 1, respectively. In addition, we also further considered the base-pairing propensity, Shannon entropy, and base-pair distance to obtain 6 features which include normalized base-pairing propensity (dP), normalized base-pairing propensity divided by its length (dP/L), normalized Shannon entropy (dQ), normalized Shannon entropy divided by its length (dQ/L), normalized base-pair distance (dD), and normalized base-pair distance divided by its length (dD/L). These features were widely used in miRNA prediction (36) and pre-miRNA prediction (37). Table 1 describes the overview of all statistics and structure features.

BiLSTM-based feature

By considering the successful application of BiLSTM in natural language processing (NLP) (38) and the timing characteristic of pre-miRNA sequence, we also applied BiLSTM to extract the deep learning original feature. Compared with the LSTM model (39, 40), BiLSTM was provided to encode information back to front when using LSTM to model the sequences. It includes two LSTMs that are used to take the input in a forward direction and a backward direction. For pre-miRNA sequences, BiLSTM can not only process sequences in temporal order but also consider the future context. LSTM is composed of a cell, an input gate, an output gate, and a forget gate. The cell remembers values over arbitrary time intervals. The three gates regulate the flow of information into and out of the cell. Based on the design characteristics of LSTM and the characteristic of time series data, LSTM is very suitable for processing text and biological sequence data.

To apply the BiLSTM to pre-miRNAs, we first also defined "word" in pre-miRNA sequences as *k*-mer nucleotide. There are 4 types of NNs (A, U, C, and G). In this study, we set *k* to be 3 based on the experiment results. Therefore, we can split a pre-miRNA sequence into an overlapping 3-mer nucleotide. For

TABLE 1 The overview of statistics and structure feature F_s .

Category	Description	Dimensionality
Base content in pre-miRNAs	The content S in pre-miRNA, $S \in U, C, G$	3
Base content in miRNA	The content S in mature-miRNA, $S \in U, C, G$	3
Base content in non-miRNA	The content S in non-mature-miRNA, $S \in U, C, G$	3
miRNA length non-mature	The sequence length of mature-miRNAs	1
miRNA length	The sequence length of non-mature miRNAs	1
Cleavage site base class	The cleave sites are assigned into 3 classes, 1: all cleavage sites of mature-miRNAs from the same pre-miRNAs are U ; 0: not all cleavage sites are U ; -1: all are <i>non-U</i>	1
Dinucleotide pairs number in pre-miRNA	The Dinucleotide pairs SZ number in pre-miRNAs, $S, Z \in U, C, G$	9
Dinucleotide pairs number in miRNA	The Dinucleotide pairs SZ number in mature-miRNAs, $S, Z \in U, C, G$	9
MFE and nMFE	The minimum free energy of pre-miRNA secondary structures and it is divided by its length	2
The base-pairing structure feature of pre-miRNAs	The normalized base-pairing propensity (dp), normalized base-pairing propensity divided by its length (dp/L), normalized Shannon entropy (dQ), normalized Shannon entropy divided its length (dQ/L), normalized base-pair distance (dD), normalized base-pair distance divided by its length (dD/L)	6

pre-miRNA “AUUGUCC...”, the 3-mer nucleotide is defined as follows:

$$“AUU”, “UUG”, “UGU”, “GUC”, “UCC”, ... \tag{1}$$

After obtaining the 3-mer nucleotide of pre-miRNA sequences, we translated them to randomly initialized embeddings (word embedding). For a pre-miRNA sequence $S = s_1, s_2, s_3, \dots, s_{|S|-1}, s_{|S|}$, the 3-mer embedding of pre-miRNA sequence can be concatenated with nucleotide embeddings, and it is defined as follows:

$$[s_1, s_2, s_3], [s_2, s_3, s_4], \dots, [s_{|S|-1}, s_{|S|-2}, s_{|S|-3}], \tag{2}$$

where $x_i = [s_i; s_{i+1}; s_{i+2}]$ and $d_b = 128$. After obtaining 3-mer embedding of sequence, we take it as input to the BiLSTM model.

The BiLSTM model is shown in Figure 2. The model includes two LSTM sub-networks for the left and right nucleotide sequenceright nucleotide sequences. They are forward and backward passes, respectively. The final output of the i -th 3-mer nucleotide sequence is computed by using an element-wise sum of forward and backward pass outputs. In this study, the input of BiLSTM is the 3-mer embedding of pre-miRNA sequence $F_{B0}^{(t)} \in R^{d_b} = x_i$ and $d_b = 128$. The output of BiLSTM is the BiLSTM-based feature $F_B = \{F_{B1}^{(t)}, F_{B2}^{(t)}, \dots, F_{B|L|}^{(t)}\}$, $|L| = |S| - 2$, and $F_{Bi} = y_i \in R^{d_b}$.

MS-based feature

By considering the application of multi-head self-attention mechanism (41) in learning tasks with contextual relationships which include drug-target interaction (42–44), prediction, etc. The multi-head self-attention mechanism can address the limits that LSTM cannot obtain long-dependent information when the sequence is long.

Figure 3 shows the overview of the multi-head self-attention mechanism model. The part of purple background block is the scaled dot-product attention model. As depicted in Figure 3, each k -mer nucleotide vector in a pre-miRNA sequence can be represented as a query(Q) and key(K)-value(V) pair by the three mapping matrices. The parameters of these matrices are learned by backward propagation. The output of each k -mer nucleotide can be obtained by mapping a query and a set of KV pairs to get the weighted sum at different locations in the pre-miRNA sequence.

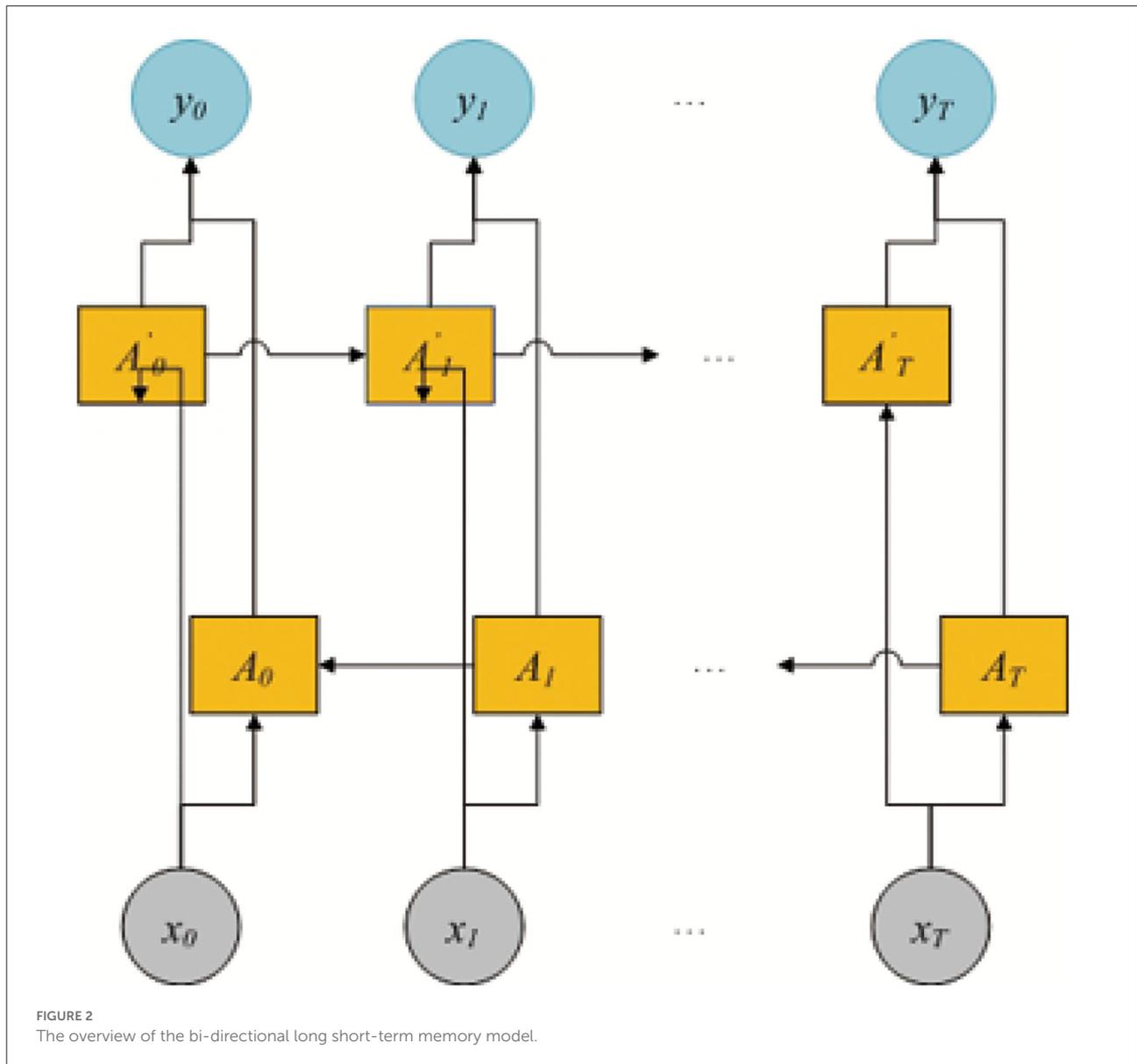
More specifically, after obtaining the BiLSTM-based feature of pre-miRNA, we can get the representation of Q and K-V matrices of each k -mer nucleotide. The dimensionality of Q and K is d_k , and the dimensionality of V is d_v . We also computed the dot products of the Q with all K, dividing each by $\sqrt{d_k}$, and applied a softmax function to obtain the weights on the values. Based on the queries, keys and values are packed into matrices Q, K, and V, the output matrix of the scaled dot-product attention model is as follows:

$$Attention(Q, K, V) = softmax(\frac{Qk^T}{\sqrt{d_k}})V \tag{3}$$

In addition, the multi-head attention mechanism can jointly attend to information from different representation subspaces at different positions. For the output multi-head attention mechanism, the computation process is defined as follows:

$$MultiHead(Q, K, V) = Concat(head_1, \dots, head_h)W^O, \tag{4}$$

where $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$,



in which the projections are mapping matrices $W_i^Q \in R^{d_{input} * d_k}$, $W_i^K \in R^{d_{input} * d_k}$, $W_i^V \in R^{d_{input} * d_v}$, and $W_0 \in R^{h d_v * d_{output}}$. The parameters setting are $d_{input} = d_b = 128$, $d_k = d_v = 64$, $d_{output} = 38$ and $h = 4$. The matrices Q, K, and V are initialized by the BiLSTM-based feature F_B . Therefore, after above process, we obtained the MS-based feature $F_M = \{F_{M1}, F_{M2}, \dots, F_{M|L} \}$ where $F_{Mi} \in R^{d_{output}}$.

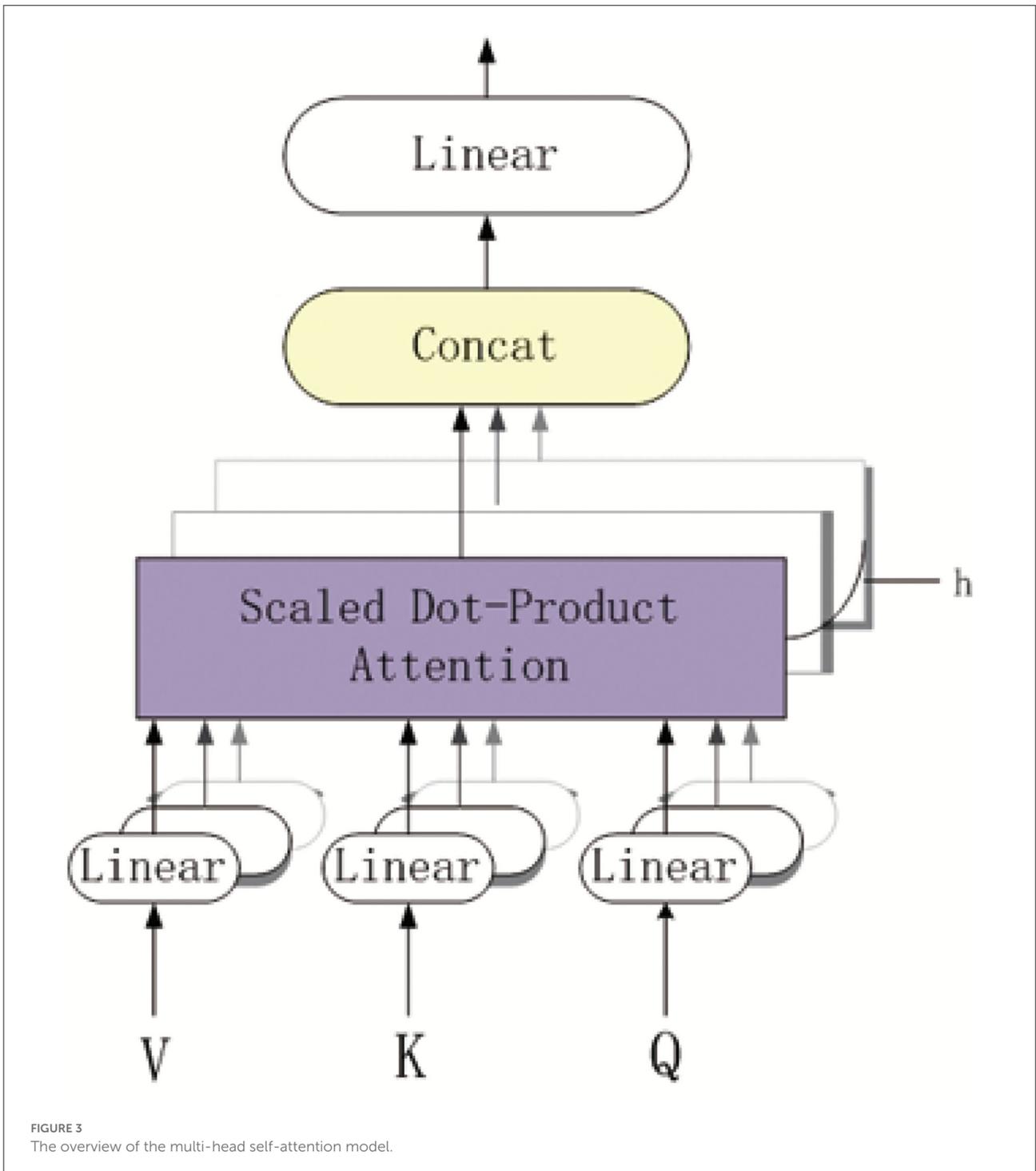
WA-based feature

After obtaining the static feature F_s and MS-based feature F_M , we further consider the importance of k-mer nucleotide to

static feature F_s . We compute which k-mer nucleotide in the pre-miRNA are more important for the static feature F_s by assigning greater weights to this k-mer nucleotide. The detail computation process of weight attention is defined as follows:

$$\begin{aligned}
 h_m &= f(w_{inter} F_s + b_{inter}), \\
 h_i &= f(w_{conv} F_{Mi}^{(i)} + b_{inter}), \\
 \alpha_i &= \sigma(h_m^T h_i),
 \end{aligned}
 \tag{5}$$

in which W_{inter} and b_{inter} are the weight matrix and bias vector, respectively. f is the rectified linear unit (ReLU) active function. Attention matrix W_{inter} represents the importance between statistic and structure feature F_s and k-mer nucleotide

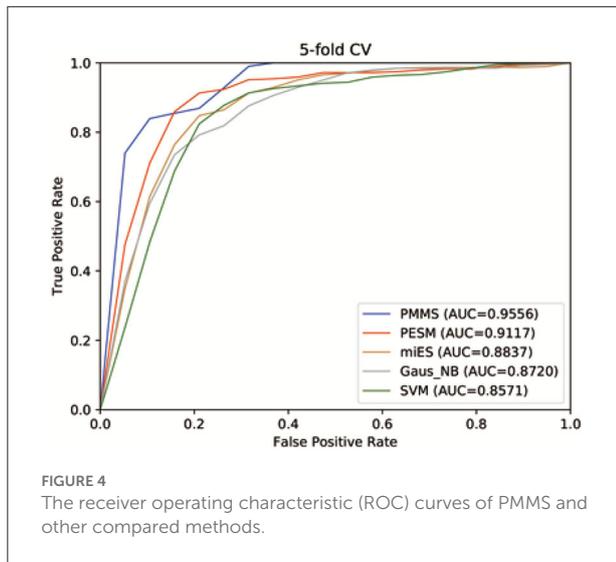


of pre-miRNA. Therefore, the WA-based feature F_W can be calculated by the weighted sum of h_i with attention and is defined as follows:

$$F_W = \sum_{i=1}^L \alpha_i h_i. \tag{6}$$

Essential miRNA prediction based on MLP

The essential miRNA prediction is a typical binary-classification problem. In this study, we used the MLP model to identify essential miRNA. After obtaining the static feature F_s and WA-based feature F_w , we concatenated them as the final



miRNA feature $F_f \in R^{d_f}$ where $d_f = 76$. Then, we took it as input for the MLP model to predict essential miRNA. The hidden vector of t -th layer can be computed as follows:

$$h_t = f(W_h h_{t-1}) + b_h, \tag{7}$$

in which W_h and b_h are the weight matrix and bias vector, and all learned by back propagation process. In addition, f is ReLU active function. Note that the input is $h_0 = F_f \in R^{d_f}$. Finally, the output vector z can be computed as follows:

$$z = (W_o h_t) + b_o, \tag{8}$$

where $W_o \in R^{2 \times d_f}$ and $b_o = R^2$ are the weight matrix and the bias vector, respectively. Based on the output vector $Z = [o_0, o_1]$, the essential miRNA probability can be computed by a softmax function and it is defined as follows:

$$p_l = \frac{\exp(o_l)}{\sum_i o_i}, \tag{9}$$

where $l \in \{0, 1\}$ is the label and p_l is the probability of label l . In addition, the cross-entropy loss is also used as the loss function and is defined as follows:

$$loss_{cross} = -\frac{1}{N} \sum_i \sum_l y_{i,l} \log(p_{i,l}), \tag{10}$$

where $y_{i,l}$ and $y_{i,l}$ are the real and predicted one-hot representation on label l of i -th sample, respectively. If the i -th sample belongs to label l , then $y_{i,l} = 1$, otherwise $y_{i,l} = 0$. N is the number of samples in the training dataset. Therefore, the

TABLE 2 The ACC, F1-score, and AUC values of five methods on the five-fold cross validation (5CV).

Method	PESM	miES	Gaus_NB	SVM	PMMS
ACC	0.8516	0.8263	0.8000	0.8206	0.9097
F1-score	0.8572	0.8326	0.8093	0.8271	0.9030
AUC	0.9117	0.8837	0.8720	0.8571	0.9556

The bold values represent the best performance.

TABLE 3 The prediction performances of predicting essential miRNAs based on the multi-head self-attention and sequences (PMMS) with different settings of k .

k	2	3	4	5	6
ACC	0.8996	0.8966	0.9097	0.8837	0.8901
F1-score	0.8856	0.8927	0.9030	0.8791	0.8832
AUC	0.9494	0.9557	0.9556	0.9421	0.9541

The bold values represent the best performance.

training objective is to minimize the function $loss$ and is defined as follows:

$$loss(\theta) = -\frac{1}{N} \sum_i \sum_l y_{i,l} \log(p_{i,l}) + \frac{\lambda}{2} \|\theta\|_2^2, \tag{11}$$

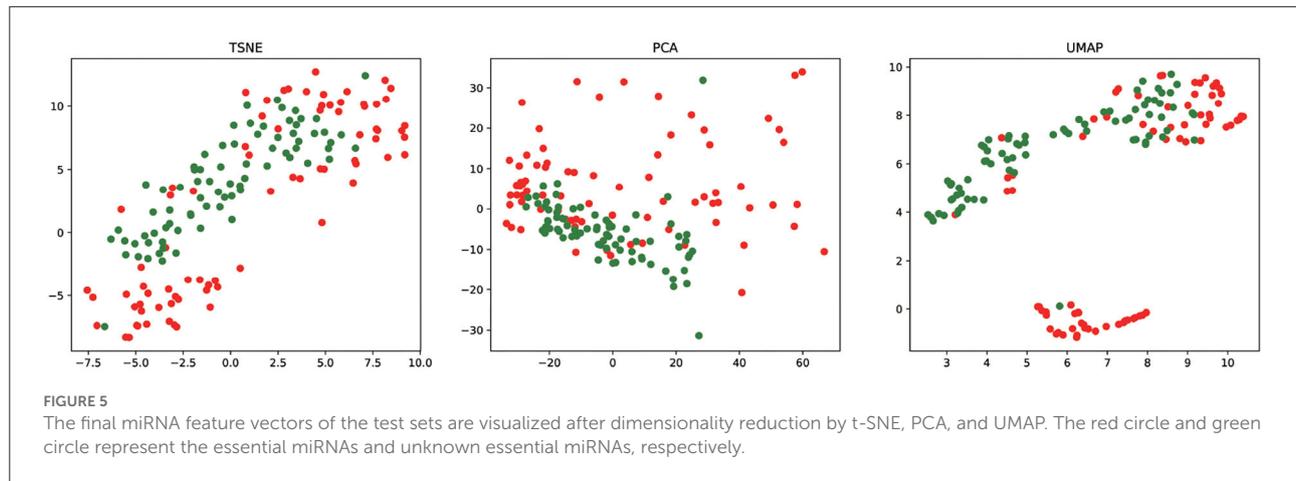
where θ is the set of all weight matrices and bias vectors. The parameter λ is the L2 regularization hyper-parameter.

Result

Comparison with previous methods

In this study, we conducted five-fold cross-validation (5CV) to evaluate the prediction performance of our methods and other compared methods, which include PESM (34), miES (27), Gaus_NB (34), and SVM (34). They are essential miRNA prediction methods. In addition, the AUC, ACC, and F1-score are used as metrics to measure the performance of prediction results. The higher the values of AUC, ACC, and F1-score are, the better the method performs. In 5CV, we randomly divided all samples into five subsets with equal size. Then, each subset is in turn considered a test sample, and the rest subsets are treated as training samples.

Figure 4 shows the receiver operating characteristics (ROC) curve and AUCs of PMMS and other compared methods. If $AUC = 1$, it would indicate that all test samples were perfectly predicted, while $AUC = 0.5$ would mean the model only had random prediction performance. We can observe from Figure 4 that PMMS obtained an AUC of 0.9556 in 5 CV. However, the AUCs of PESM, miES, Gaus_NB, and SVM are 0.9117, 0.8837, 0.8720, and 0.8571, respectively. It illustrates that our method can obtain better prediction performance than other compared methods.



Furthermore, Table 2 also demonstrates the ACC, F1-score, and AUC values of PMMS and other compared methods. PMMS obtained the ACC and F1-score values of 0.9097 and 0.9030, respectively. In addition, the best ACC and F1-score values of compared methods were 0.8516 and 0.8572, respectively. It also demonstrated that our method outperformed other compared methods according to ACC and F1-score values.

Model and parameter analysis

In this study, we also analyzed the influence of prediction performance on different parameters. Furthermore, we also analyzed the feature learning ability of our method.

The parameter k is used to obtain the BiLSTM-based feature in k -mer of pre-miRNA sequences. We can observe from Table 3 that PMMS achieved better prediction results when k is set to be 3 (AUC:0.9557) or 4 (ACC:0.9097, F-measure:0.9030) on the 5CV. Our method has stable prediction performance when k ranged from 3 to 6, the AUC values were 0.9494, 0.9557, 0.9556, 0.9421, and 0.9541, respectively. Therefore, we set the default value of parameter k to 4.

To analyze the feature learning ability of our method, we project the feature vectors into the two-dimensional feature space and visualize the result of essential miRNA classification based on the 5CV. We take the final miRNA feature as input to t-SNE (45), PCA (46), and UMAP (47) for reducing the feature dimensionality. Figure 5 shows the visualization result of essential miRNA classification based on the final miRNA feature by reducing the feature dimensionality by three methods. We can observe from Figure 5 that PMMS can distinguish essential miRNA from unknown essential miRNA in three feature dimensionality reduction methods. In addition, it also shows that compared with t-SNE and PCA dimensionality reduction methods, UMAP is relatively more obvious in distinguishing the samples.

Discussion

MiRNAs are an important class of single-stranded ncRNA molecules that have close association with human diseases. In addition, some miRNAs are also essential through knocking out gene experiments. Due to the importance of miRNA to human disease, it is very urgent to identify potential essential miRNAs. It is also very important to systematically understand the mechanisms of the etiology and pathogenesis of diseases. However, since identifying potential essential miRNAs *via* biomedical experiments is expensive and time-consuming, the effective computational methods for essential miRNAs prediction are in demand. Currently, some essential miRNAs prediction methods have been proposed by researchers. They also provide a basis for the development of new computational methods.

Conclusion

In this study, we also proposed a new computational method (PMMS) to predict essential miRNAs. PMMS first calculated the statistics and structure features based on the pre-miRNA and miRNA sequences. By considering the timing characteristic of the pre-miRNA sequence, we also obtained the original deep learning feature by the BiLSTM model. The multi-head self-attention-based feature is obtained by original deep learning feature and multi-head self-attention mechanism. Furthermore, by further considering the importance of the subsequence of pre-miRNA to statistics and structure feature of miRNA, the final deep learning feature is obtained by weight attention mechanism with the statistics and structure feature and multi-head self-attention-based feature. Finally, we concatenated the statistics and structure feature and the final deep learning feature as an input to the MLP model to predict essential miRNAs. The experiment

results demonstrated that our method outperformed other compared methods and is an effective essential miRNA prediction approach.

However, despite the effectiveness of PMMS as discussed above, some limits also exist in this method. The first limitation is that the number of known essential miRNAs is relatively small based on the limit of benchmark dataset. We would construct a new benchmark dataset by extracting the essential miRNAs from published literatures. In addition, other biological networks of miRNAs should be considered, such as miRNA-target associations. Furthermore, new deep learning model should also be considered based on added biological networks, such as GCN and other models (48–50). Therefore, we would develop new computational method to improve the ACC of essential miRNA prediction method in the future.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.cuilab.cn/mies>.

Author contributions

CY abstracted the main ideas, conducted a comparative assessment, analyzed the results, and wrote the manuscript. CD and GD conceived the subject, instructed the experimental design, analyzed the results, and revised the manuscript. All authors contributed to the article and approved the submitted version.

References

- Diederichs S, Haber DA. Dual role for argonautes in microRNA processing and posttranscriptional regulation of microRNA expression. *Cell*. (2007) 131:1097–108. doi: 10.1016/j.cell.2007.10.032
- Gu S, Jin L, Zhang F, Sarnow P, Kay MA. Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat Struct Mol Biol*. (2009) 16:144–50. doi: 10.1038/nsmb.1552
- O'Brien J, Hayder H, Zayed Y, Peng C. Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Front Endocrinol*. (2018) 9:402. doi: 10.3389/fendo.2018.00402
- Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*. (1993) 75:843–54. doi: 10.1016/0092-8674(93)90529-Y
- Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, et al. The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*. (2000) 403:901–6. doi: 10.1038/3502607
- Jopling CL, Yi M, Lancaster AM, Lemon SM, Sarnow P. Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA. *Science*. (2005) 309:1577–81. doi: 10.1126/science.1113329
- Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. (2010) 39:D152–7. doi: 10.1093/nar/gkq1027
- Vahdat Lasemi F, Mahjoubin Tehran M, Aghaee-Bakhtiari S, Jalili A, Jaafari M, Sahebkar A. Harnessing nucleic acid-based therapeutics for atherosclerotic cardiovascular disease: state of the art. *Drug Discovery Today*. (2019) 24:1116–31. doi: 10.1016/j.drudis.2019.04.007
- Cheng AM, Byrom MW, Shelton J, Ford LP. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res*. (2005) 33:1290–7. doi: 10.1093/nar/gki200
- Chen CZ, Li L, Lodish HF, Bartel DP. MicroRNAs modulate hematopoietic lineage differentiation. *Science*. (2004) 303:83–6. doi: 10.1126/science.109190
- Karp X, Ambros V. Encountering microRNAs in cell fate signaling. *Science*. (2005) 310:1288–9. doi: 10.1126/science.1121566
- Chen YJ, Guo YN, Shi K, Huang HM, Huang SP, Xu WQ, et al. Down-regulation of microRNA-144-3p and its clinical value in non-small cell lung cancer: a comprehensive analysis based on microarray, miRNA-sequencing, and quantitative real-time PCR data. *Respirat Res*. (2019) 20:1–18. doi: 10.1186/s12931-019-0994-1
- Shimono Y, Zabala M, Cho RW, Lobo N, Dalerba P, Qian D, et al. Downregulation of miRNA-200c links breast cancer stem cells with normal stem cells. *Cell*. (2009) 138:592–603. doi: 10.1016/j.cell.2009.07.011
- Li T, Leong MH, Harms B, Kennedy G, Chen L. MicroRNA-21 as a potential colon and rectal cancer biomarker. *World J Gastroenterol*. (2013) 19:5615. doi: 10.3748/wjg.v19.i34.5615

Funding

The authors would like to express their gratitude for the support from the National Natural Science Foundation of China (Nos. 61962050 and 62072473), Natural Science Foundation of Hunan Province of China (No. 2022JJ30428), Science and Technology Foundation of Guizhou Province of China under Grant No. [2020]1Y264, and Natural Science Foundation of Education of Guizhou Province (No. KY[2017]351).

Acknowledgments

The authors are very grateful to the reviewers for their constructive comments which have helped significantly in revising this work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

15. Xu L, Li Y, Yin L, Qi Y, Sun H, Sun P, et al. miR-125a-5p ameliorates hepatic glycolipid metabolism disorder in type 2 diabetes mellitus through targeting of STAT3. *Theranostics*. (2018) 8:5593. doi: 10.7150/thno.27425
16. Zhang C. MicroRNA-145 in vascular smooth muscle cell biology: a new therapeutic target for vascular disease. *Cell Cycle*. (2009) 8:3469–73. doi: 10.4161/cc.8.21.9837
17. Pers Y, Djouad F, Duroux-Richard I, Apparailly F, Noel D, Jorgensen C. The micromRNAs as biomarkers in knee osteoarthritis. *Ann Rheumat Dis*. (2010) 69(Suppl 2):A48. doi: 10.1136/ard.2010.129635n
18. Lovat F, Fassin M, Gasparini P, Rizzotto L, Cascione L, Pizzi M, et al. miR-15b/16-2 deletion promotes B-cell malignancies. *Proc Natl Acad Sci USA*. (2015) 112:11636–41. doi: 10.1073/pnas.1514954112
19. Ding L, Zhang Y, Han L, Fu L, Mei X, Wang J, et al. Activating and sustaining c-Myc by depletion of miR-144/451 gene locus contributes to B-lymphomagenesis. *Oncogene*. (2018) 37:1293–307. doi: 10.1038/s41388-017-0055-5
20. Griffiths-Jones S, Saini HK, Van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res*. (2007) 36:D154–8. doi: 10.1093/nar/gkm952
21. Nam S, Kim B, Shin S, Lee S. miRigator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res*. (2007) 36:D159–64. doi: 10.1093/nar/gkm829
22. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res*. (2009) 37:D98–104. doi: 10.1093/nar/gkn714
23. Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, et al. HMDD v2. 0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res*. (2014). 42:D1070–4. doi: 10.1093/nar/gkt1023
24. Huang Z, Shi J, Gao Y, Cui C, Zhang S, Li J, et al. HMDD v3. 0: a database for experimentally supported human microRNA-disease associations. *Nucleic Acids Res*. (2019). 47:D1013–7. doi: 10.1093/nar/gky1010
25. Barupal JK, Saini AK, Chand T, Meena A, Beniwal S, Suthar JR, et al. ExclmiRDB for translational genomics: a curated online resource for extracellular microRNAs. *Omic J Integrat Biol*. (2015) 19:24–30. doi: 10.1089/omi.2014.0106
26. Xie B, Ding Q, Han H, Wu D. miRCancer: a microRNA-cancer association database constructed by text mining on literature. *Bioinformatics*. (2013) 29:638–44. doi: 10.1093/bioinformatics/btt014
27. Song F, Cui C, Gao L, Cui Q. miES: predicting the essentiality of miRNAs with machine learning and sequence features. *Bioinformatics*. (2019) 35:1053–4. doi: 10.1093/bioinformatics/bty738
28. Chen X, Xie D, Wang L, Zhao Q, You ZH, Liu H. BNPMDA: bipartite network projection for MiRNA-disease association prediction. *Bioinformatics*. (2018) 34:3178–86. doi: 10.1093/bioinformatics/bty333
29. Yan C, Wang J, Ni P, Lan W, Wu FX, Pan Y. DNRLMF-MDA: predicting microRNA-disease associations based on similarities of microRNAs and diseases. *IEEE/ACM Trans Comput Biol Bioinform*. (2017) 16:233–43. doi: 10.1109/TCBB.2017.2776101
30. Chen X, Wang L, Qu J, Guan NN, Li JQ. Predicting miRNA-disease association based on inductive matrix completion. *Bioinformatics*. (2018) 34:4256–65. doi: 10.1093/bioinformatics/bty503
31. Chen X, Yin J, Qu J, Huang L. MDHGI: matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction. *PLoS Comput Biol*. (2018) 14:e1006418. doi: 10.1371/journal.pcbi.1006418
32. Li J, Zhang S, Liu T, Ning C, Zhang Z, Zhou W. Neural inductive matrix completion with graph convolutional networks for miRNA-disease association prediction. *Bioinformatics*. (2020) 36:2538–46. doi: 10.1093/bioinformatics/btz965
33. Yan C, Duan G, Li N, Zhang L, Wu FX, Wang J. PDMDA: predicting deep-level miRNA-disease associations with graph neural networks and sequence features. *Bioinformatics*. (2022) 38:2226–34. doi: 10.1093/bioinformatics/btac077
34. Yan C, Wu FX, Wang J, Duan G. PESM: predicting the essentiality of miRNAs based on gradient boosting machines and sequences. *BMC Bioinformatics*. (2020) 21:111. doi: 10.1186/s12859-020-3426-9
35. Min H, Xin XH, Chu-Qiao G, Wang L, Du PF. XGEM: predicting essential miRNA by the ensembles of various sequence-based classifiers with XGBoost algorithm. *Front Genet*. (2022) 13:877409. doi: 10.3389/fgene.2022.877409
36. Tseng KC, Chiang-Hsieh YF, Pai H, Chow CN, Lee SC, Zheng HQ, et al. microRPM: a microRNA prediction model based only on plant small RNA sequencing data. *Bioinformatics*. (2018) 34:1108–15. doi: 10.1093/bioinformatics/btx725
37. Stegmayer G, Yones C, Kamenetzky L, Milone DH. High class-imbalance in pre-miRNA prediction: a novel approach based on deepSOM. *IEEE/ACM Trans Comput Biol Bioinform*. (2016) 14:1316–26. doi: 10.1109/TCBB.2016.2576459
38. Zhou P, Shi W, Tian J, Qi Z, Li B, Hao H, et al. Attention-based bidirectional long short-term memory networks for relation classification. In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Berlin (2016). p. 207–12. doi: 10.18653/v1/P16-2034
39. Graves A. Long short-term memory. In: *Supervised Sequence Labelling with Recurrent Neural Networks*. Studies in Computational Intelligence, Berlin: Springer (2012). p. 37–45. doi: 10.1007/978-3-642-24797-2_4
40. Sun Y, Xiong F, Sun Y, Zhao Y, Cao Y. A miRNA target prediction model based on distributed representation learning and deep learning. *Comput Math Methods Med*. (2022) 2022:4490154. doi: 10.1155/2022/4490154
41. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: *Advances in Neural Information Processing Systems* 30. Long Beach, CA (2017). p. 5998–6008.
42. Cheng Z, Yan C, Wu FX, Wang J. Drug-target interaction prediction using multi-head self-attention and graph attention network. *IEEE/ACM Trans Comput Biol Bioinform*. (2021) 19:2208–18. doi: 10.1109/TCBB.2021.3077905
43. Zeng Y, Chen X, Luo Y, Li X, Peng D. Deep drug-target binding affinity prediction with multiple attention blocks. *Brief Bioinform*. (2021) 22:bbab117. doi: 10.1093/bib/bbab117
44. Wang J, Liu X, Shen S, Deng L, Liu H. DeepDDS: deep graph neural network with attention mechanism to predict synergistic drug combinations. *Brief Bioinform*. (2022) 23:bbab390. doi: 10.1093/bib/bbab390
45. Hinton G, van der Maaten L. Visualizing data using t-SNE. *J Mach Learn Res*. (2008) 9:2579–2605.
46. Abdi H, Williams LJ. Principal component analysis. *Wiley Interdiscipl Rev Comput Stat*. (2010) 2:433–59. doi: 10.1002/wics.101
47. McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:180203426*. (2018). doi: 10.21105/joss.00861
48. Li J, Liu J, Yue H, Cheng J, Kuang H, Bai H, et al. DARC: Deep adaptive regularized clustering for histopathological image classification. *Med Image Anal*. (2022) 80:102521. doi: 10.1016/j.media.2022.102521
49. Cheng J, Liu J, Kuang H, Wang J. A fully automated multimodal MRI-based multi-task learning for glioma segmentation and IDH genotyping. *IEEE Trans Med Imaging*. (2022) 41:1520–32. doi: 10.1109/TMI.2022.3142321
50. Gou N, Xiang Y, Zhou J, Zhang S, Zhong S, Lu J, et al. Identification of violent patients with schizophrenia using a hybrid machine learning approach at the individual level. *Psychiatry Res*. (2021) 306:114294. doi: 10.1016/j.psychres.2021.114294