MixLMI: A Multi-level Enhancement Model for Plant lncRNA-miRNA Interaction Prediction

Written by XMU Students[1]
Xinru Ruan, Xiaoling Huang, Pingjiang Long, Qihong Mao, Yan Jin
[1]School of Information, Xiamen University

Abstract

MicroRNA (miRNA) and long non-coding RNAs (lncRNAs) are non-coding RNAs (ncRNAs), and their interaction plays an important role in biological processes. Computational methods, such as machine learning and many bioinformatics tools, can predict potential miRNA-lncRNA interactions, which are of great significance for studying its mechanism and biological functions. More and more animal RNA interaction predictors have been studied, but due to differences, they are not reliable in the lncRNA between animals and plants. Establishing reliable plant predictors is a basic task, especially cross-species plant predictors. This paper proposes a deep learning model (MixLMI), which is a model based on multi-scales information enhancement for predicting miRNA-lncRNA interactions in plants. The use of complex feature fusion and multi-scale convolutional long short-term memory network to enhance sample information on features, scales and models verifies the positive impact of multi-scales information enhancement on prediction performance. Experiments show that MixLMI has good prediction performance and strong generalization ability, and can be used for cross-species prediction.

Introduction

MicroRNAs (miRNAs) are small molecules similar to SiRNA. It is non-coding RNAs like long non-coding RNAs.They regulate gene expression and play an important role in the regulation of the growth and development of animal and plant cells.At the same time, the interaction between miRNA and lncRNA also plays an important role in biological processes.It may play a regulatory role in the growth of plants,miRNA can target lncRNA and trigger the generation of phased small interfering RNAs that affect the seed germination in Triticum aestivum (T.aestivum) (Guo et al. 2018). lncRNA can accelerate the pro-liferation of primary cardiomyocytes by targeting miRNA(Yang et al. 2020a). lncRNA can inhibit the expression of miRNA in a variety of plants by adsorbing miRNA(Zhou et al. 2020). With the development of high-throughput sequencing technology, a large number of miRNAs and lncRNAs can be obtained with the existing technology, but the interaction between miRNAs and lncRNAs has been confirmed to be very limited, and many of its mechanisms and biological functions are still unclear. In order to understand the interaction mechanism between miRNA and lncRNA in plants, it is essential and important to determine the interaction between them.

Biological experiments are an important means to discover the interaction between miRNAs and lncRNAs, but this is expensive and time-consuming. With the rapid development of information technology, the emergence of machine learning and various bioinformatics tools makes it possible to predict potential RNA, and can save a lot of time and cost. RIblast is a powerful RNA interaction predictor and its last version is released in November 2019(Fukunaga and Hamada 2017). LncRRIsearch is a web server, which can quickly predict small-scale lncRNA–RNA interactions in human and mouse(Fukunaga et al. 2019).

Howeverthese methods were designed specifically for animals, not applicable to plants. Specifically, for animals, RNA polymerase II transcribes the NcRNAs, while RNA polymerase II, IV, and V are in charge of this in plants (Movahedi et al. 2015). What's more, lncRNA has low sequence conserved, especially in distant species such as animals and plants (Noviello et al. 2018). Therefore, animal predictions are unreliable in plants. It is hoped that a limited number of plant species can be used to train prediction models of plant species. Therefore, it is necessary to make cross-species prediction of plants. Four predictors of miRNA-lncRNA interactions in plants have been reported (Bouba et al. 2019),(Kang et al. 2020),(Zhang et al. 2020),(Song et al. 2020). They made important contributions, but three of them did not validate predictive performance across species. Although the published model PmliPred(Kang et al. 2020) has successfully achieved cross-species prediction, it has only been tested on two dicotyledons (Arabidopsis thaliana and lycopene). Since PmliPred was trained with dicotyledonous plant data, its generalization ability to more plants, especially monocotyledons, needs to be further verified. PmliPred still has room for improvement in terms of performance and generalization capability. In addition, we have not found any reports of predictors of miRNA-lncRNA interactions in plants. It is an urgent task to establish a reliable plant predictor, especially a cross-species plant predictor. Cross-species prediction requires predictors with strong generalization ability. Ensemble learning has strong generalization ability and has been applied in bioinformatics (Liu

et al. 2018). Deep learning model is a hot spot of current research, showing good performance (Zhang et al. 2019), and can be used as the basic model of integrated learning. There are two common approaches to training deep learning models with biological data. One is to encode the data as the input of training (Zhang et al. 2020), the other is to extract features from the data as the input of training (Peng et al. 2019). It solved the problem of low sequence preservation of lncRNA in different plants and achieved good results in cross-species prediction. Inspired by this, we think that feature extraction of different values may contain more sample information than encoded data. For example, one-hot coding uses matrices with only five numbers to represent different RNA sequences, while *k-mer* features can represent RNA sequences with normalized vectors or matrices with different values.

By introducing the method above, an integrated deep learning model MixLMI based on multi-level information enhancement is proposed to predict miRNA-lncRNA interactions in plants. It integrates complex features through nonlinear transformation of feature extraction (Dai et al. 2019). It takes into account the sequence features and structural features and enhances the information at the feature level. Conv-LSTM, as the basic model, converts the fused complex feature vectors into three matrices of different scales.(Shen, Deng, and Huang 2019).

Related Work

In recent years, considerable effort has been devoted to developing computational methods for identifying associations in multiple biological data sets. At present, in the prediction of the interaction between miRNA and lncRNA, many researchers have used shallow machine-learning methods to construct the prediction model through feature selection.

RIblast is a powerful RNA interaction predictor and its last version is released in November 2019(Fukunaga and Hamada 2017). It uses suffix arrays to discover seed regions, and then extends seed regions based on an RNA secondary structure energy model. LncRRIsearch is a web server, which can quickly predict small-scale lncRNA–RNA interactions in human and mouse(Fukunaga et al. 2019). GCLMI(Huang et al. 2019) is an end-to-end model that combines the technologies of graph convolution and autoencoder for miRNA–lncRNA interaction prediction. A hybrid sequence feature-based model, named LncMirNet, is pro-posed to predict miRNA–lncRNA interactions via deep convolutional neural networks(Yang et al. 2020b). PmliPred(Kang et al. 2020) proposed a new method based on hybrid model and fuzzy decision, PmliPred, that was applied to plant miRNA–lncRNA interactions prediction. It hybridizes CNN-BiGRU and RF, utilizes raw sequences and manually extracted features. PmliPred obtains better performance and generalization ability compared with existing methods. By the biological experiments, several new miRNA–lncRNA interactions in S.lycopersicum are successfully identified from the candidates predicted by PmliPred, which further verifies its effectiveness. Zhang *et al.*(Zhang et al. 2020) introduce the plant miRNA–lncRNA interaction prediction with the ensemble of CNN and IndRNN they uses the two-stage convolutional neural network to automatically learn sequence features and detect functional domains of nucleotide sequences, and then uses the two-layer independently recurrent neural network (IndRNN) to learn the long-term dependence in functional domains to classify data. It obtains above 96 percent accuracy on Zea mays test set and better results on other plant data sets. This shows its good performance and generalization ability.

Materials and Methodology

Datasets

As there is no public plant miRNA-lncRNA interaction database available, we plan to construct a new credible dataset of variety of plants, using previous method(Kang et al. 2020). For data sources, miRNA and lncRNA are downloaded from miRBase 22.1(Kozomara, Birgaoanu, and Griffiths-Jones 2019) and GreenNC v1.12(Gallart et al. 2016). By referring to the used species in the training set of RNAplonc(Negri et al. 2019), the samples of A. thaliana, Glycine max (G. max), Oryza sativa (O. sativa), and P. trichocarpa are selected to be a training-validation set. For constructing balanced sample test setsthe samples of Brachypodium distachyon (B. distachyon), Medicago truncatula (M. truncatula), and Solanum tuberosum (S. tuberosum) are selected, respectively. Detailed information about the constructed datasets is shown in Table 1.

Fusing Complex Features

The construction of complex features is mainly realized by nonlinear transformation of features. In 2019, Dai et al. (Dai et al. 2019) proved that the construction of complex features has the potential to be applied to LMI prediction problems when calculating and predicting ncRNA-protein interaction prediction. But there are still two problems. First of all, this article uses K-mer to construct complex features, they contain limited sample information, but also need to consider structural features. Secondly, it also mentioned four common strategies when constructing complex features: geometric average, harmonic average, PowRP and PowPR. But they are not all applicable to LMI prediction problems. For miRNA, because it is a short sequence, its characteristic value must be 0, and the same is true for lncRNA. So the last three strategies are meaningless to them. For the geometric average strategy, as long as there is a feature value of 0 in miRNA and lncRNA, the constructed complex feature is also 0. This will make the constructed feature matrix sparse and have many identical complex feature values, thereby losing a lot of Information. Here, we propose arithmetic averaging to construct complex features, so as to avoid information loss, while fusing these complex features to characterize miRNA-lncRNA pairs, which greatly enhances the ability to characterize information. The RNA sequence is composed of four bases: adenine (A), thymine (T), cytosine (C) and guanine (G). (Uracil (U) in miRNA is converted to T for the convenience of feature extraction.) K-mers are subsequences of length k included in biological sequences, and g-gap represents a discontinuous base

Table 1: Details of datasets

Dataset	Species	Type	Number of positive sample	Number of negative sample
Training-validation set				
1	A.thaliana	Dicotyledon	1200	1200
2	G.max	Dicotyledon	1200	1200
3	O.sativa	Monocotyledon	1200	1200
4	P.trichocarpa	Dicotyledon	1200	1200
Test sets				
1	B.distachyon	DMonocotyledon	500	500
2	M.truncatula	Dicotyledon	500	500
3	S.tuberosum	Monocotyledon	500	500

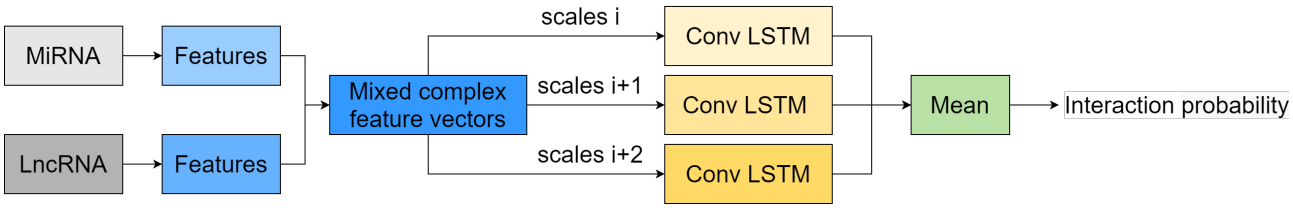

Figure 1: Overview of the MixLMI

sequence. (Zhang et al. 2021) In this article, we extracted 1-mer, 2mer and 3mer, as well as 1-gap, 2-gap and 3-gap sequences.

The secondary structure of RNA is obtained by RNAfold, which is indicated by dots and brackets(Lorenz et al. 2011). The sliding window is used to match each feature along the sequence or structure character. Calculate the frequency f of each feature and normalize it to:

$$f = \frac{c}{n^{m-l}(L-l-1)} \tag{1}$$

Among them, c is the number of matched feature types, n is the number of matched feature types, m is the maximum length of the matched feature types, L is the length of the sequence, and l is the length of the sliding window. The complex features can be constructed by the arithmetic averaging strategy as:

$$cf_{i,j} = \frac{(mf_i + lf_i)}{2} \tag{2}$$

mf_i is the i-th feature frequency of miRNA, lf_i is the j-th feature frequency of lncRNA, and $cf_{i,j}$ is the constructed complex feature value. For sequence features and structural features, their k-mer and g-gap features can be obtained respectively:

$$Kmer_{seq} = \left[cf_{ks1,1}, \ldots, cf_{ks1,84}, cf_{ks2,1}, \ldots, cf_{ks84,84}\right] \tag{3}$$

$$Ggap_{seq} = \left[cf_{gs1,1}, \ldots, cf_{gs1,48}, cf_{ks2,1}, \ldots, cf_{ks48,48}\right] \tag{4}$$

$$Kmer_{str} = \left[cf_{kt1,1}, \ldots, cf_{kt1,14}, cf_{kt2,1}, \ldots, cf_{kt14,14}\right] \tag{5}$$

$$Ggap_{str} = \left[cf_{gt1,1}, \ldots, cf_{gt1,48}, cf_{gt2,1}, \ldots, cf_{gt48,48}\right] \tag{6}$$

$cf_{ksi,j}$ and $cf_{gsi,j}$ can be obtained from sequence features, and $cf_{kti,j}$ and $cf_{gti,j}$ can Obtained from structural features. They are fused into a 11860-dimensional vector to describe the sample as

$$Fcf = \left[Kmer_{seq}, Ggap_{seq}, Kmer_{str}, Ggap_{str}\right] \tag{7}$$

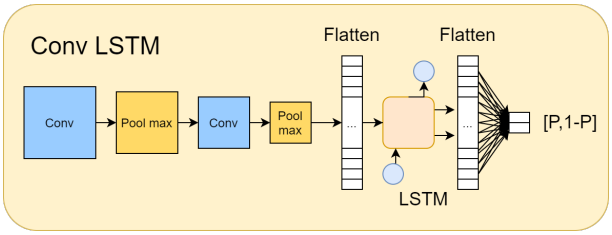

Figure 2: Architecture of ConvLSTM. Two convolution layers with 32 and 64 filters are used, followed by Pool-max, which is a pooling layer with max pooling scheme. Dropout is applied in the last fully connected layer with rate of 0.5, respectively.

Multi-scale ConvLSTM Model

LSTM is a special kind of RNN, which is used to solve the problem of explosion decomposition and explosion during training in the process of long sequence. The performance of the model combined with LSTM and CNN tends to perform better(LeCun, Bengio, and Hinton 2015),(Vidal and Kristjanpoller 2020). As shown in Fig.1. The mixed feature vector is input to the base model in the form of a matrix. Use Eq.(8) to extract feature maps

$$F_e = ReLU\left(X \bigotimes k_e\right) \tag{8}$$

F_e is the eth feature map, X is the input fusion feature matrix with different scales, k_e is the e-th convolution kernel

with different scales, and is the convolution operation. The maximum pooling is used to reduce the dimensionality of the feature map, and the flatten layer is used to compress the output of the previous layer. A fully connected layer is used as the classifier, and the "Adam" optimizer is used in the experiment, where the classification loss function uses cross-entropy loss. At the same time, we used droupout to prevent the model from overfitting. The output of the model is a 2-d [p,1-p] vector, where p is the probability of an interaction between miRNA and lncRNA in the sample. In the training process, the model parameters are updated through the backward propagation of the model.

We believe that only a single-scale feature map is not enough for model learning(Shen, Deng, and Huang 2019), so we will use multi-scale features with scales of 2×5930, 3×3954, 4×2965, where the empty elements are zero-padded. Pass the obtained features of different scales through the ConvLSTM module which its architecture is shown in Fig.2, and perform an averaging operation on the final output to obtain the final confidence probability that there has been interaction between miRNA and lncRNA in the sample.

By using multiple scales, the diversity of feature pictures can be mined, and the feature information of the data can be preserved as much as possible.

Results and Discussion

Performance Evaluation Criteria

Sensitivity, specificity, accuracy, F1 score, and geometric mean of sensitivity and specificity (GEP) are used as the performance evaluation criteria:

$$\mathbf{Sensitivity} = \frac{TP}{TP + FN} \tag{9}$$

$$\mathbf{Specificity} = \frac{TN}{TN + FP} \tag{10}$$

$$\mathbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$\mathbf{F1\ score} = \frac{\mathbf{2}TP}{\mathbf{2}TP + FP + FN} \tag{12}$$

$$\mathbf{GEP} = \sqrt{\frac{TP}{TP + FN} * \frac{TN}{TN + FP}} \tag{13}$$

where true positive (TP) is the number of correctly predicting miRNA–lncRNA interaction, false negative (FN) is the number of incorrectly predicting miRNA–lncRNA interaction, false positive (FP) is the number of incorrectly predicting that there has no interaction between miRNA and lncRNA, and true negative (TN) is the number of correctly predicting that there has no interaction between miRNA and lncRNA. In addition, AUC value from ROC curve is also used for evaluations.

Table 2: Results of 10-fold cross validation,we list the average, maximum, and minimum accuracy of models among the ten folds.

Model	Average Accuarcy(%)	Maximum Accuarcy(%)	Minimum Accuarcy(%)
MixLMI	**85.8**	87.1	**85.1**
MixLMI(kmer)	85.2	86.9	84.5
MixLMI(GM)	85.3	86.7	83.7
LMI(2-5930)	85.3	**87.3**	83.5
LMI(3-3954)	85.4	86.2	84.8
LMI(4-2965)	85.1	86.7	82.3

Table 3: Results of 5-fold cross validation, the comparison between MixLMI and MixLMI(residual) which add skip-connection in the ConvLSTM of MixLMI.

Model	Average Accuarcy(%)	Maximum Accuarcy(%)	Minimum Accuarcy(%)
MixLMI	**85.5**	87.2	**84.4**
MixLMI(residual)	85.3	**87.5**	83.6

Experiments

In order to verify the advantages of the fused complex feature module in MixLMI, we use k-mer complex feature vectors to replace the original MixLMI fused complex feature vectors, denoted as MixLMI(k-mer).At the same time, we think that the arithmetic mean strategy used to build the complex features in MixLMI can also have other options. Here we choose the typical geometric mean strategy as a replacement, denoted as MixLMI (GM).

MixLMI uses three scales of "2×5930", "3×3954", and "4×2965", and uses these three scales to fully mine the potential information in the data, and merge the three kinds of information together to make predictions. In order to verify the efectiveness of multiple scales in MixLMI, we designed three single-scale base models of MixLMI, denoted as LMI(2-5930), LMI(3-3954), LMI(4-2965), each of them use one scale to make predictions.

The experimental results are given in the table 2. It can be seen that although the peak accuracy of MixLMI (k-mer) is only slightly lower than that of MixLMI, its average accuracy is significantly lower than that of MixLMI, which verifies the advantages of fused complex feature module in MixLMI. For MixLMI (GM), the accuracy fluctuates greatly in 10 folds, and the minimum accuracy is lower than others. The overall performance is not as good as MixLMI. We think this is because arithmetic mean strategy can produce more consistent features and improve the stability of the model. For the three single-scale models, the accuracy of LMI (2-5930) fluctuates greatly. In some folds, it can exceed the performance of MixLMI, but it also has quite low accuracy in other folds. In terms of average accuracy, it is slightly higher than LMI (4-2965) and not as good as MixLMI. LMI (3-3954) has less fluctuations in accuracy, and the average accuracy is the highest among the three single-scale models. LMI (4-2965) performed poorly in both the minimum ac-

Table 4: Results of the predictors on balanced sample test sets.

Test set	Predictor	Sensitivity(%)	Specificity(%)	Accuracy(%)	F1 score(%)	AUC
B.distachyon	Riblast	53.53	82.87	67.78	62.42	NA
	LncMirNet	87.53	2.31	45.62	60.64	0.4024
	CIRNN	84.62	**84.29**	86.8	85.8	0.9357
	MixLMI	**93.73**	80.71	**86.98**	**86.95**	**0.9683**
M.truncatula	Riblast	42.81	**92.42**	68.76	58.82	NA
	LncMirNet	**87.72**	6.36	45.82	61.74	0.4572
	CIRNN	75.32	70.72	72.72	73.59	0.8063
	MixLMI	84.52	82.84	**84.81**	**84.77**	**0.9071**
S.tuberosum	Riblast	42.31	**88.32**	65.89	57.82	NA
	LncMirNet	**92.83**	3.69	46.34	63.9	0.4591
	CIRNN	85.24	63.83	75.73	77.58	0.8307
	MixLMI	80.12	81.82	**80.64**	**80.7**	**0.9011**

curacy and the average accuracy. In general, MixLMI integrates three scales for prediction, which is a certain improvement compared to the single-scale model. However, we also found that although LMI (2-5930) uses a more extreme scale setting, it can also compete with MixLMI in terms of peak accuracy.

Inspired by ResNet(He et al. 2016), we also tried to add skip-connection to the model. More specifically, we add skip-connection in the LSTM layer of ConvLSTM, denoted as MixLMI (residual).In the experiment, in order to show the effect of skip-connection, we adopted 5-fold cross validation. The experimental results are listed in Table 3. The highest accuracy of MixLMI (residual) is higher than MixLMI, and the average accuracy is comparable to MixLMI. However, the minimum accuracy of MixLML (residual) is lower than MixLMI, and we think this may be related to the training method.

MixLMI is compared with state-of-the-art predictors on a balanced sample test set. Compared predictors include RIblast(Fukunaga and Hamada 2017), LncMirNet(Yang et al. 2020b) and CIRNN(Zhang et al. 2020), which are representatives of RNA interaction, miRNA-lncRNA interaction and miRNA-lncRNA interaction in plants, respectively. To make the results more objective, MixLMI and CIRNN separately made 10 independent predictions for each test set to obtain the average results. RIblast and LncMirNet are encapsulation tools and training models respectively, which directly output prediction results with miRNA and lncRNA sequences as inputs.The results of these predictors are shown in Table 4. As can be seen from Table 4, RIblast has the highest specificity against M. amputate and M. tuberosity, followed by M. spikelet. Its sensitivity was the worst in the three test sets, which adversely affected its accuracy and F1 score. LncMirNet had the highest sensitivity to truncated spikelets and late spikelets, followed by double spikelets. Its specificity is unacceptable, which greatly reduces its accuracy and F1 scores. Its AUC value was the worst of the three test sets. In b. Diachyon's test set, CIRNN had the best specificity, with its accuracy and F1 score slightly lower than MixLMI and far better than RIblast and LncMirNet, with specificity greater than 84 percent. In the other two test sets, except for the sensitivity of tuberose test set, the results of CIRNN did not reach 80 percent. In the three test sets, the AUC value of CIRNN takes the second place. MixLMI achieved the best sensitivity on B. Diachyon's test set, and the best accuracy, F1 scores, and AUC values on all three test sets. In other cases, the results were above 80 percent, although not the best. These results suggest that RIblast tends to predict miRNA-lncRNA interactions in samples without interaction, LncMirNet predicts a large number of false positive samples, and CIRNN does well at predicting individual plants, but is not satisfied with predicting more plants. In general, MixLMI is the most reliable predictor.

Conclusion

In this paper, we propose an multi-scales information enhancement (MixLMI) for plant miRNA–lncRNA interaction prediction. With strong generalization ability, MixLMI achieves well performance than PmliPred in cross-species prediction. Being beneficial to the exploration of plant biological functions, this method may provide valuable references for related research. In the future, we will train more base models and build a powerful predictor using ensemble pruning technology. Beyond that, the contribution of our research to database construction is also worth expecting as there is currently no plant miRNA–lncRNA interaction database.

References

Bouba, I.; Kang, Q.; Luan, Y.-S.; and Meng, J. 2019. Predicting miRNA-lncRNA interactions and recognizing their regulatory roles in stress response of plants. *Mathematical biosciences*, 312: 67–76.

Dai, Q.; Guo, M.; Duan, X.; Teng, Z.; and Fu, Y. 2019. Construction of complex features for computational predicting ncRNA-protein interaction. *Frontiers in genetics*, 10: 18.

Fukunaga, T.; and Hamada, M. 2017. RIblast: an ultrafast RNA–RNA interaction prediction system based on a seed-and-extension approach. *Bioinformatics*, 33(17): 2666–2674.

Fukunaga, T.; Iwakiri, J.; Ono, Y.; and Hamada, M. 2019. LncRRIsearch: a web server for lncRNA-RNA interaction prediction integrated with tissue-specific expression and subcellular localization data. *Frontiers in genetics*, 10: 462.

Gallart, A. P.; Pulido, A. H.; De Lagrán, I. A. M.; Sanseverino, W.; and Cigliano, R. A. 2016. GREENC: a Wiki-based database of plant lncRNAs. *Nucleic acids research*, 44(Database issue): D1161.

Guo, G.; Liu, X.; Sun, F.; Cao, J.; Huo, N.; Wuda, B.; Xin, M.; Hu, Z.; Du, J.; Xia, R.; et al. 2018. Wheat miR9678 affects seed germination by generating phased siRNAs and modulating abscisic acid/gibberellin signaling. *The Plant Cell*, 30(4): 796–814.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Huang, Y.-A.; Huang, Z.-A.; You, Z.-H.; Zhu, Z.; Huang, W.-Z.; Guo, J.-X.; and Yu, C.-Q. 2019. Predicting lncRNA-miRNA interaction via graph convolution auto-encoder. *Frontiers in genetics*, 10: 758.

Kang, Q.; Meng, J.; Cui, J.; Luan, Y.; and Chen, M. 2020. PmliPred: a method based on hybrid model and fuzzy decision for plant miRNA–lncRNA interaction prediction. *Bioinformatics*, 36(10): 2986–2992.

Kozomara, A.; Birgaoanu, M.; and Griffiths-Jones, S. 2019. miRBase: from microRNA sequences to function. *Nucleic acids research*, 47(D1): D155–D162.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *nature*, 521(7553): 436–444.

Liu, B.; Li, K.; Huang, D.-S.; and Chou, K.-C. 2018. iEnhancer-EL: identifying enhancers and their strength with ensemble learning approach. *Bioinformatics*, 34(22): 3835–3842.

Lorenz, R.; Bernhart, S. H.; Höner zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P. F.; and Hofacker, I. L. 2011. ViennaRNA Package 2.0. *Algorithms for molecular biology*, 6(1): 1–14.

Movahedi, A.; Sun, W.; Zhang, J.; Wu, X.; Mousavi, M.; Mohammadi, K.; Yin, T.; and Zhuge, Q. 2015. RNA-directed DNA methylation in plants. *Plant cell reports*, 34(11): 1857–1862.

Negri, T. d. C.; Alves, W. A. L.; Bugatti, P. H.; Saito, P. T. M.; Domingues, D. S.; and Paschoal, A. R. 2019. Pattern recognition analysis on long noncoding RNAs: a tool for prediction in plants. *Briefings in bioinformatics*, 20(2): 682–689.

Noviello, T. M.; Di Liddo, A.; Ventola, G. M.; Spagnuolo, A.; D'Aniello, S.; Ceccarelli, M.; and Cerulo, L. 2018. Detection of long non–coding RNA homology, a comparative study on alignment and alignment–free metrics. *BMC bioinformatics*, 19(1): 1–12.

Peng, C.; Han, S.; Zhang, H.; and Li, Y. 2019. RPITER: a hierarchical deep learning framework for ncRNA–protein interaction prediction. *International journal of molecular sciences*, 20(5): 1070.

Shen, Z.; Deng, S.-P.; and Huang, D.-S. 2019. RNA-protein binding sites prediction via multi scale convolutional gated recurrent unit networks. *IEEE/ACM transactions on computational biology and bioinformatics*, 17(5): 1741–1750.

Song, J.; Tian, S.; Yu, L.; Yang, Q.; Xing, Y.; Zhang, C.; Dai, Q.; and Duan, X. 2020. MD-MLI: Prediction of miRNA-lncRNA Interaction by Using Multiple Features and Hierarchical Deep Learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

Vidal, A.; and Kristjanpoller, W. 2020. Gold volatility prediction using a CNN-LSTM approach. *Expert Systems with Applications*, 157: 113481.

Yang, L.; Lu, Y.; Ming, J.; Pan, Y.; Yu, R.; Wu, Y.; and Wang, T. 2020a. SNHG16 accelerates the proliferation of primary cardiomyocytes by targeting miRNA-770-5p. *Experimental and Therapeutic Medicine*, 20(4): 3221–3227.

Yang, S.; Wang, Y.; Lin, Y.; Shao, D.; He, K.; and Huang, L. 2020b. LncMirNet: predicting LncRNA–miRNA interaction based on deep learning of ribonucleic acid sequences. *Molecules*, 25(19): 4372.

Zhang, P.; Meng, J.; Luan, Y.; and Liu, C. 2020. Plant miRNA–lncRNA interaction prediction with the ensemble of CNN and IndRNN. *Interdisciplinary Sciences: Computational Life Sciences*, 12(1): 82–89.

Zhang, Y.; Jia, C.; Fullwood, M. J.; and Kwoh, C. K. 2021. DeepCPP: a deep neural network based on nucleotide bias information and minimum distribution similarity feature selection for RNA coding potential prediction. *Briefings in bioinformatics*, 22(2): 2073–2084.

Zhang, Z.; Zhao, Y.; Liao, X.; Shi, W.; Li, K.; Zou, Q.; and Peng, S. 2019. Deep learning in omics: a survey and guideline. *Briefings in functional genomics*, 18(1): 41–57.

Zhou, X.; Cui, J.; Meng, J.; and Luan, Y. 2020. Interactions and links among the noncoding RNAs in plants under stresses. *Theoretical and Applied Genetics*, 1–14.