Prediction of RNA secondary structure with pseudoknots using coupled deep neural networks

Kangkun Mao¹, Jun Wang¹, Yi Xiao¹✉

¹ School of Physics and Key Laboratory of Molecular Biophysics of the Ministry of Education, Huazhong University of Science and Technology, Wuhan 430074, China

Received: 12 June 2020 / Accepted: 3 July 2020 / Published online: 4 August 2020

Abstract Noncoding RNAs play important roles in cell and their secondary structures are vital for understanding their tertiary structures and functions. Many prediction methods of RNA secondary structures have been proposed but it is still challenging to reach high accuracy, especially for those with pseudoknots. Here we present a coupled deep learning model, called 2dRNA, to predict RNA secondary structure. It combines two famous neural network architectures bidirectional LSTM and U-net and only needs the sequence of a target RNA as input. Benchmark shows that our method can achieve state-of-the-art performance compared to current methods on a testing dataset. Our analysis also shows that 2dRNA can learn structural information from similar RNA sequences without aligning them.

Keywords RNA secondary structure prediction, Deep learning, Minimum free energy

INTRODUCTION

RNAs participate in many important biological activities (Xiyuan *et al.* 2017; Zhao *et al.* 2016). To do these, they need to form correct tertiary structures in general. Therefore, it is necessary to know the tertiary structures of RNAs to understand their functions. At present, experimental determination of RNA tertiary structures are more difficult than proteins and so many theoretical or computational methods have been proposed to predict RNA tertiary structures (Cao and Chen 2011; Das *et al.* 2010; Jain and Schlick 2017; Wang *et al.* 2017; Wang and Xiao 2017; Xu *et al.* 2014). Although these methods use different principles, their performances all depend on the accuracy of RNA secondary structures. Therefore, accurate prediction of RNA secondary structure is very important.

Traditional methods of RNA secondary structure prediction can be divided mainly into two categories: single-sequence methods and homologous sequences methods. The single-sequence methods only need the sequence of target RNA as input and most of them are based on thermodynamic model or minimum free energy principle (Bellaousov *et al.* 2013; Janssen and Giegerich 2015; Zuker 2003). Therefore, the accuracy of prediction results of these methods depends largely on the thermodynamic parameters that are difficult to determine accurately (Zhao *et al.* 2018). Comparing to the single-sequence method, homologous sequence method uses the evolution information of homologous sequences to infer their common secondary structure (Lorenz *et al.* 2011), *e.g.*, TurboFold (Tan *et al.* 2017). It was shown that the homologous sequences method usually could achieve higher accuracy than the single-sequence method (Puton *et al.* 2013). But the disadvantage of the homologous sequence method is also obvious, *i.e.*, many RNAs of interest don't have sufficient numbers of homologous sequences at present, leading to low prediction accuracy using the homologous

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s41048-020-00114-x>) contains supplementary material, which is available to authorized users.

✉ Correspondence: yxiao@hust.edu.cn (Y. Xiao)

sequence method. Furthermore, correct prediction of pseudoknots is still a challenge for both methods.

Deep learning has recently showed impressive applications across a variety of domains. Very recently deep learning has also been applied to RNA secondary structure prediction, *e.g.*, DMFold (Wang *et al.* 2019), which first predicts the dot or bracket state by bidirectional long short-term memory (LSTM) neural network and then infers the secondary structure by an algorithm based on improved base pair maximization principle. The input of DMFold is just the sequence of the target RNA. We also proposed to apply fully convolutional network (FCN) to improve RNA secondary structure prediction using direct coupling analysis of aligned homologous sequences of the target RNA (He *et al.* 2019). These methods show higher prediction accuracy than current related methods.

Here we propose a method of RNA secondary structure prediction with pseudoknots using coupled deep learning neural networks, which combines two famous neural networks architecture bidirectional LSTM (Krizhevsky *et al.* 2017) and U-net (Ronneberger *et al.* 2015). We call this method as 2dRNA. The sequence information of a target RNA is the only need for 2dRNA. Benchmarks show that 2dRNA can achieve state-of-the-art performance compared to most popular prediction methods on a testing dataset.

RESULTS

In this section, 2dRNA is benchmarked on the two testing sets and its performance is compared with several existing methods, including Mfold (Zuker 2003), RNAfold (Zuker and Stiegler 1981), cofold (Proctor and Meyer 2013), Ipknnot (Sato *et al.* 2011), Probknot (Bellaousov and Mathews 2010), and DMfold (Wang *et al.* 2019). Among them, Mfold, RNAfold and cofold could predict pseudoknot-free secondary structures, while Ipknnot, Probknot and DMfold could predict secondary structures with pseudoknots. During preparing our manuscript, Zhou's group reported a method of RNA secondary structure prediction using deep learning and transfer learning that also achieved significantly higher prediction precision than current methods (Singh *et al.* 2019). Since they used a different dataset for training, it is inappropriate to compare the performance of their method with ours directly. So, we shall not compare them here but this issue will be discussed later in more detail in the next section.

Table 1 shows the prediction results of our method and other methods on the ArchiveII testing set. The prediction results of other methods are from previous

Table 1 The comparison between 2dRNA and other methods on the ArchiveII testing set

Method	STY	PPV	MCC
Mfold	0.6658	0.6265	0.6459
RNAfold	0.5888	0.5290	0.5581
Cofold	0.5203	0.4883	0.5040
IPknot	0.5805	0.6163	0.5981
Probknot	0.5868	0.5285	0.5569
DMfold	0.7598	0.8585	0.8076
2dRNA	0.9017	0.9820	0.9353
2dRNA + DB	0.9288	0.9838	0.9519
2dRNA (< 80%)	0.8168	0.9631	0.8781
2dRNA + DB (< 80%)	0.8777	0.9787	0.9197

The prediction results of other methods other than 2dRNA are from previous work (Wang *et al.* 2019)

work (Wang *et al.* 2019). It could be obviously seen that the STY, PPV and MCC values of 2dRNA far exceed other methods. See the supporting material for the detailed performance of our method on the testing set (Supplementary Table S1).

The output of 2dRNA is a contact map, which also includes pseudoknot information. In the ArchiveII testing set there are 72 structures with total 996 pseudoknot base pairs. For comparison, we chose IPknot (Sato *et al.* 2011) and Probknot (Bellaousov and Mathews 2010), the most famous ones that can predict pseudoknots. The result is shown in Table 2 where the number of true positive (TP) is in relative to the total 1114 pseudoknot base pairs and the same is for TN, FP, FN, STY, PPV and MCC. As it can be seen, 2dRNA has a much higher performance in pseudoknot base pairs than other methods, with MCC 0.650. After applying data balancing technique, the prediction accuracy can be further promoted and MCC is up to 0.745, even for a trained model (2dRNA + DB (<80%)) with training data that only keep the sequences having similarity less than 80% with the testing set (see the following).

DISCUSSION

Better prediction using data balancing technique

The original training dataset contains total 2111 sequences with length less than 500. Among them, most sequences (1390) are less than 200 in length, accounting for 65.8% of the total. Specifically, there are 436 sequences with length less than 100, 954 in 100–200, 92 in 200–300, 565 in 300–400, 64 in 400–500. The

Table 2 Evaluation of pseudoknot prediction

Method	<i>TP</i>	<i>FP</i>	<i>TN</i>	<i>FN</i>	<i>STY</i>	<i>PPV</i>	<i>MCC</i>
IPknot	75	295	3415	924	0.071	0.125	0.089
Probknot	16	234	3436	983	0.011	0.035	0.018
2dRNA	584	47	4611	415	0.582	0.762	0.650
2dRNA + DB	726	62	5088	273	0.706	0.809	0.745
2dRNA (< 80%)	269	88	3334	730	0.169	0.373	0.242
2dRNA + DB (< 80%)	597	38	4495	402	0.600	0.735	0.653

number of the long RNA sequences is so few that our model can't learn enough knowledge from it for dealing with long sequence prediction. So here we apply the data balancing technique, not increasing new data but just duplicating the minority long sequences to let our model train with those data more times. In particular, we have duplicated the sequences with length in 200–300 by six times, in 400–500 by nine times, bring about total 3083 sequences on balanced training set. Figure 3A and B illustrate the distribution before and after data balancing.

After this data processing, our model can well deal with long RNA sequences, and the whole accuracy on testing set is also improved a lot. The *STY* has increased by approximately 3% from 0.9017 to 0.9288, while *PPV* has just improved a little from 0.9820 to 0.9838, resulting in the highest *MCC* 0.9519 among all the other methods. Especially for the long sequences, the prediction results look much better than before (Fig. 1C).

Correlation between similarity and accuracy

As it is well known that deep learning is a black box, making it hard to understand what it has learned. Here we try to give some insights about how prediction

accuracy is correlated with the sequence similarity between the testing and training sets.

We used the Levenshtein distance (Levenshtein 1966) to measure the similarity between two sequences. It is noted that this kind of similarity is different from that given by CD-HIT-EST algorithm (Fu *et al.* 2012). For each sequence in the testing set, its distances with every sequence in the training set are calculated. Figure 2 shows the correlations between the prediction accuracy (here *MCC*) of each sequence in the testing set and the number of the sequences in the training set with their distances being larger than equal to 0.7 to the former. It shows that the prediction accuracy increases with the number of the sequences in general. For shorter sequences (blue dots in Fig. 2), the *MCCs* of most of them are close to 1. This means that their structural information could be learned from those sequences with lower similarity. For the longer sequences (*e.g.*, >300 (red dots in Fig. 2)) in the testing set, they usually have lower accuracy than the shorter ones with the same number of similar sequences. This is reasonable because for long sequences, more homologous sequences are needed to learn their structural information. These results suggest that the neural network model can learn common structural features from

Fig. 1 Comparison of using data balancing technique or not. **A** The distribution of sequences length and number of base pairs on unbalanced training set. **B** Training set distribution after data balancing. **C** The result of RNaseP_Onchorhynchus-2. The *black points* are the native base pairs. The *red* is the prediction result of 2dRNA trained on balanced training set, and the *blue* is the prediction result of 2dRNA trained on unbalanced training set

Fig. 2 Correlation between accuracy (here MCC) of each sequence in the testing set and the number of the sequences in the training set with their distances being larger than or equal to 0.7 to the former. The sequence length is scaled in color

similar RNA sequences without aligning them. This is in agreement with previous study (Wang *et al.* 2019).

To see the generalization ability of 2dRNA, we also have trained a new model with training data that only keep the sequences having similarity less than 80% with the testing set using the CD-HIT-EST (sequence-identity cutoff is set as 0.8) (Fu *et al.* 2012). The result is shown in Table 1 as 2dRNA (<80%) and 2dRNA + DB (<80%). It is clear that in this case 2dRNA still significantly outperforms other methods in PPV, STY and MCC.

Performance for different RNA types

Table 3 shows the performances of 2dRNA for different types of RNAs in the testing set. The performances for 5sRNA and tRNA are higher than tmRNA and RNaseP, mainly in STY. This may be that the latter two usually have much longer lengths than the former in the training set (Supplementary Table S2). Furthermore, as discussed above, the sequence number of longer lengths is much small and this may be also a reason why the performance for tmRNA and RNaseP is relatively lower.

Our method also has been tested on the Zhou's testing set TS0 (38), which contains 1305 sequences

that include not only the four types mentioned above but also microRNA, riboswitch and so on. Table 4 shows the performances of 2dRNA for different types of RNAs in the TS0 testing set. Since our training dataset does not contain these other types of RNAs, it's not surprising to see poor performance of those. But in 5sRNA, tRNA, tmRNA and RNaseP, our method still maintains impressive performance.

CONCLUSION

In this work, we present an end-to-end coupled deep learning model 2dRNA to predict RNA secondary structure with pseudoknots. Our method achieves a state-of-the-art performance compared to other methods, including the prediction of pseudoknots. It seems

Table 3 Performance of 2dRNA for different RNA types on ArchiveII testing set

Type	Number	STY	PPV	MCC
5sRNA	109	0.9898	0.9912	0.9903
tRNA	50	0.9736	0.9900	0.9804
tmRNA	31	0.7723	0.9601	0.8544
RNaseP	44	0.8369	0.9754	0.8931

Table 4 Performance of 2dRNA for different RNA types on TS0 testing set

Type	Number	STY	PPV	MCC
5sRNA	30	0.6807	0.8028	0.7321
tRNA	128	0.8761	0.9463	0.9058
tmRNA	25	0.6807	0.9042	0.7747
RNaseP	23	0.7742	0.8501	0.8001
snRNA	148	0.1335	0.1438	0.1272
MicroRNA	202	0.0627	0.2463	0.1177
Riboswitch	48	0.0169	0.0826	0.0327
Group I Introns	10	0.0277	0.4141	0.0992
Group II	55	0.0080	0.1091	0.0287
CRISPR	7	0.0612	0.1071	0.0810
Other	629	0.0586	0.1696	0.0897

that the coupled deep learning model can use structural information learning from similar RNAs with known structures to predict the secondary structure of a target RNA. Our model was trained in the dataset with limited types of RNAs. In future we shall further apply it to larger dataset. Our method is available as a web server at <https://biophy.hust.edu.cn/new/2dRNA>. The web server is currently only for evaluating our method.

MATERIALS AND METHODS

Problem formulation

Here, the secondary structure of a RNA is defined as the pattern formed by the canonical base pairs A-U and C-G, and the wobble base pair G-U as usual. Non-canonical pairs are neglected in our secondary structure prediction. The bases in the secondary structure fulfill the following constraints: (1) a base cannot participate in more than one base pair; (2) bases that are paired with each other must be separated by at least three bases along sequence.

In this work, we used two different representations of RNA secondary structure. One is dot-bracket notation (Hofacker *et al.* 1994), which contains seven symbols. Unpaired nucleotides are indicated with the character “.”, matching pairs of parentheses “()” indicate base pairs. To indicate non-nested base pairs (pseudoknots), additional brackets are used: “[]”, “{ }”, or “< >”. Another representation is the dot-plot matrix (Maizel and Lenk 1981), where each base pair (i, j) is represented by a dot or box in row i and column j of a rectangular grid or contact matrix of the structure. Figure 3 shows the RNA secondary structure and its two representations.

Methods

Our deep learning neural network is a coupled two-stage model: coarse-grained dot-bracket prediction (CGDBP) and fine-grained dot-plot prediction (FGDPP). Motivated by DMfold (Wang *et al.* 2019), CGDBP takes one-hot encoding format of four type bases of the sequence as input and predicts the dot-bracket states of the bases. We use a two-layer bidirectional LSTM (Krizhevsky *et al.* 2017) as the encoder that encodes the input sequence information into high dimension space (the dimension of the vectors is $L \times 256$, where L is the length of input sequence, 256 is the encoding feature size), followed by a decoder to give dot-bracket prediction. This decoder is a fully connected network, taking the encoding vectors from RNA sequence to one-hot format of dot-bracket sequences ($L \times 7$, for the dot-bracket notation contains seven symbols). The FGDPP is a fully convolution network, which inputs RNA sequence encoding but outputs dot-plot matrix ($L \times L$). The output of first stage is the dot-bracket state of each base, *i.e.*, the base is in unpaired state “.” or in one of the six paired states “[, “]”, “[, “]”, “{” or “}”, which tells what state each base is at, but there is still much mismatching between brackets. That is why it's called coarse-grained prediction. The second part FGDPP of our model is created to address this problem. Since the LSTM network can only capture sequential context, we combined it with U-net (Ronneberger *et al.* 2015) to get precise pairwise base-pairing prediction and provide the ability of extracting structural information. Figure 4 displays the coupled neural network architecture of 2dRNA.

Coarse-grained dot-bracket prediction

The CGDBP uses the bidirectional LSTM architecture (Krizhevsky *et al.* 2017) as encoder and a fully connected layer as decoder. LSTM network (Hochreiter and

Fig. 3 RNA secondary structure and its two representations. The secondary structure of a RNA (A), its dot-plot matrix representation (B) and its dot-bracket representation (C)

Fig. 4 The coupled neural network architecture of 2dRNA. Coarse-grained dot-bracket prediction uses two-layer bidirectional LSTM and a fully connected layer to output dot-bracket prediction and fine-grained dot-plot prediction uses fully convolutional network (U-net) to predict pairwise base-pairing as final result

Schmidhuber 1997) is a special kind of recurrent neural network, capable of learning long-term dependency and now is widely used in natural language processing. Our model uses a 2-layers bidirectional LSTM network, the dimension of hidden vector h is 128 for all the layers. The decoder is a 3-layers fully connected network, with size of 128, 64 and 7. In both networks, the dropout rate is 0.1 and the activation function is rectified linear unit (ReLU) (Agarap 2018).

The input of CGDBP is RNA sequences, one-hot encoding by $L \times 4$ vectors. Thence the residue A is encoded as "1000", U is "0100", G is "0010", and C is "0001". Then it outputs the $L \times 7$ vectors which correspond to the one-hot vector of dot-bracket symbols, with "0001000" represents ".", "1000000" and "0000001" represent "(" and ")"; "0100000" and "0000010" for "[" and "]", and "0010000" and "0000100" for "{" and "}".

Fine-grained dot-plot prediction

The FGDPP uses the U-net architecture (Ronneberger *et al.* 2015) as decoder. U-Net is a famous network developed for biomedical image segmentation, which is based on FCN (Shelhamer *et al.* 2017) and its architecture is extended to yield more precise segmentations. U-net consists of a contracting path for downsampling and an expansive path for upsampling. During the contraction path, the spatial information is reduced while feature information is increased. Compared to the vanilla fully convolutional networks, U-net has the shortcut connection crossing convolutional layers analogous to ResNet, its expansive path concatenates the spatial information with high-resolution features from

the contracting path, which provide local information to the global. This is significant to the secondary structure problem, high-resolution features make a benefit to the local structures like stems or hairpins, global features about these local structures provide spatial information that contributes to the prediction of global structure, especially for pseudoknots.

We used a 9-blocks/layers U-net in our model. The contracting path is composed of four blocks, each block is composed of: (1) two units of 3×3 convolution layer and ReLU activation function (with batch normalization), followed by (2) 2×2 max pooling. The number of feature maps doubles at each pooling, starting with 32 feature maps for the first block to the final 256 feature maps. The parameters of all the convolution layers are the same, with kernel size being 3, stride being 1, and padding being 1. The middle part of U-net between the contracting and expanding paths is a block built from simply two convolutional layers with dropout and batch normalization. And the last two blocks are the upsampling path, each of these blocks is composed of: (1) deconvolution layer with stride 2; (2) skip connections with the corresponding cropped feature map from the contracting path; (3) two 3×3 convolution layers and ReLU activation function (with batch normalization).

We take the output of the bidirectional LSTM, which is 256-dimensional encoding vectors, and do a pairwise addition operation on it to get a tensor with size $L \times L \times 256$, as the input of the U-net. The output of the U-net is a $L \times L \times 1$ matrix, which is through the sigmoid layer to be normalized into $[0, 1]$, reflecting the pairing probability of each pair of bases. In our model,

we set the cutoff as 0.5, the score of base pairs that exceed this cutoff is considered as paired.

The original U-net uses binary cross entropy loss to optimize its parameters (Goodfellow *et al.* 2016). But for RNA secondary structure prediction problem, the native base pair is so rare relative to all possible pairing in dot-plot matrix representation, *i.e.*, positive labels are very lacking in training set, which cause inefficient training as most locations are simple negatives and contribute no useful learning signal. Mass negatives can overwhelm training and lead to degenerate models. So we used focal loss function (Lin *et al.* 2018) to address this class imbalance, which reshapes the standard cross entropy loss such that it down-weights the loss assigned to well-classified examples. Another loss function our model uses is soft dice loss, which is based on the dice coefficient (DSC) that is a measure of overlap between two samples (Milletari *et al.* 2016). Especially for RNA, we proposed a symmetric loss function according to its secondary structure. The RNA dot-plot representation is a symmetric matrix, but the predictions are often not. So, a symmetric loss as a penalty is used to prevent occurring of this:

$$SL(p) = (p - p^T)^2, \quad (1)$$

Those two parts (coarse-grained and fine-grained) are trained together as a whole, for the reason that both dot-bracket output and matrix output can help and constrain each other in the process of loss propagation. While training, we used Adam (Kingma and Ba 2014) with learning rate 0.002 to optimize loss function. The train epoch is set to 30, and batch size is 16, since L is different, forward calculate one sample at a time and then backward total loss after 16 samples, meanwhile the batch normalization in U-net is instance normalization. The PyTorch implement of U-net that we use comes from the open-source codes provided by milesial on Github (<https://github.com/milesial/Pytorch-UNet>).

Data collection and processing

The data used for training and testing contain sequence and secondary structure, is the same as DMFold (Wang *et al.* 2019) for the performance comparison. They are the ArchiveII dataset from the public database of Mathews lab (Ward *et al.* 2017). Although the RNA types of this dataset are less than the larger bpRNA dataset (Danaee *et al.* 2018), the sequence number of each RNA type has similar order of magnitude and so this dataset is better for training and testing a deep learning model (Supplementary Table S2). This dataset totally has 2345 known RNA primary sequences and their secondary structures. For comparing with DMFold,

the same 234 sequences randomly chosen is taken as testing set. The left is for training and validating. Here we used tenfold cross-validation, split the left data into tenfolds, each time use ninefolds to train the model and onefold for validation. We have trained 10 models and used the averaged results of these models as final predictions, *i.e.*, two residues are considered to form base pair if their averaged pairing probability is equal to or larger than 0.5.

All data above are adopted to remove the duplicate data by the CD-HIT tool (Fu *et al.* 2012) and the data that contain the unknown bases. Particularly, we only selected the sequences with length less than 500 and longer than 32. Finally, the training set contains totally 2111 sequences with length less than 500.

Performance metrics

To estimate the performance of 2dRNA and other methods, we can calculate the accuracy of the base pairs to represent the accuracy of prediction structures. So that, the predictions can be compared between different methods. We use the precision (PPV) and sensitivity (STY) (Matthews 1975) to measure the performance of the methods. STY measures the ability to find the positive base pairs, while PPV measures the ability of not predicting false positive base pairs. They are defined as follows:

$$PPV = \frac{TP}{TP + FP}, \quad (2)$$

$$STY = \frac{TP}{TP + FN}, \quad (3)$$

where TP denotes true positive; FP, false positive; TN, true negative; FN, false negative.

Generally, the requirements of PPV and STY could not be satisfied simultaneously when comparing the accuracy of the prediction results. Therefore, the Matthews correlation coefficients (MCC) (Parisien *et al.* 2009) is used to comprehensively evaluate the prediction results.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(FP + TN)(TN + FN)(FN + TP)}} \approx \sqrt{STY \times PPV}. \quad (4)$$

Acknowledgements This work is supported by the National Natural Science Foundation of China (31570722). The data used for initial training along with their annotated secondary structure are publicly available at <https://biophy.hust.edu.cn/new/2dRNA>.

Compliance with Ethical Standards

Conflict of interest Kangkun Mao, Jun Wang and Yi Xiao declare no conflict of interest.

Human and animal rights and informed consent This article does not contain any studies with human or animal subjects performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agarap AFM (2018) Deep learning using rectified linear units (ReLU). arXiv: neural and evolutionary computing. <https://www.arxiv-vanity.com/papers/1803.08375>
- Bellaousov S, Mathews DH (2010) ProbKnot: fast prediction of RNA secondary structure including pseudoknots. *RNA* 16(10):1870–1880
- Bellaousov S, Reuter JS, Seetin MG, Mathews DH (2013) RNAstructure: web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res* 41(W1):W471–W474
- Cao S, Chen SJ (2011) Physics-based de novo prediction of RNA 3D structures. *J Phys Chem B* 115(14):4216–4226
- Danaee P, Rouches M, Wiley M, Deng D, Huang L, Hendrix D (2018) bpRNA: large-scale automated annotation and analysis of RNA secondary structure. *Nucleic Acids Res* 46(11):5381–5394
- Das R, Karanickolas J, Baker D (2010) Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat Methods* 7(4):291–294
- Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28(23):3150–3152
- Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge, MA
- He X, Li S, Ou X, Wang J, Xiao Y (2019) Inference of RNA structural contacts by direct coupling analysis. *Commun Inf Syst* 19(3):279–297
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P (1994) Fast folding and comparison of RNA secondary structures. *Monatshefte Fur Chemie* 125(2):167–188
- Jain S, Schlick T (2017) F-RAG: generating atomic coordinates from RNA graphs by fragment assembly. *J Mol Biol* 429(23):3587–3605
- Janssen S, Giegerich R (2015) The RNA shapes studio. *Bioinformatics* 31(3):423–425
- Kingma DP, Ba JA (2014) A method for stochastic optimization. arXiv: 1412.6980
- Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. *Commun ACM* 60(6):84–90
- Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions and reversals. *Sov Phys Dokl* 10(4):707–710
- Lin TY, Goyal P, Girshick R, He K, Dollar P (2018) Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell*. <https://doi.org/10.1109/TPAMI.2018.2858826>
- Lorenz R, Bernhart SH, Siederdisen CHZ, Tafer H, Flamm C, Stadler PF, Hofacker IL (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol*. <https://doi.org/10.1186/1748-7188-6-26>
- Maizel JV Jr, Lenk RP (1981) Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc Natl Acad Sci USA* 78(12):7665–7669
- Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta* 405(2):442–451
- Milletari F, Navab N, Ahmadi SA (2016) V-net: fully convolutional neural networks for volumetric medical image segmentation. *Proceedings of 2016 Fourth International Conference on 3d Vision (3dv)*, 565–571
- Parisien M, Cruz JA, Westhof E, Major F (2009) New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA* 15(10):1875–1885
- Proctor JR, Meyer IM (2013) COFOLD: an RNA secondary structure prediction method that takes co-transcriptional folding into account. *Nucleic Acids Res* 41(9):e102. <https://doi.org/10.1093/nar/gkt174>
- Puton T, Kozłowski LP, Rother KM, Bujnicki JM (2013) CompaRNA: a server for continuous benchmarking of automated methods for RNA secondary structure prediction. *Nucleic Acids Res* 41(7):4307–4323
- Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional Networks for Biomedical Image Segmentation. *Med Image Comput Comput-Assist Interv* 9351:234–241
- Sato K, Kato Y, Hamada M, Akutsu T, Asai K (2011) IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics* 27(13):i85–93
- Shelhamer E, Long J, Darrell T (2017) Fully convolutional networks for semantic segmentation. *IEEE Trans Pattern Anal Mach Intell* 39(4):640–651
- Singh J, Hanson J, Paliwal K, Zhou Y (2019) RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat Commun* 10:5407. <https://doi.org/10.1038/s41467-019-13395-9>
- Tan Z, Fu YH, Sharma G, Mathews DH (2017) TurboFold II: RNA structural alignment and secondary structure prediction informed by multiple homologs. *Nucleic Acids Res* 45(20):11570–11581
- Wang J, Xiao Y (2017) Using 3dRNA for RNA 3-D structure prediction and evaluation. *Curr Protoc Bioinformatics*. <https://doi.org/10.1002/cpbi.21>
- Wang J, Mao K, Zhao Y, Zeng C, Xiang J, Zhang Y, Xiao Y (2017) Optimization of RNA 3D structure prediction using evolutionary restraints of nucleotide-nucleotide interactions from direct coupling analysis. *Nucleic Acids Res* 45(11):6299–6309
- Wang L, Liu Y, Zhong X, Liu H, Lu C, Li C, Zhang H (2019) DMfold: a novel method to predict RNA secondary structure with pseudoknots based on deep learning and improved base pair

- maximization principle. *Front Genet* 10:143. <https://doi.org/10.3389/fgene.2019.00143>
- Ward M, Datta A, Wise M, Mathews DH (2017) Advanced multi-loop algorithms for RNA secondary structure prediction reveal that the simplest model is best. *Nucleic Acids Res* 45(14):8541–8550
- Xiyuan L, Dechao B, Liang S, Yang W, Shuangfang F, Hui L, Haitao L, Chunlong L, Wenzheng F, Runsheng C, Yi Z (2017) Using the NONCODE database resource. *Curr Protoc Bioinform*. <https://doi.org/10.1002/cpbi.25>
- Xu X, Zhao P, Chen SJ (2014) Vfold: a web server for RNA structure and folding thermodynamics prediction. *PLoS ONE* 9(9):e107504. <https://doi.org/10.1371/journal.pone.0107504>
- Zhao Y, Li H, Fang SS, Kang Y, Wu W, Hao YJ, Li ZY, Bu DC, Sun NH, Zhang MQ, Chen RS (2016) NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res* 44(D1):D203–D208
- Zhao Y, Wang J, Zeng C, Xiao Y (2018) Evaluation of RNA secondary structure prediction for both base-pairing and topology. *Biophys Rep* 4(3):123–132
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406–3415
- Zuker M, Stiegler P (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 9(1):133–148