



## Method article

# SSRTool: A web tool for evaluating RNA secondary structure predictions based on species-specific functional interpretability



Tzu-Hsien Yang<sup>a,\*</sup>, Yu-Cian Lin<sup>a,1</sup>, Min Hsia<sup>a,1</sup>, Zhan-Yi Liao<sup>a,1</sup>

<sup>a</sup>Department of Information Management, National University of Kaohsiung, Kaohsiung University Rd, 811 Kaohsiung, Taiwan

## ARTICLE INFO

## Article history:

Received 19 February 2022

Received in revised form 13 May 2022

Accepted 13 May 2022

Available online 18 May 2022

## Keywords:

RNA secondary structure

RNA structure prediction evaluation

RNA functional interpretability

## ABSTRACT

RNA secondary structures can carry out essential cellular functions alone or interact with one another to form the hierarchical tertiary structures. Experimental structure identification approaches can show the *in vitro* structures of RNA molecules. However, they usually have limits in the resolution and are costly. *In silico* structure prediction tools are thus primarily relied on for pre-experiment analysis. Various structure prediction models have been developed over the decades. Since these tools are usually used before knowing the actual RNA structures, evaluating and ranking the pile of secondary structure predictions of a given sequence is essential in computational analysis. In this research, we implemented a web service called SSRTool (RNA Secondary Structure prediction Ranking Tool) to assist in the ranking and evaluation of the generated predicted structures of a given sequence. Based on the computed species-specific interpretability significance in four common RNA structure–function aspects, SSRTool provides three functions along with visualization interfaces: (1) Rank user-generated predictions. (2) Provide an automated streamline of structure prediction and ranking for a given sequence. (3) Infer the functional aspects of a given structure. We demonstrated the applicability of SSRTool via real case studies and reported the similar trends between computed species-specific rankings and the corresponding prediction F1 values. The SSRTool web service is available online at <https://cobisHSS0.im.nuk.edu.tw/SSRTool/>, <http://cosbi3.ee.ncku.edu.tw/SSRTool/>, or the redirecting site <https://github.com/cobisLab/SSRTool/>.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

RNAs can participate in essential cellular processes such as gene post-transcriptional regulation, translation initiation control, and mediator structures in protein complexes [1–4]. Like the well-understood protein structure hierarchy, RNAs can carry out their functions through secondary or tertiary structures formed by base-pair hydrogen bonds [5,6]. Understanding the particular structural forms of different non-coding RNAs (ncRNAs) is key to reconstructing the mechanisms of many vital cellular responses. Since the secondary structures of many RNAs alone can be involved in many biological processes and the 3D RNA structures can be derived from the interaction of secondary structure sub-domains [7], probing RNA secondary structures is the first step to unraveling the functionalities of these molecules.

Experimental methods, including X-ray crystallography, nuclear magnetic resonance (NMR) [8], cryoelectron microscopy [9], and small-angle X-ray scattering [10], have been developed to identify RNA molecular structures. However, there are limits in the resolution for these experiments, and these approaches are generally of high cost. These limitations result in the requirement for computational tools to assist in structure identification and pre-screening. For this reason, researchers have come up with various RNA secondary structure prediction tools over the past decades. These tools can be roughly categorized into two genres. Tools in the first category help predict the base pairings in RNA secondary structures by considering thermodynamic free energy minimization [11]. These tools search for the lowest-energy folding of a given RNA sequence. The second type of tools utilizes the multiple alignments of homologous sequences to reach a consensus structure of this family of RNA sequences [12]. Recently, some of the developed tools also incorporate chemical structure probing techniques, such as dimethyl sulfate (DMS) or selective 2'-hydroxyl analyzed by primer extension (SHAPE) reagents, to refine the prediction accuracy [13].

\* Corresponding author.

E-mail addresses: [c102570@gmail.com](mailto:c102570@gmail.com) (T.-H. Yang), [m1083313@mail.nuk.edu.tw](mailto:m1083313@mail.nuk.edu.tw) (Y.-C. Lin), [a1083301@mail.nuk.edu.tw](mailto:a1083301@mail.nuk.edu.tw) (M. Hsia), [a1083345@mail.nuk.edu.tw](mailto:a1083345@mail.nuk.edu.tw) (Z.-Y. Liao).

<sup>1</sup> These authors contributed equally.

Current prediction tools provide the lowest-energy or consensus structure folding under their constructed thermodynamic models. Nonetheless, RNAs are so flexible that a short sequence with only adenines and uracils can lead to more than 10 million distinct conformations for such a simple sequence [14]. Different prediction tools adopt different assumptions in obtaining the best-fit secondary structure for a given sequence. Because the actual structure of a given RNA sequence is usually not determined yet when applying these structure prediction tools, it remains computationally challenging to rank different prediction results and infer the most native-like structure of a sequence among various predictions. There is still no easy way to obtain useful rankings for various predicted structures of a given sequence. Hence, to facilitate the computational analysis of RNA structures, an online software tool satisfying the following requirements is in demand [15]: (1) A ranking criterion for users to select the most prevalent functional structure from the predicted structure pools before determining the actual RNA structure of a sequence; (2) Alleviate the burden of using different tools and automate the secondary structure prediction collection for existing tools; (3) Implement an easy-to-use interface with visualization services for the provided RNA structure predictions. We have previously proposed a novel method that helps evaluate the functional interpretability of a given structure prediction [16]. The proposed method can compute the significance scores of an RNA secondary structure in four structure–function aspects (cellular fitness, RNA–protein interaction (RPI) complex, translational regulation, and post-transcriptional regulation). And the designed algorithm has been verified to help infer the best native-like secondary structure for a given RNA sequence in yeast and humans. Based on this concept of functional interpretability, we can further extend the method in additional species and construct an automated pipeline to help provide the rankings of diverse secondary structure predictions for an RNA sequence.

In this research, we implemented an online software tool called SSRTTool (RNA Secondary Structure prediction Ranking Tool) to help rank a list of predicted structures of an RNA sequence based on functional interpretability. In SSRTTool, we extended our previous structure functional interpretability calculation algorithm to support six different model organisms (*Homo sapiens*, *Saccharomyces cerevisiae*, *Mus musculus*, *Rattus norvegicus*, *Danio rerio*, and *Arabidopsis thaliana*). And we implemented a user-friendly online software service for automated ranking generation of the predicted structures of a given RNA sequence. Three major functions are provided in SSRTTool: (1) Rank Predictions. SSRTTool computes the rankings for a given list of structure predictions of a sequence using the functional interpretability significance of each structure. Further, SSRTTool aggregates the functionally interpretable structures as a suggestion for the most prevalent secondary structure of this sequence. (2) Automated structure prediction and ranking pipeline. SSRTTool streamlines the structure prediction process of 19 publicly available tools to lessen the software learning burden on users. Then the species-specific functional interpretability rankings of these predictions are calculated to help identify the most native-like structure. (3) Infer species-specific functional aspects for a given structure. We demonstrated the biological applicability of SSRTTool by considering the predicted structures of the human spliceosome mediator RNA sequence. Then we showed that despite the sparsity of reference data in some species, SSRTTool can still identify the most native-like structures with at least top-five performance compared with existing tools. Finally, we reported that the calculated rankings based on interpretability could indicate the similarities between structure predictions and their corresponding prevalent structures. SSRTTool is available online at <https://cobisHSS0.im.nuk.edu.tw/SSRTTool/>, <http://cosbi3.ee.ncku.edu.tw/SSRTTool/>, or the redirecting site <https://github.com/cobisLab/SSRTTool/>.

## 2. Construction and Contents

### 2.1. The overview of the SSRTTool workflow

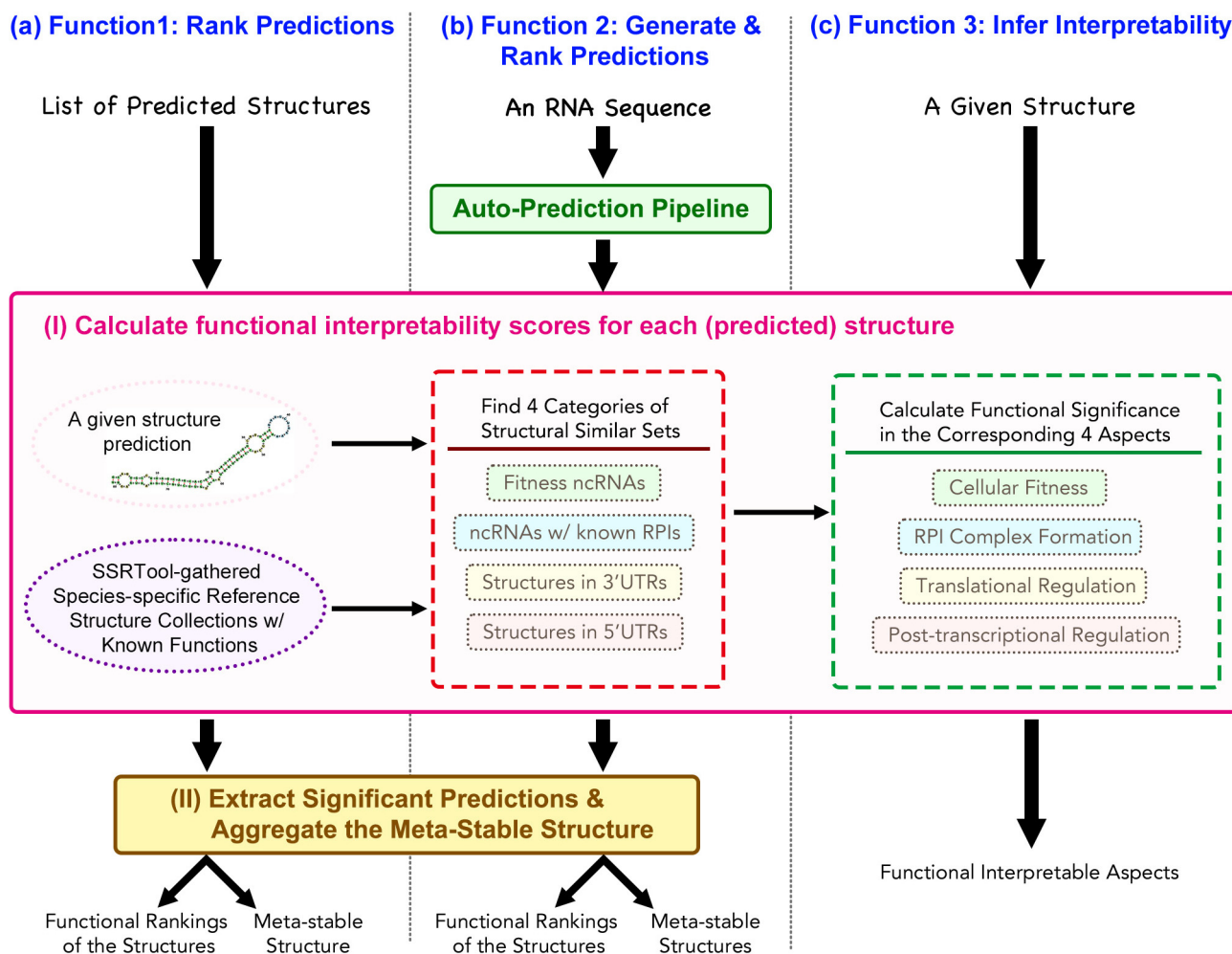
SSRTTool is designed to provide confidence rankings of a list of structure predictions based on functional interpretability. The overall workflow of SSRTTool is depicted in Fig. 1. There are three ways for users to provide the inputs and use the SSRTTool pipeline: (a) A given list of structure predictions of an RNA sequence (Fig. 1-a). Users can first run any structure prediction tools by themselves to get the predicted structures of a sequence. Then SSRTTool applies the implemented prediction ranking algorithm to each predicted structure in the given list (Fig. 1-I). In the algorithm, the designed biological significance tests calculate the interpretability scores of a predicted structure with the help of SSRTTool-gathered species-specific reference structure collections in four RNA structure–function aspects (cellular fitness, RPI complex formation, translational regulation, and post-transcriptional regulation). Then, the functionally interpretable predictions are extracted to generate the meta-stable structure, or the most prevalent secondary structures in the cellular contents, for the given structure prediction list based on functional weighted aggregation (Fig. 1-II). (b) A given RNA sequence (Fig. 1-b). SSRTTool also implements an automated pipeline to obtain default structure predictions of 19 existing tools and then perform the ranking algorithm on these auto-generated structure prediction results. (c) A given RNA structure (Fig. 1-c). Finally, SSRTTool can also infer the functional aspects of a user-given structure. The details of the implemented prediction ranking algorithm are described in Section 2.2, and the gathered knowledge-base datasets to support these functions are portrayed in Section 2.3 and 2.4.

### 2.2. The implemented prediction ranking algorithm

We implemented our previously designed RNA secondary structure prediction ranking method [16] as the core algorithm in SSRTTool and extended it to support up to six organisms in this current version. Furthermore, an easy-to-use interface was constructed to facilitate the structure prediction analysis. The extended structure ranking algorithm can take one predicted structure or a list of structure predictions to calculate its/their functional significance scores. These structure predictions can be generated by any prediction tools. The algorithm procedure is roughly divided into two parts (Fig. 1-I and 1-II): (I) Find four structural similar sets for each structure prediction from the corresponding species-specific reference structure collections. And then calculate the four functional interpretability scores for each structure prediction based on these four structural similar sets. (II) Extract significant structure predictions and aggregate the functional meta-stable secondary structure. The details of the algorithm can be found in our previous work ([16]). A brief sketch of each part is provided in the following subsections.

#### 2.2.1. Part I: find the structural similar sets and calculate functional interpretability scores

The first part of the prediction ranking algorithm computes the functional interpretability scores for every given secondary structure. It is known that similar structures can carry out similar functions [4]. Hence for each structure prediction, the algorithm first finds its four structural similar sets from the species-specific reference structure collection using the ExpaRNA [17] distance. Notice



**Fig. 1.** The workflow of SSRTTool. (a), (b), and (c) refer to the three different ways to provide inputs for the corresponding three webtool functions of SSRTTool. (a)(b) Users can either provide a list of predicted structures of an RNA sequence or a raw sequence (for prediction generation) into SSRTTool. SSRTTool first calculates the functional interpretability score rankings for each given prediction. And then, those functionally interpretable predictions are extracted and aggregated to suggest the most native-like candidate structure. (c) A specific structure of interest can also be input into SSRTTool to identify the potential functional aspects of this structure.

that ExpaRNA considers only the local plain structures in calculating the structure distances. For provided pseudoknotted predictions, the pseudoknot pairings are ignored when calculating functional significance for these pseudoknotted predictions. The species-specific reference structure collections encompass experimentally verified structures for RNA sequences found in individual species. And there are four categories comprising the species-specific reference collections: the structural fitness ncRNAs, the structural ncRNAs with known RNA–protein interactions (RPIs), the regulatory structures in 5'UTRs, and the regulatory substructures in 3'UTRs. We have gathered these reference structure collections in six species in SSRTTool. And the preparation of them is described in Section 2.3. The structural similar sets of a structure in these four parts can be intermediates to help check if this given structure prediction is prevalent in cellular fitness, RPI complex formation [18], translational regulation [19], and post-transcriptional regulation [20], respectively. Based on the structural similar sets, the functional interpretability scores of the given predicted secondary structure are calculated.

To calculate the interpretability scores in the four RNA functional aspects (cellular fitness, RPI complex formation, translational regulation, and post-transcriptional regulation), we refer to the following species-specific tests: the functional profiling coherence test [21], the translation efficiency test [22], the protein-complex prevalence test [23], and the gene ontology enrichment

test [23]. We have prepared these tests in six species in SSRTTool. And the datasets that support these species-specific tests are described in Section 2.4. The algorithm first considers the functional profiling coherence of the extracted fitness ncRNA similar set to estimate the cellular fitness significance of the given predicted structure. And to evaluate the functional significance for the predicted structure to be involved in an RPI complex, we consider if the protein set having interactions with the ncRNA structural similar set is prevalent in protein-complex or enriched in gene ontology. For evaluating the translational regulation interpretability, the translation efficiency of mRNAs for the 5'UTR structural similar set is estimated. Finally, to calculate the significance of the predicted structure to participate in post-transcriptional regulation, the 3'UTR structural similar set is tested if the mRNAs with these 3'UTRs are prevalent in protein interaction complex or enriched in gene ontology. The significance results were represented in  $q$ -values, or the calibrated test  $p$ -values via the FDR (false discovery rate) multiple hypotheses correction procedure. Details of these tests can be found in [16].

### 2.2.2. Part II: extract significant predictions and aggregate the most native-like candidate secondary structure

After calculating the four functional interpretability scores for each structure prediction, the prediction ranking algorithm assigns the best functional score (the lowest interpretability  $q$ -value of the

four aspects) for each predicted structure. These assigned best functional interpretability scores are then used as the confidence rankings for the given list of predicted structures. Furthermore, functionally interpretable secondary structures are extracted against a user-defined significance threshold. Based on the ensemble of functionally interpretable structures, the meta-stable structure is obtained based on weighted aggregation using the significance scores (See [16] for the detailed formula). The meta-stable structure is also suggested as the most native-like candidate secondary structure for users.

### 2.3. Generation of the species-specific reference structure collections

In Part I of the implemented prediction ranking algorithm, species-specific reference structure collections with known functionality are needed. We BLASTed the BRALibaseII RNA structure benchmark dataset [24] with the following datasets to obtain the 4 parts of the species-specific reference structure collections in SSRTTool (E-value < 1e-6, percent-identity > 70%): (1) cellular fitness-related ncRNAs. We downloaded 443 ncRNAs that have undergone fitness functional profiling experiments from the work of Parker et al. [25]. (2) ncRNAs with RPIs. The literature-curated RPI data for *Saccharomyces cerevisiae* was gathered from the work of Panni et al. [26]. And the literature-curated RPI data for *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus* were gathered from the RNAInter v3.0 database [27]. In order to eliminate noises, we enforced that the confidence of each RPI relation should be larger than 0.7 to be used in SSRTTool. Since there are only few ncRNAs with RPI information in *Arabidopsis thaliana* and *Danio rerio* thus far, the RPI complex significance estimation was unavailable and thus omitted in these two species. The sequences of the ncRNAs in different species were downloaded from the RefSeq database [28]. (3) The 5'UTR and 3'UTR sequence datasets. We downloaded the transcript datasets of *Arabidopsis thaliana* and *Saccharomyces cerevisiae* from the TAIR database [29] and the SGD database [30], respectively. And the transcript information of *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, and *Danio rerio* were adopted from the UCSC genome browser [31]. Finally, the following genome assembly versions were used: *Saccharomyces cerevisiae* (S288C), *Homo sapiens* (hg38), *Mus musculus* (mm39), *Rattus norvegicus* (rn7), *Arabidopsis thaliana* (TAIR10), and *Danio rerio* (danRer11). We have also considered the possible reference structure collections for *Drosophila melanogaster* and *Caenorhabditis elegans*. However, there are insufficient sequences with known structures that can be matched in these two species. These two species were thus not included in SSRTTool. The numbers of reference sequences with known structures collected and used for different species in SSRTTool are summarized in Table 1.

### 2.4. Datasets for computing the species-specific functional interpretability significance

SSRTTool utilizes the prediction ranking algorithm to consider four functional aspects for every candidate structure: cellular fit-

**Table 1**

The summary of sequences with known structures for the species-specific reference collections gathered in SSRTTool.

Species	ncRNA Structures w/ Known RPIs	Structures in 5'UTR	Structures in 3'UTR
<i>Homo sapiens</i>	709	208	552
<i>Saccharomyces cerevisiae</i>	323	109	351
<i>Mus musculus</i>	128	55	191
<i>Rattus norvegicus</i>	6	20	388
<i>Danio rerio</i>	–	54	134
<i>Arabidopsis thaliana</i>	–	407	340

ness, RPI complex formation, translational regulation, and post-transcriptional regulation. The functional interpretability scores for each structure prediction in these four aspects are estimated using combinations of the following four biological significance tests and datasets: (1) Fitness profiling coherence test: used to evaluate the cellular fitness aspect of the candidate secondary structure. In this test, the fitness profiling scores of a fitness ncRNA similar set are compared to those of the fitness ncRNA reference collection using the rank-sum test [22]. To perform the fitness coherence test, we adopted the fitness functional profiling data from the work of Parker et al. [25]. (2)/(3) Interaction complex prevalence test and the gene ontology enrichment test: designed to estimate the RPI complex formation and the post-transcriptional regulation interpretability. Based on the Fisher exact test [23], we test the interaction complex prevalence and gene ontology enrichment of the proteins interacting with the RPI ncRNA similar set for RPI complex interpretability. Similar, these tests are performed on the genes transcribed to contain the 3'UTR structural similar set for post-transcriptional regulation interpretability. For all species, the physical protein–protein interaction data were gathered from the BioGRID database [32]. And the gene ontology annotation terms in different species were all downloaded from the Gene Ontology Consortium database [33]. (4) Ribosome profiling coherence test: adopted to calculate the translational interpretability significance. In this test, the ribosome profiling coherence of the 5'UTR structural similar set is compared with that of the species-specific 5'UTR regulatory reference structure collection via the rank-sum test [22]. For *Homo sapiens*, we obtained the preprocessed ribosome profiling datasets from the HRPDViewer database [34]. And for *Saccharomyces cerevisiae*, 5 different ribo-seq datasets were downloaded and merged to be used in the test: methionine-restricted cells [35], ribonuclease-treated strains [36], rich/starvation conditions [37], cells in the elongation phase [38], and localization identification environments [39]. For *Mus musculus*, *Rattus norvegicus*, *Danio rerio*, and *Arabidopsis thaliana*, we downloaded the species-specific ribosome profiling datasets from the RPFdb database [40].

### 2.5. Implementation of SSRTTool

The core prediction ranking algorithm of SSRTTool is implemented using the Python programming language (version 3.6.12). And the online web software is implemented using the PHP Model-View-Controller (MVC) framework CodeIgniter (version 2.1.3) as the back-end and the JavaScript framework JQuery (version 3.31) as the front-end. The RNA secondary structure visualization is presented using Forna [41].

## 3. Utility and Discussion

### 3.1. Web software interface

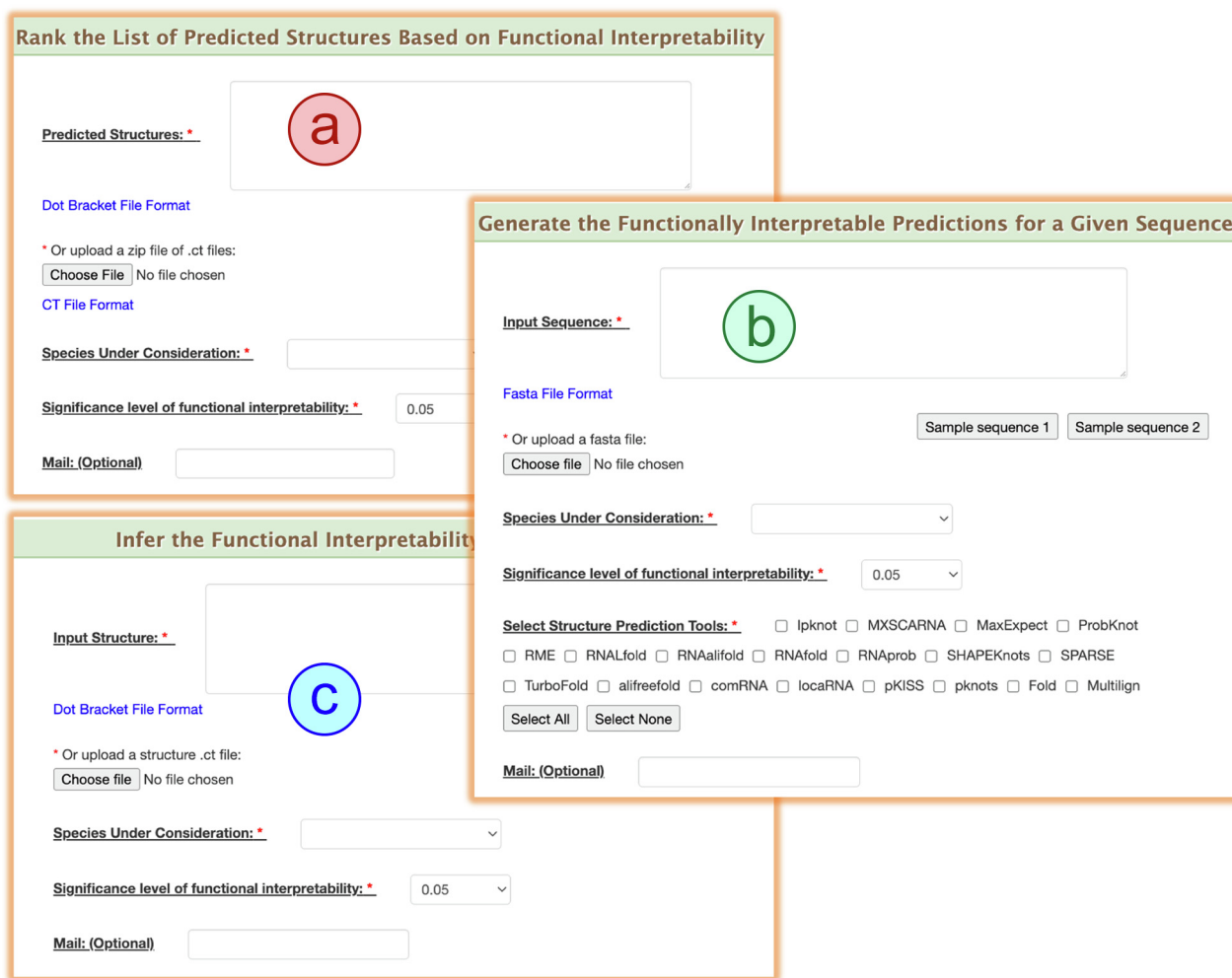
Since structure prediction for a given RNA sequence is usually performed before its actual secondary structure is clear, it is vital to evaluate the confidence of different predictions and infer the most prevalent structure among them. We constructed SSRTTool to tackle this obstacle. The prediction ranking algorithm based on functional interpretability was implemented with an easy-to-use interface. Three functions were implemented in SSRTTool to assist the evaluation of different predictions: (1) Function 1: Rank Predictions. In this function, users can provide a list of candidate structure predictions for a given RNA sequence, and SSRTTool can help evaluate the confidence rankings of these predictions based on species-specific functional interpretability. (2) Function 2: Generate and Rank Predictions. Users can also provide only the RNA

sequence of interest. In SSRTool, we implemented an automated pipeline for carrying out the prediction process for a given sequence to facilitate easy utilization of existing tools. The automation pipeline in SSRTool integrates 19 different tools and calculates the species-specific functional ranking for each prediction. SSRTool includes tools that are publicly available under the GNU General Public License or are declared to be free for academic usage: RNA-Fold [42], RNALfold [43], RNAprob [44], MaxExpect [45], RNAalifold [46], TurboFold [47], SPARSE [48], MXSCARNA [49], aliFreeFold [50], LocARNA [51], comRNA [52], RME [53], Ipknnot [7], ProbKnot [54], ShapeKnots [55], pKISS [56], pknnots [57], Fold [58], and Multilign [59]. In SSRTool, the prediction tools were applied to the given input sequence using default parameters suggested by the authors. And for tools that require homologous sequence information, the top five similar ones to the given

sequence provided by the Rfam `rfam_scan.pl` tool [60] are fed into these homology-based tools. (3) Function 3: Infer Interpretability. An extra function was implemented in SSRTool to help users infer the functional aspects of an RNA structure of interest. Users first input the RNA structure and then select the targeted species. SSRTool will compute the species-specific functional significance of the structure in four functional aspects. For each of these three SSRTool webtool functions, the estimated execution time (for a 352-base-long sequence processed when the server is not heavily loaded) and fundamental limitations in the maximum allowed sequence number and sequence/structure length are summarized in Table 2. The allowed maximum sequence length for each function of SSRTool is 2000 bases. However, in Function 2 (Generate and Rank Predictions) of SSRTool, the Multilign/Ipknnot prediction tools by design can only accept sequences with lengths

**Table 2**  
The computation time estimation (for a 352-base-long sequence processed when the server is not heavily loaded) and limitations of the functions in SSRTool.

SSRTool Functions	Estimated Execution Time	Allowed Sequence Number	Allowed Maximum Sequence/Structure Length
Function 1: Rank Predictions	~ 4.5 min (for 19 structures)	≥ 1 predicted structures	< 2000
Function 2: Generate & Rank Predictions	~ 10 min (using 19 tools)	1 sequence	< 2000
Function 3: Infer Interpretability	~ 1 min	1 structure	< 2000



**Fig. 2.** The interfaces of the three functions provided in SSRTool. (a) Function 1: Rank Predictions. Users can use this function to rank the significance of the provided structure predictions from one sequence based on functional interpretability. (b) Function 2: Generate and Rank Predictions. Users can also provide the RNA sequence of interest, and SSRTool will execute the automated structure prediction pipeline of 19 tools. Then these predictions will be ranked based on functional interpretability. (c) Function 3: Infer Interpretability. Users can input a secondary structure of interest, and SSRTool will compute the functional significance of the structure in four functional aspects.

less than 600/1500 bases. Since some of the tools can only accept canonical nucleotides, we restrict the input sequence to contain only canonical nucleotides (AUGC) for consistency in SSRTool.

In the input page of SSRTool, users need to specify different input data and parameters (Fig. 2). In Function 1 (Rank Predictions) of SSRTool (Fig. 2-a), users need to upload a .zip file of predicted structures in the .ct format or a list of predicted structures in the dot-bracket format. In Function 2 (Generate and Rank Predictions) of SSRTool (Fig. 2-b), users can either type in the RNA sequence of interest or upload the sequence in the .fasta format. Then the prediction tools intended to be included in the automated prediction process should be chosen. And a user-interested structure (either in the .ct format or in the dot-bracket format) is required in Function 3 (Infer Interpretability) of SSRTool (Fig. 2-c). Notice that the species under consideration and the threshold of functional significance should be selected in all three functions. Since the

web service may take a while to generate the rankings and the functional significance results, which depends on the server workload and the input sequence length, SSRTool allows users to leave an email for further notification when the job is finished. After completing the job, users will be directed to the detailed result page, which contains the rankings and functional significance scores of each predicted secondary structure (See Fig. 3 and the "A walk-through example" section). All the predicted structures can be visualized and downloaded in SSRTool.

### 3.2. A walk-through example

We provide a walk-through example of the human U11 snRNA (small nuclear RNA) to demonstrate the results generated by SSRTool (See Fig. 3). We utilize Function 2 (Generate and Rank Predictions) and input the RNA sequence of the U11 snRNA to SSRTool.

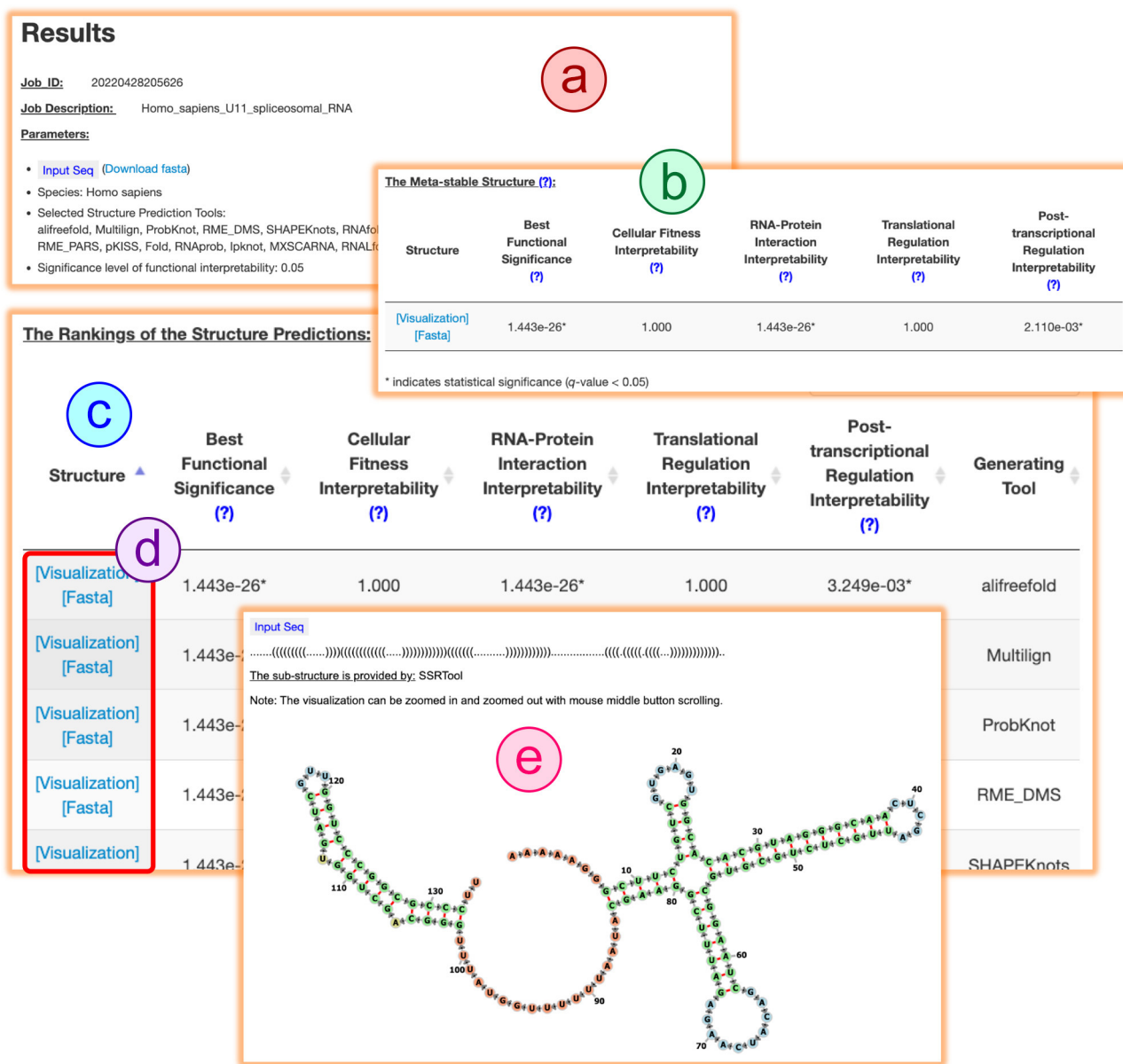


Fig. 3. The result page of prediction rankings and the suggested most native-like candidate structure by SSRTool. (a) The user-specified input and algorithm parameters. (b) The suggested most native-like secondary structure aggregated from the functionally interpretable predictions. (c) The individual significance rankings of the predicted structures. (d) The structure listed in the result page can all be downloaded in .fasta format or visualized by Forna. (e) The visualization of the selected secondary structure prediction.

In this example, all 19 prediction tools are included in the prediction process, and the significance threshold is set to be 0.05. As shown in the result page, the user-specified input sequence and the chosen parameters are first listed as a reminder for the user (Fig. 3-a). Then the meta-stable secondary structure aggregated from functionally interpretable structure predictions is given as the suggested most native-like candidate secondary structure. And the functional interpretability significance scores in four aspects are tabulated for this meta-stable secondary structure (Fig. 3-b). Finally, the detailed significance rankings for the predicted structures of the given RNA sequence are provided in a tabular form (Fig. 3-c). Users can visualize or download all structure predictions by clicking the links under the "Structure" column (Fig. 3-d and 3-e). By examining the result page of SSRTTool, users can get an idea of the most native-like structure prediction and the possible functional aspects of the given RNA. And users can also investigate the confidence ranking of each predicted structure. Similar result pages are provided in Function 1 (Rank Predictions) and Function 3 (Infer Interpretability) of SSRTTool.

### 3.3. Case study

SSRTTool is designed to help rank the prediction results and suggest the most native-like candidate secondary structure of a given RNA sequence. We first demonstrate the biological applicability of SSRTTool using the human U11 snRNA sequence that participates in the spliceosome. We input the U11 sequence into SSRTTool Function 2 (Generate and Rank Predictions) to automatically perform the prediction process using the 19 integrated tools and then rank

these predictions based on functional significance. The results reveal that the most native-like candidate structure is enriched in RPI complex formation (interpretability  $q$ -value =  $1.443e-26^*$ ) and post-transcriptional regulation (interpretability  $q$ -value =  $2.110e-03^*$ , see Fig. 3-b). The experimental secondary structure of U11 has now been verified to contain a 5'-ended stem-loop as the splice site and four other stem-loops [61]. By comparing the most native-like candidate secondary structure suggested by SSRTTool (See Fig. 3-e) with the experimentally verified structure,  $F1 = 0.988$  is achieved. The high  $F1$  value indicates the applicability of SSRTTool in ranking predictions based on functional significance. And in previous experiments, the functional roles of U11 within the spliceosome RPI complex have also been confirmed to target mRNA introns by the 5'-ended stem-loop for subsequent splicing [62], matching the identified significant RNA-protein interaction and post-transcriptional regulation aspects of U11. These calculated functional significance results in SSRTTool are consistent with the experimental findings. Therefore, SSRTTool can suggest the most native-like structure and the functionally significant aspects of an RNA sequence for users.

We provide a biochemically probed RNA secondary structure as another example for utilizing SSRTTool. The human immunodeficiency virus (HIV) Rev protein responsive element (RRE) is the essential molecule in HIV gene regulation. And the dominant conformation structure of HIV-1 RRE was probed by the SHAPE chemical reagents in previous works [63]. We input the HIV-1 RRE sequence into SSRTTool Function 2 (Generate and Rank Predictions) to perform the automatic structure prediction and ranking using the human reference species. The results are summarized in

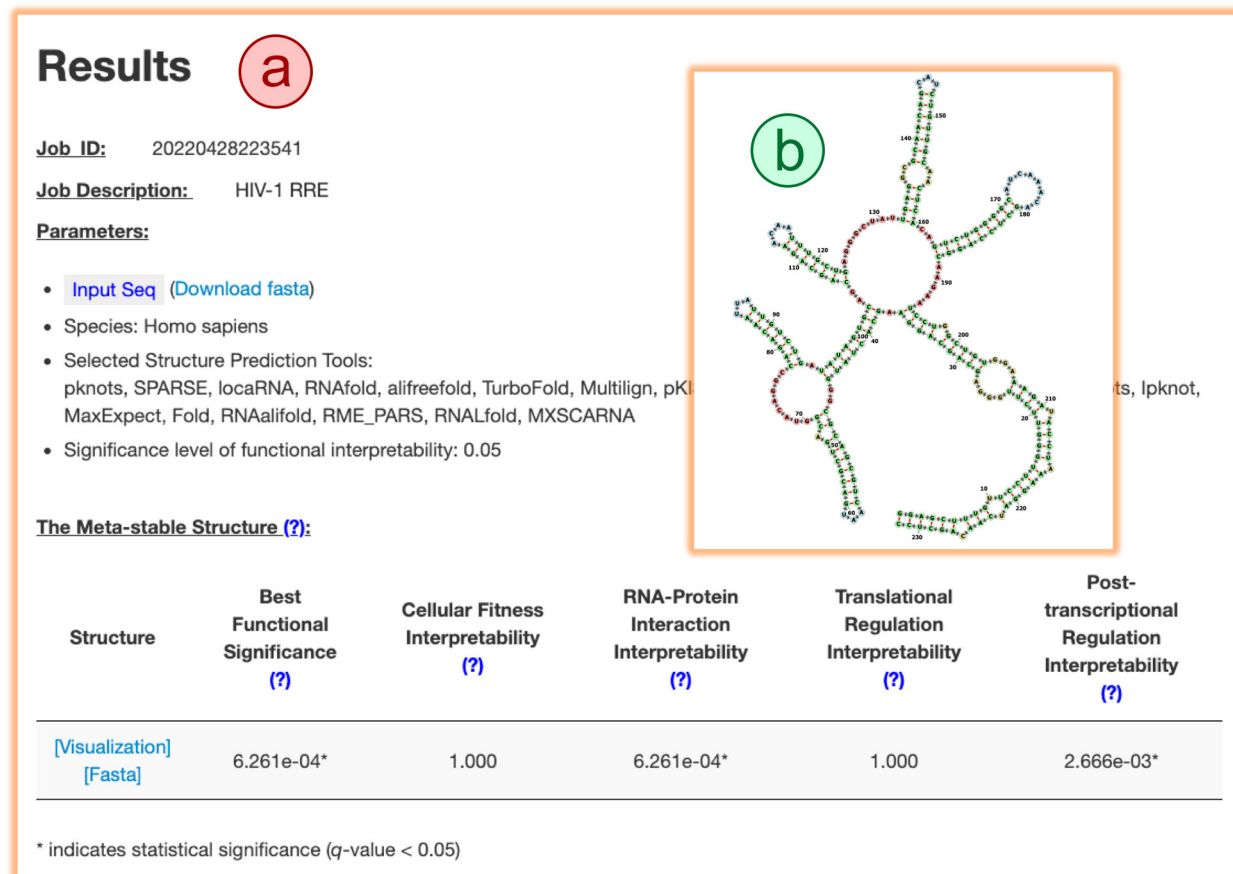


Fig. 4. The confidence ranking results of the HIV-1 RRE sequence. (a) The input parameters, the aggregated HIV meta-stable structure, and the computed functional significance for the meta-stable structure. (b) The visualization of the aggregated meta-stable structure.

**Fig. 4.** The results show that the aggregated meta-stable structure prediction bears five stem-loops and obtains an F1 value of 0.994 to the more functionally active five stem-loop structure identified by SHAPE experiments [63]. This meta-stable structure also demonstrates functional significance in RNA–protein interaction interpretability ( $q = 6.261e-04^*$ ) and post-transcriptional regulation interpretability ( $q = 2.666e-03^*$ ). Researchers have now verified that the HIV-1 RRE can bind the Rev protein to control HIV RNA post-transcriptional trafficking [64,63], showing consistency with the SSRTTool-calculated significant functional aspects. This extra example further validates the biological applicability of SSRTTool.

### 3.4. Comparison of SSRTTool-aggregated results with other prediction tools in different species

SSRTTool provides an automated pipeline for ranking and extracting functionally interpretable structures of a given RNA sequence in six different model organisms. It also aggregates the meta-stable secondary structure based on the calculated functional significance of the predictions. It is of interest if the aggregated meta-stable structures help suggest the most native-like candidate secondary structures. We thus compared the SSRTTool-aggregated secondary structures with the results of 22 existing RNA secondary structure prediction tools: RNAFold [42], RNALfold [43], RNAprob [44], MaxExpect [45], RNAalifold [46], TurboFold [47], SPARSE [48], MXSCARNA [49], aliFreeFold [50], LocARNA [51], RME (both with the PARS and DMS models) [53], RNAspa [65], Ipknnot [7], HotKnots [66], IterativeHfold [67], ProbKnot [54], ShapeKnots [55], pKISS [56], pknnots [57], Fold [58], and Multilign [59]. Our previous study has demonstrated that the prediction ranking algorithm can provide meta-stable secondary structures for given RNA sequences with high matches to verified structures in humans and yeast [16]. SSRTTool implements and extends the prediction ranking algorithm to support four additional model organisms (*Mus musculus*, *Rattus norvegicus*, *Danio rerio*, and *Arabidopsis thaliana*). We compared the meta-stable secondary structures aggregated by SSRTTool with the 22 existing RNA structure prediction tools in these four additional model organisms. The accuracy of a predicted structure can be evaluated using the F1 metric [68]:

$$\text{Recall} = \frac{TP}{TP + FN},$$

$$\text{Precision} = \frac{TP}{TP + FP},$$

$$F1 = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}},$$

where TP counts the matching pairings between the predicted structure and the actual structure, FP sums up the wrongly paired base pairs in the predicted structure, and FN represents the number of missing true pairings in the predicted structure. A high F1 value of a structure prediction result indicates that this prediction resembles the actual structure.

To gather the ground-truth structure test set for *Mus musculus*, *Rattus norvegicus*, *Danio rerio*, and *Arabidopsis thaliana*, we downloaded the RNAStralign structure benchmark dataset [12] and BLASTed these benchmarks to species-specific known RNAs annotated by RefSeq [28] to obtain the species-related structures (E-value < 1e-6, percent-identity > 70%). We randomly picked 114, 119, 107, and 138 RNA sequences with verified known structures in *Mus musculus*, *Danio rerio*, *Arabidopsis thaliana*, and *Rattus norvegicus*, respectively, as the test sets. The comparison results for different species are summarized in Table 3, Table 4, Table 5, and Table 6. Similar to the results in humans and yeast, the aggrega-

tion results from SSRTTool outperform most of the existing tools in the test sets. Although there are fewer available known structures in the reference collections of *Danio rerio*, *Arabidopsis thaliana*, and *Rattus norvegicus* to precisely estimate the functional interpretability, considering the functional interpretability of structure predictions still helps users identify the most native-like candidate secondary structures (performance in at least top 5). On the contrary, no one existing tool is guaranteed to provide the top five best structure predictions than others among sequences from different species. These comparison results conclude that based on functional interpretability aggregation, SSRTTool can help better suggest the native-like candidate structures than existing tools.

**Table 3**

The structure prediction performance comparison in *Mus musculus*. The values are represented in the form of average  $\pm$  relative standard error.

Prediction Tool	F1	Precision	Recall
<b>SSRTTool aggregation</b>	0.650 $\pm$ 2.4%	0.650 $\pm$ 2.5%	0.672 $\pm$ 2.5%
RNAfold	0.605 $\pm$ 2.6%	0.572 $\pm$ 2.6%	0.656 $\pm$ 2.5%
HotKnots	0.605 $\pm$ 2.7%	0.575 $\pm$ 2.8%	0.654 $\pm$ 2.7%
IterativeHfold	0.602 $\pm$ 2.7%	0.573 $\pm$ 2.8%	0.651 $\pm$ 2.8%
MaxExpect	0.597 $\pm$ 2.6%	0.570 $\pm$ 2.6%	0.645 $\pm$ 2.6%
RNAalifold	0.586 $\pm$ 2.6%	0.548 $\pm$ 2.6%	0.651 $\pm$ 2.6%
ProbKnot	0.585 $\pm$ 2.5%	0.549 $\pm$ 2.5%	0.646 $\pm$ 2.6%
SHAPEKnots	0.576 $\pm$ 2.6%	0.545 $\pm$ 2.6%	0.627 $\pm$ 2.6%
alifold	0.573 $\pm$ 2.5%	0.537 $\pm$ 2.6%	0.634 $\pm$ 2.6%
RME_DMS	0.560 $\pm$ 2.8%	0.533 $\pm$ 2.8%	0.607 $\pm$ 2.9%
TurboFold	0.559 $\pm$ 2.7%	0.535 $\pm$ 2.8%	0.607 $\pm$ 2.8%
Multilign	0.559 $\pm$ 2.8%	0.529 $\pm$ 2.8%	0.610 $\pm$ 2.9%
RME_PARS	0.557 $\pm$ 2.8%	0.539 $\pm$ 2.9%	0.595 $\pm$ 2.9%
locARNA	0.548 $\pm$ 2.8%	0.511 $\pm$ 2.8%	0.612 $\pm$ 2.9%
pKISS	0.544 $\pm$ 2.7%	0.533 $\pm$ 2.8%	0.574 $\pm$ 2.7%
SPARSE	0.542 $\pm$ 2.8%	0.505 $\pm$ 2.8%	0.604 $\pm$ 3.0%
Ipknnot	0.517 $\pm$ 2.4%	0.569 $\pm$ 2.6%	0.489 $\pm$ 2.4%
RNAspa	0.501 $\pm$ 3.1%	0.476 $\pm$ 3.1%	0.543 $\pm$ 3.3%
RNALfold	0.418 $\pm$ 2.9%	0.479 $\pm$ 3.2%	0.402 $\pm$ 3.1%
RNAprob	0.388 $\pm$ 2.3%	0.364 $\pm$ 2.3%	0.432 $\pm$ 2.6%
Fold	0.388 $\pm$ 2.3%	0.364 $\pm$ 2.3%	0.432 $\pm$ 2.6%
MXSCARNA	0.341 $\pm$ 2.4%	0.355 $\pm$ 2.5%	0.337 $\pm$ 2.5%
pknnots	0.296 $\pm$ 2.2%	0.277 $\pm$ 2.1%	0.330 $\pm$ 2.5%

**Table 4**

The structure prediction performance comparison in *Danio rerio*. The values are represented in the form of average  $\pm$  relative standard error.

Prediction Tool	F1	Precision	Recall
SHAPEKnots	0.712 $\pm$ 2.5%	0.695 $\pm$ 2.4%	0.733 $\pm$ 2.6%
<b>SSRTTool aggregation</b>	0.706 $\pm$ 2.3%	0.715 $\pm$ 2.3%	0.708 $\pm$ 2.4%
RNAfold	0.705 $\pm$ 2.4%	0.697 $\pm$ 2.4%	0.717 $\pm$ 2.4%
MaxExpect	0.698 $\pm$ 2.4%	0.694 $\pm$ 2.4%	0.708 $\pm$ 2.5%
RME_DMS	0.694 $\pm$ 2.4%	0.683 $\pm$ 2.4%	0.708 $\pm$ 2.5%
RME_PARS	0.693 $\pm$ 2.4%	0.693 $\pm$ 2.4%	0.700 $\pm$ 2.5%
ProbKnot	0.692 $\pm$ 2.4%	0.676 $\pm$ 2.4%	0.717 $\pm$ 2.5%
RNALfold	0.671 $\pm$ 2.8%	0.666 $\pm$ 2.8%	0.679 $\pm$ 2.9%
Ipknnot	0.655 $\pm$ 1.6%	0.702 $\pm$ 1.7%	0.625 $\pm$ 1.6%
HotKnots	0.639 $\pm$ 2.4%	0.630 $\pm$ 2.3%	0.651 $\pm$ 2.4%
IterativeHfold	0.639 $\pm$ 2.4%	0.630 $\pm$ 2.3%	0.651 $\pm$ 2.4%
pKISS	0.625 $\pm$ 2.8%	0.627 $\pm$ 2.8%	0.627 $\pm$ 2.9%
SPARSE	0.608 $\pm$ 3.2%	0.594 $\pm$ 3.1%	0.627 $\pm$ 3.3%
RNAprob	0.606 $\pm$ 2.3%	0.595 $\pm$ 2.3%	0.618 $\pm$ 2.4%
Fold	0.606 $\pm$ 2.3%	0.595 $\pm$ 2.3%	0.618 $\pm$ 2.4%
locARNA	0.601 $\pm$ 3.2%	0.587 $\pm$ 3.1%	0.620 $\pm$ 3.3%
RNAalifold	0.570 $\pm$ 3.0%	0.565 $\pm$ 2.9%	0.579 $\pm$ 3.0%
alifold	0.563 $\pm$ 3.0%	0.567 $\pm$ 3.0%	0.566 $\pm$ 3.0%
TurboFold	0.563 $\pm$ 3.0%	0.572 $\pm$ 3.1%	0.559 $\pm$ 3.0%
Multilign	0.552 $\pm$ 2.9%	0.553 $\pm$ 2.9%	0.555 $\pm$ 2.9%
pknnots	0.472 $\pm$ 2.1%	0.482 $\pm$ 2.3%	0.472 $\pm$ 2.1%
MXSCARNA	0.252 $\pm$ 2.5%	0.279 $\pm$ 2.6%	0.236 $\pm$ 2.5%
RNAspa	0.239 $\pm$ 3.1%	0.228 $\pm$ 2.9%	0.252 $\pm$ 3.2%



**Table 5**

The structure prediction performance comparison in *Arabidopsis thaliana*. The values are represented in the form of average  $\pm$  relative standard error.

Prediction Tool	F1	Precision	Recall
RNAalifold	0.656 $\pm$ 1.7%	0.638 $\pm$ 1.8%	0.682 $\pm$ 1.7%
alifold	0.644 $\pm$ 1.8%	0.629 $\pm$ 1.8%	0.665 $\pm$ 1.8%
TurboFold	0.638 $\pm$ 1.9%	0.623 $\pm$ 2.0%	0.659 $\pm$ 1.9%
<b>SSRTool aggregation</b>	0.605 $\pm$ 2.2%	0.608 $\pm$ 2.3%	0.611 $\pm$ 2.3%
SPARSE	0.579 $\pm$ 2.1%	0.558 $\pm$ 2.1%	0.605 $\pm$ 2.2%
Multilign	0.575 $\pm$ 2.2%	0.560 $\pm$ 2.2%	0.598 $\pm$ 2.2%
locarNA	0.567 $\pm$ 2.2%	0.546 $\pm$ 2.2%	0.593 $\pm$ 2.2%
SHAPEKnots	0.541 $\pm$ 2.1%	0.522 $\pm$ 2.1%	0.567 $\pm$ 2.1%
RNAfold	0.532 $\pm$ 2.1%	0.514 $\pm$ 2.1%	0.555 $\pm$ 2.2%
MaxExpect	0.530 $\pm$ 2.0%	0.519 $\pm$ 2.0%	0.546 $\pm$ 2.0%
ProbKnot	0.512 $\pm$ 2.1%	0.496 $\pm$ 2.0%	0.534 $\pm$ 2.1%
HotKnots	0.504 $\pm$ 2.5%	0.492 $\pm$ 2.5%	0.522 $\pm$ 2.6%
IterativeHFold	0.501 $\pm$ 2.6%	0.489 $\pm$ 2.5%	0.519 $\pm$ 2.6%
pKISS	0.490 $\pm$ 2.3%	0.487 $\pm$ 2.3%	0.498 $\pm$ 2.3%
RME_PARS	0.469 $\pm$ 2.4%	0.465 $\pm$ 2.4%	0.478 $\pm$ 2.5%
RME_DMS	0.461 $\pm$ 2.5%	0.450 $\pm$ 2.4%	0.477 $\pm$ 2.6%
Ipknot	0.457 $\pm$ 1.8%	0.498 $\pm$ 1.9%	0.430 $\pm$ 1.8%
RNAAspa	0.414 $\pm$ 3.0%	0.396 $\pm$ 2.9%	0.437 $\pm$ 3.2%
pknots	0.366 $\pm$ 1.8%	0.351 $\pm$ 1.7%	0.386 $\pm$ 1.9%
MXSCARNA	0.295 $\pm$ 1.8%	0.321 $\pm$ 1.8%	0.278 $\pm$ 1.8%
RNAprob	0.275 $\pm$ 1.7%	0.267 $\pm$ 1.6%	0.288 $\pm$ 1.8%
Fold	0.275 $\pm$ 1.7%	0.267 $\pm$ 1.6%	0.288 $\pm$ 1.8%
RNAfold	0.256 $\pm$ 2.3%	0.340 $\pm$ 3.0%	0.212 $\pm$ 2.0%

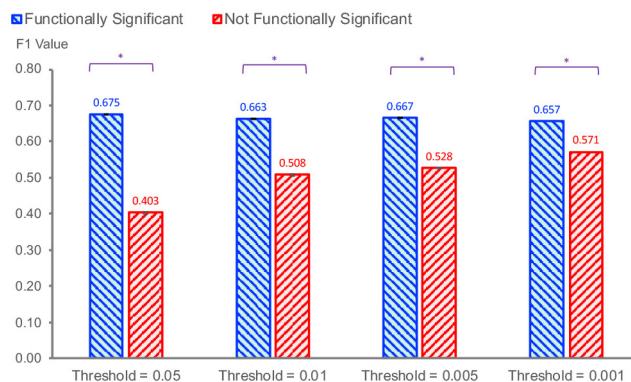
**Table 6**

The structure prediction performance comparison in *Rattus norvegicus*. The values are represented in the form of average  $\pm$  relative standard error.

Prediction Tool	F1	Precision	Recall
HotKnots	0.626 $\pm$ 3.1%	0.621 $\pm$ 3.2%	0.638 $\pm$ 2.9%
IterativeHFold	0.626 $\pm$ 3.1%	0.621 $\pm$ 3.2%	0.638 $\pm$ 2.9%
RNAalifold	0.595 $\pm$ 2.2%	0.577 $\pm$ 2.3%	0.627 $\pm$ 2.0%
alifold	0.585 $\pm$ 2.2%	0.568 $\pm$ 2.4%	0.614 $\pm$ 2.0%
<b>SSRTool aggregation</b>	0.566 $\pm$ 2.7%	0.602 $\pm$ 2.9%	0.548 $\pm$ 2.6%
locarNA	0.559 $\pm$ 2.5%	0.540 $\pm$ 2.6%	0.590 $\pm$ 2.4%
SPARSE	0.558 $\pm$ 2.5%	0.539 $\pm$ 2.6%	0.587 $\pm$ 2.4%
TurboFold	0.537 $\pm$ 2.7%	0.532 $\pm$ 2.8%	0.553 $\pm$ 2.5%
Multilign	0.510 $\pm$ 2.8%	0.504 $\pm$ 2.9%	0.524 $\pm$ 2.7%
RNAAspa	0.489 $\pm$ 2.9%	0.475 $\pm$ 3.0%	0.512 $\pm$ 2.9%
MaxExpect	0.472 $\pm$ 2.8%	0.468 $\pm$ 2.8%	0.486 $\pm$ 2.7%
ProbKnot	0.466 $\pm$ 2.7%	0.456 $\pm$ 2.8%	0.488 $\pm$ 2.7%
RNAfold	0.442 $\pm$ 2.9%	0.427 $\pm$ 2.9%	0.469 $\pm$ 2.9%
pKISS	0.442 $\pm$ 2.6%	0.449 $\pm$ 2.7%	0.444 $\pm$ 2.6%
SHAPEKnots	0.427 $\pm$ 3.1%	0.413 $\pm$ 3.1%	0.452 $\pm$ 3.1%
RME_PARS	0.426 $\pm$ 3.0%	0.429 $\pm$ 3.1%	0.430 $\pm$ 2.9%
RME_DMS	0.423 $\pm$ 2.9%	0.418 $\pm$ 3.0%	0.436 $\pm$ 2.9%
Ipknot	0.386 $\pm$ 1.9%	0.449 $\pm$ 2.2%	0.351 $\pm$ 1.7%
RNAprob	0.252 $\pm$ 1.7%	0.247 $\pm$ 1.7%	0.264 $\pm$ 1.8%
Fold	0.252 $\pm$ 1.7%	0.247 $\pm$ 1.7%	0.264 $\pm$ 1.8%
MXSCARNA	0.248 $\pm$ 1.9%	0.272 $\pm$ 2.1%	0.234 $\pm$ 1.8%
pknots	0.200 $\pm$ 1.4%	0.190 $\pm$ 1.3%	0.222 $\pm$ 1.6%
RNAfold	0.191 $\pm$ 2.2%	0.273 $\pm$ 3.2%	0.149 $\pm$ 1.6%

### 3.5. Functional significance thresholds help identify native-like predictions

In calculating the functional interpretability of the given structures, users can adjust the significance threshold in the implemented prediction ranking algorithm to control the confidence level. We next provide a thorough analysis of the impact of the choice of the significance thresholds. In this analysis, we first collected all structure predictions and their corresponding species-specific functional interpretability significance values from the test sets in *Homo sapiens*, *Saccharomyces cerevisiae*, *Mus musculus*, *Danio rerio*, *Arabidopsis thaliana*, and *Rattus norvegicus*. Then we compared the average F1 values of the predicted structures having interpretability  $q$ -values lower than the specified threshold with those having  $q$ -values higher than the specified threshold. We adopted the thresholds of 0.05, 0.01, 0.005, and 0.001 to estimate



**Fig. 5.** The functional significance threshold comparison in identifying the native-like secondary structures using SSRTool.

the threshold effect. As shown in Fig. 5, under all chosen threshold values, the average F1 values of predictions having interpretability  $q$ -values lower than the selected threshold are always statistically larger than the average F1 values of those predictions having interpretability  $q$ -values larger than the selected threshold (one-tailed  $t$ -test  $p$ -value  $<$  0.0001). And from the analysis, the discrimination using functional interpretability scores performs best when threshold = 0.05. The threshold of 0.05 is thus suggested for users. In conclusion, the functional interpretability significance can help relate the similarities between structure predictions of an RNA sequence and its actual structure.

## 4. Conclusions

RNA structure prediction evaluation is currently an urgent task in functional biology. An online software tool called SSRTool was implemented in this research to help provide prediction rankings based on the concept of functional interpretability. In SSRTool, users can compute the four functional significance scores (cellular fitness, RPI complex formation, translational regulation, and post-transcriptional regulation) in six different species (*Homo sapiens*, *Saccharomyces cerevisiae*, *Mus musculus*, *Rattus norvegicus*, *Danio rerio*, and *Arabidopsis thaliana*) as confidence rankings for the given RNA structures or the automatic generated structure predictions of an RNA sequence. And a user-friendly interface for using SSRTool was also constructed for convenient usage and visualization. We reported that the prediction interpretability rankings calculated by SSRTool could indirectly indicate the similarities between the predicted structures and the native prevalent structures. We believe that this online software tool can enhance the computational analysis of RNA secondary structures and broaden the understanding of RNA function-structure relations.

## Web Service Availability

The online web software SSRTool is freely available online at available online at <https://cobisHSS0.im.nuk.edu.tw/SSRTool/>, <http://cosbi3.ee.ncku.edu.tw/SSRTool/>, or the redirecting site <https://github.com/cobisLab/SSRTool/>.

## CRedit authorship contribution statement

**Tzu-Hsien Yang:** Conceptualization, Investigation, Supervision, Project administration, Software, Formal analysis, Writing - original draft, Writing - review & editing. **Yu-Cian Lin:** Investigation, Software, Visualization, Writing - original draft. **Min Hsia:** Software, Visualization, Writing - review & editing. **Zhan-Yi Liao:** Software, Visualization, Writing - review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This study was supported by the National University of Kaohsiung and the Ministry of Science and Technology of Taiwan (MOST 107–2218-E-390–009-MY3, MOST 110–2222-E-390–001).

## References

- Cech TR, Steitz JA. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* 2014;157(1):77–94.
- Kwok CK, Tang Y, Asmann SM, Bevilacqua PC. The RNA structurome: transcriptome-wide structure probing with next-generation sequencing. *Trends in Biochemical Sciences* 2015;40(4):221–32.
- Mortimer SA, Kidwell MA, Doudna JA. Insights into RNA structure and function from genome-wide studies. *Nature Reviews Genetics* 2014;15(7):469.
- Spitale RC, Flynn RA, Zhang QC, Crisalli P, Lee B, Jung J-W, Kuchelmeister HY, Batista PJ, Torre EA, Kool ET, et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* 2015;519(7544):486.
- He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews Genetics* 2004;5(7):522.
- Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. Understanding the transcriptome through RNA structure. *Nature Reviews Genetics* 2011;12(9):641.
- Sato K, Kato Y, Hamada M, Akutsu T, Asai K. IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinformatics* 2011;27(13):i85–93.
- Dagenais P, Girard N, Bonneau E, Legault P. Insights into RNA structure and dynamics from recent NMR and X-ray studies of the *Neurospora Varkud* satellite ribozyme. *Wiley Interdisciplinary Reviews: RNA* 2017;8(5):e1421.
- Kappel K, Zhang K, Su Z, Watkins AM, Kladwang W, Li S, Pintlilic G, Topkar VV, Rangan R, Zheludev IN, et al. Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nature Methods* 2020;17(7):699–707.
- Chen Y, Pollack L. SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdisciplinary Reviews: RNA* 2016;7(4):512–26.
- Yao J, Reinhartz V, Major F, Waldspühl J. RNA-MolP: prediction of RNA secondary structure and local 3d motifs from sequence data. *Nucleic Acids Research* 2017;45(W1):W440–4.
- Tan Z, Fu Y, Sharma G, Mathews DH. TurboFold II: RNA structural alignment and secondary structure prediction informed by multiple homologs. *Nucleic Acids Research* 2017;45(20):11570–81.
- Lorenz R, Luntzer D, Hofacker IL, Stadler PF, Wolfinger MT. Shape directed RNA folding. *Bioinformatics* 2015;32(1):145–7.
- Karplus M. The Levinthal paradox: yesterday and today. *Folding and Design* 1997;2:569–75.
- A. Wienecke, A. Laederach, A novel algorithm for ranking rna structure candidates, *Biophysical Journal*.
- Yang T-H. An aggregation method to identify the RNA meta-stable secondary structure and its functionally interpretable structure ensemble. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2022;19(1):75–86.
- Heyne S, Will S, Beckstette M, Backofen R. Lightweight comparison of RNAs based on exact sequence–structure matches. *Bioinformatics* 2009;25(16):2095–102.
- Long Y, Wang X, Youmans DT, Cech TR. How do lncRNAs regulate transcription? *Science*. *Advances* 2017;3(9):eao2110.
- Johnson AG, Grosely R, Petrov AN, Puglisi JD. Dynamics of IRES-mediated translation. *Philosophical Transactions of the Royal Society B* 2017;372(1716):20160177.
- Kertesz M, Wan Y, Mazor E, Rinn JL, Nutter RC, Chang HY, Segal E. Genome-wide measurement of RNA secondary structure in yeast. *Nature* 2010;467(7311):103.
- Yang T-H. Transcription factor regulatory modules provide the molecular mechanisms for functional redundancy observed among transcription factors in yeast. *BMC Bioinformatics* 2019;20(23):1–16.
- Yang T-H, Wang C-Y, Tsai H-C, Liu C-T. Human IRES Atlas: an integrative platform for studying IRES-driven translational regulation in humans. *Database* 2021.
- Yang T-H, Wu W-S. Inferring functional transcription factor-gene binding pairs by integrating transcription factor binding data with transcription factor knockout data. *BMC Systems Biology* 2013;7(S6):S13.
- Gardner PP, Wilm A, Washietl S. A benchmark of multiple sequence alignment programs upon structural RNAs. *Nucleic Acids Research* 2005;33(8):2433–9.
- Parker S, Fraczek MG, Wu J, Shamsah S, Manousaki A, Dungtrattanalert K, de Almeida RA, Invernizzi E, Burgis T, Omara W, et al. Large-scale profiling of noncoding RNA function in yeast. *PLoS Genetics* 2018;14(3):e1007253.
- Panni S, Prakash A, Bateman A, Orchard S. The yeast noncoding RNA interaction network. *RNA* 2017;23(10):1479–92.
- Lin Y, Liu T, Cui T, Wang Z, Zhang Y, Tan P, Huang Y, Yu J, Wang D. RNAInter in 2020: RNA interactome repository with increased coverage and annotation. *Nucleic Acids Research* 2020;48(D1):D189–97.
- O’Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* 2016;44(D1):D733–45.
- Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al. The Arabidopsis information resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* 2012;40(D1):D1202–10.
- Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T, Schroeder M, et al. SGD: *Saccharomyces* genome database. *Nucleic Acids Research* 1998;26(1):73–9.
- Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, et al. The UCSC genome browser database: 2019 update. *Nucleic Acids Research* 2019;47(D1):D853–8.
- Oughtred R, Stark C, Breitkreutz B-J, Rust J, Boucher L, Chang C, Kolas N, O’Donnell L, Leung G, McAdam R, et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Research* 2019;47(D1):D529–41.
- Consortium GO. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Research* 2019;47(D1):D330–8.
- Wu W-S, Jiang Y-X, Chang J-W, Chu Y-H, Chiu Y-H, Tsao Y-H, Nordling TE, Tseng Y-Y, Tseng JT. HRPDviewer: human ribosome profiling data viewer. *Database* 2018.
- Zou K, Ouyang Q, Li H, Zheng J. A global characterization of the translational and transcriptional programs induced by methionine restriction through ribosome profiling and RNA-seq. *BMC Genomics* 2017;18(1):189.
- Gerashchenko MV, Gladyshev VN. Ribonuclease selection for ribosome profiling. *Nucleic Acids Research* 2016;45(2). e6–e6.
- Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 2009;324(5924):218–23.
- Lareau LF, Hite DH, Hogan GJ, Brown PO. Distinct stages of the translation elongation cycle revealed by sequencing ribosome-protected mRNA fragments. *Elife* 2014;3:e01257.
- B.M. Zid, E.K. O’shea, Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast, *Nature* 514 (7520) (2014) 117..
- H. Wang, L. Yang, Y. Wang, L. Chen, H. Li, Z. Xie, RPFdb v2. 0: an updated database for genome-wide information of translated mRNA generated from ribosome profiling, *Nucleic Acids Research* 47 (D1) (2019) D230–D234..
- Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics* 2015;31(20):3377–9.
- Lorenz R, Bernhart SH, Zu Siederdisen CH, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA package 2.0. *Algorithms for Molecular Biology* 2011;6(1):26.
- Hofacker IL, Priwitzer B, Stadler PF. Prediction of locally stable RNA secondary structures for genome-wide surveys. *Bioinformatics* 2004;20(2):186–90.
- F. Deng, M. Ledda, S. Vaziri, S. Aviran, Data-directed RNA secondary structure prediction using probabilistic modeling, *RNA*.
- Lu ZJ, Gloor JW, Mathews DH. Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA* 2009;15(10):1805–13.
- Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* 2008;9(1):474.
- Reuter JS, Mathews DH. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 2010;11(1):129.
- Will S, Otto C, Miladi M, Möhl M, Backofen R. SPARSE: quadratic time simultaneous alignment and folding of RNAs without sequence-based heuristics. *Bioinformatics* 2015;31(15):2489–96.
- Tabai Y, Kiryu H, Kin T, Asai K. A fast structural multiple alignment method for long RNA sequences. *BMC Bioinformatics* 2008;9(1):33.
- Glouzon J-PS, Ouagraoua A. alifreeFold: an alignment-free approach to predict secondary structure from homologous RNA sequences. *Bioinformatics* 2018;34(13):i70–8.
- Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R. LocARNA-P: accurate boundary prediction and improved detection of structural RNAs. *RNA* 2012;18(5):900–14.
- Ji Y, Xu X, Stormo GD. A graph theoretical approach for predicting common RNA secondary structure motifs including pseudoknots in unaligned sequences. *Bioinformatics* 2004;20(10):1591–602.
- Wu Y, Shi B, Ding X, Liu T, Hu X, Yip KY, Yang ZR, Mathews DH, Lu ZJ. Improved prediction of RNA secondary structure by integrating the free energy model with restraints derived from experimental probing data. *Nucleic Acids Research* 2015;43(15):7247–59.
- Bellaousov S, Mathews DH. ProbKnot: fast prediction of RNA secondary structure including pseudoknots. *RNA* 2010;16(10):1870–80.
- Hajdin CE, Bellaousov S, Huggins V, Leonard CW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proceedings of the National Academy of Sciences* 2013;110(14):5498–503.
- Theis C, Janssen S, Giegerich R. Prediction of RNA secondary structure including kissing hairpin motifs. In: *International Workshop on Algorithms in Bioinformatics*. Springer; 2010. p. 52–64.

- [57] Rivas E, Eddy SR. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *Journal of Molecular Biology* 1999;285(5):2053–68.
- [58] Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology* 1999;288(5):911–40.
- [59] Xu Z, Mathews DH. Multilign: an algorithm to predict secondary structures conserved in multiple RNA sequences. *Bioinformatics* 2011;27(5):626–32.
- [60] Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Research* 2017;46(D1):D335–42.
- [61] Russell AG, Charette JM, Spencer DF, Gray MW. An early evolutionary origin for the minor spliceosome. *Nature* 2006;443(7113):863–6.
- [62] Kolosova I, Padgett RA. U11 snRNA interacts in vivo with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. *RNA* 1997;3(3):227.
- [63] Sherpa C, Rausch JW, Le Grice SF, Hammarskjöld M-L, Rekosh D. The hiv-1 rev response element (rre) adopts alternative conformations that promote different rates of virus replication. *Nucleic Acids Research* 2015;43(9):4676–86.
- [64] Bartel DP, Zapp ML, Green MR, Szostak JW. Hiv-1 rev regulation involves recognition of non-watson-crick base pairs in viral rna. *Cell* 1991;67(3):529–36.
- [65] Horesh Y, Doniger T, Michaeli S, Unger R. RNAspa: a shortest path approach for comparative prediction of the secondary structure of ncRNA molecules. *BMC Bioinformatics* 2007;8(1):366.
- [66] Ren J, Rastegari B, Condon A, Hoos HH. HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *RNA* 2005;11(10):1494–504.
- [67] Jabbari H, Condon A. A fast and robust iterative algorithm for prediction of RNA pseudoknotted secondary structures. *BMC Bioinformatics* 2014;15(1):147.
- [68] Yang T-H, Yang Y-C, Tu K-C. regCNN: identifying drosophila genome-wide cis-regulatory modules via integrating the local patterns in epigenetic marks and transcription factor binding motifs. *Computational and Structural Biotechnology Journal* 2022;20:296–308.