Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A fully open structure-guided RNA foundation model for robust structural and functional inference

S. Kevin Zhou

s.kevin.zhou@gmail.com

University of Science and Technology of China

Heqin Zhu

University of Science and Technology of China https://orcid.org/0000-0001-9469-950X

Ruifeng Li

Zhejiang University https://orcid.org/0000-0002-4373-5811

Feng Zhang

University of Science and Technology of China

Fenghe Tang

School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China https://orcid.org/0009-0009-6193-4855

Tong Ye

University of Science and Technology of China

Xin Li

University of Science and Technology of China

Yunjie Gu

School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China

Peng Xiong

University of Science and Technology of China

Biological Sciences - Article

Keywords:

Posted Date: August 29th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7321052/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature on November 4th, 2026. See the published version at https://doi.org/10.1038/s41586-026-11210-2.

A fully open structure-guided RNA foundation model for robust structural and functional inference

Heqin Zhu[1,2], Ruifeng Li[3], Feng Zhang[1,2], Fenghe Tang[1,2], Tong Ye[1,2], Xin Li[1,2], Yunjie Gu[1,2], Peng Xiong[1,2*], S. Kevin Zhou[1,2,4,5*]

[1*]School of Biomedical Engineering, Division of Life Sciences and Medicine, University of Science and Technology of China (USTC), Hefei, Anhui, 230026, China.

[2*]Center for Medical Imaging, Robotics, Analytic Computing & Learning (MIRACLE), Suzhou Institute for Advance Research, USTC, Suzhou, Jiangsu, 215123, China.

[3]College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang, 310027, China.

[4]Jiangsu Provincial Key Laboratory of Multimodal Digital Twin Technology, Suzhou, Jiangsu, 215123, China.

[5]State Key Laboratory of Precision and Intelligent Chemistry, USTC, Hefei, Anhui, 230026, China.

*Corresponding author(s). E-mail(s): xiongxp@ustc.edu.cn; skevinzhou@ustc.edu.cn;

Abstract

RNA language models have achieved strong performances across diverse downstream tasks by leveraging large-scale sequence data. However, RNA function is fundamentally shaped by its hierarchical structure, making the integration of structural information into pre-training essential. Existing methods often depend on noisy structural annotations or introduce task-specific biases, limiting model generalizability. Here, we introduce structRFM, a structure-guided RNA foundation model that is pre-trained on millions of RNA sequences and secondary structures data by integrating base pairing interactions into masked language modeling through a novel pair matching operation. The structure-guided mask and nucleotide-level mask are further balanced by a dynamic masking ratio. structRFM learns joint knowledge of sequential and structural data, producing

versatile representations, including classification-level, sequence-level, and pairwise matrix features, that support a broad spectrum of downstream adaptations. structRFM ranks among the top models in zero-shot homology classification across fifteen biological language models, and sets new benchmarks for secondary structure prediction. structRFM further derives Zfold, which enables robust and reliable tertiary structure prediction, with consistent improvements in estimating 3D structures and their accordingly extracted 2D structures, achieving a pronounced 19% performance gain compared with AlphaFold3 on RNA Puzzles dataset. In functional tasks such as internal ribosome entry site identification, structRFM achieves a whopping 49% performance gain in F1 score. These results demonstrate the effectiveness of structure-guided pre-training and highlight a promising direction for developing multi-modal RNA language models in computational biology. To support the broader scientific community, we have made the 21-million sequence-structure dataset and the pre-trained structRFM model fully open-source, facilitating the development of multimodal foundation models in biology.

1 Introduction

Ribonucleic acid (RNA), which plays an essential role in molecular biology, not only serves as a carrier of genetic information to bridge the central dogma of deoxyribonucleic acid (DNA) transcription and protein translation, but also folds into various structures and functions as a versatile architect of cellular processes [1]. Beyond its classical roles such as ribosomal RNA (rRNA) and transfer RNA (tRNA) in gene expression, numerous non-coding RNAs (ncRNAs) have emerged as essential regulators across cellular contexts, such as the internal ribosome entry site (IRES) for 5' cap-independent translation in circular RNAs (circRNA) [2], small nuclear RNA (snRNA) for splicing pre-mRNA to produce mature mRNA [3]. These profound functions of ncRNAs are encoded within their distinguished sequence patterns and structures, which demonstrate hierarchical relationships: primary sequence motifs determine base pair patterns and folding tertiary structures, which further dictate specific biochemical activities and interactions. Unraveling these complex relationships, linking sequence to structure and ultimately to biological function, is crucial for understanding fundamental RNA biology and advancing the development of next-generation RNA-centric therapeutics.

Decades of innovation in nucleic acid sequencing have produced an unprecedented volume of RNA sequence data [4], revolutionizing the human ability to map the complexity of the transcriptome [5]. Inspired by advances in natural language processing (NLP), RNA language models, pre-trained in a task-agnostic, self-supervised manner, have demonstrated great potential in extracting representative patterns and capturing contextual relationships from large-scale RNA sequential data, offering a considerable promise for systematically decoding RNA's molecular language, predicting tertiary structures [6, 7], specializing in functional tasks [8, 9], and advancing both fundamental research and translational applications.

Most existing RNA language models are primarily pre-trained on non-coding RNA sequences for general-purpose applications [10]. Examples include RNA-FM [11] and RiNALMo [12], which employ standard BERT architectures; RNAErnie [13], which incorporates motif patterns during pre-training; and AIDO.RNA [14], which scales the model to 1.6B parameters. These models exhibit strong generalizability across diverse downstream tasks. In contrast, other models focus on specialized RNA sequences for specific objectives, such as PlantRNA-FM [15] to identify functional RNA motifs in plants, UTR-LM [8] to analyze untranslated regions and related functions of mRNA, and SpliceBERT [9] to improve prediction of RNA splicing sites.

As RNA structure plays a central role in defining molecular properties, mediating interactions with other biomolecules, and ultimately determining biological function [16, 17], incorporating structural information into the pre-training of language models holds substantial potential beyond sequence alone. Several recent RNA language models integrate encoded structural data. For instance, RNABERT [18] incorporates aligned seed structures from the Rfam database into base embeddings to capture structural differences across RNA families. ERNIE-RNA [19] introduces structural features via a pairwise position matrix embedded within the attention mechanism. MP-RNA [20] and UTR-LM [8] adopt an auxiliary loss during pre-training to reconstruct masked secondary structures. While these approaches mark significant progress, they also present key limitations. RNABERT depends on potentially inaccurate seed alignments, which may hinder RNA family discrimination. ERNIE-RNA's pairwise matrix, based on simplified interaction scores (e.g., 3 for GC, 2 for AU, 0.8 for GU pairs), may inadequately capture the complexity of RNA structural features. MP-RNA and UTR-LM rely on ground-truth structures from ViennaRNA [21] and incorporate task-specific supervised losses, diverging from task-agnostic self-supervised learning and potentially biasing the model toward secondary structure prediction. To overcome these limitations and support robust and generalizable downstream adaptation, it is essential to develop a new pre-training strategy that incorporates accurate structural information while maintaining task agnosticism.

Motivated by this, we introduce **structRFM**, a fully open structure-guided RNA foundation model, pre-trained using an elaborately designed structure-guided masked language modeling (SgMLM) strategy. This approach implicitly encodes base-pairing interactions through informed masking patterns and jointly models structural and sequential dependencies. structRFM effectively harnesses structural priors while maintaining task-agnostic pre-training, enabling robust generalization across a broad spectrum of downstream tasks, including zero-shot prediction, secondary and tertiary structure prediction, and multiple function inferences (Fig. 1a). Our work makes three key contributions toward integrating structural knowledge into RNA foundation modeling.

First, we design SgMLM, a structure-guided pre-training strategy (Fig. 1c), featuring two core components: structure-guided masking and dynamic masking balance. The first component selectively masks input tokens corresponding to canonical base pairs within local structural contexts, encouraging the model to recover base-pair interactions based on neighboring loop regions. This implicitly guides structRFM to capture both RNA sequential patterns and hierarchical structural

regularities without relying on task-specific objectives, uncovering diverse nucleotide-to-nucleotide relationships, facilitating the transfer of sequential and structural knowledge to downstream structure and function prediction tasks. The secondary component dynamically increases the proportion of structure-guided masking within the overall mask during pre-training, gradually shifting the model's focus from sequence-level to structure-informed representations, and finally balances nucleotide-wise and structure-wise masking.

Second, we comprehensively evaluate structRFM across zero-shot assessments, downstream-adapted structure and function inference tasks. structRFM exhibits strong generalization ability, ranking among the top-performing models in zero-shot homology classification and zero-shot secondary structure prediction (Fig. 2c) within fifteen biological language models. For secondary structure prediction, structRFM attains the best F1 score of 0.873 and 0.641 on ArchiveII600 [22] and bpRNA-TS0 [23] datasets (Fig. 3c and d), respectively. In tertiary structure prediction, structRFM (derived method termed "Zfold") outperforms trRosettaRNA [24] with RMSD improvements of 24.6% and 6.80% on the CASP15 [25]-natural and RNA Puzzles datasets (Fig. 3h), respectively. These gains are corroborated by consistent 15.8% and 2.9% increases in the F1 score of secondary structures extracted from the predicted 3D structures (Fig. 3i), affirming the effectiveness and robustness of SgMLM strategy for structural adaptation. Notably, structRFM also surpasses AlphaFold3 [26], yielding RMSD improvements of 6.08% on CASP15-natural and 19.0% on RNA Puzzles (Extended Data Fig. 2a and b). Furthermore, structRFM shows remarkable ability for function prediction (Fig. 5), achieving state-of-the-art performances on splice site prediction, IRES identification, and ncRNA classification, benchmarking a top-k accuracy of 0.420 on SpliceAI dataset [3], an F1 score of 0.708 on IRES dataset [27, 28], and an F1 score of 0.973 on nRC dataset [29], respectively.

Third, we release the entire suite of resources to the community, including a 21-million sequence-structure dataset with according base-pairing interactions, the pre-trained structRFM model, and the fine-tuned state-of-the-art models for structural and functional tasks. These open resources are intended to democratize the development of RNA foundation models and support future work in multi-modal biological modeling.

Conclusively, structRFM demonstrates its effectiveness and versatility as a fully open RNA foundation model for robust downstream adaptation. Moreover, the SgMLM pre-training strategy and the collected sequence-structure dataset introduced here are readily extensible and can be seamlessly integrated into other pre-training frameworks, marking a substantial step toward empowering RNA foundation models with multi-modal structural knowledge and advancing broader biological research.

2 Results

In this section, we conduct a series of evaluations to demonstrate the versatility and effectiveness of the proposed structRFM foundation model that is pre-trained on 21 millions RNA sequences and secondary structures. We begin with feature visualization and zero-shot tasks, where the model is tested without labeled training data.

structRFM produces RNA sequence feature embeddings that encode family-level and base-pair interaction information, which are validated on zero-shot RNA homology classification and zero-shot secondary structure prediction. We further explore the model's capability in structure prediction, showing that structRFM consistently and reliably improves the prediction accuracy of RNA secondary and tertiary structures. Finally, we fine-tune structRFM on several supervised function prediction tasks, including splice site prediction, IRES identification, and ncRNA sequence classification, demonstrating its strong transferability across nucleotide-level and sequence-level applications. We denote structRFM without SgMLM and structRFM with freezing encoder as structRFM^{-} and structRFM$_F$, respectively. The overall results on these tasks are displayed in Extended Data Table 1.

2.1 structRFM extracts category-discriminative and structure-encoded features

The task-agnostic self-supervised pre-training enables structRFM to capture the intrinsic characteristic of any RNA sequence, while the SgMLM strategy allows structRFM to extract structure-encoded feature embeddings. For an input RNA sequence of length L, structRFM extracts classification feature $F_{\text{cls}} \in \mathbb{R}^{1\times D}$ and sequence feature $F_{\text{seq}} \in \mathbb{R}^{L\times D}$ that represent the RNA category and nucleotide-wise sequential embeddings, respectively. Based on this, we can further obtain the matrix feature $F_{\text{mat}} \in \mathbb{R}^{L\times L}$ by applying outer product to F_{seq}, which encodes the nucleotide-to-nucleotide relationships.

To qualitatively assess the learned sequential and structural knowledge, we visualize the classification feature and matrix feature using data from the pre-training dataset. In detail, we randomly sample 23,994 RNA sequences across 19 RNA families and feed them into structRFM to extract the sequence features and matrix features. Using the t-SNE [30] algorithm, we project the classification feature F_{cls} of structRFM into two dimensions, contrasting it with a randomly initialized model without SgMLM pre-training. As illustrated in Fig. 2a and Extended Data Table 2, structRFM exhibits strong category-discriminative capability, effectively clustering RNA families, projecting different categories of RNA features into distinct clusters, whereas the random model fails to do so, projecting the RNA sequences into mixed vector representations of latent space.

Furthermore, we visualize the matrix feature as a heatmap, comparing it to the native contact map. The high-response regions (in blue) highlight structRFM's ability to capture structural information, particularly base-pairing interactions. Collectively, these results demonstrate that structRFM successfully learns RNA representations that encode both sequential and structural properties, enhancing its adaptability for structure- and function-related tasks.

2.2 structRFM enables effective zero-shot prediction

Following previous work [10], we perform zero-shot assessments, *i.e.*, zero-shot homology classification and zero-shot secondary structure prediction to evaluate whether the pre-trained model acquires generalizable structural and functional knowledge

from unlabeled data, such as structural regularities, functional motif information, and base-pairing formation rules.

To evaluate whether the learned sequence embeddings encode homology-level information, we conduct zero-shot RNA homology classification using the Rfam and ArchiveII datasets. Specifically, we randomly sample 100,000 homologous and 100,000 non-homologous sequence pairs from each dataset. In Rfam, homologous pairs come from the same RNA family, where non-homologous pairs are from different families. For ArchiveII, homologous sequences share the same RNA type. We extract the sequence embeddings and compute the cosine similarity between each pair of sequences. The cosine similarity distributions are visualized, and the overlap ratio (OR) of these two distributions is computed for comparison, where a lower OR indicates less overlap between distributions and thus better discriminative ability. As shown in Figure 2b and c, structRFM achieves state-of-the-art results, ranking top 4 among 15 DNA/RNA/protein language models on the Rfam dataset (OR=0.10) and ArchiveII dataset (OR=0.48), showing a strong separation between homologous and non-homologous embeddings, despite AIDO.RNA costs 10 times more of the parameters. More distributions of other models are visualized in Supplementary Figs. 1 and 2. In contrast, DNA language models (e.g., DNABERT and NT) and specialized RNA language models (e.g., SpliceBERT and UTR-LM) perform poorly, particularly on ArchiveII (OR $>$ 0.7), suggesting that general task-agnostic pre-training better captures of family-level semantics.

Beyond classification, we further assess whether the model captures RNA secondary structures. Inspired by previous findings in RNA-MSM [31], where 2D attention maps encode base-pairing patterns, we construct a test set of 116 RNA sequences from the PDB database and extract corresponding features using structRFM for zero-shot evaluation. As shown in Figure 2d, we calculate the F1 score over a range of decision thresholds, from 0 to 1 with an interval of 0.001, to assess pairing prediction. The model achieves the highest F1 score of 0.44 at a threshold of 0.545, indicating a moderate ability to capture secondary structure features without any task-specific supervision or labeled data. The sharp peak around the optimal threshold highlights the sensitivity of prediction to threshold choice, especially in zero-shot settings, where no calibration is available. We also perform a grid search from each layer and each head to explore whether the attention map captures structural interactions that corresponds to the contact map of secondary structure, which is evaluate under F1 score. As Extended Data Fig. 1 shows, only a limited number of layers and heads are effective for zero-shot secondary structure prediction, while other layers and heads may focus on the sequential data and joint modeling.

2.3 structRFM achieves new benchmarks for secondary structure prediction

As shown in Fig. 3a, for secondary structure prediction, we utilize structRFM as the backbone to extract the sequence feature $F_{\mathbf{seq}}$. We concatenate extracted feature with the one-hot embedding of the input RNA sequence, and further adopt a structure module - MXfold2 [32] to predict four predefined stem-loop structural patterns. The

final predicted secondary structure is obtained by processing the predicted patterns using Zuker-style dynamic programming approach.

Following RNAErnie [13], we fine-tune structRFM and baseline methods on RNAStrAlign [33] (19,313 RNAs) and bpRNA-1m [23] (12,114 RNAs), and evaluate on ArchiveII600 [22] (3,611 RNAs) and bpRNA-TS0 [23] (1,305 RNAs), respectively. All models are evaluated under the same configuration as RNAErnie for a fair comparison. As Figs. 3c, 3d and Supplementary Table 2 demonstrate, structRFM achieves the best performance across all methods, including fine-tuned RNA language models, specialized learning-based methods, and traditional non-learning methods. Specifically, structRFM reaches an F1 score of 0.873 on ArchiveII600 and an F1 score of 0.641 on bpRNA-TS0, much better than RNA-FM (0.847 on ArchiveII600, $p = 2.59 \times 10^{-8}$, two-sided t-test; 0.639 on bpRNA-TS0), also outperforming other non-LM deep learning methods that are tailored for secondary structure prediction, such as MXfold2 [32] (0.768 on ArchiveII600, 0.558 on bpRNA-TS0), Contextfold [34] (0.842 on ArchiveII600, 0.546 on bpRNA-TS0).

We further compare structRFM with its ablated variant, structRFM^{-}, which lacks the structure-guided pre-training component. structRFM consistently outperforms structRFM^{-} on ArchiveII600 by 2.7% and bpRNA-TS0 by 3.2% with statistical significance ($p = 2.19 \times 10^{-7}$ for ArchiveII600, two-sided t-test; $p = 0.05$ for bpRNA-TS0, one-sided t-test). These results indicate that structure-guided pre-training effectively imparts prior knowledge that facilitates downstream secondary structure prediction. When comparing structRFM with structRFM$_F$ that freezes the encoder during fine-tuning, we observe that structRFM$_F$ achieves comparable performances (0.861 on ArchiveII600 and 0.641 on bpRNA-TS0), suggesting that the structure module and extended fine-tuning compensate for the fixed encoder, and that the learned representations are robust.

In addition, we evaluate family-level performance on the nine detailed families within the ArchiveII600 dataset (Fig. 3e, Supplementary Table 3). structRFM achieves superior results across eight families with the largest number of samples. Interestingly, structRFM^{-} slightly outperforms structRFM on the 23S rRNA family, which may be attributed to the small number of test sequences (only 21) and their long lengths, extending beyond the structural interaction range (512 nt) considered during SgMLM pre-training.

2.4 structRFM enables robust and reliable tertiary structure prediction

RNA structure folds in a hierarchical manner, where tertiary structure is largely determined by its secondary structure. We hypothesize that accurate base pairing interactions can boost tertiary structure prediction, and the secondary structure extracted from predicted tertiary structures provides a complementary assessment of their reliability. Moreover, unlike sparse representations of secondary structure (e.g., dot bracket notation or binary contact maps), feature embeddings extracted from RNA language models encode rich sequential and structural information that may better support tertiary structure modeling.

Despite this potential, existing RNA language models have not validated the utility of their learned features in tertiary structure prediction. To address this gap, we evaluate whether structRFM pre-trained with base-pairing interactions can consistently improve the accuracy and reliability of both tertiary structures and secondary structures derived from them, showing remarkable robustness in structural adaptation. As illustrated in Fig. 3b, we adopt trRosettaRNA [24] as the structure prediction framework and use structRFM as the backbone for extracting a matrix feature representing base-to-base relationships. The original trRosettaRNA pipeline takes multiple sequence alignment (MSA) generated by rMSA [35] of input RNA sequence and the contact map of secondary structure predicted by SPOT-RNA [36, 37] as input to predict 1D and 2D geometries (e.g., 1D orientations, 2D contacts, distances, and orientations). These are then converted into structural restraints for 3D folding via energy minimization. In our variant, we replace the input contact maps with matrix features from structRFM to assess the impact of structural embeddings on tertiary structure prediction.

We follow the original trRosettaRNA setup by using a training set comprising 3,633 RNA chains from the PDB [38, 39], filtering with a cutoff date of 2022-01 and excluding self-distillation data. We adopt trRosettaRNA framework and structRFM backbone for fine-tuning, where the resulting method is termed "Zfold" (we continue to use "structRFM" in this paper for clarity and comparison). Evaluation is performed on two benchmark datasets, namely CASP15 [25] (12 RNAs, 8 natural and 4 synthetic) and RNA Puzzles [40] (20 RNAs). Secondary structures are extracted from predicted tertiary structures using RNAVIEW [41], and visualized using tools R2DT [42] and Forna [43]. For evaluation metrics, we utilize aligned all-atom root mean square deviation (RMSD, Å, lower is better) calculated by TMscore tool [44] to evaluate the predicted tertiary structure, while the extracted secondary structures are evaluated by F1 score (higher is better).

As shown in Figs. 3h, 3i, and Supplementary Table 4, structRFM achieves RMSDs of 24.1Å, 14.8Å, 6.82Å, and F1 scores of 0.74, 0.76, 0.66 on CASP15, CASP15-natural, RNA Puzzles datasets, respectively, outperforming both trRosettaRNA and ablated variants of our model. Notably, structRFM$_F$ obtains comparable performance with an RMSD of 23.6Åand an F1 score of 0.74 on CASP15 dataset, suggesting that the structure module contributes significantly to prediction accuracy even though the freezing encoder not fine-tuned on specific task. In contrast, structRFM$^-$, pre-trained without the SgMLM strategy and lacks structural awareness, falling to supply knowledge of base pairing interactions for tertiary structure prediction, performs worse with RMSDs of 26.0Å, 15.9Å, 8.83Å, and F1 scores of 0.67, 0.71, 0.66 on CASP15, CASP15-natural, RNA Puzzles datasets. Compared to trRosettaRNA, structRFM improves secondary structure F1 scores by 19.4% on CASP15 and 1.5% on RNA-Puzzles, particularly effective under no homology and difficult prediction scenarios.

To further validate consistent improvements in extracted secondary structures achieved by structure-guided pre-training, we conduct head-to-head comparison between structRFM and other methods, including trRosettaRNA and structRFM$^-$. As illustrated in Figs. 3f and 3g, structRFM consistently outperforms its ablated

variant and trRosettaRNA on most CASP15 and RNA Puzzles targets. Representative examples (R1107 from CASP15 and PZ10 from RNA Puzzles) are visualized in Figs. 3j and 3k, showing that structRFM achieves more accurate predictions, especially when homology or template is available, while other models tend to fail under such conditions.

To provide a comprehensive case-level analysis, we visualize all predicted 3D and extracted 2D structures from structRFM, together with native structures from CASP15 in Fig. 4. Notably, structRFM benchmarks an RMSD of 14.8Åon the 8 natural targets of CASP15, much better than that (42.9Å) of the 4 synthetic targets. This performance gap is likely due to synthetic RNAs having longer sequences and denser long-range interactions, which are underrepresented in the training distribution. While structRFM achieves strong overall performance, the extracted 2D structures still exhibit some deficiencies, particularly in folding short local stems (green), occasionally forming larger, incorrect multi-loops (red).

Finally, we compare structRFM with state-of-the-art methods, including trRosettaRNA and AlphaFold3 [26] under RMSD measurement. As shown in Extended Data Fig. 2a, b, structRFM achieves RMSDs of 14.8Åon CASP15-natural and 6.8Åon RNA Puzzles, better than AlphaFold3 (15.6Åon CASP15-natural and 8.1Åon RNA Puzzles, respectively), obtaining improvements of 5.4% and 19.1%, regardless of the poorer performance on synthetic RNA targets. Extended Data Fig. 2c further shows that structRFM is superior to both structRFM^{-} and trRosettaRNA on most targets with lower RMSD.

In summary, structRFM leverages structure-guided pre-training to enhance the robustness and reliability of RNA structural predictions, consistently improving the accuracy of both tertiary and secondary structure modeling. Challenges remain in modeling synthetic RNAs, long-range interactions, and local stem formation, presenting opportunities for future improvement.

2.5 structRFM transfers knowledge for splice site prediction

Splicing is a key step in processing pre-mRNAs into mature mRNAs and is essential for gene expression and protein translation. The splice site prediction task is to classify each nucleotide of a pre-mRNA into one of three categories: donor site, acceptor site, or non-site. As shown in Fig. 5a, we construct the neural network that uses structRFM as the backbone to extract sequence features, followed by a simple linear layer that maps the concatenated input features to nucleotide-level splice site scores, where each category is represented as 0 for non-site, 1 for donor, and 2 for acceptor.

We adopt SpliceAI dataset [3] to fine-tune and evaluate structRFM and structRFM^{-}, which consists of 162,706 RNA sequences for training and 16,505 RNA sequences for testing, with all sequences in a uniform length of 100. Since these sequences are pre-mRNAs, substantially different from ncRNA sequences for pre-training structRFM, we enable trainable encoder of structRFM to make effective fine-tuning and knowledge transfer for splice site prediction. We follow the same configurations as BEACON [45] and adopt the top-k accuracy as measurement. As Fig. 5c demonstrates and BEACON [45] benchmarks, structRFM achieves a top-k accuracy

of 0.42, surpassing most existing RNA language models, such as RNA-MSM, UTR-LM, RNA-FM, and falls a little behind the specialized SpliceBERT that is tailored for splice site prediction, thus validating the efficiency of structRFM. Moreover, structRFM outperforms structRFM^{-} with an improvement of 7.6% in accuracy, showing the robustness and versatility of structRFM and SgMLM strategy for transferring the learned sequential and structural knowledge from ncRNAs to mRNAs.

2.6 structRFM generalizes well for IRES identification

Internal ribosome entry site (IRES) is a special structure that mediates the 5' cap-independent translation in circular RNAs, which produces a stable and low-free-energy structure and leads to promising applications in RNA therapy. Therefore, accurate identification of IRES is pivotal in developing RNA-centric therapeutics and biological innovations. Here we adopt structRFM to extract the classification feature F_{cls} for IRES identification, illustrated in Fig. 5b. The classifier is a simple feed-forward network that takes the concatenated input features and outputs the distribution probability of each class. To deal with the imbalanced data, we further introduce oversampling and undersampling strategies to sample different training datasets and fine-tune models on these datasets. The final model is an ensemble of these high-precision and high-recall models.

We evaluate structRFM, its variants and two state-of-the-art methods, *i.e.* IRESfinder [46], DeepIRES [47], on an IRES dataset [27, 28] that contains 12,998 RNA sequences for training (3,949 positive and 9,034 negative samples) and 1,164 RNA sequences for testing (582 positive and 582 negative samples), with a uniform length of 174. As shown in Fig. 5d, Fig. 5e, and Supplementary Table 5, the precision-recall curve (PRC) and receiver operating characteristic curve (ROC) indicate that structRFM achieves the best performances, with an area under PRC (AUPRC) of 0.6720 and an area under ROC (AUROC) of 0.6972, surpassing structRFM^{-} which obtains 0.6417 in AUPRC and 0.6840 in AUROC. This highlights the effectiveness of the SgMLM pre-training in capturing structural patterns relevant to IRES functionality. Moreover, structRFM performs much better than structRFM$_F$ with an improvement of 9.34% in AUPRC, which is different from the case for structure prediction - the trainable parameters in a linear layer are fewer than the structure module's, thus leading to poorer performances after freezing the backbone encoder of structRFM. Furthermore, we visualize seven classification metrics, *i.e.*, F1 score, recall, precision, accuracy, AUPRC, AUROC, Matthews correlation coefficient (MCC), in Fig. 5f and Extended Data Fig. 3a, for an overall comparison against other methods. As visualized in the radar plot, structRFM consistently outperforms all baselines methods across most metrics, particularly excelling against non-language model methods - IRESfinder and DeepIRES. For instance, structRFM benchmarks an F1 score of 0.708 with an improvement of 49% than IRESfinder (F1 = 0.475) and 149% over DeepIRES (F1 = 0.284).

2.7 structRFM achieves highly accurate ncRNA classification

As demonstrated in the previous subsection, structRFM has shown a remarkable ability for zero-shot homology classification without relying on labeled training data. To further assess its downstream adaptability, we evaluate structRFM on a supervised non-coding ncRNA classification task using the nRC dataset [29], which comprises 13 primary ncRNA classes. The model architecture also consists of two components: the structRFM backbone for extracting classification features, and a linear classifier for outputting the class probability distribution, where the predicted category is selected based on the maximum probability. All models are fine-tuned under the same configurations as RNAErnie for fair comparison.

As demonstrated in Fig. 5g and Extended Data Fig. 3b, structRFM reaches the best performance and highest accuracy for ncRNA classification, achieving new benchmarks of F1 score=0.973, MCC=0.970, AUROC=0.999, and AUPRC=0.993, outperforming structRFM$_F$, structRFM$^-$, and structRFM$_F^-$ by a huge gap, suggesting that SgMLM strategy in the pre-training and trainable encoder in the fine-tuning largely enhance the performance of structRFM by encoding and transferring structural information, respectively. RNABERT and RNA-MSM extract less representative features for classifying different categories of ncRNAs, obtaining an F1 score of 0.727 and 0.934, respectively. Notably, RNA-FM attains comparable performance with structRFM, regardless of its weaker performances for secondary structure prediction and splice site prediction than structRFM, which indicates it may capture classification-relevant features more effectively than sequential or structural features. More benchmark results of other methods are displayed in Supplementary Table 6, where RNAErnie [13] and tailored non-language model method ncRDense [48] are ahead of other methods, reaching an F1 score of 0.964 and 0.951, respectively.

Furthermore, we visualize the F1 scores of the seven models across the thirteen detailed RNA categories. As Fig. 5h demonstrates, structRFM performs consistently accurate classifications on all RNA types, while structRFM$^-$ obtains lower F1 scores on rare categories (in contrast with rRNA that weights 66.3% of the pre-training dataset), such as miRNA, HACA-box, and scaRNA, showing the generalizability and transferability of the sequential and structural knowledge learned by structRFM through the structure-guided pre-training strategy. Extended Data Fig. 3c further displays the PR curves and ROC curves, confirming robust classification across all RNA families.

3 Discussion

In this paper, we introduce structRFM, a model designed to integrate structural information into RNA language modeling, enabling the joint pre-training of sequential and structural knowledge. This integration produces discriminative features that incorporate structural insights, enhancing both structural and functional predictions. structRFM takes an important step towards explicitly exploring multimodal structural data without introducing task-specific supervised objectives that may bias the model. Our experimental results demonstrate the model's effectiveness, robustness, and transferability across diverse downstream tasks, achieving top-tier performance

in zero-shot prediction, tertiary structure prediction, and splice site prediction. Additionally, structRFM sets new benchmarks in secondary structure prediction, internal ribosome entry site (IRES) identification, and non-coding RNA (ncRNA) classification. Notably, the multimodal 21-million-sequence-structure dataset we constructed, along with our structure-guided pre-training strategy, can be adapted to other biological language models, improving the understanding of hierarchical structures and the relationship between sequence and structure. The constructed large-scale multimodal dataset, code, and pre-trained structRFM are publicly available and fully open-source.

Despite these achievements, there are several avenues for improvement. First, structRFM currently supports RNA sequences up to a maximum length of 512, which covers most RNA sequences but entails a trade-off in computational resources. To extend structRFM to accommodate longer sequences, such as long non-coding RNAs, one approach would be to fine-tune a new positional embedding for these longer sequences while preserving the original structRFM encoder. Alternatively, we could pre-train structRFM again with longer target RNA sequences. For sequence-wise classification tasks, a simple and effective strategy is to split the long sequence into shorter sequences and make aggregated predictions. Second, while we use BPfold [49] to extract base-pairing information and construct the sequence-structure pre-training dataset, we recognize that incorporating additional thermodynamic methods could enhance the accuracy of structural data. An ensemble approach that captures the dynamic nature of RNA structures and includes non-canonical base pairs would further improve base-pairing interactions and tertiary structure predictions. Third, our structure-guided pre-training strategy, although effective, could be enhanced by incorporating different weights between structure-guided masks and random masks when optimizing task-agnostic objectives.

Looking ahead, an exciting direction is the development of multimodal language models for biological research. structRFM employs structure-guided pre-training by implicitly integrating base-pairing interactions within a single backbone encoder, jointly extracting sequential and structural features. A promising extension would be to explicitly integrate multimodal structural data by using separate encoders for each modality, enabling contrastive sequence-structure pre-training. Such an approach would align RNA sequence and structural data through large-scale contrastive learning, enhancing generalization across downstream tasks. Contrastive learning in this context could facilitate robust predictions of RNA structure, function, and interactions, even in data-scarce conditions.

Furthermore, another emerging direction is to build unified frameworks that integrate diverse biomolecular modalities. For instance, AlphaFold3 [26] expands structural prediction to proteins, nucleic acids, and small molecules within a single architecture, while models such as evo2 [50] and LucaOne [5] aim to unify molecular language modeling across biomolecular types. These efforts reflect a growing trend toward general-purpose, multimodal foundation models capable of capturing the complexity of biological systems. We anticipate that the convergence of contrastive learning, structure-guided pre-training, and unified multi-modal modeling will catalyze the next wave of advances in computational biology, offering deeper interpretability and broader applications.

4 Methods

4.1 Overview of structRFM

In this work, we present structRFM (Fig. 1a), an RNA foundation model that is pre-trained on multi-modal non-coding RNA sequences together with corresponding secondary structures for general-purpose downstream adaptation. We first construct the pre-training sequence-structure dataset by filtering RNA sequences from the RNAcentral dataset [51] and collecting secondary structures with confidence index generated by BPfold [49]. Next, as illustrated in Fig. 1b, we design a structure-guided masked language modeling (SgMLM) strategy to integrate RNA secondary structures into RNA sequence modeling when pre-training task-agnostically. After structRFM is well pre-trained, it is capable of extracting representative structural feature of an input RNA sequence to further play a vital role in various downstream tasks, including structure and function tasks, such as secondary structure prediction, tertiary structure prediction, non-coding RNA classification, IRES identification in circular RNAs, and splice site prediction.

4.2 Pre-training sequence-structure dataset construction

Starting with the largest non-coding RNA dataset RNAcentral version 24.0 [51], similar to existing RNA language models [11, 13], we filter 21,477,078 RNA sequences from the active part of the RNAcentral dataset with a maximum sequence length of 512, which is different from some methods [14, 52] that collect RNA sequences from both active part and inactive part that are no longer maintained. After collecting RNA sequences, we apply the RNA secondary structure prediction method BPfold [49], a base pair motif energy-integrated [53] deep learning method, which is robust and accurate by means of high-quality and full-coverage canonical base pair motif library, for forecasting secondary structure with confidence index to reveal the potential interactions between canonical base pairs, including G-C, C-G, A-U, U-A, G-U, and U-G. Based on these vast sequence-structure pairs, we randomly split them into training and test datasets, with the number of samples being 20,403,225 and 1,073,855, respectively. This multi-modal dataset consists of a large number of RNA sequences and RNA secondary structures, relieving the shortage of structural data in a computational way at the secondary structure level and facilitating the development of multi-modal structure-enhanced RNA foundation models.

4.3 Structure-guided masked language modeling strategy

To improve the ability of learning base pairing interactions and capturing structural patterns for RNA language model, we elaborately design a structure-guided masked language modeling strategy (SgMLM) to replace the standard masked language modeling for pre-training structRFM with inspirations from base pair interactions. By doing so, we can leverage the structure-level knowledge and extract discriminative representations of RNA sequences to serve the downstream tasks.

Specifically, given an RNA sequence $\mathbf{X} = \{x_1, x_2, \ldots, x_L\} \in \{\text{A, U, G, C, N}\}^L$, the standard masked language modeling strategy generates a random mask $\mathbf{M_r} \in$

$\{0,1\}^L$ on this RNA sequence in a Bernoulli distribution with a probability of 15%, where $\mathbf{M_r}(i) = 1$ indicates the token of base i is masked as "[MASK]" and $\mathbf{M_r}(i) = 0$ indicates the token of base i is kept the same as original.

Based on this, we introduce the canonical base pairs of the RNA secondary structure into this procedure. The corresponding secondary structure of RNA sequence with a length of L can be represented in format of a binary adjacent matrix $\mathbf{C} \in \{0,1\}^{L\times L}$, where $\mathbf{C}(i,j) = 1$ indicates that bases i and j form a base pair and $\mathbf{C}(i,j) = 0$ indicates that bases i and j do not form a base pair. As Fig. 1b shows, we can obtain a secondary structure mask $\mathbf{M_s} \in \{0,1\}^L$ from the secondary structure where $\mathbf{M_s}(i) = 1$ indicates that base i is a part of a base pair and vice versa. The structure-guided masking strategy is to mask RNA bases with more paired interactions instead of isolated bases. Therefore, we apply an operation of matching canonical base pairs, denoted as "PairMatch", to make up for these isolated bases that have formed base pairing interactions in the random mask $\mathbf{M_r}$. The final structured-guided mask $\mathbf{M_f}$ comprises two parts: 1. "PairMatch" of the "logical and" of random mask $\mathbf{M_r}$ and secondary structure mask $\mathbf{M_s}$, 2. the "logical and" of random mask $\mathbf{M_r}$ and the "logical not" of secondary structure mask $\mathbf{M_s}$, which can be formulated as follows:

$$\mathbf{M_f} = (\text{PairMatch}(\mathbf{M_r} \wedge \mathbf{M_s}, \mathbf{C})) \vee (\mathbf{M_r} \wedge \neg \mathbf{M_s}). \tag{1}$$

$$\text{PairMatch}(\mathbf{M}, \mathbf{C})(i) = \begin{cases} 1, & \mathbf{M(i)} = 1, \\ 1, & \mathbf{M(i)} = 0 \ \& \ \exists j(\mathbf{C}(i,j) = 1 \| \mathbf{C}(j,i) = 1), \\ 0, & \mathbf{M(i)} = 0 \ \& \ \forall j(\mathbf{C}(i,j) = 0 \& \mathbf{C}(j,i) = 0). \end{cases} \tag{2}$$

In the whole training process, to conduct data augmentation and enhance the diversity of training data, among the structure-guided masked tokens, we replace the "[mask]" with random tokens at 10% of the time and keep the masked input tokens unchanged at another 10% of the time. Furthermore, to enhance the pre-training stability and enable structRFM to learn both token-wise and pair-wise knowledge, we introduce a warm-up mechanism by gradually increasing the ratio r of structure-guided masking against standard masking from 0 to 0.5, which can be formulated as follows:

$$r(t) = \begin{cases} 0, & t < 0.15T, \\ 0.5 * (t - 0.15T)/(0.7T), & 0.15T < t < 0.85T, \\ 0.5, & t > 0.85T. \end{cases} \tag{3}$$

where T is the total training steps where t is current training steps. Therefore, the language model is finally pre-trained with sequence-level masking and structure-level masking, jointly modeling the relationship of RNA sequences with base pairing interactions, resulting in enhanced capability of understanding RNA structural patterns. During masked language modeling, the proxy task is to recover the masked tokens as the original sequence base tokens with a cross-entropy loss. By applying structure-guided masking, structRFM can infer the masked matched base pairs in the unmasked unpaired surrounding base loops, intrinsically capturing the base pair interactions according to input RNA sequences, which is beneficial for structural prediction.

4.4 Pre-training structRFM with secondary structures

In the pre-training process, structRFM makes use of the constructed sequence-structure dataset. For an input RNA sequence, structRFM generates a structured-guided mask for it according to its secondary structure, tokenizes the input sequence, and embeds it as feature vectors. Consequently, structRFM utilizes transformer [54] as its backbone, effectively models the long-range relationship of RNA sequences, and aims at learning the sequential and structural representations by the proxy task of recovering the masked tokens.

Specifically, given an RNA sequence $\mathbf{X} = \{x_1, x_2, \ldots, x_L\} \in \{\text{A,U,G,C,N}\}^L$ that is composed of "A", "U", "G", and "C" base nucleotides (other unknown base is denoted as "N"), after applying structure-guided masking, the masked tokens are replaced with "[mask]" token. Additionally, we add a "[CLS]" token and an "[SEP]" token at the beginning and end of the sequence, respectively, which is helpful for fine-tuning on classification tasks and separating multiple input RNA sequences in situations of dealing with RNA-RNA interaction and multiple chains. To keep the sequence length uniform in a batch, we pad it with "[PAD]" token at the end of the sequence. These tokens are further encoded as numbers and processed by an embedding layer to form the input sequence embedding $F_0 \in \mathbb{R}^{L \times D}$.

As demonstrated in Fig. 1a, structRFM employs N consecutive transformer neural network blocks as its backbone to process the input sequence embedding. The transformer block consists of a multi-head self attention module (MSA) and a feed-forward network (FFN), which is composed of linear layers and GELU activation. After the input feature is processed by the MSA and FFN module, respectively, it is further processed by a residual operation and a LayerNorm neural network layer. The whole process can be formulated as follows:

$$\begin{aligned} Q_i &= \mathrm{Q}_{\theta_i}(F_i), K_i = \mathrm{K}_{\theta_i}(F_i), V_i = \mathrm{V}_{\theta_i}(F_i), \\ A_i &= \mathrm{softmax}(\frac{Q_i K_i^T}{\sqrt{D_k}} V_i), \\ \tilde{A}_i &= \mathrm{LN}_{\theta_i}(A_i + F_i), \\ F_{i+1} &= \mathrm{LN}_{\theta_i}(\tilde{A}_i + \mathrm{FFN}_{\theta_i}(\tilde{A}_i)). \end{aligned} \tag{4}$$

where $i = 0, 1, 2, \ldots, N-1$ denotes the i-th transformer block with learnable parameter θ_i.

4.5 structRFM feature embeddings for multi-task inferences

After pre-training, given an arbitrary RNA sequence of length L as input, structRFM is capable of outputting a representative embedding F in the shape of $\mathbb{R}^{(1+L)\times D}$ with structural information encoded, which consists of a classification feature and a sequence feature. Classification feature $F_{\text{cls}} \in \mathbb{R}^{1\times D}$ is derived from the "[CLS]" token placed in the first row of the output embedding F, helpful for sequence-level classification tasks. Sequence feature $F_{\text{seq}} \in \mathbb{R}^{L\times D}$ encodes each nucleotide of RNA sequence, powerful for nucleotide-level classification and regression tasks. Moreover,

we derive a matrix feature embedding $F_{\text{mat}} \in \mathbb{R}^{L \times L}$ by applying the outer product of the sequence feature F_{seq} and the transpose of it F_{seq}^T, which implies the nucleotide-to-nucleotide relationships into an adjacent matrix in a latent feature space and plays a vital role in modeling interactions between nucleotide pairs, especially for structure prediction tasks. Formally, F_{mat} can be constructed as follows:

$$F_{\text{mat}} = F_{\text{seq}} \times F_{\text{seq}}^T. \tag{5}$$

The above three kinds of feature embeddings, with discriminative structural information, are powerful representations as the transitional intermediates to be further processed by a simple multiple-layer perceptron or other downstream modules for various RNA downstream tasks, including structural and functional predictions with task-specific fine-tuning. For sequence-level classification tasks, such as non-coding RNA classification, the classification feature F_{cls} is used. For sequence-level regression tasks, the sequence feature F_{seq} is adopted by calculating its mean value in the sequence-wise dimension. Besides, the sequence feature F_{seq} is intrinsically suitable for nucleotide-level regression tasks, such as splice site prediction. For RNA secondary structure and tertiary structure prediction, regrading the difficulties of modeling interactions between base pairs, we employ the matrix feature F_{mat} for providing sufficient base-to-base prior knowledge.

4.6 Fine-tuning structRFM for downstream adaptation

After pre-training structRFM with RNA sequences and corresponding secondary structures in a task-agnostic self-supervised manner, we are able to adapt it for various downstream tasks by zero-shot prediction or fine-tuning it with labeled data in a task-specific supervised manner. Here, we demonstrate the supervised fine-tuning strategies in terms of fine-tuning different parts of task-specific neural networks and the ways to construct neural architectures for various structural and functional tasks.

Fine-tuning strategy. The learnable parameters of structRFM for downstream task adaptations consist of two parts: the pre-trained backbone (including sequence embedding layers and feature encoder) and task-specific heads or modules, which naturally derives two fine-tuning strategies for transferring the learned knowledge of structRFM. The first fine-tuning strategy is to freeze the pre-trained backbone and fine-tune the task-specific heads and modules, which costs few computational resources and training time. The other fine-tuning strategy is to fine-tune whole neural network including backbone and other parts, which is usually resource-consuming but more powerful. In our experiments, regrading the distribution shift of input RNA sequences, we only freeze the feature encoder and fine-tune both sequence embedding layers and task-specific layers as the first fine-tuning strategy, termed as "structRFM$_F$", while another fine-tuning strategy is marked as vanilla "structRFM" for complementary comparison. Besides, we pre-train structRFM without secondary structures and SgMLM pre-training strategy for ablation study, which is fine-tuned by the above two kinds of fine-tuning strategies and denoted as "structRFM$_F^-$" and "structRFM$^-$", respectively.

Neural architecture for downstream tasks. Given an input RNA sequence, structRFM, as a foundation model, processes and extracts sequence-related and encoded structural feature embeddings, which are further fed into the prediction head for customizing different output representations. Tasks such as RNA sequence classification only need a task-specific head - a fully connected layer while some tasks outputting complex dense predictions like secondary structure prediction and tertiary structure prediction need post-processing procedures and structural modules.

Secondary structure prediction. RNA secondary structure demonstrates base pair patterns and stem-loop regions, playing vital roles in modeling RNA tertiary structure and exploration of functional RNA motifs within non-coding genomic regions. Most existing methods predict the canonical base pairs (*i.e.*, A-U, G-C, G-U) of a given input RNA sequence of length L. The secondary structure can be represented as an adjacent matrix in the shape of $L \times L$, in which each element stores the pairing relationship of base i and base j, *i.e.*, 0 for unpaired and 1 for paired. For predicting such dense connections in a scheme of nucleotide-wise classification, complex post-process procedures such as symmetry operation, sharp loop, and overlap removal are applied [55, 56], and thermodynamic priors are adopted [32, 49] in previous methods. In our experiments, similar to RNAErnie [13], we adopt the light-weighted MXfold2 neural network [32] as the output structural module. Specifically, as demonstrated in Fig. 3a, structRFM extracts sequence feature from the input RNA sequence and then concatenates it with the one-hot encoding of the RNA sequence, forming the enhanced feature embeddings, which further utilizes the MXfold2 neural networks to process feature embeddings and computes four kinds of folding scores at base pair level: helix stacking, unpaired region, helix opening and helix closing. At last, a Zuker-style dynamic programming approach [57] is applied for postprocessing to predict the secondary structures.

Tertiary structure prediction. Regarding the widely-participant cellular activities such as protein translation and gene regulation, RNA tertiary structures are essential for understanding their functions in biological processes. Predicting RNA tertiary structure is still challenging due to its dynamic nature and lack of high-quality experimental data. More and more methods [6, 7] take use of RNA language model to generate representative sequential feature as the input formation of tertiary structure prediction methods. In our experiments, as illustrated in Fig. 3a, we employ trRosettaRNA [24] as the structure module to replace the direct secondary structure input with the matrix feature produced by structRFM, aiming to examine the effectiveness of the structural matrix feature generated by structRFM. The structural module takes multiple sequence alignments (MSA) and matrix feature as inputs and output 1D and 2D geometries, which is further refined by PyRosetta [58] to generate the final tertiary structure.

Splice site prediction. Pre-mRNA splicing is an essential step in eukaryotic gene expression. Introns (non-coding regions) must be accurately removed, and exons (coding regions) are mergeed together by the spliceosome complex to form mature mRNA, which is then translated into functional protein. Therefore, it is important to accurately detect the exon-intron boundaries - the splice sites. Splice site prediction is a nucleotide-wise classification task that aims to recognize whether each base is a donor

site, an acceptor site, or non-site. As Fig. 5a shows, we simply add a linear layer as the prediction head to the structRFM backbone to construct the neural network architecture for splice site prediction. To compare with RNA language models benchmarked in BEACON [45], we follow the same configuration by passing the sequence feature of the first layer of structRFM to a linear layer to make nucleotide-wise predictions.

IRES identification. The internal ribosome entry site (IRES) mediates the cap-independent translation process in circular RNAs, widely found in viral genomes and eukaryotic cells, exhibiting high stability and tissue specificity. Therefore, accurate identification of IRES aids in developing RNA therapeutics and targeted cancer interventions. As illustrated in Fig. 5a, the neural network architecture for IRES identification is composed of the structRFM backbone and a simple classifier consisting of a linear fully connected layer. Regarding the sequence-wise classification task, we adopt the classification feature extracted from the RNA sequence by structRFM, and concatenate it with the one-hot encoding of the input sequence as the input feature for the classifier, making it neat and effective. Furthermore, to deal with the imbalanced data, we fine-tune different models on oversampling (IRES) dataset and undersampling (Non-IRES) dataset, respectively, and ensemble these models to make soft voting predictions.

Sequence classification. Non-coding RNAs play various roles and demonstrate diverse functions in gene regulation and cellular processes. Clear classification of ncRNA is meaningful for advancing biological research and developing RNA-based therapy, and improving disease diagnosis and treatment. In this downstream task, we simply use the classification feature provided by structRFM and a light-weighted classifier head - linear fully connected layer as the neural network architecture to predict varieties of ncRNA categories, such as rRNA, tRNA, mRNA, and so on.

4.7 Experimental configurations

At the pre-training stage, we pre-train structRFM for 3,238,154 steps with a batch size of 64 on two NVIDIA Tesla A100 80GB graphics processing units for approximately 400 hours. structRFM adopts the vanilla BERT architecture [59] with hidden dimension being 768 and the number of transformer layers being 12, and is pre-trained on our constructed sequence-structure dataset with a maximum sequence length of 512, which is derived from RNAcentral [51] and RNA secondary structure prediction method - BPfold [49]. We utilize PyTorch [60] framework to implement structRFM, and the optimizer is an AdamW optimizer with a learning rate of 2×10^{-4}. We design the SgMLM strategy and utilize the cross-entropy loss to enable structRFM to capture sequential and structural patterns.

We evaluate structRFM on two zero-shot assessments and five downstream fine-tuning tasks, namely: zero-shot homology classification, zero-shot secondary structure prediction, secondary structure prediction, tertiary structure prediction, splice site prediction, IRES identification, and ncRNA classification. The overall performances are displayed in Extended Data Table 1. These experimental datasets include Rfam [61, 62], PDB [38], ArchiveII [22], bpRNA-1m [23], CASP15 [25], RNA Puzzles [40, 63], Splice [3], IRES [27, 28] and nRC [29], detailed in Supplementary Table 1.

Data availability

The pre-training sequence-structure dataset, pre-trained, and fine-tuned structRFM are publicly available at Zenodo [64]. The training and testing datasets are publicly available as follows: RNAcentral [51]:https://ftp.ebi.ac.uk/pub/databases/RNAcentral/releases/24.0/;Rfam [61, 62]:https://rfam.org/;ArchiveII [22]:https://rna.urmc.rochester.edu/publications.html;RNAStrAlign [33]:https://github.com/mxfold/mxfold2/releases/tag/v0.1.0;bpRNA [23]:https://bprna.cgrb.oregonstate.edu/download.php#bpRNA;PDB [38, 39]:https://rcsb.org/;CASP15 [25]:https://predictioncenter.org/download_area/CASP15;RNA Puzzles [40]:https://www.rnapuzzles.org/puzzles/list/;SpliceAI [3]:https://github.com/Illumina/SpliceAI; IRES [27, 28]:https://bitbucket.org/alexeyg-com/irespredictor/src/v2/data/;nRC [29]:http://tblab.pa.icar.cnr.it/public/nRC/paper_dataset/. Source data are provided with this paper.

Code availability

The source code of structRFM is publicly available at Zenodo [65] and GitHub (https://github.com/heqin-zhu/structRFM).

References

[1] Cooper, T. A., Wan, L. & Dreyfuss, G. Rna and disease. *Cell* **136**, 777–793 (2009).

[2] Fan, X., Yang, Y., Chen, C. & Wang, Z. Pervasive translation of circular rnas driven by short ires-like elements. *Nature Communications* **13**, 3751 (2022).

[3] Jaganathan, K. *et al.* Predicting splicing from primary sequence with deep learning. *Cell* **176**, 535–548 (2019).

[4] Ozsolak, F. & Milos, P. M. Rna sequencing: advances, challenges and opportunities. *Nature reviews genetics* **12**, 87–98 (2011).

[5] He, Y. *et al.* Generalized biological foundation model with unified nucleic acid and protein language. *Nature Machine Intelligence* **7**, 942–953 (2025). URL https://doi.org/10.1038/s42256-025-01044-4.

[6] Li, Y., Feng, C., Zhang, X. & Zhang, Y. Ab initio rna structure prediction with composite language model and denoised end-to-end learning. *bioRxiv* 2025–03 (2025).

[7] Shen, T. *et al.* Accurate rna 3d structure prediction using a language model-based deep learning approach. *Nature Methods* 1–12 (2024).

[8] Chu, Y. *et al.* A 5 utr language model for decoding untranslated regions of mrna and function predictions. *Nature Machine Intelligence* **6**, 449–460 (2024).

[9] Chen, K. *et al.* Self-supervised learning on millions of primary rna sequences from 72 vertebrates improves sequence-based rna splicing prediction. *Briefings in bioinformatics* **25**, bbae163 (2024).

[10] Wang, H., Zhang, Y., Chen, J., Zhan, J. & Zhou, Y. A comparative review of rna language models. *arXiv preprint arXiv:2505.09087* (2025).

[11] Chen, J. *et al.* Interpretable rna foundation model from unannotated data for highly accurate rna structure and function predictions. *arXiv preprint arXiv:2204.00300* (2022).

[12] Penić, R. J., Vlašić, T., Huber, R. G., Wan, Y. & Šikić, M. Rinalmo: General-purpose rna language models can generalize well on structure prediction tasks. *Nature Communications* **16**, 5671 (2025).

[13] Wang, N. *et al.* Multi-purpose rna language modelling with motif-aware pre-training and type-guided fine-tuning. *Nature Machine Intelligence* **6**, 548–557 (2024).

[14] Zou, S. *et al.* A large-scale foundation model for rna function and structure prediction. *bioRxiv* 2024–11 (2024).

[15] Yu, H. *et al.* An interpretable rna foundation model for exploring functional rna motifs in plants. *Nature Machine Intelligence* **6**, 1616–1625 (2024).

[16] Cao, X., Zhang, Y., Ding, Y. & Wan, Y. Identification of rna structures and their roles in rna functions. *Nature Reviews Molecular Cell Biology* **25**, 784–801 (2024).

[17] Mortimer, S. A., Kidwell, M. A. & Doudna, J. A. Insights into rna structure and function from genome-wide studies. *Nature Reviews Genetics* **15**, 469–479 (2014).

[18] Akiyama, M. & Sakakibara, Y. Informative rna base embedding for rna structural alignment and clustering by deep representation learning. *NAR genomics and bioinformatics* **4**, lqac012 (2022).

[19] Yin, W. *et al.* Ernie-rna: An rna language model with structure-enhanced representations. *bioRxiv* 2024–03 (2024).

[20] Yang, H. & Li, K. Al-Onaizan, Y., Bansal, M. & Chen, Y. (eds) *MP-RNA: unleashing multi-species RNA foundation model via calibrated secondary structure prediction.* (eds Al-Onaizan, Y., Bansal, M. & Chen, Y.) *Findings of the Association for Computational Linguistics: EMNLP 2024, Miami, Florida, USA, November 12-16, 2024*, 5278–5296 (Association for Computational Linguistics, 2024).

[21] Hofacker, I. L. *et al.* Fast folding and comparison of rna secondary structures. *Monatshefte fur chemie* **125**, 167–167 (1994).

[22] Sloma, M. F. & Mathews, D. H. Exact calculation of loop formation probability identifies folding motifs in rna secondary structures. *RNA* **22**, 1808–1818 (2016).

[23] Danaee, P. *et al.* bprna: large-scale automated annotation and analysis of rna secondary structure. *Nucleic acids research* **46**, 5381–5394 (2018).

[24] Wang, W. *et al.* trrosettarna: automated prediction of rna 3d structure with transformer network. *Nature communications* **14**, 7266 (2023).

[25] Das, R. *et al.* Assessment of three-dimensional rna structure prediction in casp15. *Proteins: Structure, Function, and Bioinformatics* **91**, 1747–1770 (2023).

[26] Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with alphafold 3. *Nature* **630**, 493–500 (2024).

[27] Chen, C.-K. *et al.* Structured elements drive extensive circular rna translation. *Molecular Cell* **81**, 4300–4318.e13 (2021).

[28] Weingarten-Gabbay, S. *et al.* Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* **351**, aad4939 (2016).

[29] Fiannaca, A., La Rosa, M., La Paglia, L., Rizzo, R. & Urso, A. nrc: non-coding rna classifier based on structural features. *BioData mining* **10**, 1–18 (2017).

[30] Van der Maaten, L. & Hinton, G. Visualizing data using t-sne. *Journal of machine learning research* **9**, 2579–2605 (2008).

[31] Zhang, Y. *et al.* Multiple sequence alignment-based rna language model and its application to structural inference. *Nucleic Acids Research* **52**, e3–e3 (2024).

[32] Sato, K., Akiyama, M. & Sakakibara, Y. Rna secondary structure prediction using deep learning with thermodynamic integration. *Nature communications* **12**, 941 (2021).

[33] Tan, Z., Fu, Y., Sharma, G. & Mathews, D. H. Turbofold ii: Rna structural alignment and secondary structure prediction informed by multiple homologs. *Nucleic acids research* **45**, 11570–11581 (2017).

[34] Zakov, S., Goldberg, Y., Elhadad, M. & Ziv-Ukelson, M. Rich parameterization improves rna structure prediction. *Journal of Computational Biology* **18**, 1525–1542 (2011).

[35] Zhang, C., Zhang, Y. & Pyle, A. M. rmsa: a sequence search and alignment algorithm to improve rna structure modeling. *Journal of Molecular Biology* **435**, 167904 (2023).

[36] Singh, J., Hanson, J., Paliwal, K. & Zhou, Y. Rna secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nature communications* **10**, 5407 (2019).

[37] Singh, J. *et al.* Improved rna secondary structure and tertiary base-pairing prediction using evolutionary profile, mutational coupling and two-dimensional transfer learning. *Bioinformatics* **37**, 2589–2600 (2021).

[38] Berman, H. M. *et al.* The protein data bank. *Nucleic acids research* **28**, 235–242 (2000).

[39] Rose, P. W. *et al.* The rcsb protein data bank: integrative view of protein, gene and 3d structural information. *Nucleic acids research* gkw1000 (2016).

[40] Cruz, J. A. *et al.* Rna-puzzles: a casp-like evaluation of rna three-dimensional structure prediction. *Rna* **18**, 610–625 (2012).

[41] Yang, H. *et al.* Tools for the automatic identification and classification of rna base pairs. *Nucleic Acids Research* **31**, 3450–3460 (2003).

[42] Sweeney, B. A. *et al.* R2dt is a framework for predicting and visualising rna secondary structure using templates. *Nature Communications* **12**, 3494 (2021).

[43] Kerpedjiev, P., Hammer, S. & Hofacker, I. L. Forna (force-directed rna): Simple and effective online rna secondary structure diagrams. *Bioinformatics* **31**, 3377–3379 (2015).

[44] Xu, J. & Zhang, Y. How significant is a protein structure similarity with tm-score= 0.5? *Bioinformatics* **26**, 889–895 (2010).

[45] Ren, Y. *et al.* Beacon: Benchmark for comprehensive rna tasks and language models. *Advances in Neural Information Processing Systems* **37**, 92891–92921 (2024).

[46] Zhao, J. *et al.* Iresfinder: Identifying rna internal ribosome entry site in eukaryotic cell using framed k-mer features. *Journal of genetics and genomics= Yi chuan xue bao* **45**, 403–406 (2018).

[47] Zhao, J. *et al.* Deepires: a hybrid deep learning model for accurate identification of internal ribosome entry sites in cellular and viral mrnas. *Briefings in Bioinformatics* **25**, bbae439 (2024).

[48] Chantsalnyam, T., Siraj, A., Tayara, H. & Chong, K. T. ncrdense: a novel computational approach for classification of non-coding rna family by deep learning. *Genomics* **113**, 3030–3038 (2021).

[49] Zhu, H. *et al.* Deep generalizable prediction of RNA secondary structure via base pair motif energy. *Nature Communications* **16**, 5856 (2025). URL https:

//doi.org/10.1038/s41467-025-60048-1.

[50] Brixi, G. *et al.* Genome modeling and design across all domains of life with evo 2. *BioRxiv* 2025–02 (2025).

[51] Sweeney, B. A. *et al.* Rnacentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res.* **49**, D212–D220 (2021).

[52] Zhang, Z. *et al.* Rnagenesis: Foundation model for enhanced rna sequence generation and structural insights. *bioRxiv* 2024–12 (2024).

[53] Xiong, P., Wu, R., Zhan, J. & Zhou, Y. Pairing a high-resolution statistical potential with a nucleobase-centric sampling algorithm for improving rna model refinement. *Nature communications* **12**, 2777 (2021).

[54] Vaswani, A. *et al.* Guyon, I. *et al.* (eds) *Attention is all you need.* (eds Guyon, I. *et al.*) *In 31th Conference on Neural Information Processing Systems*, 5998–6008 (NeurIPS, 2017).

[55] Fu, L. *et al.* Ufold: fast and accurate rna secondary structure prediction with deep learning. *Nucleic acids research* **50**, e14–e14 (2022).

[56] Chen, X., Li, Y., Umarov, R., Gao, X. & Song, L. Wallach, H. M. *et al.* (eds) *RNA secondary structure prediction by learning unrolled algorithms.* (eds Wallach, H. M. *et al.*) *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*, 8024–8035 (OpenReview.net, 2020). URL https://openreview.net/forum?id=S1eALyrYDH.

[57] Zuker, M. & Stiegler, P. Optimal computer folding of large rna sequences using thermodynamics and auxiliary information. *Nucleic acids research* **9**, 133–148 (1981).

[58] Chaudhury, S., Lyskov, S. & Gray, J. J. Pyrosetta: a script-based interface for implementing molecular modeling algorithms using rosetta. *Bioinformatics* **26**, 689–691 (2010).

[59] Devlin, J., Chang, M., Lee, K. & Toutanova, K. Burstein, J., Doran, C. & Solorio, T. (eds) *BERT: pre-training of deep bidirectional transformers for language understanding.* (eds Burstein, J., Doran, C. & Solorio, T.) *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186 (Association for Computational Linguistics, 2019).

[60] Paszke, A. *et al.* Wallach, H. M. *et al.* (eds) *Pytorch: An imperative style, high-performance deep learning library.* (eds Wallach, H. M. *et al.*) *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information*

Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, 8024–8035 (2019). URL https://proceedings.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html.

[61] Kalvari, I. *et al.* Non-coding rna analysis using the rfam database. *Current protocols in bioinformatics* **62**, e51 (2018).

[62] Kalvari, I. *et al.* Rfam 14: expanded coverage of metagenomic, viral and microrna families. *Nucleic acids research* **49**, D192–D200 (2021).

[63] Miao, Z. *et al.* Rna-puzzles round iv: 3d structure predictions of four ribozymes and two aptamers. *Rna* **26**, 982–995 (2020).

[64] Zhu, H. Data and checkpoints for structrfm (2025). URL https://doi.org/10.5281/zenodo.16754363.

[65] Zhu, H. heqin-zhu/structrfm: v0.0.3 (2025). URL https://doi.org/10.5281/zenodo.16754337.

Acknowledgements

We thank Dr. Zaixi Zhang (Postdoctoral researcher, Princeton University) for valuable comments and discussions that improved this work. This work is supported by National Natural Science Foundation of China (62271465 to S.K.Z., 32370581 to P.X.) and Suzhou Basic Research Program (SYG202338 to S.K.Z.).

Author contribution

S.K.Z and P.X conceived the idea, supervised the study. H.Z. designed the framework and experiments, conducted pre-training, structure prediction, function prediction, analyzed the results and wrote the paper. R.L. conducted zero-shot experiments, wrote the corresponding section, and revised the overall writing. F.Z. conducted the IRES identification experiments. F.T. improved the overview of structRFM. T.Y. visualized the tertiary structures. X.L. designed the experiment of evaluating model on various sequence lengths. Y.G improved the t-SNE visualization. All authors read, contributed to the discussion, and approved the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information. The supplementary for this paper is publicly available.

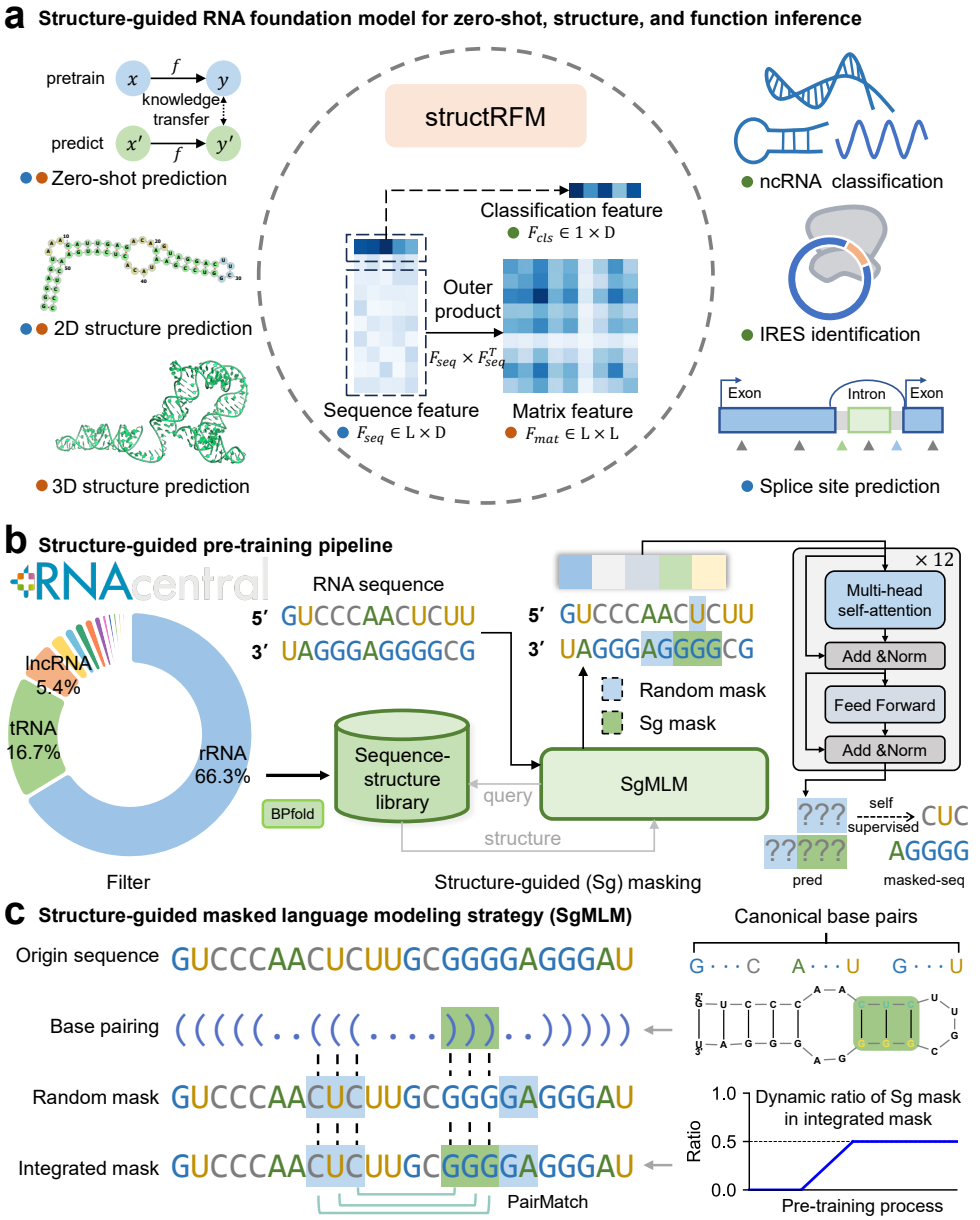


Fig. 1: **Overview of structRFM**. **a** structRFM as an RNA foundation model for zero-shot or downstream-adapted structural and functional inferences. The model outputs classification feature, sequence feature, and pairwise matrix feature, supporting tasks such as zero-shot homology classification, zero-shot 2D structure prediction, 2D and 3D structure prediction, splice site prediction, IRES identification, and ncRNA classification. **b** Structure-guided pre-training workflow of structRFM. RNA sequences from RNAcentral are filtered by length, and their secondary structures are predicted using BPfold to construct a sequence-structure library for SgMLM pre-training. **c** The structure-guided masked language modeling (SgMLM) strategy. RNA base-pairing information is used to generate an integrated mask that combines random nucleotide-level and structure-guided masking, with the proportion of structure-guided masking dynamically increasing during pre-training.

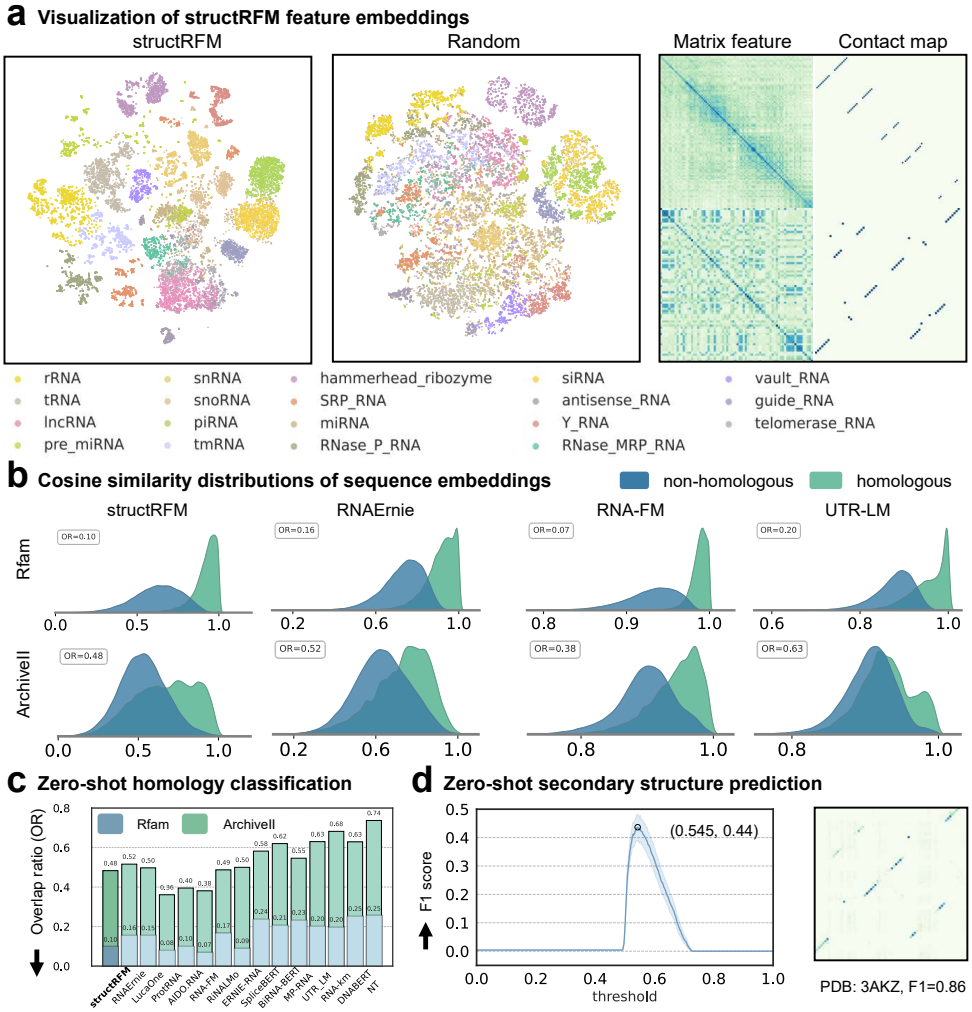


Fig. 2: **Feature visualization and zero-shot evaluation of structRFM**. **a** Visualization of RNA embeddings from structRFM and a random baseline using t-SNE on a subset of RNAcentral (23,994 RNAs), showing that structRFM makes clear separations of RNA families. Also shown are two examples of matrix features extracted from structRFM and corresponding contact maps. **b** Cosine similarity distributions between homologous and non-homologous RNA pairs for different models on Rfam (24,523 RNAs) and ArchiveII600 (3,611 RNAs) datasets, and **c** the overlap ratio (OR) of the two distributions is demonstrated, showing remarkable homology clarification ability of structRFM. The mean value is annotated for each method and each dataset. **d** Zero-shot secondary structure prediction with structRFM, showing F1 score across thresholds from 0 to 1 with an interval of 0.001 and an example heatmap.

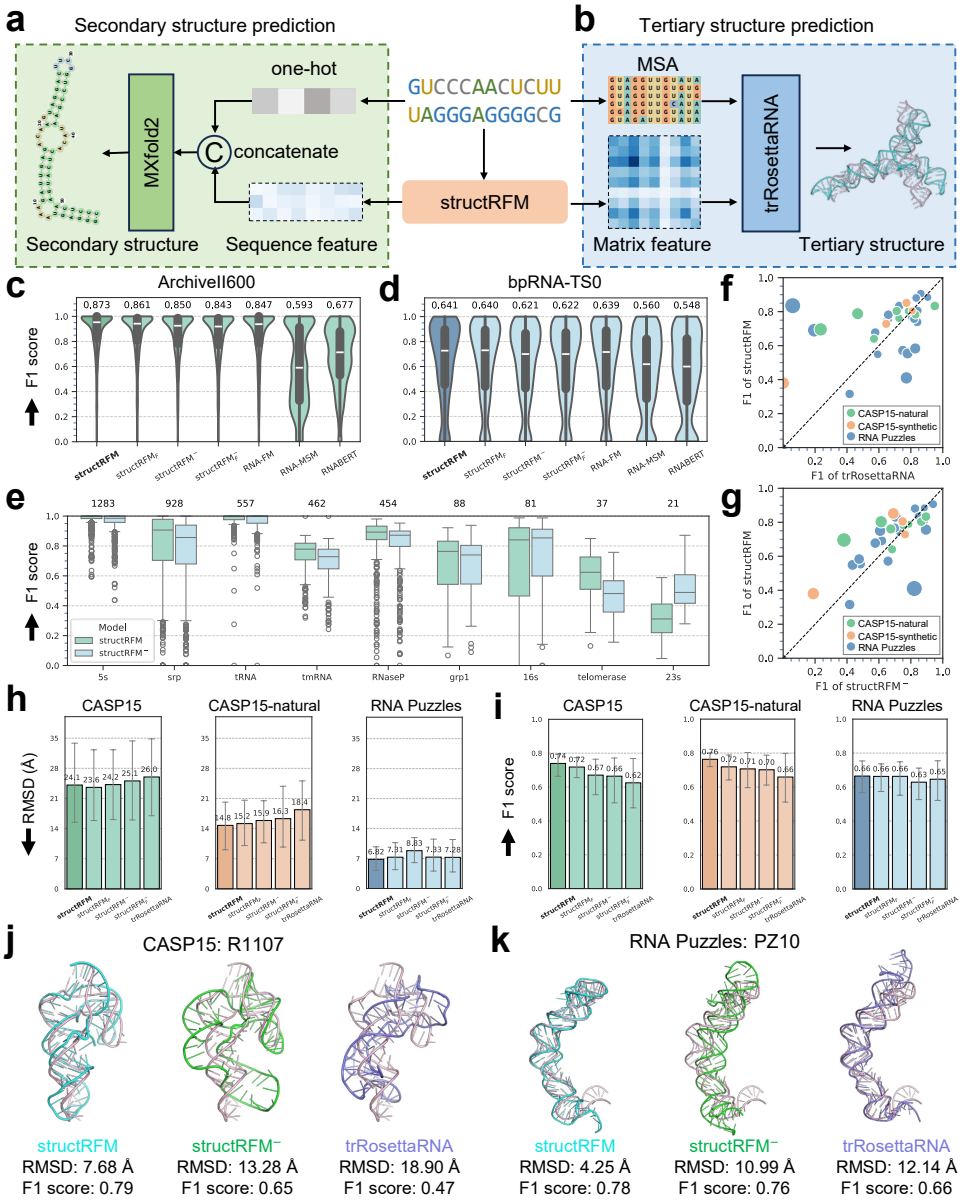


Fig. 3: **structRFM for structure prediction. a, b** Neural network architecture for secondary and tertiary structure prediction, respectively. **c, d** The F1 score of secondary structure prediction among structRFM and other models on ArchiveII600 (3,611 RNAs) and bpRNA-TS0 (1,305 RNAs) datasets. The white line in each violin denotes the medium number, while the mean value is displayed above each violin. **e** F1 score comparison of structRFM and structRFM$^-$ on the nine families of ArchiveII600. The number of samples of each RNA family is displayed at the top of the figure. **f, g** Head-to-head comparison of extracted secondary structure from predicted tertiary structures on CASP15 (12 RNAs) and RNA Puzzles (20 RNAs) datasets. The size of the circle depends on the absolute difference value of the two methods. **h, i** RMSD, F1 score comparison for tertiary structure prediction on CASP15 (12 RNAs), CASP15-natural (8 RNAs), and RNA Puzzles (20 RNAs), where the error bar in gray represents 95% confidence index. **j, k** Examples from predicted tertiary structures on CASP15 and RNA Puzzles, respectively. Native structure is in pink.

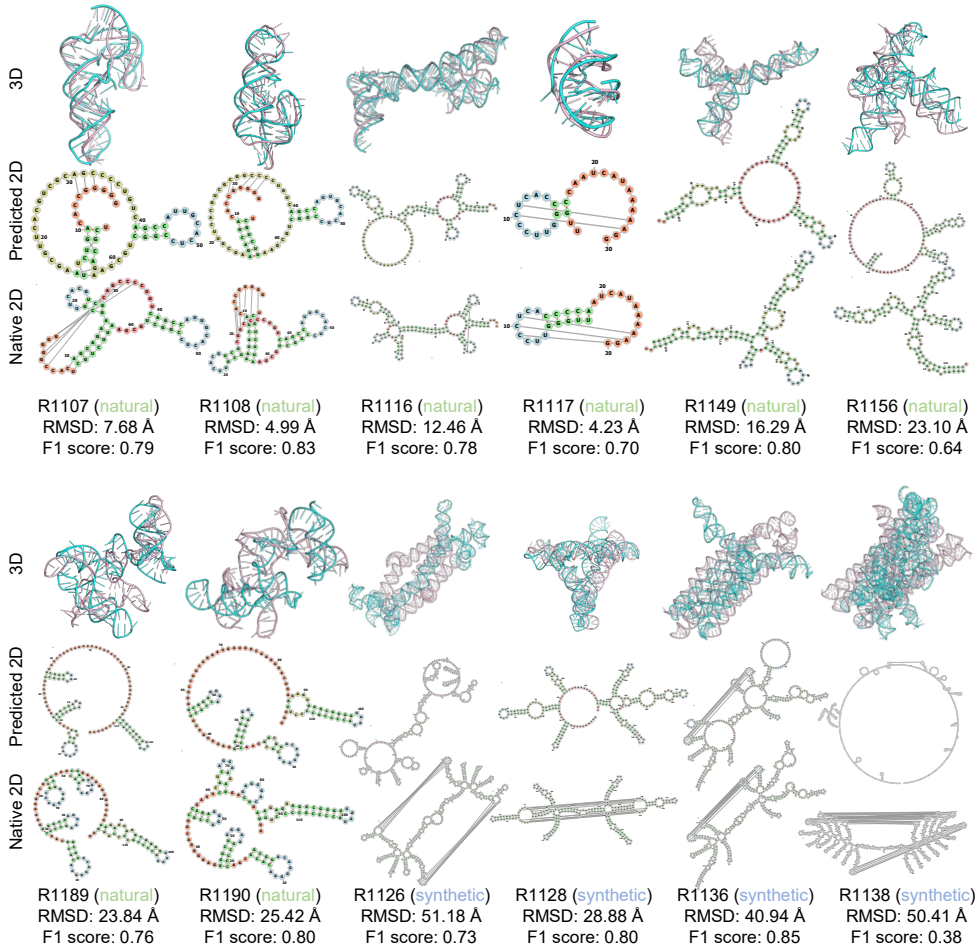


Fig. 4: **Visualization of structRFM (trRosettaRNA framework) predicted structures and native structures of the CASP15 datasets.** The RMSD of tertiary structures and the F1 score of secondary structures extracted from tertiary structures are listed. In the secondary structure, stems are in green, multiple loops are in red, interior loops are in yellow, hairpin loops are in blue, while 5' and 3' unpaired regions are in orange. Generally, structRFM performs better on natural targets than synthetic targets that consist of dense long-distance interactions, but sometimes fails to predict local short stems (in green) and result in large multiple loops (in red).

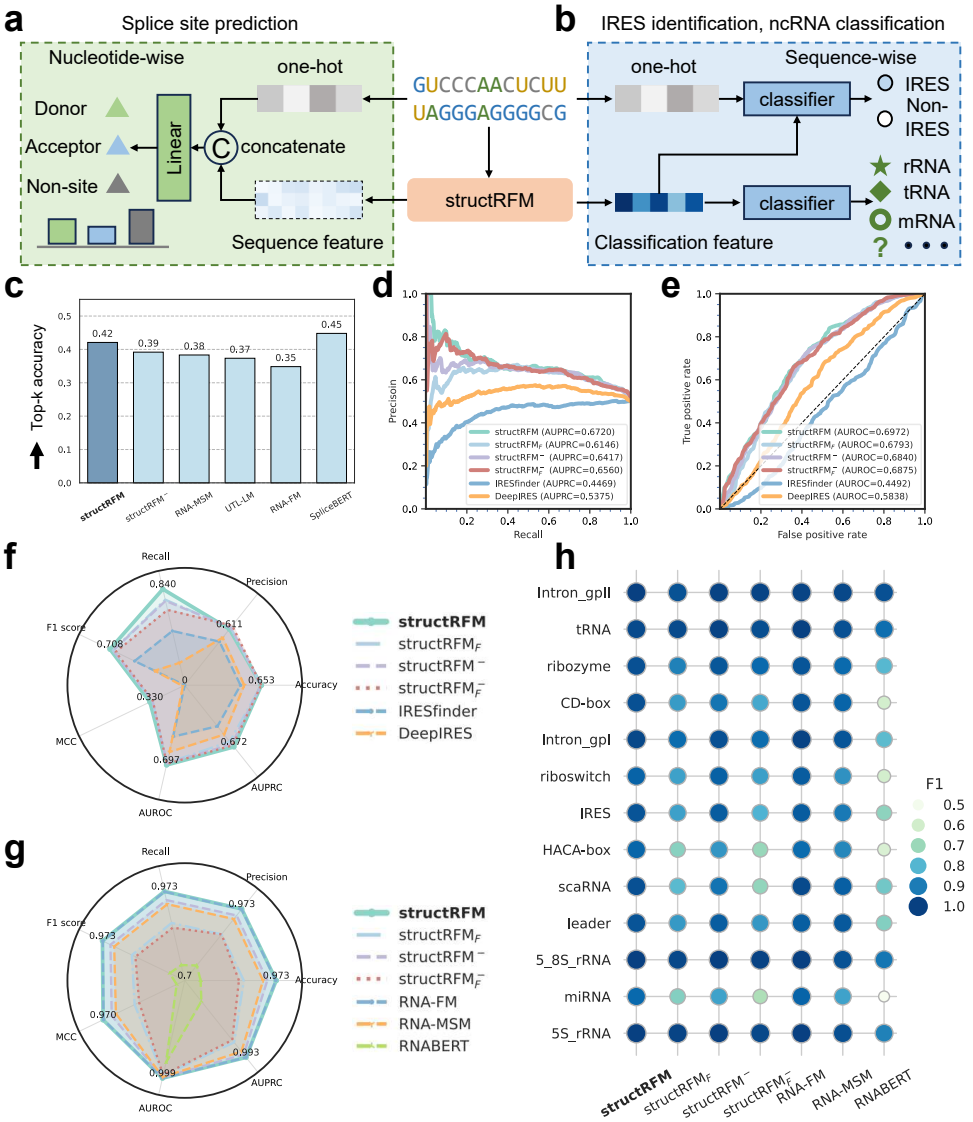
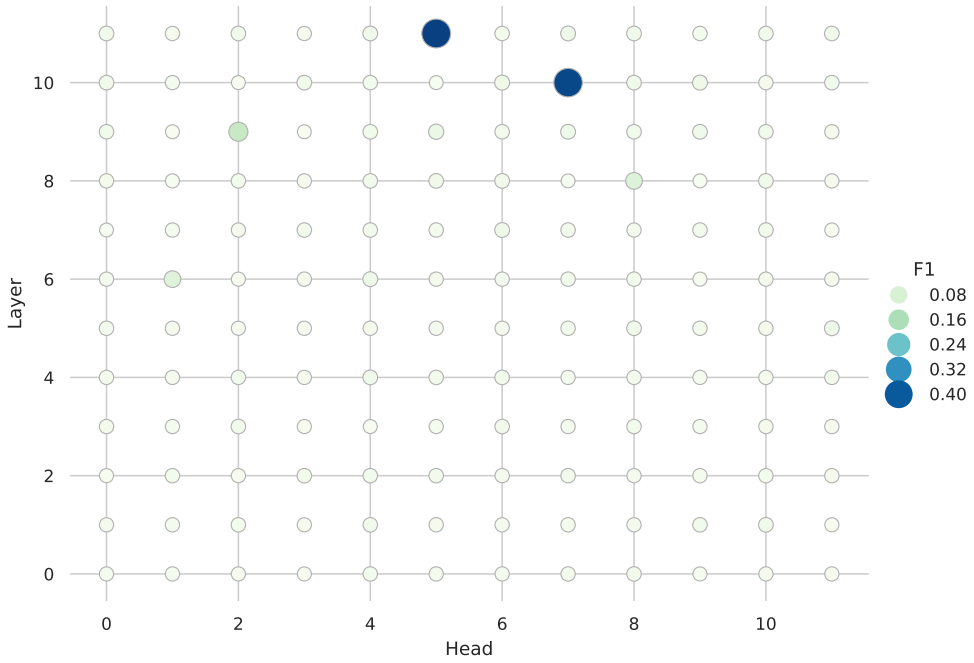
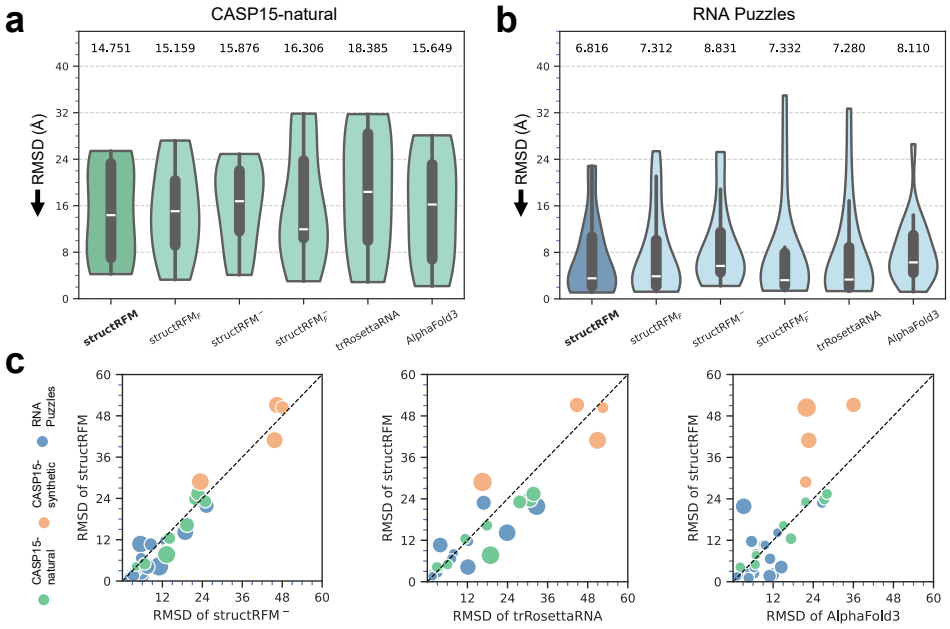


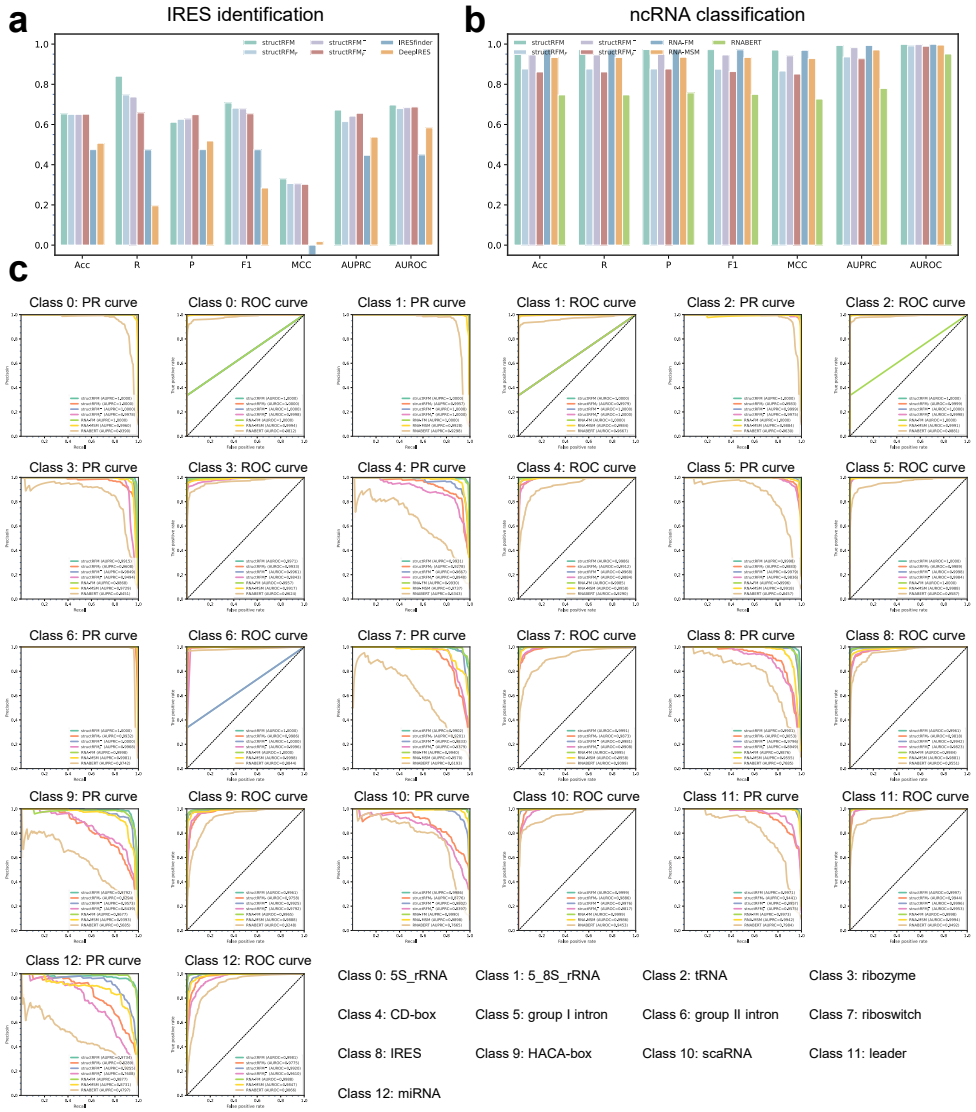
Fig. 5: **structRFM for function prediction. a, b** Neural network architecture for splice site prediction, IRES identification, and RNA sequence classification. **c** The top-k accuracy of structRFM, its variants, and SpliceBERT for splice site prediction (16,505 RNAs). **d, e** PR curve and ROC curve for IRES identification (1,164 RNAs). The radar visualization under seven metrics for **f** IRES identification (1,164 RNAs) and **g** ncRNA classification on nRC dataset (2,600 RNAs), where the metrics of structRFM are displayed. **h** F1 comparison of seven models on nRC dataset of 13 ncRNA categories. structRFM and RNA-FM achieve state-of-the-art performances, leading ahead of other models by a huge gap.



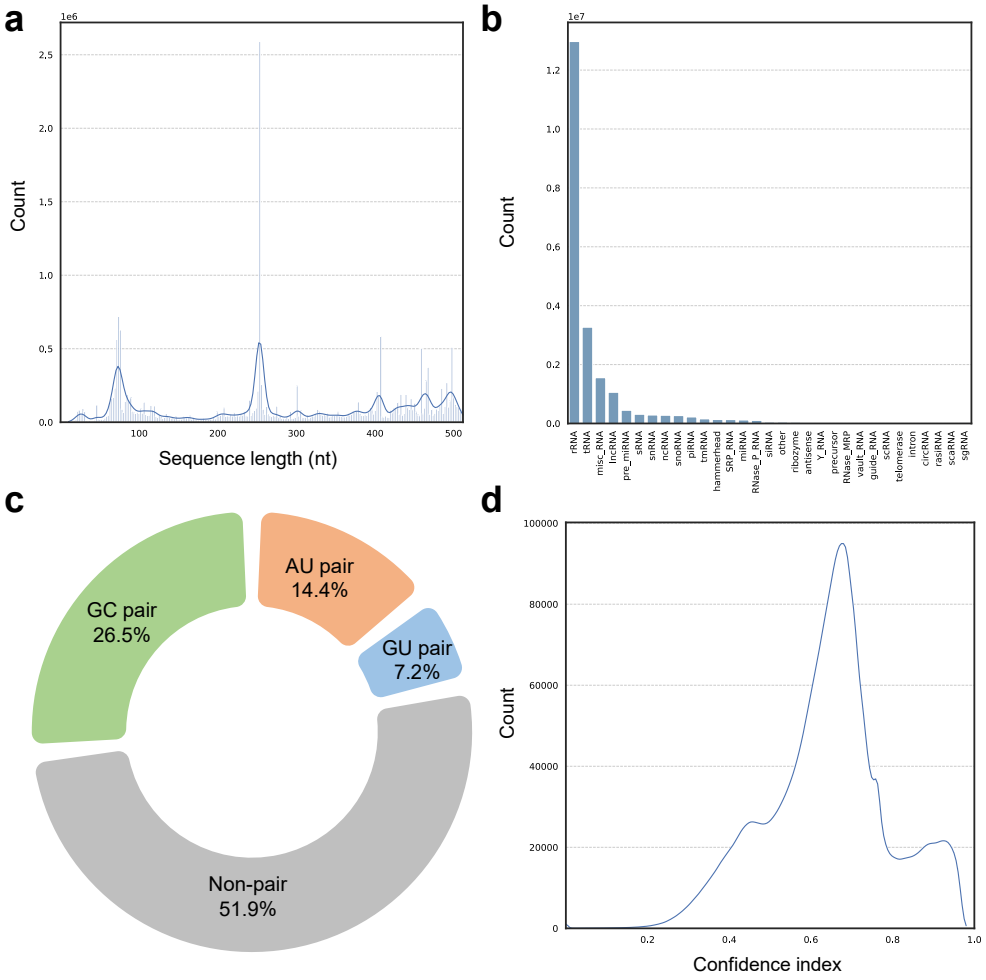
Extended Data Fig. 1: Grid searching across layer and head of attention maps to explore the best threshold of capturing structural contact information for zero-shot secondary structure prediction.



Extended Data Fig. 2: Comparison of structRFM, trRosettaRNA, AlphaFold3 [26] for tertiary structure prediction on CASP15 and RNA Puzzles datasets. a RMSD evaluation on CASP15-natural dataset (8 RNAs). **b** RMSD evaluation on RNA Puzzles dataset (20 RNAs). **c** Head-to-head comparison of structRFM with structRFM^{-}, trRosettaRNA, and AlphaFold3 on CASP15-natural (8 RNAs), CASP15-synthetic (4 RNAs), and RNA Puzzles (20 RNAs). The size of the circle depends on the absolute difference value of the two methods.



Extended Data Fig. 3: **Detailed visualization of IRES identification and ncRNA classification.** **a** Overall comparison of methods for IRES identification under seven metrics. **b** Overall comparison of methods for ncRNA classification under seven metrics. **c** The PR curve and ROC curve of the detailed 13 families in nRC dataset for ncRNA classification.



Extended Data Fig. 4: Overall statistics of our constructed pre-training sequence-structure dataset. a Distribution of sequence lengths, with most RNA sequences centered around 250 nucleotides. **b** Number count of RNA families (categories). Distribution of RNA families. rRNA and tRNA dominate the dataset, accounting for 66.3% and 16.7%, respectively. **c** Composition of base-pairing interactions in secondary structures. Canonical base pairs represent approximately half of all interactions, including GC (26.5%), AU (14.4%), and GU (7.2%) pairs. **d** Distribution of confidence scores for predicted secondary structures. The majority exhibit high confidence, supporting the reliability of base-pairing annotations.

Extended Data Table 1: Overall performance of evaluating structRFM on zero-shot, structure, and function inference tasks.

Task	Dataset	N	Metric	Method	Mean
Zero-shot homology classification	Rfam	24,523	OR ↓	RNAErnie	0.156
				MP-RNA	0.234
				UTR-LM	0.202
				RNA-FM	**0.071**
				structRFM	0.097
	ArchiveII600	3,611	OR ↓	RNAErnie	0.516
				MP-RNA	0.546
				UTR-LM	0.630
				RNA-FM	**0.381**
				structRFM	0.483
Zero-shot structure prediction	PDB	116	F1 ↑	structRFM	**0.436**
Secondary structure prediction	ArchiveII600	3,611	F1 ↑	RNAErnie	**0.873**
				RNA-FM	0.847
				structRFM	**0.873**
	bpRNA-TS0	1,305	F1 ↑	RNAErnie	0.608
				RNA-FM	0.639
				structRFM	**0.641**
Tertiary structure prediction	CASP15-natural	8	RMSD (Å) ↓	trRosettaRNA	18.4
				AlphaFold3	15.6
				structRFM	**14.8**
	RNA Puzzles	20	RMSD (Å) ↓	trRosettaRNA	7.28
				AlphaFold3	8.11
				structRFM	**6.82**
	CASP15	12	F1 ↑	trRosettaRNA	0.625
				structRFM	**0.739**
	RNA Puzzles	20	F1 ↑	trRosettaRNA	0.646
				structRFM	**0.665**
Splice site prediction	SpliceAI	16,505	Accuracy ↑	RNA-FM	0.350
				structRFM	**0.416**
IRES identification	IRES	1,164	F1 ↑	DeepIRES	0.284
				IRESfinder	0.475
				structRFM	**0.708**
ncRNA classification	nRC	2,600	F1 ↑	RNAErnie	0.964
				RNA-FM	0.972
				structRFM	**0.973**

Extended Data Table 2: **Comparative overview of structRFM with eight RNA language models.** Seq. is short for sequence and Stru. is short for secondary structure. The performances of overlap ratio (OR, lower is better) for zero-shot homology classification on Rfam [61, 62] dataset are displayed.

Model name	Model size(M)	Pre-training dataset	Structure information	Task-agnostic pre-training	Zero-shot (OR↓)
RiNALMo	33	∼36M Seq.	No	Yes	0.167
RiNALMo	148	∼36M Seq.	No	Yes	0.145
RiNALMo	651	∼36M Seq.	No	Yes	0.115
RNAErnie	86	∼23M Seq.	No	Yes	0.156
BiRNA-BERT	116	∼37M Seq	No	Yes	0.207
AIDO.RNA	650	∼42M Seq.	No	Yes	0.102
AIDO.RNA	1,600	∼42M Seq.	No	Yes	0.110
SpliceBERT	20	∼2M Seq.	No	Yes	0.239
RNA-FM	100	∼23M Seq.	No	Yes	0.071
ERNIE-RNA	86	∼20M Seq.	Pairwise score	Yes	0.090
MP-RNA	186	-	ViennaRNA	No	0.234
UTR_LM	1	∼2.3M Seq., Stru.	ViennaRNA	No	0.202
structRFM	86	∼21M Seq., Stru.	BPfold	Yes	0.097

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- structRFMsupplementary.pdf