# A transformer-based model to predict micro RNA interactions

Marco Nicolini[1,*][0009−0008−5137−2361], Federico Stacchietti[1][0009−0009−1945−134X], Carlos Cano[2][0000−0002−0181−2444], Elena Casiraghi[1,3,4,5][0000−0003−2024−7572], and Giorgio Valentini[1,3,*][0000−0002−5694−3919]

[1] AnacletoLab - Dipartimento Informatica, Università degli Studi di Milano, Italy
[2] Ciencias de la Computacion e Inteligencia Artificial, Universidad de Granada, Spain
[3] ELLIS - European Lab for Learning and Intelligent Systems, Milan Unit
[4] Department of Computer Science, Aalto University, Espoo, Finland
[5] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, United States
*{name.surname}@unimi.it

**Abstract.** The prediction of the interactions of micro RNA (miRNA) for the regulation of the cellular biological processes represents a challenging bioinformatics problem, with important implications for the design of new RNA-based drugs. We present *miRInter-Trans*, a model that integrates the RNA-FM foundation model pre-trained on a large corpus of non coding RNA (ncRNA) data with a feed-forward neural network trained on the RNA-FM hidden representations of ncRNA sequences. The model is able to successfully predict miRNA interactions using only the sequence of the ncRNA pairs. To our knowledge, this is the first work addressing ncRNA–ncRNA interaction prediction using sequence alone and embeddings from an RNA foundational model. The proposed approach demonstrates superior performance compared to a state-of-art Minimum Free Energy method.

## 1 Introduction

Deciphering RNA-RNA interactions is crucial for understanding the regulatory networks underlying gene expression and cellular biological processes [7]. However, experimental mapping of ncRNA interactions remains challenging due to limitations in current experimental methods and computational models [21].

Predictions of RNA-RNA interactions have relied on approaches such as Minimum Free Energy (MFE) calculations and accessibility-based models. Tools like IntaRNA evaluate the interaction energy between ncRNA and messenger RNA (mRNA) to predict their interactions [9]. Accessibility-based methods can effectively distinguish true interactions from background noise [18], but these techniques are limited by their reliance on predefined parameters and simplified energy assumptions. On the wet-experimental side, methods such as RNA Antisense Purification (RAP-RNA) are limited by high costs and working-intense demands [5].

Recently deep learning and transformer-based methods have been proposed for modeling biological sequences directly from raw data [12, 19, 11]. In particular deep learning methods have been successfully applied to RNA-protein inter-

actions [2, 16], and network-based contrastive learning approaches have shown promise in related association tasks [14].

Deep learning architectures and graph-based methods have indeed demonstrated potential in predicting biomolecular interactions [1, 17]. Convolutional neural networks (CNNs) and deep forests excel in detecting local sequence patterns, while graph-based learning methods leverage topological relationships to infer associations. Nonetheless, these approaches often depend on manual feature curation, predefined graph structures, or supervised training on specific datasets, limiting their applicability to novel ncRNA sequences.

In contrast, Large Language Model (LLM)-based methods can learn directly from RNA public databases without predefined energy parameters or explicit structural graphs, using self-supervised learning, and can also learn dependencies and relationships between motifs and functional patterns through the attention mechanism [22, 13, 20]. Unlike thermodynamic models that rely on simplified interaction assumptions, LLMs capture complex interaction motifs—including non-canonical base pairings and contextual structural cues—often missed by energy-based or graph-based approaches. By leveraging deep contextual embeddings, these models can infer interaction probabilities based on sequence patterns and latent structural signals, offering a flexible, data-driven strategy for ncRNA interaction prediction that is advantageous when structural annotations are incomplete or lacking.

We propose a method for predicting interactions between miRNAs and various ncRNAs—namely long non coding RNAs (lncRNAs), micro RNAs (miRNAs), and small nucleolar RNAs (snoRNAs)— that leverages deep sequence embeddings generated by the transformer based model RNA-FM [4]. RNA-FM is a foundational encoder-only model pre-trained using self-supervised learning on a vast collection of over 23 million non-coding RNA sequences, capturing sequential and evolutionary features without relying on annotated data. In our approach, RNA-FM embeddings are pooled using two strategies—maximum and average pooling (evaluated both independently and in combination)—to extract robust representations from RNA sequences. These pooled features are then used to train a feed-forward neural network (FFNN) for classifying potential interactions, offering a framework for uncovering novel regulatory relationships in the ncRNA realm. By bypassing the need for explicit thermodynamic parameterization and manually crafted features, this design presents a scalable and efficient approach for identifying novel regulatory interactions [6].

## 2 Methods

### 2.1 Dataset

Our dataset comprises a curated selection of multispecies ncRNA interaction pairs extracted from RNA-KG [3][6]. In this work, we focus on miRNA interactions (miRNA–lncRNA, miRNA–miRNA, and miRNA–snoRNA), as detailed in Tables 1 and 2.

[6] RNA sequences and interacting pairs were retrieved using scripts provided at https://github.com/AnacletoLAB/RNA-KG; we kept pairs where we successfully obtained the RNA sequences for both interacting molecules.

To comply with the constraints of the ncRNA Language Model (RNA-FM [4]), in our pipeline we retained only those molecules with a sequence length of at most 1022 nucleotides. After filtering, the datasets comprise a total of 13104 unique interaction pairs and 4374 unique sequences.

For notational convenience, we define the set of length-filtered molecules as $\mathcal{R} = \{r_i\}, \quad i = 1, \ldots, |\mathcal{R}|$, and a mapping $\phi : \mathcal{R} \to \mathcal{T}$, where $\mathcal{T} = \{miRNA, lncRNA, snoRNA\}$ represents the types of ncRNA sequences considered in our experiments. The identity of an interaction pair is determined solely by its constituent molecules, independent of their order:$(r_i, r_j) = (r_j, r_i)$, and analogously, the interaction type is defined as $(\phi(r_i), \phi(r_j)) = (\phi(r_j), \phi(r_i))$.

Table 1: Distribution of miRNA interactions in the training set across three processing stages. The table reports counts for miRNA–lncRNA, miRNA–miRNA, and miRNA–snoRNA interaction couples. The "Initial couples (dataset)" provides the raw counts from the initial dataset. The "Filter for seq. length (dataset)" displays the counts of the dataset after applying a nucleotide length threshold, (< 1022 nucleotides). The "After data aug. (train/test set)" rows show the increased counts resulting from data augmentation.

| **Couples/interaction types** | miRNA-lncRNA | miRNA-miRNA | miRNA-snoRNA |
|---|---|---|---|
| Initial couples (dataset) | 81891 | 2132 | 1927 |
| Filter for seq. length (dataset) | 9058 | 2071 | 1975 |
| Filter for seq. length (train set) | 8152 | 1863 | 1777 |
| Filter for seq. length (test set) | 906 | 208 | 198 |
| After data aug. (test set) | 3624 | 832 | 792 |
| After data aug. (train set) | 32608 | 7452 | 7108 |

Table 2: Dataset unique sequences: counts of unique sequences across different interacting couple types. For miRNA–lncRNA and miRNA–snoRNA pairs, unique counts are reported separately for each partner (type1/type2). For miRNA–miRNA interactions, sequence uniqueness is computed on the entire set of interacting sequences, as the same set serves for both partners. The "Initial unique seq." lists the counts prior to any filtering, while the "Filter for seq. length" shows the counts after applying a nucleotide length threshold (< 1022).

| **Unique Sequences/types** | miRNA-lncRNA | miRNA-miRNA | miRNA-snoRNA |
|---|---|---|---|
| Initial unique seq. | 4004/4251 | 1322 | 428/553 |
| Filter for seq. length | 1343/730 | 1322 | 427/552 |

### 2.2 Data Augmentation and generation of negative interactions

We adopt a data augmentation strategy that increases the effective size of the training set by a factor of four. For each original training instance represented as $(r_i, r_j)$ we generate three additional instances:

1. **Molecule Order Reversal:** Swap the order of molecules, yielding: $(r_j, r_i)$

2. **Sequence Flipping:** Reverse the nucleotide order in each molecule, denoted by the superscript $F$: $(r_i^F, r_j^F)$
3. **Combined Augmentation:** Apply both order reversal and sequence flipping: $(r_j^F, r_i^F)$

Thus, if the original dataset comprises $N$ instances, the augmented dataset contains $N_{\text{aug}} = 4N$. This approach introduces invariance to both molecule order and sequence orientation, enhancing the model's ability to capture underlying biological patterns and improving robustness against input variability. Sequence flipping is a common data augmentation strategy in protein language models [11]. Initial experiments with validation sets indicated that this augmentation scheme substantially improves performance.

*Negative Interactions.* Since our dataset includes only validated "positive" miRNA interactions, we generate negative examples by assuming that any pair of distinct ncRNA sequences not observed as a positive interaction can be a valid negative candidate.

Let $\mathcal{R} = \{r_1, r_2, \ldots, r_N\}$ denote the set of unique ncRNA sequences. Define the set of known positive interactions as

$$\mathcal{P} = \{(r_i, r_j) \mid \text{an interaction between } r_i \text{ and } r_j \text{ is observed}\}.$$

Then, the complete set of possible ncRNA pairs (excluding self-interactions) is given by $\mathcal{R} \times \mathcal{R}$, and the set of potential negatives is

$$\mathcal{N}_{\text{potential}} = \{(r_i, r_j) \in \mathcal{R} \times \mathcal{R} \mid r_i \neq r_j\} \setminus \mathcal{P}.$$

*Negative Sampling.* For each positive interaction pair $(r_i, r_j)$ with type $(\phi(r_i), \phi(r_j))$, we retain $r_i$ and replace $r_j$ by sampling a candidate $r' \in \mathcal{R}$ that satisfies:

$$r' \neq r_i, \quad \phi(r') = \phi(r_j), \quad (r_i, r') \notin \mathcal{P}.$$

For each positive instance, a parameter $n$ controls the number of negative samples generated, to control the imbalance of negative vs positive examples. In our experiments we set $n = 20$, in order to partially reproduce the imbalance between negative and positive interactions in real data.

### 2.3 Model Architecture

Our model is implemented as a two-stage pipeline (see Figure 1). In the first stage, ncRNA sequences are converted into dense representations (embedded sequences) using a pre-trained ncRNA language model (RNA-FM). These representations are then aggregated and fed into a feed-forward neural network (FFNN) to predict interaction probabilities.

*RNA-FM Overview.* RNA-FM is based on a transformer architecture and was pre-trained on millions of ncRNA sequences using a self-supervised learning paradigm [4]. The model comprises 12 bidirectional encoder layers with multi-head self-attention, where each token is embedded into a vector of dimension $H$. For an RNA sequence of length $L$, an initial embedding layer converts each nucleotide into a vector in $\mathbb{R}^H$, resulting in an $L \times H$ representation.

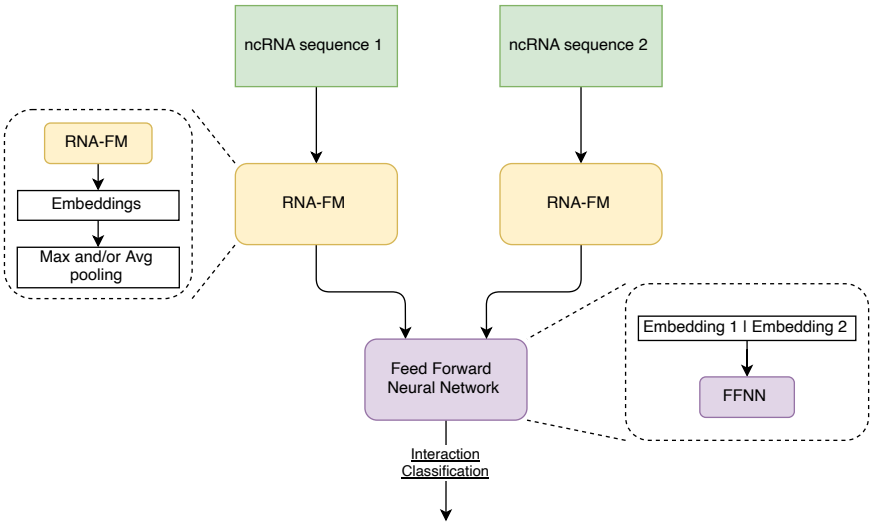



Fig. 1: *miRInter-Trans* architecture schema. In the first stage, a pre-trained ncRNA language model (RNA-FM) transforms ncRNA sequences into dense, embedded representations. These RNA-FM embeddings~~embeddings~~[MN] are subsequently aggregated and processed by a feed-forward neural network to compute interaction probabilities.

*Tokenization and Feature Extraction.* The tokenization strategy is designed to capture the intrinsic properties of RNA sequences. Each nucleotide is mapped directly to a token. The transformer processes these tokens and outputs a matrix $\mathbf{X} \in \mathbb{R}^{L \times H}$, where the $i$-th row $\mathbf{x}_i$ is the latent representation of the $i$-th token. To derive a fixed-length sequence embedding, we consider two pooling methods:

$$\mathbf{e}_{\text{avg}} = \frac{1}{L} \sum_{i=1}^{L} \mathbf{x}_i, \qquad \mathbf{e}_{\text{max}} = \max_{i=1}^{L} \mathbf{x}_i.$$

These can be used individually or concatenated:

$$\mathbf{e} = [\mathbf{e}_{\text{avg}}; \mathbf{e}_{\text{max}}] \in \mathbb{R}^{2H}.$$

To train and evaluate ncRNA pairs, the pooled embeddings from each molecule are concatenated, resulting in a vector of dimension $2H$ (or $4H$ if both pooling methods are combined).

*Downstream Prediction.* The concatenated embeddings serve as input to the FFNN, which outputs a classification label and an interaction probability. The choice of $H$ depends on the specific ncRNA language model employed; RNA-FM [4] uses $H = 640$.

The FFNN consists of 4 layers, each with a hidden dimension of 1024. The FFNN processes the embedded features through successive linear transformations interleaved with non-linear activations, with the learning goal of discriminating between interacting and non-interacting ncRNA pairs.

*Mini-batch balancing.* To mitigate class imbalance, our training strategy constructs mini-batches that incorporate a controlled mix of positive and negative examples. Let $D$ be the complete dataset with $n$ examples, where $P$ and $N$ denote the sets of positives and negatives, with $|P| = n_p$ and $|N| = n_n$, respectively. Each mini-batch $B$ of size $m$ is formed by randomly drawing $m_p$ positives (with replacement) and $m_n$ negatives (without replacement), such that $m_p + m_n = m$. The negative proportion is defined as

$$k = \frac{m_n}{m},$$

so that $k = 0.5$ implies 50% of the mini-batch comprises negatives. Positive samples are drawn with replacement due to their lower frequency, while negatives are sampled without replacement to ensure wider coverage of the available examples.

### 2.4 Experimental Setup

*Data Splitting.* We partition the dataset by assigning 90% for training and 10% for testing. Within the training set, 80% is used for model training and 20% for validation. The validation subset is employed for early stopping and for optimizing the classification threshold by maximizing the Matthews correlation coefficient [10].

*Training Parameters.* Training is conducted with the following settings: a learning rate $\eta = 5 \times 10^{-4}$ with a warm-up phase lasting 4 epochs, followed by cosine decay; 50 epochs with early stopping (patience of 10 epochs); batch size of 512; dropout rate of 0.2; and the Adam optimizer minimizing a binary cross-entropy loss. The network features a hidden dimension of 1024 across 4 layers. Balanced batch sampling is implemented with each batch containing 70% negatives, and an epoch is defined as one complete pass through the negative examples. These hyperparameter choices were informed by initial validation performance, with a preference for a simple architecture to ensure computational efficiency. Training and validation loss trajectories are monitored to assess convergence and check for overfitting.

## 3 Results and discussion

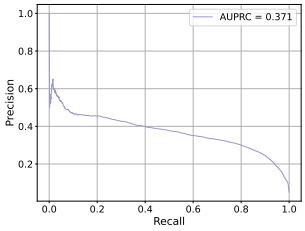
We evaluated *miRInter-Trans* on 3 different interaction prediction tasks, i.e. miRNA-lncRNA, miRNA-miRNA and miRNA-snoRNA, using Area under the receiver operating curve (AUROC) and Area under the precision-recall curve (AUPRC) metrics. IntaRNA [9], originally developed to predict ncRNA and mRNA interactions, was run on the same prediction tasks for a reference baseline.

Data augmentation was applied to both training and test sets. *miRInter-Trans* was trained using 3 different pooling strategies:

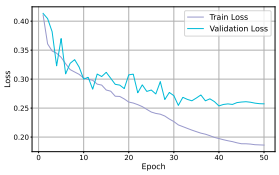
- ***miRInter-Trans*-concat**: Embedded representations of the sequences are obtained by concatenation of the average and max pooling of the amino acid embeddings.

Table 3: Comparison of AUROC and AUPRC results on the test set between *miRInter-Trans* with a random baseline and IntaRNA for the prediction of miRNA-lncRNA, miRNA-miRNA and miRNA-snoRNA interactions.

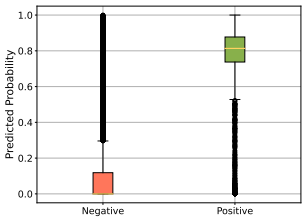
| **Method** | **AUROC** | | |
|---|---|---|---|
| | miRNA-lncRNA | miRNA-miRNA | miRNA-snoRNA |
| Random baseline | 0.500 | 0.500 | 0.500 |
| *miRInter-Trans*-concat | 0.859 | 0.547 | 0.870 |
| *miRInter-Trans*-AVG | **0.903** | **0.809** | **0.906** |
| *miRInter-Trans*-MAX | 0.859 | 0.559 | 0.826 |
| IntaRNA | 0.571 | 0.632 | 0.66 |
| **Method** | **AUPRC** | | |
| | miRNA-lncRNA | miRNA-miRNA | miRNA-snoRNA |
| Random baseline | 0.047 | 0.047 | 0.047 |
| *miRInter-Trans*-concat | 0.170 | 0.054 | 0.231 |
| *miRInter-Trans*-AVG | **0.255** | **0.227** | **0.350** |
| *miRInter-Trans*-MAX | 0.166 | 0.055 | 0.166 |
| IntaRNA | 0.051 | 0.123 | 0.100 |



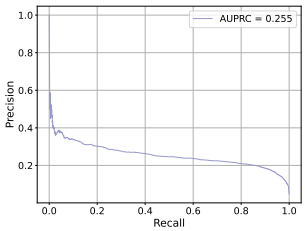

(a) Val. PRC Curve



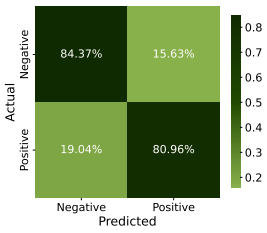

(b) Training-val. loss



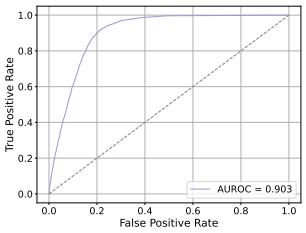

(c) Test boxplot probabilities




(d) Test PRC curve




(e) Test confusion matrix




(f) Test ROC curve

Fig. 2: Results for miRNA-lncRNA average pooling. (a) Overall precision recall curve (PRC) on the validation set; (b) Training and validation loss across epochs; (c) Distribution of the *miRInter-Trans* predicted probabilities on negative and positive examples on the test set; (d) Precision recall curve (PRC) on the test set; (e) Confusion matrix on the test set; (f) Receiver operating curve (ROC) on the test set.

- ***miRInter-Trans*-AVG**: Average pooling is applied to represent ncRNA sequences.
- ***miRInter-Trans*-MAX**: Max pooling is applied to represent ncRNA sequences.

A random baseline, which represents the expected performance of a random classifier, is also included in Table 3. The AUPRC of the random baseline method is given by: $\frac{N_+}{N_+ + N_-}$ where $N_+$ is the number of positive samples, and $N_-$ is the number of negative samples. Given a positive-to-negative ratio of 1:20, the random baseline AUPRC is: $\frac{1}{1+20} = \frac{1}{21} \approx 0.0476$.

These results show that *miRInter-Trans*-AVG achieves the best performance for all the interaction prediction tasks for the two metrics under consideration (Wilcoxon rank-sum test, $\alpha = 10^{-5}$). *miRInter-Trans*-concat and *miRInter-Trans*-MAX outperform IntaRNA on miRNA-lncRNA and miRNA-snoRNA prediction tasks, but IntaRNA performs better on miRNA-miRNA predictions.

The poor results of IntaRNA on miRNA-lncRNA and miRNA-snoRNA interactions are probably due to the fact that this method has been designed to predict ncRNA-mRNA interaction in bacteria, while our dataset includes a larger set of ncRNA interactions, involving also eukaryotic ncRNA. Fig. 5, 6, 7 provide results about IntaRNA predictions for each considered miRNA interaction prediction task.

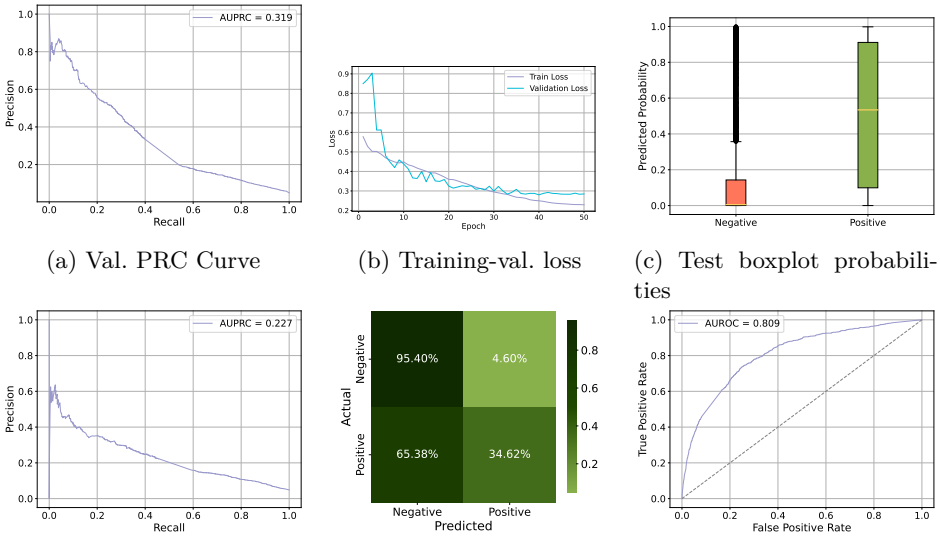
Detailed results obtained by *miRInter-Trans*-AVG on each of the considered miRNA interaction prediction tasks are available in Fig. 2, 3, 4, including precision-recall curves (PRC), receiver operating curves (ROC), confusion matrices, training and validation loss curves, and distribution of predicted probabilities for positive and negative examples. These results allow to identify potential issues for further experimentation and improvement, such as the high ratio of negative predictions for the miRNA-miRNA prediction task and the significant volume of false negatives for the miRNA-snoRNA task.

However, the obtained results show that *miRInter-Trans* can successfully predict miRNA interactions, achieving AUROC above 0.8 for the prediction of different types of miRNA interactions with *miRInter-Trans* AVG.

In practical applications, false positives (FP) may lead to the waste of experimental resources on ncRNA pairs that do not interact, while false negatives (FN) risk overlooking genuine regulatory relationships. For example, an FP miRNA–lncRNA call could prompt costly pull-down assays without yielding a true binding event, and an FN could exclude a critical regulatory axis from downstream functional studies. To mitigate this, we recommend calibrating the decision threshold to prioritize either sensitivity or precision depending on the context: for target discovery pipelines, a higher sensitivity may be preferred (lower threshold), whereas for validation-driven workflows, higher precision (higher threshold) will reduce experimental burden [5].

We plan to train the model also with other types of ncRNA interactions, using data available from large databases such as RNAInter [8] and RNAcentral [15]. A limitation of our approach is the maximum length of the allowed input sequence, which is due to the constraints of the underlying RNA-FM model. We also plan to explore new pooling strategies, since max and average pooling may overly compress the RNA-FM embedding matrices. Further, we will experiment with alternative classifiers beyond the current feed-forward neural network applied to pooled embeddings.

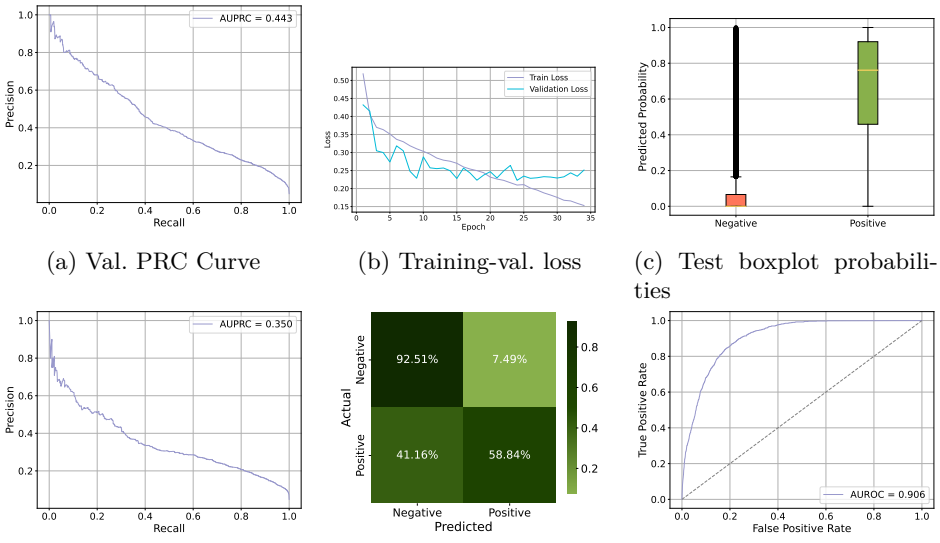
In perspective, experiments with other RNA foundation models (e.g. GenerRNA [22] or RNAErnie [20]), could circumvent this limitation, allowing us to obtain embedded representations of longer ncRNA, and thus extending the applicability of our proposed approach to a broader ncRNA context.




(a) Val. PRC Curve (b) Training-val. loss (c) Test boxplot probabilities

(d) Test PRC curve (e) Test confusion matrix (f) Test ROC curve

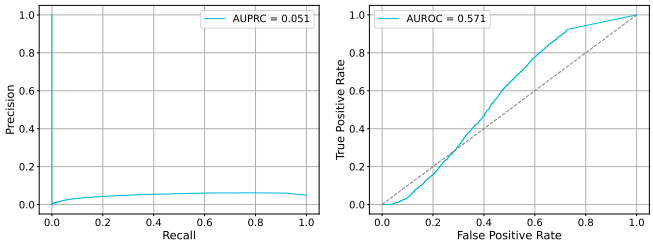
Fig. 3: Results for miRNA-miRNA average pooling. (a) Overall precision recall curve (PRC) on the validation set; (b) Training and validation loss across epochs; (c) Distribution of the *miRInter-Trans* predicted probabilities on negative and positive examples on the test set; (d) Precision recall curve (PRC) on the test set; (e) Confusion matrix on the test set; (f) Receiver operating curve (ROC) on the test set.




(a) Val. PRC Curve (b) Training-val. loss (c) Test boxplot probabilities

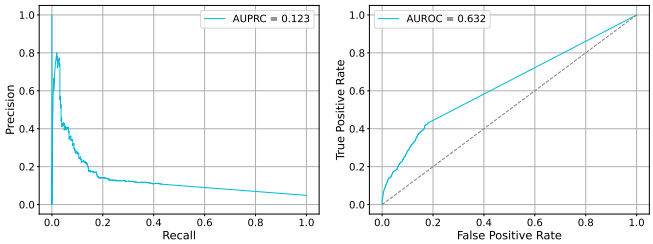
(d) Test PRC curve (e) Test confusion matrix (f) Test ROC curve

Fig. 4: Results for miRNA-snoRNA average pooling. (a) Overall precision recall curve (PRC) on the validation set; (b) Training and validation loss across epochs; (c) Distribution of the *miRInter-Trans* predicted probabilities on negative and positive examples on the test set; (d) Precision recall curve (PRC) on the test set; (e) Confusion matrix on the test set; (f) Receiver operating curve (ROC) on the test set.



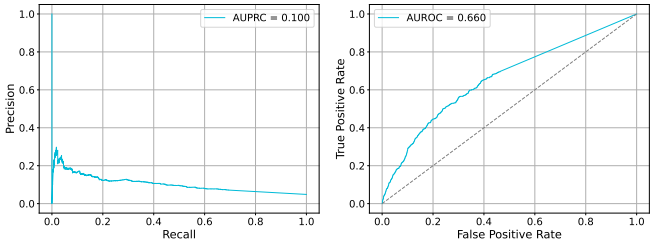

(a) Test PRC curve (b) Test ROC curve

Fig. 5: Results for miRNA-lncRNA IntaRNA classification. (a) Precision recall curve (PRC) on the test set; (b) Receiver operating curve (ROC) on the test set.




(a) Test PRC curve (b) Test ROC curve

Fig. 6: Results for miRNA-miRNA IntaRNA classification. (a) Precision recall curve (PRC) on the test set; (b) Receiver operating curve (ROC) on the test set.




(a) Test PRC curve (b) Test ROC curve

Fig. 7: Results for miRNA-snoRNA IntaRNA classification. (a) Precision recall curve (PRC) on the test set; (b) Receiver operating curve (ROC) on the test set.

**Acknowledgments.** This work was supported by National Center for Gene Therapy and Drugs Based on RNA Technology—MUR (Project no. CN_00000041) funded by NextGeneration EU program, and by FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence) - AIDH – FAIR - PE0000013.

We thank Emanuele Cavalleri for providing the initial ncRNA-ncRNA interaction datasets.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

**Supplementary material** The *miRInter-Trans* code, and the scripts to reproduce the experiments and tutorials are available from GitHub:
https://github.com/AnacletoLAB/miRInter-Trans.

## References


1. Abramson, J., Adler, J., Dunger, J., et al.: Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature **630**, 493–500 (2024). https://doi.org/https://doi.org/10.1038/s41586-024-07487-w
2. Alipanahi, B., Delong, A., Weirauch, M.T., Frey, B.J.: Predicting the sequence specificities of dna-and rna-binding proteins by deep learning. Nature biotechnology **33**(8), 831–838 (2015)
3. Cavalleri, E., Cabri, A., Soto-Gomez, M., Bonfitto, S., Perlasca, P., Gliozzo, J., Callahan, T.J., Reese, J., Robinson, P.N., Casiraghi, E., et al.: An ontology-based knowledge graph for representing interactions involving rna molecules. Scientific Data **11**(1), 906 (2024)
4. Chen, J., Hu, Z., Sun, S., Tan, Q., Wang, Y., Yu, Q., Zong, L., Hong, L., Xiao, J., Shen, T., et al.: Interpretable rna foundation model from unannotated data for highly accurate rna structure and function predictions. arXiv preprint arXiv:2204.00300 (2022)
5. Engreitz, J.M., Sirokman, K., McDonel, P., Shishkin, A.A., Surka, C., Russell, P., Grossman, S.R., Chow, A.Y., Guttman, M., Lander, E.S.: Rna-rna interactions enable specific targeting of noncoding rnas to nascent pre-mrnas and chromatin sites. Cell **159**(1), 188–199 (2014)
6. Fabbri, M., Girnita, L., Varani, G., Calin, G.A.: Decrypting noncoding rna interactions, structures, and functional networks. Genome research **29**(9), 1377–1388 (2019)
7. Hombach, S., Kretz, M.: Non-coding RNAs: Classification, Biology and Functioning, p. 3–17. Springer International Publishing (2016). https://doi.org/http://dx.doi.org/10.1007/978-3-319-42059-2_1
8. Kang, J., Tang, Q., He, J., Li, L., Yang, N., Yu, S., Wang, M., Zhang, Y., Lin, J., Cui, T., et al.: Rnainter v4. 0: Rna interactome repository with redefined confidence scoring system and improved accessibility. Nucleic acids research **50**(D1), D326–D332 (2022)
9. Mann, M., Wright, P.R., Backofen, R.: Intarna 2.0: enhanced and customizable prediction of rna–rna interactions. Nucleic acids research **45**(W1), W435–W439 (2017)
10. Matthews, B.W.: Comparison of the predicted and observed secondary structure of t4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Protein Structure **405**(2), 442–451 (1975)
11. Nicolini, M., Saitto, E., Jimenez-Franco, R., Cavalleri, E., Galeano, A., Malchiodi, D., Paccanaro, A., Robinson, P., Casiraghi, E., Valentini, G.: Fine-tuning of conditional Transformers improves in silico enzyme prediction and generation. Computational and Structural Biotechnology Journal **27**, 1318–1334 (2025). https://doi.org/10.1016/j.csbj.2025.03.037
12. Sapoval, N., Aghazadeh, A., Nute, M., et al.: Current progress and open challenges for applying deep learning across the biosciences. Nat Commun **13**(1728) (2022). https://doi.org/http://doi.org/10.1038/s41467-022-29268-7
13. Shen, T., Hu, Z., Sun, S., Liu, D., Wong, F., Wang, J., Chen, J., Wang, Y., Hong, L., Xiao, J., et al.: Accurate rna 3d structure prediction using a language model-based deep learning approach. Nature Methods pp. 1–12 (2024)

14. Sun, S.L., Jiang, Y.Y., Yang, J.P., Xiu, Y.H., Bilal, A., Long, H.X.: Predicting noncoding rna and disease associations using multigraph contrastive learning. Scientific Reports **15**(1), 230 (2025)
15. Sweeney, B., Petrov, A., Ribas, C., et al.: RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. Nucleic Acids Research **49**(D1), D212–D220 (Oct 2020). https://doi.org/http://dx.doi.org/10.1093/nar/gkaa921
16. Tian, X., Shen, L., Wang, Z., Zhou, L., Peng, L.: A novel lncrna–protein interaction prediction method based on deep forest with cascade forest structure. Scientific reports **11**(1), 18881 (2021)
17. Torgano, F., Soto Gomez, M., Zignani, M., Gliozzo, J., Cavalleri, E., Mesiti, M., Casiraghi, E., Valentini, G.: Rna knowledge-graph analysis through homogeneous embedding methods. Bioinformatics Advances **5**(1), vbaf109 (05 2025). https://doi.org/10.1093/bioadv/vbaf109
18. Umu, S.U., Gardner, P.P.: A comprehensive benchmark of rna–rna interaction prediction tools for all domains of life. Bioinformatics **33**(7), 988–996 (2017)
19. Valentini, G., Malchiodi, D., Gliozzo, J., Mesiti, M., Soto-Gomez, M., Cabri, A., Reese, J., Casiraghi, E., Robinson, P.: The promises of large language models for protein design and modeling. Frontiers in Bioinformatics **3**(1304099) (2023)
20. Wang, N., Bian, J., Li, Y., et al.: Multi-purpose RNA language modelling with motif-aware pretraining and type-guided fine-tuning. Nat Mach Intell **6**, 548–557 (2024). https://doi.org/http://dx.doi.org/10.1093/nar/gkaa921
21. Winkle, M., El-Daly, S.M., Fabbri, M., Calin, G.A.: Noncoding rna therapeutics — challenges and potential solutions. Nature Reviews Drug Discovery **20**(8), 629–651 (Jun 2021). https://doi.org/http://dx.doi.org/10.1038/s41573-021-00219-z
22. Zhao, Y., Oono, K., Takizawa, H., Kotera, M.: Generrna: A generative pre-trained language model for de novo rna design. PLoS One **19**(10), e0310814 (2024)