From Base Pairs to Functions: Rich RNA Representations via Multimodal Language Modeling

Joel Bonnie,[1,2] Tin Vlašić,[1] and Mile Šikić[1,3]

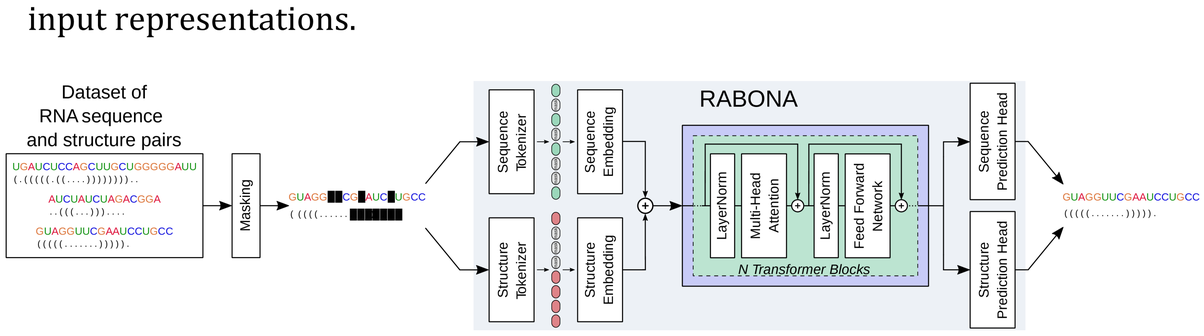
[1] Genome Institute of Singapore, A*STAR, Singapore
[2] University of British Columbia, Canada
[3] University of Zagreb, Croatia

Motivation

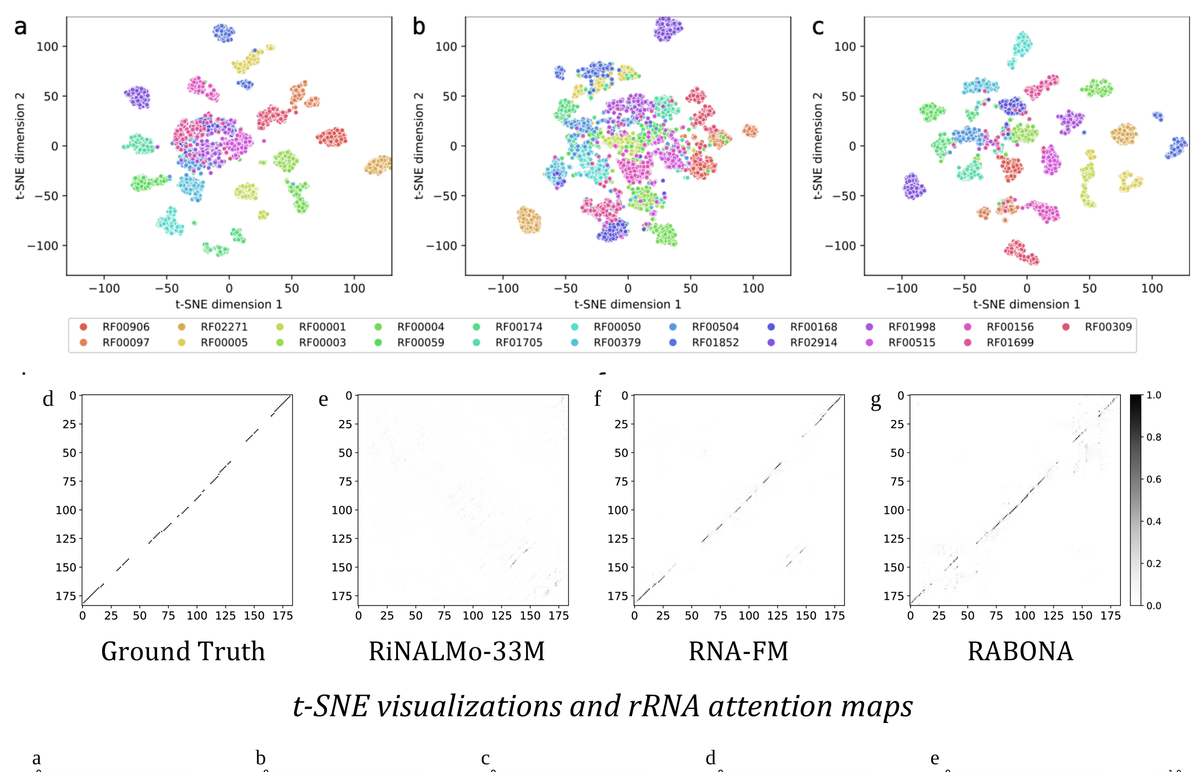
In recent years, RNA foundation models have become essential for predictive and generative tasks. However, most current RNA models are trained exclusively on sequence data, with representations that often fail to capture strong structural information. This inhibits predictive performances on downstream tasks, necessitating more robust representations. We present RABONA, a multimodal RNA foundation model that integrates both primary and secondary structural knowledge to deliver richer and more informative representations.

Model

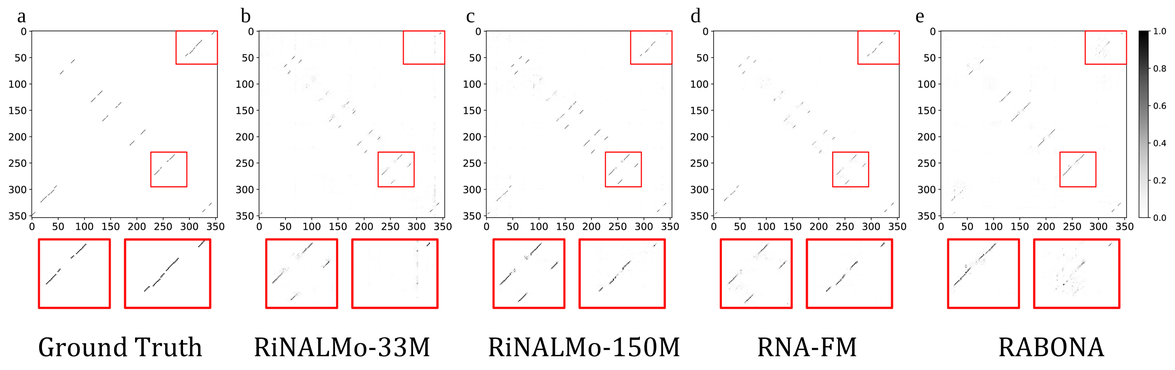
RABONA is a 33.5M parameter language model with 12 Transformer blocks. Pre-training employed masked language modelling (MLM) over sequence-structure pairs, with each modality masked independently. The pre-training dataset consisted of ~1M sequences. Following pre-training, RABONA's embeddings can be harnessed in several diverse downstream tasks as RNA input representations.



Pre-training RABONA



t-SNE visualizations and rRNA attention maps



tmRNA attention maps

Acknowledgements

This work was supported in part by the Agency for Science, Technology and Research (A*STAR) through the Singapore International Pre-Graduate Award (SIPGA).

Results

We evaluated RABONA's embeddings under two training paradigms: fine-tuning and linear probing. Comparisons were made against other state-of-the-art foundational models such as RNA-FM, RiNALMo-33M, and RiNALMo-150M. Across downstream tasks such as OpenVaccine and reactivity predictions, RABONA's embeddings shine, outperforming not only models of similar scale but also substantially larger models.

Table 1: ncRNA Classification - Finetuning

MODEL	ACC_0 ↑	ACC_200 ↑
RABONA	0.976	0.979
RABONA_MASK	0.975	0.974
RiNALMo-33M	0.980	0.977
RiNALMo-150M	0.982	0.985
RNA-FM	0.919	0.951

Table 2: ncRNA Classification - Linear Probing

MODEL	ACC_0 ↑	ACC_200 ↑
RABONA	0.909	0.488
RABONA_MASK	0.921	0.537
RiNALMo-33M	0.861	0.541
RiNALMo-150M	0.896	0.541
RNA-FM	0.791	0.462

Table 3: Reactivity - Finetuning

MODEL	RMSE ↓	MAE ↓
RABONA	0.413	0.255
RABONA_MASK	0.427	0.262
RiNALMo-33M	0.438	0.271
RiNALMo-150M	0.427	0.264
RNA-FM	0.471	0.287
One-Hot Sequence	0.551	0.336
One-Hot Sequence + SS	0.538	0.324

Reactivity Error Decrease vs RiNALMo-33M

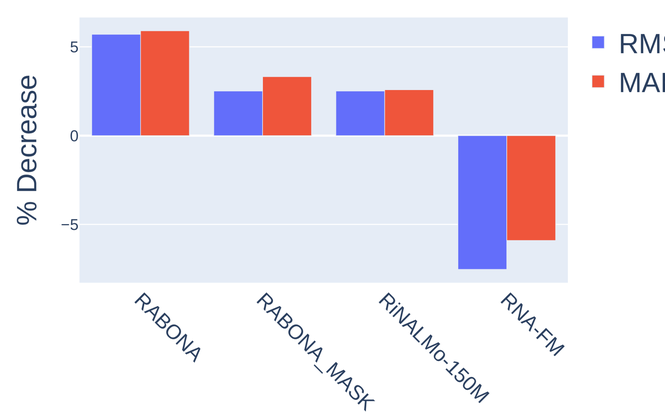


Table 4: OpenVaccine - Finetuning

MODEL	RMSE ↓	MAE ↓
RABONA	0.306	0.169
RABONA_MASK	0.376	0.215
RABONA_FROZEN	0.365	0.198
RiNALMo-33M	0.395	0.230
RiNALMo-150M	0.375	0.215
RNA-FM	0.408	0.231
One-Hot Sequence	0.518	0.304
One-Hot Sequence + SS	0.468	0.263

OpenVaccine Error Decrease vs RiNALMo-33M

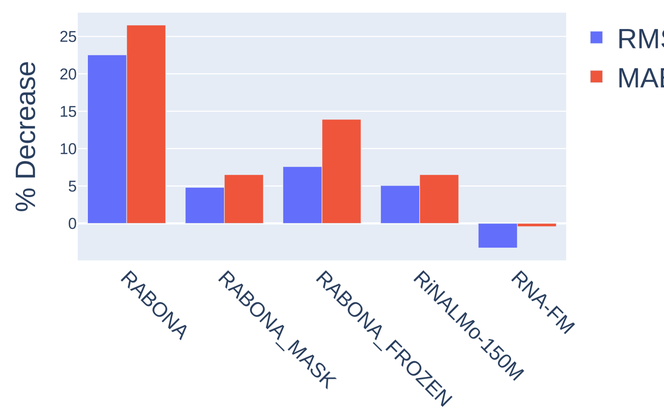


Table 5: Ribosome Loading - Finetuning

MODEL	R^2 ↑	MAE ↓
RABONA	0.769	0.409
RABONA_MASK	0.772	0.406
RiNALMo-33M	0.811	0.377
RiNALMo-150M	0.844	0.342
RNA-FM	0.719	0.455

References

1. R. J. Penić, T. Vlašić, R. G. Huber, Y. Wan, and M. Šikić. RiNALMo: General-purpose RNA language models can generalize well on structure prediction tasks. Nature Communications, 16(1):5671, 2025.
2. J. Chen, Z. Hu, S. Sun, Q. Tan, Y. Wang, Q. Yu, L. Zong, L. Hong, J. Xiao, T. Shen, et al. Interpretable RNA foundation model from unannotated data for highly accurate RNA structure and function predictions. arXiv preprint arXiv:2204.00300, 2022.
3. R. Lorenz, S. H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L.Hofacker. ViennaRNA Package 2.0. Algorithms for molecular biology, 6(1):26, 2011.