Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- [ ] n/a
- [ ] Confirmed
- [ ] The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- [ ] A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- [ ] The statistical test(s) used AND whether they are one- or two-sided
  - *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- [ ] A description of all covariates tested
- [ ] A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- [ ] A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- [ ] For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - *Give P values as exact values whenever suitable.*
- [ ] For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- [ ] For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- [ ] Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

| Data collection | cmbuild, cmmcalibrate, cmscan and cmssearch from INFERNAL 1.1.3, cd-hit-est from CD-HIT 4.8.1, Infernal 1.1.4, nhmmer from HMMER 3.3, HH-suite 2.0.15 |

| Data analysis | numpy=1.21.2, an open-source python package for numerical calculations. pytorch=1.10.2, an open-source deep learning framework in python, pandas>=1.3.1, an open-source python package for data analysis. matplotlib>=3.4, an open-source python package for visualization. sci-kit-learn>=0.24, an open-source python package for machine learning. scipy=1.7.1, an open-source python package for mathematics, science, and engineering. bioalign=1.79, an open-source python package for biological computation. pytorch-ignite=0.4.6, an open-source python package for training and evaluating neural networks in PyTorch flexibly and transparently. openmm=7.7, a high-performance toolkit for molecular simulation including AMBER relaxation. tensorboard=2.6.0, an open-source python package for visualizing training process. RNA-FM, our custom code (open source) for (https://github.com/mi4bio/RNA-FM) pretrained, rMSA (https://github.com/pyrelab/rMSA), an open source code for RNA MSA construction. RhoFold, our primary model code (open source) for (https://github.com/RFold/RhoFold) structure prediction. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data. All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data used in our work were obtained from related public datasets. We obtained all the training RNA 3D structures using the data list arranged by BGSU RNA Representative Sets (Ver.04-13.2022) [http://rna.bgsu.edu/rna3dhub/nrlist/release/3.126], all train and test data by BGSU RNA Representative Sets (Ver.02-15.2023) [http://rna.bgsu.edu/rna3dhub/nrlist/release/3.270] and downloaded them from Protein Data Bank [https://www.rcsb.org].

For RNA-FM pre-training, we downloaded the unannotated RNA sequences from RNAcentral [https://rnacentral.org]. For RNA MSA, we built the database upon Nucleotide database [ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nt.gz], Rfam [https://rfam.xfam.org] and RNAcentral [https://rnacentral.org] and use RMFA (https://github.com/pylelab/RMFA) for searching and construction tools.

For RNA-Puzzles, we downloaded native structures and submissions of other methods from http://www.mapuzzles.org/results/ and https://github.com/RNA-Puzzles/standardized_dataset. Similarly, CASP15 data is downloaded via https://predictioncenter.org/casp15/index.cgi

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

**Sample size**
We use 16,349 single chain RNA 3D structures extracted from 4213 RNA 3D structures in the PDB database for training. RNA-FM is trained on 23,735,169 sequences extracted from the RNAcentral database in an unsupervised manner to generate rich information to benefit structural prediction. We used 2,4183 RNA secondary structure data in bprna90-1m for self-distillation training.

**Data exclusions**
All data were used following previous researches, no exclusion was done prior to analysis.

**Replication**
The performance of our model could be reproduced, and we also offer codes, on-line service, and tutorials for the key downstream tasks.

**Randomization**
There are 3 cross-validation experiments in our tests. 1. 10-fold: After obtain 782 sequence clusters from CD-HIT, we randomly masked 80 sequence clusters for validation, leaving 702 sequence clusters for training. 2. cross-type/cross-fam: RNA types and families are obtained via Rfam. We then randomly selected an RNA type/fam and masked all structures (structure number depending on the RNA type) associated that type/fam, trained the model using the remaining types/families, and evaluated the model on the masked type/fam.

**Blinding**
Not applicable. All experiments, data collection, and analysis are performed in an unbiased way.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.
| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a                             | n/a     |
| ✗ Involved in the study         | Involved in the study |
| ✗ Antibodies                    | ✗ ChIP-seq |
| ✗ Eukaryotic cell lines         | ✗ Flow cytometry |
| ✗ Palaeontology and archaeology | ✗ MRI-based neuroimaging |
| ✗ Animals and other organisms   |         |
| ✗ Human research participants   |         |
| ✗ Clinical data                 |         |
| ✗ Dual use research of concern  |         |