nature portfolio

Corresponding author(s): Qiwei Ye, Yaqing Zhang

Last updated by author(s): Apr 28, 2025

# Reporting Summary

Nature Portfolio 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 Portfolio 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<br>*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<br>*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 | All scientific datasets used for generating training and evaluation inputs are freely accessible from public sources (see Data section below). No additional data was collected. |
| Data analysis | Data analysis was performed using Python v3.8.5, Numpy v1.24.4, Scipy v1.6.0, Scikit-learn v0.24.1, and Matplotlib v3.4.2. All data, model source code, and pre-trained models are freely available at the CodeOcean (https://codeocean.com/capsule/1240603/tree/v1) as well as our webserver (https://www.ultrarnalab.com/). |

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All scientific datasets used for generating inputs for training and evaluation are freely available from public sources. UltraSelex HTS raw data supporting the findings

of this study are available for academic use in the online repository (https://www.ultrarnalab.com). A mirror data repository can be accessed on NCBI Sequence Read Archive under BioProject (PRJNA1216547). Other HTS data utilized in this study are available on Zenodo (https://doi.org/10.5281/zenodo.15294875). Twelve benchmark SELEX datasets, CLIP datasets, m6A methylation datasets were obtained from corresponding published articles, as detailed in Supplementary Table 1-11. Millions of 3'-untranslated RNA species across 22 human tissues were acquired from the APASdb database (http://genome.bucm.edu.cn/utr/), with the last access in October 2023. All data underwent initial processing post-collection to ensure quality and consistency. These data, including both raw and processed datasets, are publicly available for download in the UltraGen panel, part 2 Dataset, at UltraRNALab (www.ultrarnalab.com).

Pathogenic human RNA virus genomes, including their 3'UTR annotations, were retrieved from the NCBI Viruses RefSeq release (https://ftp.ncbi.nlm.nih.gov/refseq/release/viral/). SARS-CoV-2 variant genomes were obtained from the NCBI Virus database (https://www.ncbi.nlm.nih.gov/labs/virus/), collected between the end of 2019 and 2023. These data have been compiled in the Supplementary Materials.

## Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

| | |
|---|---|
| Reporting on sex and gender | N/A |
| Population characteristics | N/A |
| Recruitment | N/A |
| Ethics oversight | N/A |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# 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 | No specific calculation for the sample size. |
| Data exclusions | No data were excluded without mention. |
| Replication | Code and method details were meticulously reviewed for completeness and replicability. |
| Randomization | No specific sample randomization for this study. |
| Blinding | No specific use for blinding on this study. |

# 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

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

### Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |