RASP: an atlas of transcriptome-wide RNA secondary structure probing data

Pan Li [ORCID][†], Xiaolin Zhou[†], Kui Xu [ORCID] and Qiangfeng Cliff Zhang [ORCID][*]

MOE Key Laboratory of Bioinformatics, Beijing Advanced Innovation Center for Structural Biology & Frontier Research Center for Biological Structure, Center for Synthetic and Systems Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing, 100084, China

ABSTRACT

RNA molecules fold into complex structures that are important across many biological processes. Recent technological developments have enabled transcriptome-wide probing of RNA secondary structure using nucleases and chemical modifiers. These approaches have been widely applied to capture RNA secondary structure in many studies, but gathering and presenting such data from very different technologies in a comprehensive and accessible way has been challenging. Existing RNA structure probing databases usually focus on low-throughput or very specific datasets. Here, we present a comprehensive RNA structure probing database called RASP (RNA Atlas of Structure Probing) by collecting 161 deduplicated transcriptome-wide RNA secondary structure probing datasets from 38 papers. RASP covers 18 species across animals, plants, bacteria, fungi, and also viruses, and categorizes 18 experimental methods including DMS-seq, SHAPE-Seq, SHAPE-MaP, and icSHAPE, etc. Specially, RASP curates the up-to-date datasets of several RNA secondary structure probing studies for the RNA genome of SARS-CoV-2, the RNA virus that caused the ongoing COVID-19 pandemic. RASP also provides a user-friendly interface to query, browse, and visualize RNA structure profiles, offering a shortcut to accessing RNA secondary structures grounded in experimental data. The database is freely available at http://rasp.zhanglab.net.

INTRODUCTION

RNA is critical across biological processes and a range of cellular mechanisms act upon it to carefully regulate and refine gene expression (1). The specific secondary structures formed by non-coding RNAs (ncRNAs) are central to their regulation and functions (2–4). Recent studies have also found that mRNA secondary structures influence gene transcription, translation and decay (5). The secondary structure of many RNA viruses also have important functions. For example, the 3′UTRs of Flaviviruses produce highly structured noncoding RNAs that are resistant to host nucleases (6). As more and more functions for RNA secondary structure are discovered, deciphering the structures themselves has become a priority.

During the past few decades, many computational methods predicting RNA secondary structure have been developed (7–9). These methods only work well on shorter RNA sequences, and has been a major source of information for RNA structure studies (10). However, computational prediction and modeling usually cannot take into consideration the complex cellular environments and thus lack of the resolution for *in vivo* studies. Small molecule approaches have been long developed to quantitatively measure RNA conformation (11). In the last few years, thanks to development of high-throughput sequencing technology, RNA structure probing has entered the omics era, allowing simultaneous, transcriptome-wide measurement of RNA (leads to the results of the so-called 'structuromes'), both in test tubes and in cellular conditions (12–14).

The principles underlying RNA probing largely fall into two categories: nuclease cleavage and small molecule-based probing (Figure 1). RNase P1, S1 and RNase V1 nucleases cut the single and double-stranded RNA respectively (15,16). During reverse transcription, cDNA synthesis stops at the cleavage site, revealing information about single-stranded and double-stranded nucleotides upon high-throughput sequencing. Small chemicals such as 1M7, DMS, N_3-kethoxal and NAI-N_3 can be used to specifically probe single-stranded RNA bases (17–22). Upon reverse transcription, the RT enzymes terminate at the modified site (17,18,20,21) or mis-incorporate nucleotides resulting from the chemical modification (19,22). By normalizing RT stop values or mutation rates, a structural score can be assigned to each base, measuring the likelihood of that base being single-stranded or double-stranded.

*To whom correspondence should be addressed. Tel: +86 10 627 96439; Email: qczhang@mail.tsinghua.edu.cn
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
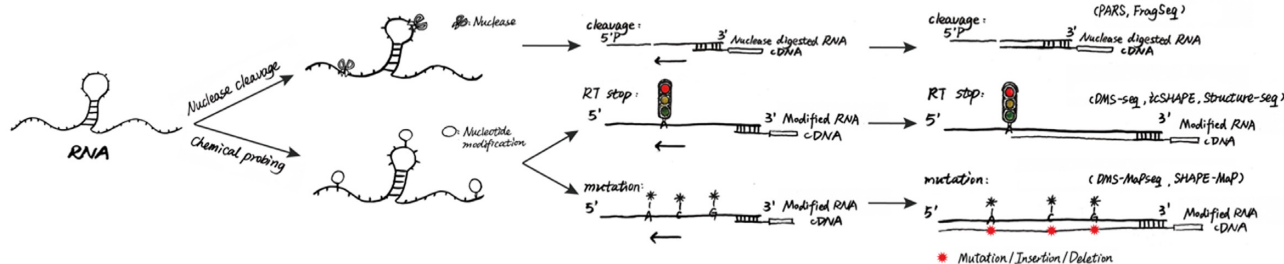
Figure 1. Schematic for RNA probing technology. Approaches for RNA secondary structure probing are based on either RNA nuclease digestion or small chemical modification.

RNA secondary structure databases such as RMDB (23) and RSVdb (24) have been developed, but they normally focus on specific datasets of a very limited coverage. For example, RMDB contains diverse RNA structural mapping experiments, but focuses on low-throughput experiments (23). RSVdb collects RNA structure data, is limited to DMS reagent-based datasets (24). Given the increasing volume of experimental RNA structure data (especially those using high-throughput technologies), and their broad relevance to biological processes, a comprehensive database is highly desired.

Here we describe a database, RASP, that collects 161 datasets from 38 papers (Table 1). RASP spans, categorizes and organizes 18 species across animals, plants, bacteria, fungi and viruses and 18 different experimental methods. RASP contains almost all currently published transcriptome-scale data, including the most recent studies that probed the RNA secondary structures of the genome of the SARS-CoV-2 RNA virus, with technologies varying from PARS (16,25), DMS-seq (17), Structure-seq (18), to SHAPE (26), icSHAPE (21) and SHAPE-MaP (27) etc. RASP provides a user-friendly interface to query, browse and download data. In addition, RASP implements analytical functions such as multiple sequence alignment, along with RNA secondary structure prediction and visualization. This dataset will greatly expand accessibility and cross-comparison of RNA secondary structures, empowering relevant researches across fields.

DATA COLLECTION AND PROCESSING

We retrieved published papers containing high-throughput RNA structure probing data from PubMed. For experiments containing multiple conditions, we collected datasets for each condition separately. We classified datasets according by species and experimental technology used (Table 1). Detailed information including species, cell line, reagents used in the experiment, and publication information was also collected, and used for classification during construction of the RASP database (Supplementary Table S1).

In principle, all RNA secondary structure probing methods generate a structure score (called with different names) to provide a measure of the pairing probability of each nucleotide (14). We directly downloaded the structural scores from publications and integrated these data into RASP if the processed structural score is provided (Figure 2, right). If not, we downloaded raw data and used the same pipeline as the original paper to calculate the scores (Figure 2, left). This data processing information, along with the transcript numbers with structural scores is listed in Supplementary Table S1.

DATABASE CONTENT AND USAGE

RASP provides a rich number of approaches to interact with the data stored in the database, including search, browse, download and functions including structure prediction, multiple sequences alignment and cross-dataset comparison (Figure 2, bottom).

Search interface for retrieving gene and sequence

RASP provides two kinds of inquiry modes: 'search gene' and 'search sequence' (Figure 3A). In the gene-based inquiry mode, a user first selects one or more species, then inputs a gene symbol (such as *GAPDH*) or *ensembl* gene ID or transcript ID of interest. Clicking 'Search Gene' gives the user a match list (Figure 3B). Each match item includes the organism name, the genome location, the match string (gene or transcript), the gene symbol, a transcript list and a genome browser link. The user has two options here: click any specific transcript to visualize the transcript sequence and structure score (see 'Visualize sequence and structure data for transcript' section below) or click 'Go' to visualize in genome browser (see 'Genome browser' section). In the sequence-based inquiry mode, a user can select a species and input a DNA or RNA sequence, and click the 'Search Sequence', to search for the inquiry sequence in the genome by using *blastn* (28). The action will return the user a hit list (Figure 3C). Each hit item includes the organism name, the genome location, the E-value, the match between query sequence and target sequence, the gene symbol, a transcript list and a genome browser link. As described above, the user can visualize data in the genome or transcript by clicking the 'Go' button or any transcript.

Genome browser

RASP integrates *Jbrowse* (29) which allows users to visualize and compare structure scores in the genome (Figure 4A). On the 'Browse' page, users select a species and input the gene symbol or *ensembl* gene ID of interest. Clicking 'Go' refreshes the browser to display the gene region. There are two options for users to load structure data: (i) a 'Click

Table 1. The data collection table. All the papers are classified with species. Please refer to Supplementary Table S1 for for more details

| Category	Species	References
Animals	Human	(2014, Nature, DMS-seq, 6 samples) (2014, Nature, PARS, 5 samples) (2016, Science, DMS-seq & SHAPE, 4 samples) (2016, Cell, icSHAPE, 2 samples) (2017, Cell, DIM-2P-seq, 1 sample) (2017, NatureMethods, DMS-MaPseq, 1 sample) (2019, Nature Structural & Molecular Biology, icSHAPE, 6 samples) (2020, Nature Chemical Biology, Keth-seq, 4 samples)
	Mouse	(2016, Science, DMS-seq & SHAPE, 11 samples) (2010, Nature Methods, FragSeq, 6 samples) (2014, Genome Biology, CIRS-seq, 1 sample) (2015, Nature, icSHAPE, 3 samples) (2015, Biochemistry, SHAPE-MaP, 2 samples) (2019, Nature Structural & Molecular Biology, icSHAPE, 6 samples) (2020, Nature Chemical Biology, Keth-seq, 6 smaples)
	Zebrafish	(2018, Nature Structural & Molecular Biology, DMS-seq, 7 samples) (2020, Genome Biology, icSHAPE, 8 samples)
Plants	Rice	(2017, Nucleic Acids Research, Structure-seq2, 4 samples)
	Arabidopsis thaliana	(2014, Nature, Structure-seq, 1 sample)
Bacteria & Fungi	E. coli	(2016, Science, DMS-seq & SHAPE, 5 samples) (2017, eLife, DMS-seq, 7 samples) (2018, Cell, SHAPE-MaP, 3 smaples)
	Yeast	(2010, Nature, PARS, 1 sample) (2012, Molecular Cell, PARS, 5 samples) (2014, Nature, DMS-Seq, 3 samples) (2016, Science, DMS-seq & SHAPE, 5 samples) (2017, Nature Methods, DMS-MaPseq, 6 samples) (2018, Nature, DMS-MaPseq, 2 smaples) (2016, Science, DMS-seq & SHAPE, 5 samples) (2016, Science, DMS-seq & SHAPE, 4 samples)
	P. putida	
	Synechococcus	
	Y. pseudotuberculosis	(2020, Nucleic Acids Research, Lead-seq, 2 samples)
Virus	HIV	(2009, Nature, SHAPE, 1 sample) (2014, Nature Methods, SHAPE-MaP, 3 samples) (2015, Plos Computational Biology, SHAPE-MaP, 1 sample)
	Dengue	(2019, Nature Communications, NAI-MaP, 1 sample) (2018, PNAS, SHAPE-MaP, 6 samples)
	Zika	(2018, Cell Host & Microbe, icSHAPE, 2 sample) (2019, Nature Communications, NAI-MaP, 1 sample)
	HCV	(2015, PNAS, SHAPE-MaP, 1 sample) (2016, Molecular Cell, SHAPE, 1 sample)
	STMV	(2013, BioChemistry, SHAPE, 2 samples)
	CMV	(2018, Nucleic Acids Research, SHAPE-seq, 2 samples)
	IAV	(2019, Nucleic Acids Research, RAPiD-MaPseq, 2 samples)
	SARS-CoV-2	(2020, bioRxiv, SHAPE-MaP, 1 samples) (2020, bioRxiv, icSHAPE, 1 samples) (2020, bioRxiv, SHAPE-MaP, 2 samples)
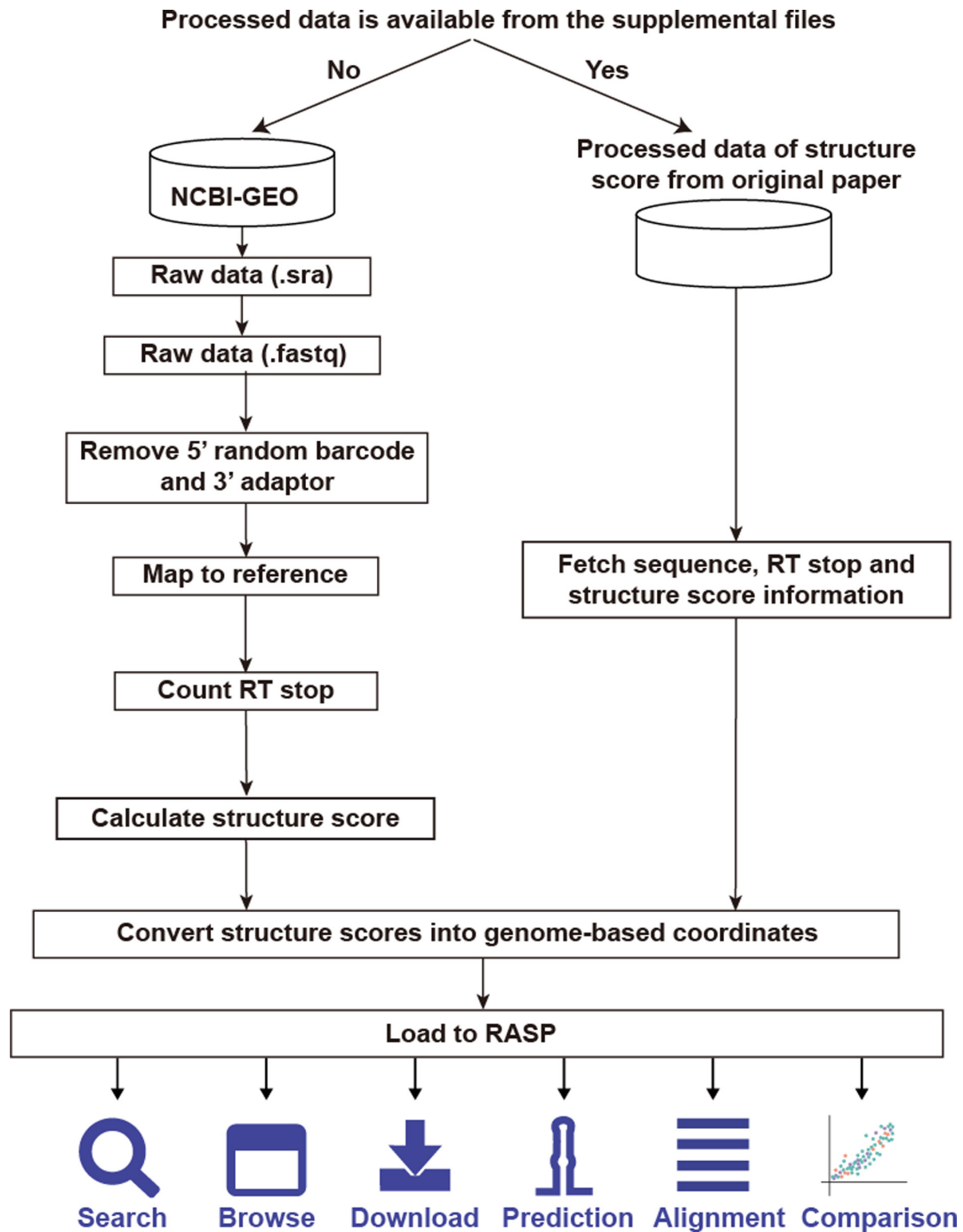
Figure 2. Flowchart of RASP data collection.

here to select datasets' button to expand the selection panel; (ii) a 'Select tracks' button on the top left of the browser region to shift the browser selection panel. Users can filter the data based on all kinds of criteria including the name of the technology, the reagents used in the experiment, the journal where the paper was published, specific experimental conditions, the cell line, the strand, the experimental principle, etc. Users can filter out those tracks without structural data coverage in the genomic region by clicking the 'Only show tracks with structural data coverage' checkbox.

Jbrowse also allows users to download a small amount of data by highlighting a region of interest, click 'save track data' in the track menu to save the structure score as a *bedGraph* file. Users can also rearrange tracks by dragging the track labels, and visualize their custom tracks by uploading local files or providing hyperlinks. *Jbrowse* also provides convenient ways for structure score comparison. Figure 4B shows two examples of the structure scores of the human *GAPDH* and yeast *RPL32* transcripts obtained by different technologies. Users can also eas-
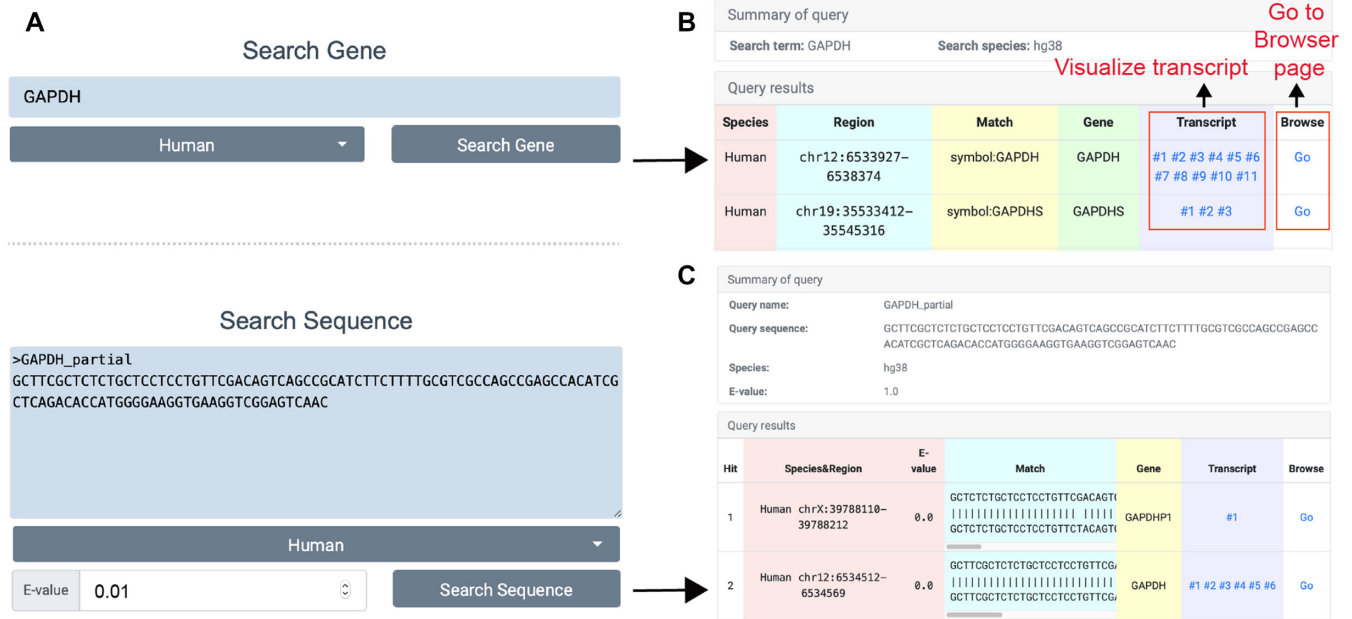
Figure 3. Gene-based and sequence-based inquiry mode. (**A**) Search interface for retrieving gene and sequence. (**B, C**) Result interface for gene search (**B**) and sequence search (**C**).

ily compare the structural differences of the same RNA under different conditions, which may help to visually identify the influence of different conditions on RNA structure.

Visualize sequence and structure data for a transcript

By clicking the transcript on the search results page (see 'Search interface for retrieving gene and sequence' section) (Figure 3B, C), users can skip to a new page to visualize the sequence and structure score (Figure 4C). There are five panels on this page. 'Summary of transcript' panel displays basic transcript information including genome location, the transcript biotype, etc. The 'Selection' panel contains buttons that allow selection of sequence regions and performs various operations on this area including copy, download, alignment and structure prediction. The 'Sequence' panel displays the full sequence of transcript, with the UTR highlighted in orange. Users can click any two bases to select a region and copy or download this region. To search a subsequence, users input the subsequence to the text area and click the 'Search' button. If successful, the subsequence will be selected. 'Load probing data' panel provides the function of loading probing data. The structure score is displayed with base-specific colors in the 'Sequence' panel. High structure scores are indicated in red, and low scores is in blue. The 'Compare probing data' panel draws scatter plots that compare structure scores between any two selected datasets. The Pearson correlation efficient is shown. If the user selects a region, only the data in the region will be displayed. The 'statistics' panel stores statistics and draw a distribution plot for the structure score of the full transcript or selected region.

Multiple sequences alignment and score comparison

If a user is interested in the structure of homologous sequences, they can compare the structure score of homologous sequences through the 'Alignment' page. Multiple sequences and corresponding structure scores can be input (Figure 5A), and upon clicking 'Submit', the RASP server uses *muscle* (30) to align the sequences and return a new page including the aligned sequences and the aligned structure scores (Figure 5B). The user can then click the 'Download' button to download the aligned sequences and structure scores. Users can also select a region of a transcript and click 'Add' button in the transcript summary page (Figure 4C) to save the data, and then directly load the data in the 'Alignment' page.

Structure prediction

The 'Predict' page allows users to fill in the sequence and structure scores in text boxes, and provide the *intercept* and *slope* parameters (Figure 5C). The structure score can be converted into pseudo free energy through the following formula (31):

$$\Delta G_{\text{score}}(i) = \text{slope} * \ln[\text{score}(i) + 1] + \text{intercept}$$

Deigan *et al.* used grid search to fit slope and intercept based on the secondary structure and SHAPE scores of *E. coli* rRNA (31). we took the values they obtained (slope = 1.8, intercept = −0.6) as default parameters, but users can adjust these parameters to change the weight of the structure score. Users can also select a region of a transcript and click the 'Predict' button in the transcript summary page (Figure 4C) to directly jump to the 'Predict' page for structural prediction.

Figure 4. Structural score visualization. (**A**) Browser page. (**B**) Structure scores of the *GAPDH* and *RPL32* transcripts. (**C**) Visualize the structure score on a transcript with nucleotides colored by the scores. Users can select a sequence region to copy scores, download data, or predict structure. In 'Compare probing data' panel, users can compare two datasets by a scatter plot. The structure score distribution is displayed in 'Statistics' panel.

Clicking the 'Submit' button returns a page with the predicted structures (Figure 5D). 'Summary of query information' panel contains input sequences, constraints and parameters. 'Prediction results' panel present a list of predicted structures ranked by free energy. Users can visualize structure with *forna* (32) by clicking the 'Go' button, or copying a *java* command to visualize the structure with *VARNA* (33).

Download

The Download page provides genome reference sequence files, annotation files and structure score files. Processed structure data are saved in the *bigWig* format and *bed* format. The *Bigwig* format is a binary file format with genomic

coordinates. *bigWig* files can be converted to text-format *bedGraph* files using the *bigWigToBedGraph* program from *UCSC* tools package (34). Users can refer to the 'Help' page for detailed operation steps. Users can also download the *bed* text-format files.

Special collection: RNA structure probing data of SARS-CoV-2

The recent outbreak COVID-19 has rapidly spread to the whole world and caused tremendous damage to our society and economy. COVID-19 is caused by a single-stranded, highly infectious RNA virus SARS-CoV-2. As previous studies on the other RNA viruses like HCV, HIV, Dengue virus, Zika virus have demonstrated that the secondary
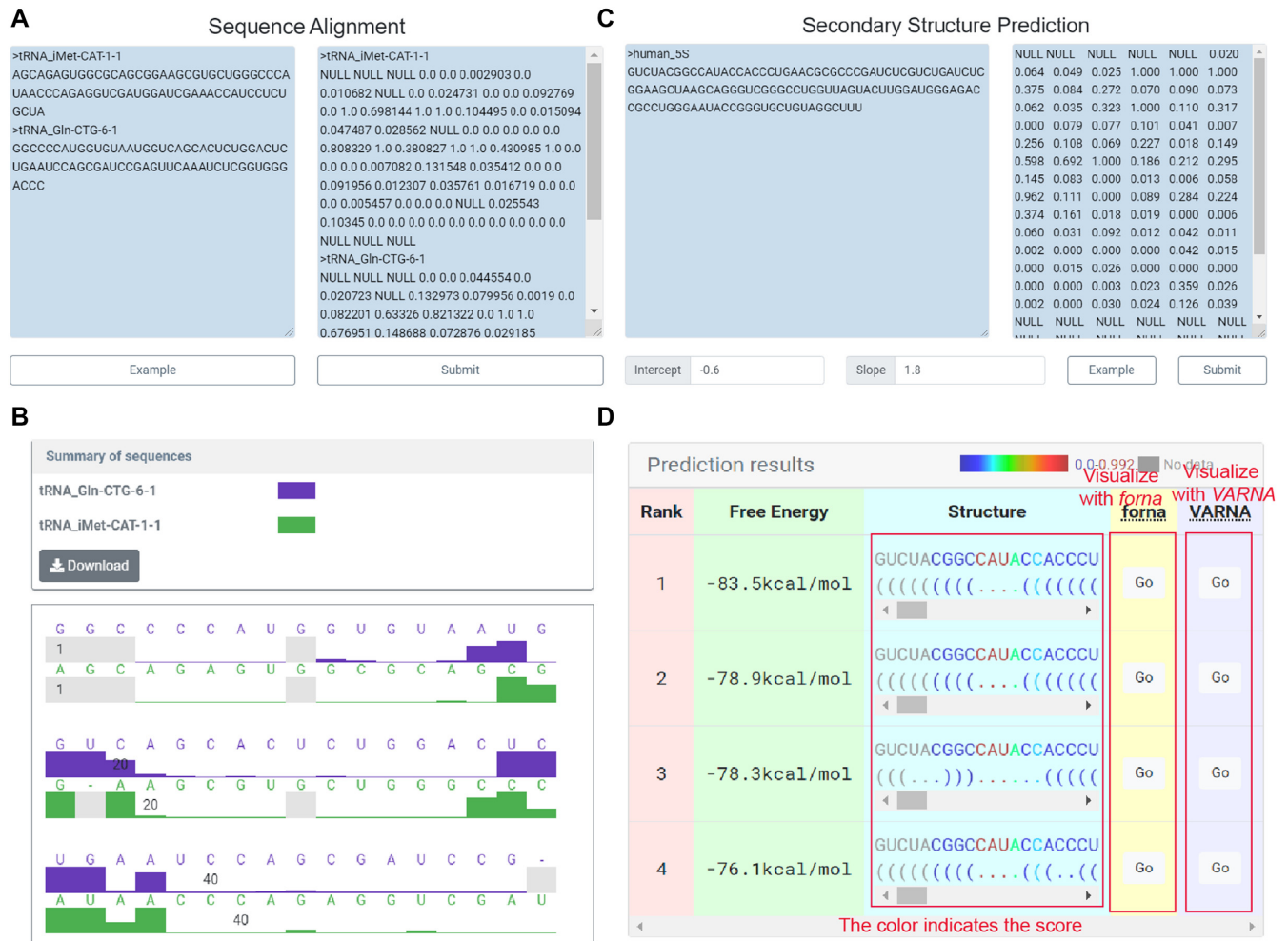
Figure 5. Sequence alignment and structure prediction. (A, B) The query (**A**) and result (**B**) interface of sequence alignment. (C, D) The query (**C**) and result (**D**) interface interface of structure prediction.

structures of their RNA genome are important for the life cycle of these viruses (26,35–37), much research efforts have been dedicated to determining the secondary structure of the SARS-CoV-2 RNA genome, by using RNA structure probing methods including SHAPE-MaP, DMS-MaPseq and icSHAPE (38–43).

Called on by the urgent need fighting against the on-going pandemic, we made a special effort for the collection of existing RNA secondary structure of the SARS-CoV-2 genome (Figure 6). We explored datasets associated with six manuscripts deposited on bioRxiv, including one study from our own laboratory (38) and found that two of the six studies disclosed their processed structural data (39,40). We have thus collected these data and integrated them into RASP. Users can easily visualize, compare and download the data through our database server. We have been actively monitoring the progress in the field and will continue to collect and update with the new results in the future.

DISCUSSION

High-throughput RNA secondary structure probing has raised a great amount of research interest (44), and large-

scale RNA secondary structure datasets are accumulating rapidly. These datasets are helpful for modeling RNA secondary structure, and analyzing correlations between structural and cellular activities including transcription rate, translation efficiency, etc. Based on the current need to comprehensively collect and process RNA probing data, we developed RASP that cover 161 manually curated RNA probing datasets from 18 species with 18 different experimental technologies. In contrast to existing databases, RASP provides a comprehensive data and analysis platform.

The current version of RASP mainly includes datasets of large-scale RNA secondary structure studies, integrated with enabling RNA structurome analysis tools. However, previous low-throughput experiments have also generated a lot of RNA structural data. For comprehensiveness, in the future RASP will include these low-throughput data, as well as results of studies focused on certain RNA targets such as the lncRNA, Hotair (45). We also aim to integrate more analysis tools, such as RNA covariation analysis, conserved RNA structure elements discovery, and multiple methods for RNA secondary structure prediction and modeling. We expect that RASP should greatly aid researchers, and accel-

Special: SARS-CoV-2 probing data

Datasets

Manuscript	Technology	Condition	Treatment reagent	Control sample	Cell line	Data accessibility	Load data
(Lei Sun, et al.)	icSHAPE	*in vivo / in vitro refolded*	NAI-N₃	DMSO	Huh7	RASP	☐
(Ilaria Manfredonia, et al.)	SHAPE-MaP	*in vitro refolded*	NAI	DMSO	VeroE6	INCARNATO lab	☑
(Nicholas C. Huston, et al.)	SHAPE-MaP	*in vivo*	NAI	DMSO	VeroE6	Github	☑
(Wes Sanders, et al.)	SHAPE-MaP	*in virion*	1M7	DMSO+Denatured	VeroE6	Not published	☐
(Tammy C. T. Lan, et al.)	DMS-MaPseq	*in vivo / in vitro*	DMS	Untreated	Vero cells	Not published	☐
(Christiane Iserman, et al.)	SHAPE-MaP	*in vitro transcribed fragments*	5NIA	DMSO	N/A	Not published	☐

Related tracks

Sequence ☑   Gene annotation ☑   RNA-Seq ☐   RNA-Seq splices ☐   Human Cov ☐   Bat Cov ☐

Genome browser

Figure 6. A special page for the secondary structure data of the SARS-CoV-2 RNA genome.

erate RNA structure research, as well as allow follow up on more data from current large datasets.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.
2. Glotz,C., Zwieb,C., Brimacombe,R., Edwards,K. and Kossel,H. (1981) Secondary structure of the large subunit ribosomal RNA from Escherichia coli, Zea mays chloroplast, and human and mouse mitochondrial ribosomes. *Nucleic Acids Res.*, **9**, 3287–3306.
3. Holley,R.W., Apgar,J., Everett,G.A., Madison,J.T., Marquisee,M., Merrill,S.H., Penswick,J.R. and Zamir,A. (1965) Structure of a ribonucleic acid. *Science*, **147**, 1462–1465.
4. Ganot,P., Caizergues-Ferrer,M. and Kiss,T. (1997) The family of box ACA small nucleolar RNAs is defined by an evolutionarily conserved secondary structure and ubiquitous sequence elements essential for RNA accumulation. *Genes Dev.*, **11**, 941–956.
5. Sun,L., Fazal,F.M., Li,P., Broughton,J.P., Lee,B., Tang,L., Huang,W., Kool,E.T., Chang,H.Y. and Zhang,Q.C. (2019) RNA structure maps across mammalian cellular compartments. *Nat. Struct. Mol. Biol.*, **26**, 322–330.
6. Pijlman,G.P., Funk,A., Kondratieva,N., Leung,J., Torres,S., van der Aa,L., Liu,W.J., Palmenberg,A.C., Shi,P.Y., Hall,R.A. *et al.* (2008) A highly structured, nuclease-resistant, noncoding RNA produced by flaviviruses is required for pathogenicity. *Cell Host Microbe*, **4**, 579–591.
7. Nussinov,R. and Jacobson,A.B. (1980) Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **77**, 6309–6313.
8. Zuker,M. and Stiegler,P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.
9. Rivas,E. (2013) The four ingredients of single-sequence RNA secondary structure prediction. A unifying perspective. *RNA Biol*, **10**, 1185–1196.
10. Doshi,K.J., Cannone,J.J., Cobaugh,C.W. and Gutell,R.R. (2004) Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics*, **5**, 105.

11. Kubota,M., Tran,C. and Spitale,R.C. (2015) Progress and challenges for chemical probing of RNA structure inside living cells. *Nat. Chem. Biol.*, **11**, 933–941.

12. Strobel,E.J., Yu,A.M. and Lucks,J.B. (2018) High-throughput determination of RNA structures. *Nat. Rev. Genet.*, **19**, 615–634.

13. Wan,Y., Kertesz,M., Spitale,R.C., Segal,E. and Chang,H.Y. (2011) Understanding the transcriptome through RNA structure. *Nat. Rev. Genet.*, **12**, 641–655.

14. Piao,M., Sun,L. and Zhang,Q.C. (2017) RNA regulations and functions decoded by transcriptome-wide RNA structure probing. *Genomics Proteomics Bioinformatics*, **15**, 267–278.

15. Underwood,J.G., Uzilov,A.V., Katzman,S., Onodera,C.S., Mainzer,J.E., Mathews,D.H., Lowe,T.M., Salama,S.R. and Haussler,D. (2010) FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat. Methods*, **7**, 995–1001.

16. Kertesz,M., Wan,Y., Mazor,E., Rinn,J.L., Nutter,R.C., Chang,H.Y. and Segal,E. (2010) Genome-wide measurement of RNA secondary structure in yeast. *Nature*, **467**, 103–107.

17. Rouskin,S., Zubradt,M., Washietl,S., Kellis,M. and Weissman,J.S. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*, **505**, 701–705.

18. Ding,Y., Tang,Y., Kwok,C.K., Zhang,Y., Bevilacqua,P.C. and Assmann,S.M. (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature*, **505**, 696–700.

19. Mustoe,A.M., Busan,S., Rice,G.M., Hajdin,C.E., Peterson,B.K., Ruda,V.M., Kubica,N., Nutiu,R., Baryza,J.L. and Weeks,K.M. (2018) Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell*, **173**, 181–195.

20. Weng,X., Gong,J., Chen,Y., Wu,T., Wang,F., Yang,S., Yuan,Y., Luo,G., Chen,K., Hu,L. *et al.* (2020) Keth-seq for transcriptome-wide RNA structure mapping. *Nat. Chem. Biol.*, **16**, 489–492.

21. Spitale,R.C., Flynn,R.A., Zhang,Q.C., Crisalli,P., Lee,B., Jung,J.W., Kuchelmeister,H.Y., Batista,P.J., Torre,E.A., Kool,E.T. *et al.* (2015) Structural imprints in vivo decode RNA regulatory mechanisms. *Nature*, **519**, 486–490.

22. Zubradt,M., Gupta,P., Persad,S., Lambowitz,A.M., Weissman,J.S. and Rouskin,S. (2017) DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods*, **14**, 75–82.

23. Cordero,P., Lucks,J.B. and Das,R. (2012) An RNA mapping dataBase for curating RNA structure mapping experiments. *Bioinformatics*, **28**, 3006–3008.

24. Yu,H., Zhang,Y., Sun,Q., Gao,H. and Tao,S. (2020) RSVdb: a comprehensive database of transcriptome RNA structure. *Brief. Bioinform.*, https://doi.org/10.1093/bib/bbaa071.

25. Wan,Y., Qu,K., Zhang,Q.C., Flynn,R.A., Manor,O., Ouyang,Z., Zhang,J., Spitale,R.C., Snyder,M.P., Segal,E. *et al.* (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature*, **505**, 706–709.

26. Pirakitikulr,N., Kohlway,A., Lindenbach,B.D. and Pyle,A.M. (2016) The coding region of the HCV genome contains a network of regulatory RNA structures. *Mol. Cell*, **62**, 111–120.

27. Mauger,D.M., Golden,M., Yamane,D., Williford,S., Lemon,S.M., Martin,D.P. and Weeks,K.M. (2015) Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, 3692–3697.

28. Morgulis,A., Coulouris,G., Raytselis,Y., Madden,T.L., Agarwala,R. and Schaffer,A.A. (2008) Database indexing for production MegaBLAST searches. *Bioinformatics*, **24**, 1757–1764.

29. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoz-Torres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E., Stein,L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.

30. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.

31. Deigan,K.E., Li,T.W., Mathews,D.H. and Weeks,K.M. (2009) Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 97–102.

32. Kerpedjiev,P., Hammer,S. and Hofacker,I.L. (2015) Forna (force-directed RNA): simple and effective online RNA secondary structure diagrams. *Bioinformatics*, **31**, 3377–3379.

33. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.

34. Kent,W.J., Zweig,A.S., Barber,G., Hinrichs,A.S. and Karolchik,D. (2010) BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics*, **26**, 2204–2207.

35. Watts,J.M., Dang,K.K., Gorelick,R.J., Leonard,C.W., Bess,J.W. Jr, Swanstrom,R., Burch,C.L. and Weeks,K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature*, **460**, 711–716.

36. Huber,R.G., Lim,X.N., Ng,W.C., Sim,A.Y.L., Poh,H.X., Shen,Y., Lim,S.Y., Sundstrom,K.B., Sun,X., Aw,J.G. *et al.* (2019) Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat. Commun.*, **10**, 1408.

37. Li,P., Wei,Y., Mei,M., Tang,L., Sun,L., Huang,W., Zhou,J., Zou,C., Zhang,S., Qin,C.F. *et al.* (2018) Integrative analysis of Zika virus genome RNA structure reveals critical determinants of viral infectivity. *Cell Host Microbe*, **24**, 875–886.

38. Sun,L., Li,P., Ju,X., Rao,J., Huang,W., Zhang,S., Xiong,T., Xu,K., Zhou,X., Ren,L. *et al.* (2020) *In vivo* structural characterization of the whole SARS-CoV-2 RNA genome identifies host cell target proteins vulnerable to re-purposed drugs. bioRxiv doi: https://doi.org/10.1101/2020.07.07.192732, 08 July 2020, preprint: not peer reviewed.

39. Manfredonia,I., Nithin,C., Ponce-Salvatierra,A., Ghosh,P., Wirecki,T.K., Marinus,T., Ogando,N.S., Snider,E.J., van Hemert,M.J., Bujnicki,J.M. *et al.* (2020) Genome-wide mapping of therapeutically-relevant SARS-CoV-2 RNA structures. bioRxiv doi: https://doi.org/10.1101/2020.06.15.151647, 15 June 2020, preprint: not peer reviewed.

40. Huston,N.C., Wan,H., de Cesaris Araujo Tavares,R., Wilen,C. and Pyle,A.M. (2020) Comprehensive in-vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. bioRxiv doi: https://doi.org/10.1101/2020.07.10.197079, 10 July 2020, preprint: not peer reviewed.

41. Sanders,W., Fritch,E.J., Madden,E.A., Graham,R.L., Vincent,H.A., Heise,M.T., Baric,R.S. and Moorman,N.J. (2020) Comparative analysis of coronavirus genomic RNA structure reveals conservation in SARS-like coronaviruses. bioRxiv doi: https://doi.org/10.1101/2020.06.15.153197, 16 June 2020, preprint: not peer reviewed.

42. Lan,T.C.T., Allan,M.F., Malsick,L.E., Khandwala,S., Nyeo,S.S.Y., Bathe,M., Griffiths,A. and Rouskin,S. (2020) Structure of the full SARS-CoV-2 RNA genome in infected cells. bioRxiv doi: https://doi.org/10.1101/2020.06.29.178343, 30 June 2020, preprint: not peer reviewed.

43. Iserman,C., Roden,C., Boerneke,M., Sealfon,R., McLaughlin,G., Jungreis,I., Park,C., Boppana,A., Fritch,E., Hou,Y.J. *et al.* (2020) Specific viral RNA drives the SARS CoV-2 nucleocapsid to phase separate. bioRxiv doi: https://doi.org/10.1101/2020.06.11.147199, 12 June 2020, preprint: not peer reviewed.

44. Bevilacqua,P.C., Ritchey,L.E., Su,Z. and Assmann,S.M. (2016) Genome-wide analysis of RNA secondary structure. *Annu. Rev. Genet.*, **50**, 235–266.

45. Somarowthu,S., Legiewicz,M., Chillon,I., Marcia,M., Liu,F. and Pyle,A.M. (2015) HOTAIR forms an intricate and modular secondary structure. *Mol. Cell*, **58**, 353–361.