

# RPS: a comprehensive database of RNAs involved in liquid–liquid phase separation

Mengni Liu<sup>1,†</sup>, Huiqin Li<sup>1,†</sup>, Xiaotong Luo<sup>1,†</sup>, Jieyi Cai<sup>1</sup>, Tianjian Chen<sup>1</sup>, Yubin Xie<sup>2,\*</sup>, Jian Ren<sup>1,\*</sup> and Zhixiang Zuo<sup>1,\*</sup>

<sup>1</sup>School of Life Sciences, State Key Laboratory of Oncology in South China, Cancer Center, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University, Guangzhou 510060, China and <sup>2</sup>Precision Medicine Institute, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510060, China

Received August 15, 2021; Revised September 24, 2021; Editorial Decision October 07, 2021; Accepted October 09, 2021

## ABSTRACT

Liquid–liquid phase separation (LLPS) is critical for assembling membraneless organelles (MLOs) such as nucleoli, P-bodies, and stress granules, which are involved in various physiological processes and pathological conditions. While the critical role of RNA in the formation and the maintenance of MLOs is increasingly appreciated, there is still a lack of specific resources for LLPS-related RNAs. Here, we presented RPS (<http://rps.renlab.org>), a comprehensive database of LLPS-related RNAs in 20 distinct biomolecular condensates from eukaryotes and viruses. Currently, RPS contains 21,613 LLPS-related RNAs with three different evidence types, including ‘Reviewed’, ‘High-throughput’ and ‘Predicted’. RPS provides extensive annotations of LLPS-associated RNA properties, including sequence features, RNA structures, RNA–protein/RNA–RNA interactions, and RNA modifications. Moreover, RPS also provides comprehensive disease annotations to help users to explore the relationship between LLPS and disease. The user-friendly web interface of RPS allows users to access the data efficiently. In summary, we believe that RPS will serve as a valuable platform to study the role of RNA in LLPS and further improve our understanding of the biological functions of LLPS.

## INTRODUCTION

Liquid-liquid phase separation (LLPS) is a reversible process driving the formation of membraneless organelles (MLOs) such as nucleoli, P-bodies and stress granules (SGs). During the LLPS process, many biomolecules in-

cluding protein and RNA aggregate together into biomolecular condensates, which play critical roles in the regulation of many biological processes such as cellular stress responses (1), homeostasis maintenance (2) and development (3). Meanwhile, increasing evidence has shown that the dysregulation of LLPS is closely associated with a variety of diseases such as amyotrophic lateral sclerosis (ALS) (4–6), frontotemporal dementia (FTD) (4–6), Alzheimer’s disease (AD) (7), cancer (8), as well as infectious diseases (9).

Deciphering the components of LLPS is vital to understand the roles of LLPS in physiological and pathological processes. Previous studies primarily focused on the protein components of LLPS. It has been demonstrated that G3BP protein was required for SG formation in mammalian cells during oxidative stress (10), and the MEG1 and MEG3 proteins are required for P-granule formation in *C. elegans* (11). These LLPS proteins generally contain intrinsically disordered regions (IDRs) or low-complexity domains (LCDs), which contribute to phase separation via multivalent weak interactions (12). To facilitate the study of proteins involved in LLPS, several databases have been developed to record proteins associated with LLPS, such as LLPSDB (13), PhaSePro (14), PhaSepDB (15), DrLLPS (16), RNAgranuleDB (17) and HUMAN CELL MAP (18).

In recent years, emerging evidences have proved that RNA also has fundamental roles in the regulation of LLPS. It has been recognized that RNAs act as a buffer in the nucleus where high RNA concentrations keep RBPs soluble (19). Besides, RNA can phase separate without protein and promote or inhibit phase separation (5). Similar to proteins, RNA can also serve as a seed for biomolecular assemblies, such as lncRNA NEAT1, which functions as a scaffold by interacting with other RBPs for paraspeckles construction (20). In addition, numerous studies have suggested that RNA-dependent condensates are tightly regulated by various RNA properties, such as RNA sequence, structure, RNA modifications, RNA–RNA interactions and

\*To whom correspondence should be addressed. Tel: +86 02087342325; Email: zuozhx@sysucc.org.cn

Correspondence may also be addressed to Jian Ren. renjian@sysucc.org.cn

Correspondence may also be addressed to Yubin Xie. Email: xieyb6@mail.sysu.edu.cn

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

RNA–protein interactions. For example, Khong *et al.* (1) suggested that relatively long transcripts are preferentially involved in SGs formation, which harbors more sites for possible interactions with RNA-binding proteins (RBPs) and/or RNAs. Moreover, compared with less structured RNAs, highly structured RNAs can rearrange the composition of protein aggregates for having more interactions with proteins (21). G-quadruplex (GQ), a specific RNA tertiary structure motif, can trigger RNA phase separation under physiological conditions *in vitro* (22). Other features like RNA modifications are also critical in phase separation. For instance, N6-methyladenosine (m<sup>6</sup>A), as the most prevalent mRNA modification, was considered as a multivalent scaffold for binding YTHDF proteins in mammalian cells (23). Furthermore, RNA expression levels are essential for the formation and maintenance of condensates. A recent study of condensate transcriptome has revealed that SGs assembly relied on the condensation of poorly translated mRNAs in mammalian and yeast cells (1). Of note, it is reported that mutations can impact RNA-dependent LLPS. For example, the disease-causing G4C2 repeat expansion in the C9orf72 gene has been shown to mediate LLPS both *in vitro* and *in vivo* (5,24). Collectively, these features encoded in RNA confer specific condensate biophysical properties, which are essential for condensate functions in homeostasis.

Although numerous studies have paid attention to the role of RNAs in LLPS, an integrative data resource of LLPS-related RNAs is still not available. To fill this gap, we present RPS (<http://rps.renlab.org>), a comprehensive database of LLPS-related RNAs in 20 distinct biomolecular condensates across eukaryotes and viruses. RPS contains 21 613 LLPS-related RNAs derived from literature mining, high-throughput analysis and predictions based on interaction network (Figure 1). RPS provides basic information of LLPS-related RNAs and LLPS processes, as well as plentiful annotations of RNAs, including sequences, RNA secondary structures, RNA–RNA/RBP binding sites and modifications. Additionally, disease annotations of LLPS-related RNA are also available. With these features, we anticipate that RPS will be helpful to investigate the role of RNA in LLPS and provide new insights into human disease.

## MATERIALS AND METHODS

### Data collection

RPS collected RNA-associated LLPS data with three evidence types. (i) Reviewed: The literature mining was performed by searching NCBI PubMed using keywords combinations ‘((phase separation[Title/Abstract]) OR (phase transition[Title/Abstract]) OR (biomolecular condensates[Title/Abstract]) OR (MLOs[Title/Abstract]) OR (membraneless organelle[Title/Abstract])) AND (RNA)’. We retrieved a total of 1507 papers published up to 30 June 2021, then screened out 131 publications that describe biomolecular condensates and related RNAs by rechecking abstracts manually. Both natural and synthetic RNAs with this evidence type were curated. (ii) High-throughput: RNA composition of MLOs under different conditions can be identified by RNA sequencing methods.

By retrieving related keywords (same with ‘Reviewed’ type) from the GEO (25), EMBL–EBI (26) and DDBJ (27) databases, we collected seven MLO-associated datasets (including 82 samples) from humans, mice and yeast (Table S1). Further analyses were performed on these datasets to identify RNAs that participate in the formation of MLOs. (iii) Predicted: RNA–protein and RNA–RNA interactions underlie the ability of particular RNAs to undergo LLPS (28). Therefore, we anticipated that RNA interactors with known-related RNAs/RBPs are likely to undergo LLPS. Briefly, we obtained LLPS-related proteins from six resources, including LLPSDB (13), PhaSePro (14), PhaSepDB (15), DrLLPS (16), RNAgranuleDB (17) and Human cell Map (18). Then, we filtered proteins that were only relevant with RNA-free condensates (such as receptor cluster and Z granule) and retained 1766 LLPS-related proteins for prediction. After that, based on experimentally validated RNA-associated interactions obtained from NPInter 4.0 (29) and RNAInter (30), we identified 1358 RNA interactors with these retained LLPS-related proteins and ‘Reviewed’ LLPS-related RNAs.

### Analysis of RNA-Seq datasets

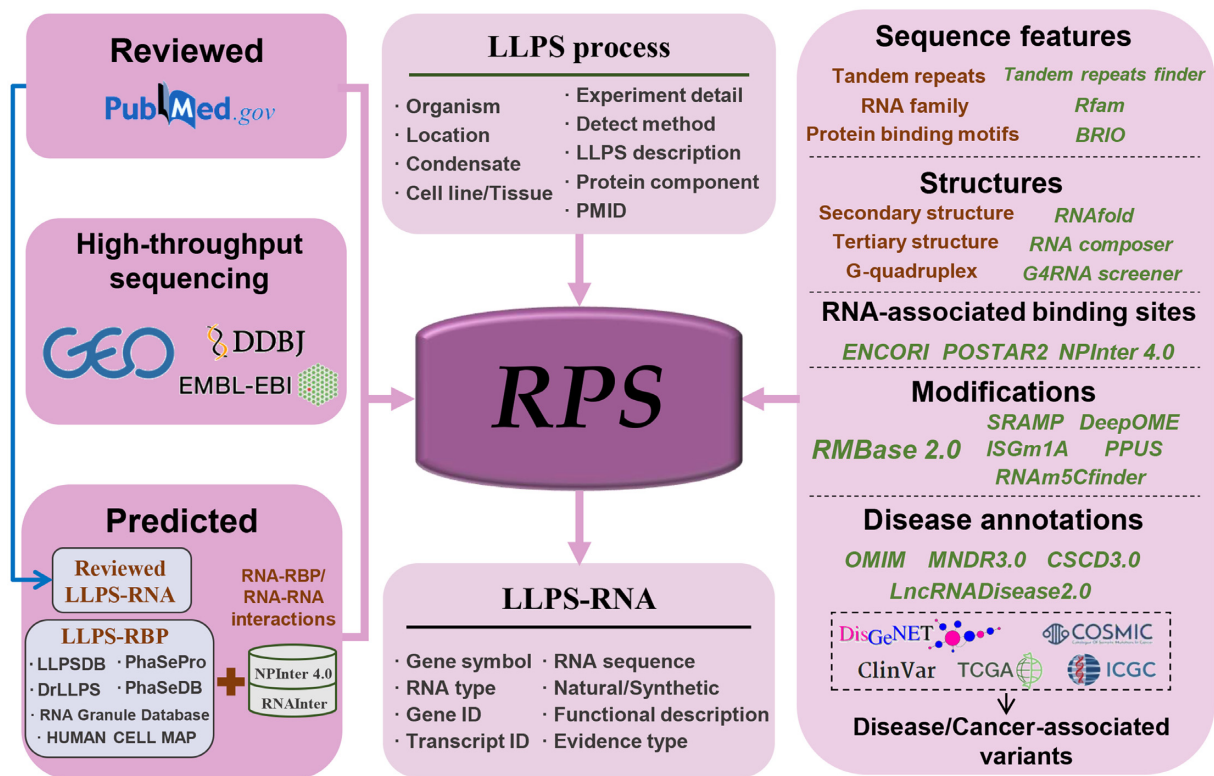
All collected RNA-seq datasets were processed with a uniform pipeline described below. Raw reads were first trimmed by TrimGalore (v2.10, <https://github.com/FelixKrueger/TrimGalore>) to remove low-quality bases (Phred score < 25) and adapters contamination. The trimmed reads were aligned to the reference genome (human: GRCh38; mouse: GRCm38; yeast: R64-1-1) using STAR (v.2.7.6) with default parameters (31) and further counted by featureCounts (v.2.0.1) (32). After that, we performed pairwise comparisons between the condensate group versus the control group (e.g. RNA granule vs cytoplasm) or condensate groups under distinct conditions (e.g. heat shock-induced RNA granule vs unstressed RNA granule). Normalization and differential expression analysis were then performed using DESeq2 (v1.28.1) (33). To determine phase-separated condensates enriched transcripts, we calculated the fold enrichment over the control group and applied a threshold of two-fold enrichment with a *P*-value < 0.05. RNA transcripts that meet this criterion were considered to be enriched in this condensate.

### Curation of LLPS processes

We extracted detailed information of the LLPS system, including detect methods, such as fluorescence recovery after photobleaching (FRAP), electron microscopy (EM), etc., experiment conditions such as RNA/salt concentrations, salt, pH, pressure, and temperature, protein components. It should be noted that ‘Predicted’ LLPS-related RNAs are not assigned to a particular LLPS process, therefore having no corresponding LLPS information.

### Annotation of LLPS-related RNAs

RPS provides various basic information of LLPS-related RNAs, such as gene symbol, gene ID, transcript ID, sequence, RNA type, functional description. These data were



**Figure 1.** Overall design and construction of RPS. LLPS-related RNAs deposited in RPS were derived from literature mining, high-throughput sequencing experiments and predictions based on interaction networks. RPS provides detailed information of LLPS-related RNAs and LLPS processes. In addition, RPS contains extensive annotations of RNA properties for each LLPS-related RNA, such as RNA sequence, RNA structure, RNA-RBP interactions, RNA-RNA interactions, RNA modifications, as well as disease annotations.

preferentially extracted from the original publications, otherwise, they were acquired from Ensembl (34) or UCSC (35) databases. For part ‘reviewed’ LLPS-related RNAs, sequences were curated from the original publications. Other RNA sequences were either represented by corresponding canonical transcripts obtained via UCSC table browser or obtained from existed databases, including LNCipedia (36), miRbase (37) and NONCODE (38).

To improve our understanding of how RNA contributes to phase separation, RPS integrated multiple resources and tools to annotate distinct LLPS-associated properties of RNAs. Sequence-specific features, including tandem repeats and RNA sequence families, were detected separately using tandem repeat finder (39) and Rfam web server (40). For structure annotation, we first extracted information of LLPS-associated structures from the original publications. Then, we predicted RNA secondary structures based on a minimum free energy (MFE) algorithm using RNAfold (v2.4.1) from ViennaRNA package 2.0 (41). For RNAs within 500 nt, we further used RNAComposer (42) to assemble the 3D modeling according to their predicted secondary structures. GQ structures were predicted using the G4RNA screener web server (43) with the default settings. Moreover, the experimentally verified and predicted RNA-protein and RNA-RNA binding sites were acquired from ENCORI (44), POSTAR2 (45) and NPinter 4.0 (29). Besides, we identified known sequence and secondary structure protein binding motifs in RNAs from humans and mice

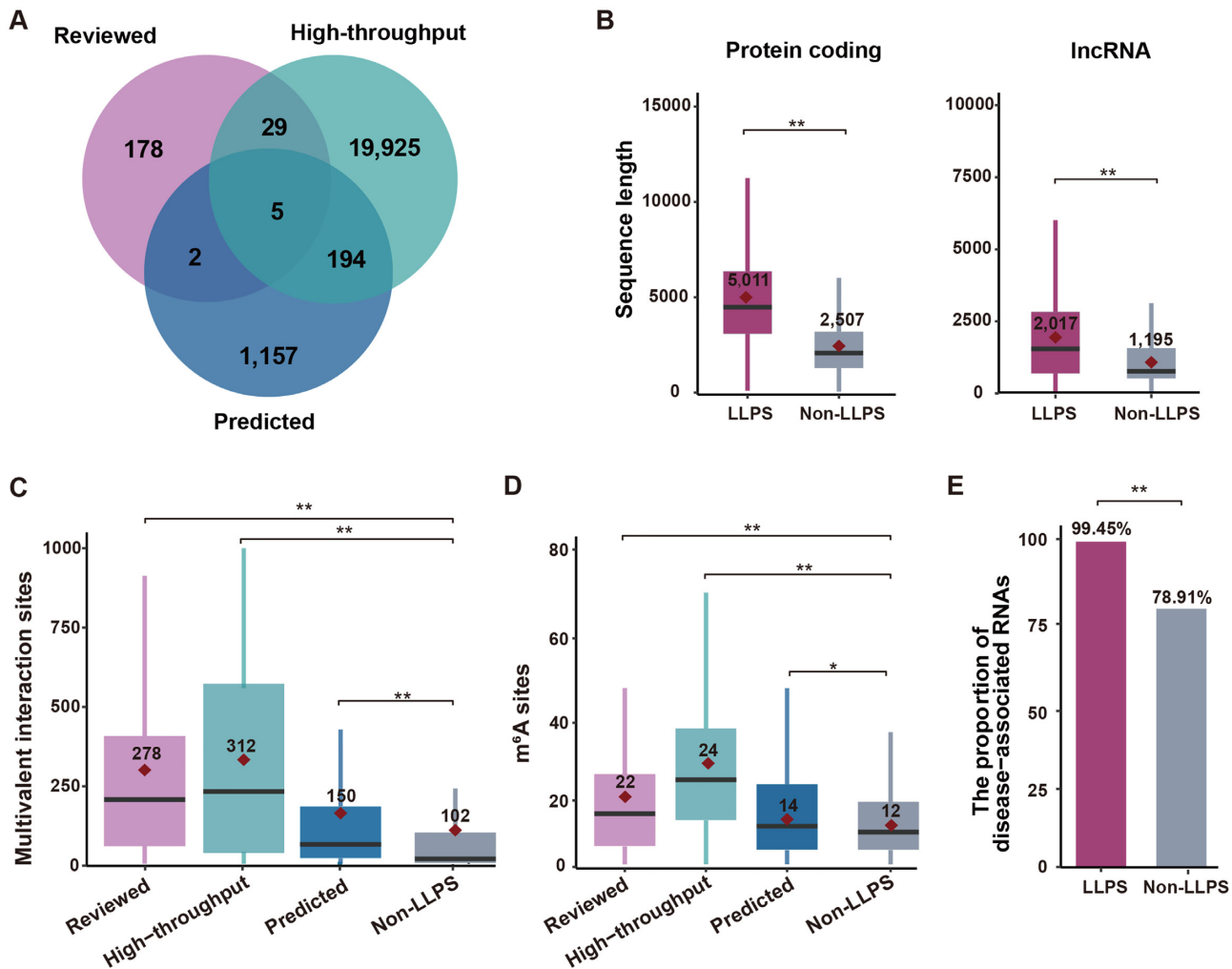
via BRIO web server (46). As another LLPS-associated property, RNA modification can also potentially contribute to the features of native condensate. In addition to integrating the RNA modifications sites with experimental validation from RMBase 2.0 (47), we employed SRAMP (48) for m<sup>6</sup>A sites prediction, DeepOME (49) for 2'-O-methylation (Nm) sites prediction, RF-PseU (50) and PPUS (51) for pseudouridine ( $\Psi$ ) sites prediction, RNAm5Cfinder (52) for 5-methylcytidine (m<sup>5</sup>C) sites prediction and ISGm1A for N<sup>1</sup>-methyladenosine (m<sup>1</sup>A) sites prediction.

To discover potential relations between LLPS-related RNAs and diseases, we collected RNA-disease associations with experimental evidence from OMIM (53), DisGeNET (54), MNDR 3.0 (55), LncRNADisease 2.0 (56) and CSCD 2.0 databases (57). Furthermore, we also integrated disease-associated variants from DisGeNET (54) and ClinVar (58), as well as cancer-associated variants taken from COSMIC (59), ICGC (60) and TCGA (61). In RPS, we considered RNAs having RNA-disease associations or disease/cancer-associated variants as disease-associated RNAs. The genomic coordinates of all data resources were further converted to GRCh38 or GRCm38 using the LiftOver program (35).

### Classification of phase-separated condensates

To better curate the deposited data, we categorized all biomolecular condensates into three classes: (i) nucleus,





**Figure 2.** Statistical analysis of LLPS-RNAs in RPS. (A) The overlap between LLPS-related RNAs with different evidence types. (B) Comparison of sequence lengths of LLPS-related RNAs and non-LLPS-related RNAs for protein-coding RNA and lncRNA, separately. Wilcoxon signed-rank test, \*\*  $P$ -value < 0.01. (C) Comparison of the multivalent interaction sites (RNA–RNA interaction and RNA–RBP interaction) of LLPS-related RNAs and non-LLPS-related RNAs in humans. Wilcoxon signed-rank test, \*\*  $P$ -value < 0.01. (D) Comparison of experimentally verified m<sup>6</sup>A sites of LLPS-RNAs and non-LLPS-related RNAs in humans. (E) The proportion of disease-associated RNAs in LLPS-related RNAs and non-LLPS-related RNAs. Two-proportions  $z$ -test, \*\*  $P$ -value < 0.01.

including paraspeckle, Cajal body, DNA damage foci, histone locus body, nuclear body, nuclear speckle, nucleolus, PML nuclear body, PcG body; (ii) cytoplasm: centrosome, P-body, cytoplasmic granule, G body, neuronal granule, stress granule and TIS granule; (iii) others, including germ cell condensates (P granule, Nuage and Balbiani body) and RNP granules in unknown locations, as well as coacervate droplets. These phase-separated condensates were further annotated with Gene Ontology (GO) cellular component terms.

### Database and web interface implementation

All data in RPS were stored and managed by MySQL tables. The server-backend development was based on java and the web-frontend interfaces were implemented in Hyper Text Markup Language (HTML), Cascading Style Sheets (CSS) and JavaScript (JS). In order to present data more efficiently and intuitively, multiple statistical diagrams were embedded

in the website. The interactive heat maps showing the expression abundance and differential expression were

constructed by Ant Design toolkit. The boxplots showing the differential expression and the charts presenting the interaction network of LLPS-related RNAs were drawn by Echarts. The RNA tertiary structures are displayed using 3Dmol.js (62). Furthermore, RPS implemented a genome browser to present genomic annotations using UCSC Genome Browser (<http://genome.ucsc.edu/>) (35).

## RESULTS

### Database content

Currently, RPS contains 42 417 entries for 21 613 unique LLPS-related RNAs from 13 organisms. There are three different evidence types of LLPS data deposited in RPS: (i) ‘Reviewed’: 523 entries for 337 RNAs were validated to participate in LLPS through *in vitro* or *in vivo* exper-

**Table 1.** LLPS-related RNAs in distinct condensates with different evidence types

Condensate	Location	Reviewed	High-throughput	Predicted	Total
Cytoplasmic granule	Cytoplasm	0	10 788	3	10 791
P-body	Cytoplasm	23	8431	747	9080
Stress granule	Cytoplasm	35	4736	954	5662
Paraspeckle	Nucleus	52	787	67	903
Nucleolus	Nucleus	5	0	791	796
Coacervate droplet	Others	174	0	163	336
G body	Cytoplasm	5	238	0	243
Nuclear speckle	Nucleus	1	0	242	242
TIS granule	Cytoplasm	6	0	106	112
Nuclear body	Nucleus	16	0	75	91
PML nuclear body	Nucleus	0	0	33	33
P granule	Others	28	0	3	31
Cajal body	Nucleus	1	0	2	3
DNA damage foci	Nucleus	1	0	2	3
Histone locus body	Nucleus	1	0	1	2
Neuronal granule	Cytoplasm	0	0	2	2
Nuage	Others	0	0	2	2
Balbiani body	Others	0	0	1	1
Centrosome	Cytoplasm	0	0	1	1
PcG body	Nucleus	0	0	1	1

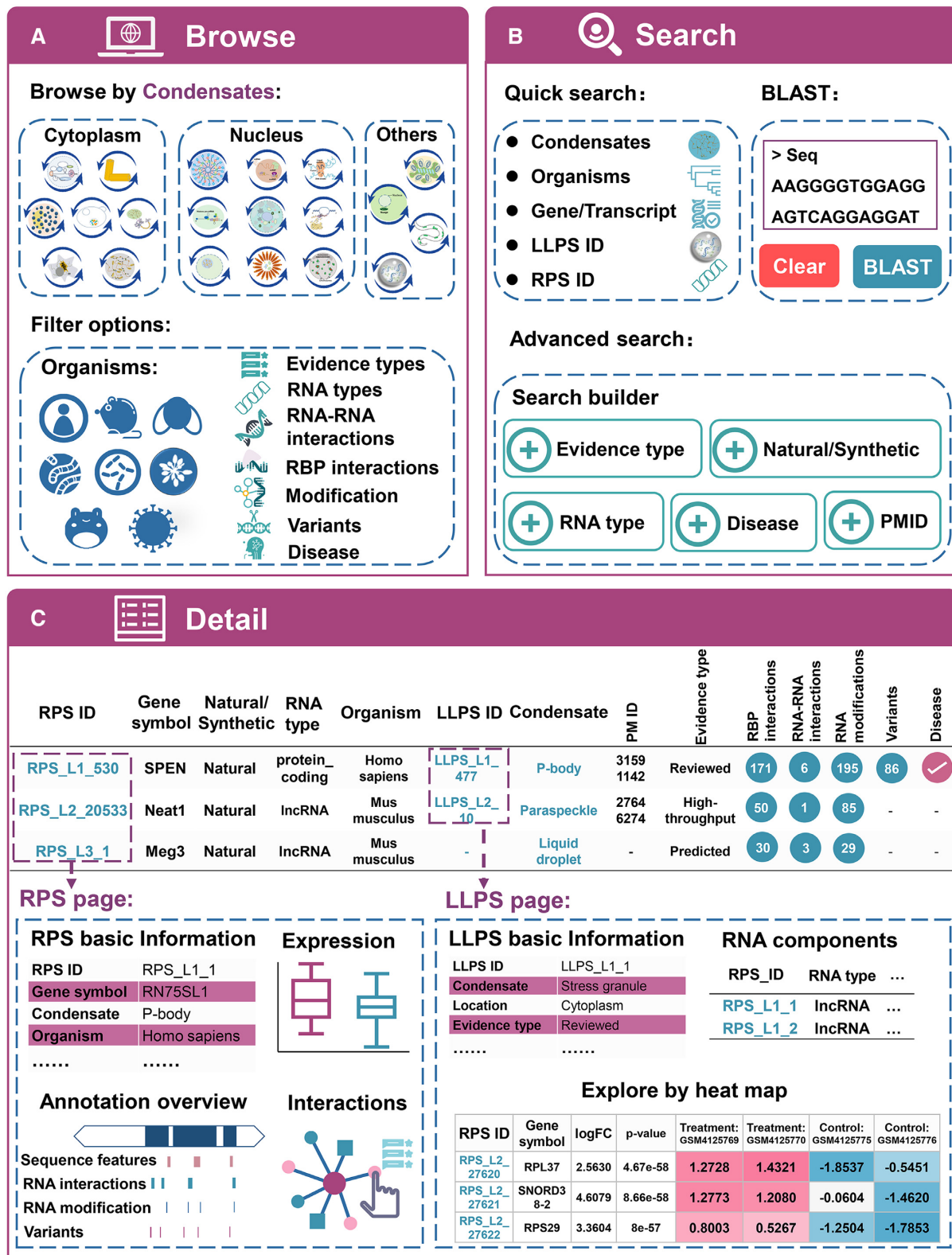
**Table 2.** LLPS-related RNAs of different RNA types from distinct organisms

RNA type	<i>Homo sapiens</i>	<i>Mus musculus</i>	<i>Saccharomyces cerevisiae</i>	Other eukaryotes	Virus	Synthetic
protein_coding	7269	6250	1928	33	0	0
lncRNA	2116	1002	0	0	0	0
pseudogene	738	366	58	0	0	0
miRNA	422	47	0	2	8	0
snoRNA	20	50	55	0	0	0
snRNA	31	22	0	0	0	0
virus ssRNA	0	0	0	0	13	0
others	219	835	8	2	0	0
synthetic nucleotides	0	0	0	0	0	123

iments. Among them, 328 entries are from 214 natural RNAs. The other 123 synthetic nucleotides with 195 entries were designed for condensate reconstruction experiments, such as homotypic RNA polymers. (ii) ‘*High-throughput*’: 20 153 RNAs involved in 18 different LLPS processes were identified by high-throughput analyses. These RNAs were considered as RNA components of MLOs such as SGs, paraspeckles, P-body, G-body and other RNA granules (Table S1). (iii) ‘*Predicted*’: 1358 RNAs were predicted as potential LLPS-related RNAs based on RNA-associated interactions, where the transcripts of NEAT1, BRCA1, NORAD, MALAT1 and ACTB were supported by three types of evidence. Besides, total 199 ‘*Predicted*’ LLPS-related RNAs were validated either by ‘*Reviewed*’ or ‘*High-throughput*’ evidence (Figure 2A), demonstrating the validity of the interaction-based prediction method. Regarding the subcellular localization, 20 598 RNAs are localized in the cytoplasm, 1943 RNAs participate in the formation of biomolecular condensates in nuclear, and 369 RNAs belong to other condensates in germ cells or unknown locations (Table 1). The majority of LLPS-related RNAs are derived from humans, mice and yeast. Besides, protein-coding RNA and lncRNA are the most abundant RNA types stored in RPS (Table 2). In addition, RPS provides a plentiful of annotations for the LLPS-related RNAs, such as RNA–RNA/RBP binding sites, RNA modification sites and disease-associations.

### Characteristics of LLPS-related RNAs

To characterize the LLPS-related RNAs, we systematically compared several RNA properties of LLPS-related and non-LLPS-related RNAs in humans. The non-LLPS-related RNAs were defined as human canonical transcripts that were not deposited in RPS. Consistent with previous studies (1,63), the sequences of LLPS-related RNAs are significantly longer than those of non-LLPS-related RNAs for both types (Figure 2B, *P*-value < 0.01, Wilcoxon signed-rank test), indicating the promoting effect of RNA length on LLPS. In addition, comparisons of multivalent interaction sites (RNA–RBP and RNA–RNA binding sites) showed that LLPS-related RNAs of all three evidence types have significantly more interactions sites than those in non-LLPS-related RNAs (Figure 2C, *P*-value < 0.01, Wilcoxon signed-rank test). This suggests that increased multivalent interactions are likely to promote LLPS, which is in line with previous findings (64). Moreover, LLPS-related RNAs have significantly more experimentally verified m<sup>6</sup>A sites than non-LLPS-related RNAs (Figure 2D), which is consistent with previous findings that m<sup>6</sup>A plays a critical role in driving LLPS in mammalian cells (23). In term of the relevance to disease, LLPS-related RNAs show a higher proportion of disease-associated RNAs than that in non-LLPS-related RNAs (Figure 2E, *P*-value < 0.01, two-proportions *z*-test), indicating that RNA-mediated phase separation is closely related to human diseases.



## Web interface and usage

RPS provides a user-friendly web interface, allowing users to explore LLPS-related RNAs or processes of interest easily and interactively (Figure 3).

**Browse**, users can browse the entries by condensates. For each condensate, the browse page presents a statistical chart, as well as diagrams for the LLPS-related RNAs distribution in distinct organisms and RNA types. The query results can be filtered by extra options, including Organism, Evidence type, RNA type and other RNA features (Figure 3A).

**Search**, RPS provides two ways to search the database (Figure 3B). First, an advanced search engine is developed to enable precise queries, providing various options including Gene symbol, Gene ID, Transcript ID, Organism, Condensate, Evidence type, RNA type, Natural/Synthetic, RPS ID, LLPS ID and PMID. Second, users can perform a sequence similarity search against RPS by an online BLAST server (v 2.7.1), which was implemented on the ‘BLAST’ page.

**Detail**, the details of each entry are displayed on the ‘Detail’ page by clicking any RPS ID or LLPS ID (Figure 3C). The detail for a LLPS-related RNA contains the evidence that it participates in LLPS, annotations of various RNA properties such as sequence features, structures, RNA–protein/RNA interactions, and RNA modifications, and annotations of disease associations. Furthermore, RPS allows users to browse genomic features of interested RPS entries by integrating all the RPS data into the UCSC genome browser. The detail page of the LLPS processes records experiment conditions, detecting methods, description of phase behavior, protein components and RNA components. Notably, an interactive heatmap was implemented explicitly for a high-throughput experiments-derived LLPS process on the LLPS page, showing the differential expression of the RNA components between condensates and controls.

All data deposited in RPS are available on the ‘Download’ page. Detailed guidance on the usage of RPS can be found on the ‘Help’ page.

## SUMMARY AND PERSPECTIVES

Phase separation has expanded our understanding of biochemical reactions and biological processes in MLOs. With the advancement of phase separation research technology, more and more RNAs have been found to participate in LLPS process and regulate the assembly of MLOs. Emerging evidence has shown that RNA is crucial in sensing stress stimulations, signal transduction, and maintenance of phase separation. Despite that, existed LLPS-associated databases primarily focus on proteins (13–18), while a comprehensive, curated database of LLPS-related RNAs is still lacking. To our knowledge, RPS is the first comprehensive database for specifically hosting the LLPS-related RNAs. It records both basic information of all LLPS-related RNAs and details of the corresponding LLPS system, such as phase behavior and experimental conditions. For humans, mice and yeast, RPS offers additional annotations of RNA features that confer condensate biophysical properties, including RNA structures, RNA-associated interactions and

modifications. Multiple comparisons of these RNA features between LLPS-related RNAs and non-LLPS-related RNAs revealed that RNA sequence, m<sup>6</sup>A modifications, as well as multivalent interactions with promote LLPS (Figure 2A–D). These observations are in line with previous findings (1,23,63,64), illustrating the ability of RPS to reveal the contribution of RNAs in mediating LLPS. Remarkably, LLPS-related RNAs exhibit a stronger correlation with disease than non-LLPS-related RNAs (Figure 2E), suggesting that a deep understanding of the contribution of RNA to LLPS will be beneficial for disease diagnosis and treatment.

Taken together, we anticipate that RPS will support investigations into the potential of RNA in condensate biology and develop better therapeutic treatments for phase-separation-related diseases. With the growing interest in the role of RNA in LLPS, there will be a rapidly increasing number of relevant studies and more LLPS-related RNAs will be discovered. Therefore, we are dedicated to ensuring the long-term maintenance and reliability of the RPS database by continuous updates and careful validation.

## DATA AVAILABILITY

RPS is a comprehensive online database available at <http://rps.renlab.org>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Natural Science Foundation of China [81772614, 91753137, 31771462, U1611261, 31801105]; Program for Guangdong Introducing Innovative and Entrepreneurial Teams [2017ZT07S096]; Guangdong Basic and Applied Basic Research Foundation [2020A1515011219, 2021B1515020108]. Funding for open access charge: National Natural Science Foundation of China [81772614].  
*Conflict of interest statement.* None declared.

## REFERENCES

- Khong,A., Matheny,T., Jain,S., Mitchell,S.F., Wheeler,J.R. and Parker,R. (2017) The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. *Mol. Cell*, **68**, 808–820.
- Audas,T.E., Jacob,M.D. and Lee,S. (2012) Immobilization of proteins in the nucleolus by ribosomal intergenic spacer noncoding RNA. *Mol. Cell*, **45**, 147–157.
- Zhang,B., Arun,G., Mao,Y.S., Lazar,Z., Hung,G., Bhattacharjee,G., Xiao,X., Booth,C.J., Wu,J., Zhang,C. *et al.* (2012) The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep.*, **2**, 111–123.
- Boeynaems,S., Bogaert,E., Kovacs,D., Konijnenberg,A., Timmerman,E., Volkov,A., Guharoy,M., De Decker,M., Jaspers,T., Ryan,V.H. *et al.* (2017) Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. *Mol. Cell*, **65**, 1044–1055.
- Jain,A. and Vale,R.D. (2017) RNA phase transitions in repeat expansion disorders. *Nature*, **546**, 243–247.
- Balendra,R. and Isaacs,A.M. (2018) C9orf72-mediated ALS and FTD: multiple pathways to disease. *Nat. Rev. Neurol.*, **14**, 544–558.
- Han,M., Liu,Z., Xu,Y., Liu,X., Wang,D., Li,F., Wang,Y. and Bi,J. (2020) Abnormality of m6A mRNA Methylation Is Involved in Alzheimer's Disease. *Front Neurosci.*, **14**, 98.



8. Bouchard,J.J., Otero,J.H., Scott,D.C., Szulc,E., Martin,E.W., Sabri,N., Granata,D., Marzahn,M.R., Lindorff-Larsen,K., Salvatella,X. *et al.* (2018) Cancer mutations of the tumor suppressor SPOP disrupt the formation of active, phase-separated compartments. *Mol. Cell*, **72**, 19–36.
9. Iserman,C., Roden,C.A., Boerneke,M.A., Sealfon,R.S.G., McLaughlin,G.A., Jungreis,I., Fritch,E.J., Hou,Y.J., Ekena,J., Weidmann,C.A. *et al.* (2020) Genomic RNA elements drive phase separation of the SARS-CoV-2 nucleocapsid. *Mol. Cell*, **80**, 1078–1091.
10. Kedersha,N., Panas,M.D., Achorn,C.A., Lyons,S., Tisdale,S., Hickman,T., Thomas,M., Lieberman,J., McInerney,G.M., Ivanov,P. *et al.* (2016) G3BP-Caprin1-USP10 complexes mediate stress granule condensation and associate with 40S subunits. *J. Cell Biol.*, **212**, 845–860.
11. Wang,J.T., Smith,J., Chen,B.C., Schmidt,H., Rasoloson,D., Paix,A., Lambrus,B.G., Calidas,D., Betzig,E. and Seydoux,G. (2014) Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically disordered proteins in *C. elegans*. *Elife*, **3**, e04591.
12. Nott,T.J., Petsalaki,E., Farber,P., Jervis,D., Fussner,E., Plochowitz,A., Craggs,T.D., Bazett-Jones,D.P., Pawson,T., Forman-Kay,J.D. *et al.* (2015) Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell*, **57**, 936–947.
13. Li,Q., Peng,X., Li,Y., Tang,W., Zhu,J., Huang,J., Qi,Y. and Zhang,Z. (2020) LLPSDB: a database of proteins undergoing liquid-liquid phase separation in vitro. *Nucleic Acids Res.*, **48**, D320–D327.
14. Meszaros,B., Erdos,G., Szabo,B., Schad,E., Tantos,A., Abukhairan,R., Horvath,T., Murvai,N., Kovacs,O.P., Kovacs,M. *et al.* (2020) PhaSePro: the database of proteins driving liquid-liquid phase separation. *Nucleic Acids Res.*, **48**, D360–D367.
15. You,K., Huang,Q., Yu,C., Shen,B., Sevilla,C., Shi,M., Hermjakob,H., Chen,Y. and Li,T. (2020) PhaSepDB: a database of liquid-liquid phase separation related proteins. *Nucleic Acids Res.*, **48**, D354–D359.
16. Ning,W., Guo,Y., Lin,S., Mei,B., Wu,Y., Jiang,P., Tan,X., Zhang,W., Chen,G., Peng,D. *et al.* (2020) DrLLPS: a data resource of liquid-liquid phase separation in eukaryotes. *Nucleic Acids Res.*, **48**, D288–D295.
17. Youn,J.Y., Dyakov,B.J.A., Zhang,J., Knight,J.D.R., Vernon,R.M., Forman-Kay,J.D. and Gingras,A.C. (2019) Properties of stress granule and P-body proteomes. *Mol. Cell*, **76**, 286–294.
18. Fazal,F.M., Han,S., Parker,K.R., Kaewsapsak,P., Xu,J., Boettiger,A.N., Chang,H.Y. and Ting,A.Y. (2019) Atlas of subcellular RNA localization revealed by APEX-Seq. *Cell*, **178**, 473–490.
19. Van Treeck,B. and Parker,R. (2018) Emerging roles for intermolecular RNA–RNA interactions in RNP assemblies. *Cell*, **174**, 791–802.
20. Clemson,C.M., Hutchinson,J.N., Sara,S.A., Ensminger,A.W., Fox,A.H., Chess,A. and Lawrence,J.B. (2009) An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol. Cell*, **33**, 717–726.
21. Strom,A.R., Emelyanov,A.V., Mir,M., Fyodorov,D.V., Darzacq,X. and Karpen,G.H. (2017) Phase separation drives heterochromatin domain formation. *Nature*, **547**, 241–245.
22. Zhang,Y., Yang,M., Duncan,S., Yang,X., Abdelhamid,M.A.S., Huang,L., Zhang,H., Benfey,P.N., Waller,Z.A.E. and Ding,Y. (2019) G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Res.*, **47**, 11746–11754.
23. Ries,R.J., Zaccara,S., Klein,P., Olererin-George,A., Namkoong,S., Pickering,B.F., Patil,D.P., Kwak,H., Lee,J.H. and Jaffrey,S.R. (2019) m(6)A enhances the phase separation potential of mRNA. *Nature*, **571**, 424–428.
24. Lee,Y.B., Chen,H.J., Peres,J.N., Gomez-Deza,J., Attig,J., Stalekar,M., Troakes,C., Nishimura,A.L., Scotter,E.L., Vance,C. *et al.* (2013) Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep.*, **5**, 1178–1186.
25. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.
26. Madeira,F., Park,Y.M., Lee,J., Buso,N., Gur,T., Madhusoodanan,N., Basutkar,P., Tivey,A.R.N., Potter,S.C., Finn,R.D. *et al.* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res.*, **47**, W636–W641.
27. Fukuda,A., Kodama,Y., Mashima,J., Fujisawa,T. and Ogasawara,O. (2021) DDBJ update: streamlining submission and access of human data. *Nucleic Acids Res.*, **49**, D71–D75.
28. Alberti,S., Gladfelter,A. and Mittag,T. (2019) Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell*, **176**, 419–434.
29. Teng,X., Chen,X., Xue,H., Tang,Y., Zhang,P., Kang,Q., Hao,Y., Chen,R., Zhao,Y. and He,S. (2020) NPInter v4.0: an integrated database of ncRNA interactions. *Nucleic Acids Res.*, **48**, D160–D165.
30. Lin,Y., Liu,T., Cui,T., Wang,Z., Zhang,Y., Tan,P., Huang,Y., Yu,J. and Wang,D. (2020) RNAInter in 2020: RNA interactome repository with increased coverage and annotation. *Nucleic Acids Res.*, **48**, D189–D197.
31. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
32. Liao,Y., Smyth,G.K. and Shi,W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**, 923–930.
33. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, **15**, 550.
34. Howe,K.L., Achuthan,P., Allen,J., Allen,J., Alvarez-Jarreta,J., Amode,M.R., Armean,I.M., Azov,A.G., Bennett,R., Bhai,J. *et al.* (2021) Ensembl 2021. *Nucleic Acids Res.*, **49**, D884–D891.
35. Navarro Gonzalez,J., Zweig,A.S., Speir,M.L., Schmelter,D., Rosenbloom,K.R., Raney,B.J., Powell,C.C., Nassar,L.R., Maulding,N.D., Lee,C.M. *et al.* (2021) The UCSC Genome Browser database: 2021 update. *Nucleic Acids Res.*, **49**, D1046–D1057.
36. Volders,P.J., Helsen,K., Wang,X., Menten,B., Martens,L., Gevaert,K., Vandesompele,J. and Mestdagh,P. (2013) LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.*, **41**, D246–D251.
37. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
38. Zhao,Y., Li,H., Fang,S., Kang,Y., Wu,W., Hao,Y., Li,Z., Bu,D., Sun,N., Zhang,M.Q. *et al.* (2016) NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.*, **44**, D203–D208.
39. Benson,G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.*, **27**, 573–580.
40. Kalvari,I., Nawrocki,E.P., Ontiveros-Palacios,N., Argasinska,J., Lamkiewicz,K., Marz,M., Griffiths-Jones,S., Toffano-Nioche,C., Gautheret,D., Weinberg,Z. *et al.* (2021) Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res.*, **49**, D192–D200.
41. Lorenz,R., Bernhart,S.H., Honer Zu Siederdisen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol Biol*, **6**, 26.
42. Biesiada,M., Purzycka,K.J., Szachniuk,M., Blazewicz,J. and Adamiak,R.W. (2016) Automated RNA 3D structure prediction with RNAComposer. *Methods Mol. Biol.*, **1490**, 199–215.
43. Garant,J.M., Perreault,J.P. and Scott,M.S. (2018) G4RNA screener web server: user focused interface for RNA G-quadruplex prediction. *Biochimie*, **151**, 115–118.
44. Li,J.H., Liu,S., Zhou,H., Qu,L.H. and Yang,J.H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.*, **42**, D92–D97.
45. Zhu,Y., Xu,G., Yang,Y.T., Xu,Z., Chen,X., Shi,B., Xie,D., Lu,Z.J. and Wang,P. (2019) POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res.*, **47**, D203–D211.
46. Guarracino,A., Pepe,G., Ballesio,F., Adinolfi,M., Pietrosanto,M., Sangiovanni,E., Vitale,I., Ausiello,G. and Helmer-Citterich,M. (2021) BRIO: a web server for RNA sequence and structure motif scan. *Nucleic Acids Res.*, **49**, W67–W71.
47. Xuan,J.J., Sun,W.J., Lin,P.H., Zhou,K.R., Liu,S., Zheng,L.L., Qu,L.H. and Yang,J.H. (2018) RMBase v2.0: deciphering the map of RNA modifications from epitranscriptome sequencing data. *Nucleic Acids Res.*, **46**, D327–D334.



48. Zhou, Y., Zeng, P., Li, Y.H., Zhang, Z. and Cui, Q. (2016) SRAMP: prediction of mammalian N6-methyladenosine (m6A) sites based on sequence-derived features. *Nucleic Acids Res.*, **44**, e91.
49. Li, H., Chen, L., Huang, Z., Luo, X., Li, H., Ren, J. and Xie, Y. (2021) DeepOMe: a web server for the prediction of 2'-O-Me sites based on the hybrid CNN and BLSTM architecture. *Front. Cell Dev. Biol.*, **9**, 686894.
50. Lv, Z., Zhang, J., Ding, H. and Zou, Q. (2020) RF-PseU: a random forest predictor for RNA pseudouridine sites. *Front. Bioeng. Biotechnol.*, **8**, 134.
51. Li, Y.H., Zhang, G. and Cui, Q. (2015) PPUS: a web server to predict PUS-specific pseudouridine sites. *Bioinformatics*, **31**, 3362–3364.
52. Li, J., Huang, Y., Yang, X., Zhou, Y. and Zhou, Y. (2018) RNAm5Cfinder: a web-server for predicting RNA 5-methylcytosine (m5C) sites based on random forest. *Sci. Rep.*, **8**, 17299.
53. Amberger, J.S., Bocchini, C.A., Scott, A.F. and Hamosh, A. (2019) OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.*, **47**, D1038–D1043.
54. Pinero, J., Ramirez-Angueta, J.M., Sauch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F. and Furlong, L.I. (2020) The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res.*, **48**, D845–D855.
55. Ning, L., Cui, T., Zheng, B., Wang, N., Luo, J., Yang, B., Du, M., Cheng, J., Dou, Y. and Wang, D. (2021) MNDR v3.0: mammal ncRNA-disease repository with increased coverage and annotation. *Nucleic Acids Res.*, **49**, D160–D164.
56. Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q. and Dong, D. (2019) LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.*, **47**, D1034–D1037.
57. Xia, S., Feng, J., Chen, K., Ma, Y., Gong, J., Cai, F., Jin, Y., Gao, Y., Xia, L., Chang, H. *et al.* (2018) CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res.*, **46**, D925–D929.
58. Landrum, M.J., Lee, J.M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J. *et al.* (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.*, **44**, D862–D868.
59. Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E. *et al.* (2019) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.*, **47**, D941–D947.
60. Zhang, J., Bajari, R., Andric, D., Gerthoffert, F., Lepsa, A., Nahal-Bose, H., Stein, L.D. and Ferretti, V. (2019) The International Cancer Genome Consortium Data Portal. *Nat. Biotechnol.*, **37**, 367–369.
61. Cancer Genome Atlas Research, N., Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C. and Stuart, J.M. (2013) The cancer genome atlas pan-cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.
62. Rego, N. and Koes, D. (2015) 3Dmol.js: molecular visualization with WebGL. *Bioinformatics*, **31**, 1322–1324.
63. Chen, H., Cui, Y., Han, X., Hu, W., Sun, M., Zhang, Y., Wang, P.H., Song, G., Chen, W. and Lou, J. (2020) Liquid-liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res.*, **30**, 1143–1145.
64. Shin, Y. and Brangwynne, C.P. (2017) Liquid phase condensation in cell physiology and disease. *Science*, **357**, eaaf4382