

# RNALocate v2.0: an updated resource for RNA subcellular localization with increased coverage and annotation

Tianyu Cui<sup>1,†</sup>, Yiyang Dou<sup>1,†</sup>, Puwen Tan<sup>1,†</sup>, Zhen Ni<sup>2</sup>, Tianyuan Liu<sup>1</sup>, DuoLin Wang<sup>3</sup>, Yan Huang<sup>4</sup>, Kaican Cai<sup>2</sup>, Xiaoyang Zhao<sup>5</sup>, Dong Xu<sup>6,3,\*</sup>, Hao Lin<sup>6,\*</sup> and Dong Wang<sup>1,7,\*</sup>

<sup>1</sup>Department of Bioinformatics, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China, <sup>2</sup>Department of Thoracic Surgery, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China, <sup>3</sup>Department of Electrical Engineering and Computer Science, Bond Life Sciences Center, University of Missouri, Columbia, Missouri 65211, USA, <sup>4</sup>Shunde Hospital, Southern Medical University (The First People's Hospital of Shunde Foshan), Foshan 528308, China, <sup>5</sup>State Key Laboratory of Organ Failure Research, Department of Developmental Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China, <sup>6</sup>Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 611731, China and <sup>7</sup>Dermatology Hospital, Southern Medical University, Guangzhou 510091, China

Received July 07, 2021; Revised September 03, 2021; Editorial Decision September 06, 2021; Accepted September 07, 2021

## ABSTRACT

**Resolving the spatial distribution of the transcriptome at a subcellular level can increase our understanding of biology and diseases. To facilitate studies of biological functions and molecular mechanisms in the transcriptome, we updated RNALocate, a resource for RNA subcellular localization analysis that is freely accessible at <http://www.rnalocate.org/> or <http://www.rna-society.org/rnalocate/>. Compared to RNALocate v1.0, the new features in version 2.0 include (i) expansion of the data sources and the coverage of species; (ii) incorporation and integration of RNA-seq datasets containing information about subcellular localization; (iii) addition and reorganization of RNA information (RNA subcellular localization conditions and descriptive figures for method, RNA homology information, RNA interaction and ncRNA disease information) and (iv) three additional prediction tools: DM3Loc, iLoc-lncRNA and iLoc-mRNA. Overall, RNALocate v2.0 provides a comprehensive RNA subcellular localization resource for researchers to deconvolute the highly complex architecture of the cell.**

## INTRODUCTION

The subcellular localization of RNA plays an important role in cell growth and development, cell differentiation and inflammation, cell signal transduction and transcriptional regulation (1,2). At the cellular level, where an RNA is located likely determines whether it will be stored, processed, translated or degraded (3–5). Although the importance of RNA subcellular localization has been widely recognized, the related bioinformatics resources are limited compared to those available for protein localization. For example, a subcellular map of the human proteome, Human Protein Atlas (HPA), records detailed information about protein subcellular localization (6). Moreover, many resources also provide information about the subcellular localization of proteins, such as UniProt, PSORTdb, SubCellBarCode, MiCroKiTS 4.0 and SUBA4 (7–12). Corresponding protein subcellular localization technology and prediction tools include FLIRT, SUBCons, BUSCA and DeepMito (13–18).

There are already some databases for collecting the information of RNA subcellular localization at the transcriptome-wide level. For example, IncSLdb (19) stores subcellular localization of long noncoding RNAs (lncRNAs) from literature mining, LncATLAS (20) collects subcellular localization of lncRNAs from RNA-seq data, and EVmiRNA (21) involves the information of microRNAs (miRNAs) in extracellular vesicles. Several computational prediction tools, including DM3Loc (22), mRNAloc (23),

\*To whom correspondence should be addressed. Tel: +86 20 61648279; Fax: +86 20 61648279; Email: wangdong79@smu.edu.cn  
Correspondence may also be addressed to Hao Lin. Tel: +86 13678168394; Fax: +86 28 83208232; Email: hlin@uestc.edu.cn  
Correspondence may also be addressed to Dong Xu. Tel: +1 573 882 2299; Fax: +1 573 882 8318; Email: xudong@missouri.edu

<sup>†</sup>The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

lncLocator (24) and iLoc-lncRNA (25), were developed based on the first version of RNALocate (26). Of note, many experimental techniques for detecting RNA subcellular localizations have been developed in recent years, including APEX-Seq (27), proximity RNA-seq (28), MERFISH (29) and subRNA-seq (30), together with extensive new data. In view of these, this is the right time to update our database to RNALocate v2.0 (<http://www.rnalocate.org/> or <http://www.rna-society.org/rnalocate/>), that is the collection of RNA subcellular localization data from literatures, other databases and RNA-seq datasets.

RNALocate v2.0 is a repository of integrated experimentally validated information on subcellular localization of RNA through manual curation of the literature and five other resources, along with analyses of 35 datasets from the Gene Expression Omnibus (GEO) (31) under a common framework (Figure 1). It also supports three RNA subcellular localization prediction tools: DM3Loc, iLoc-lncRNA and iLoc-mRNA (32). In total, RNALocate v2.0 integrates more than 213 000 RNA subcellular localization entries at 171 locations across 104 species. This resource will provide a valuable resource for better understanding the subcellular localization of the transcriptome.

## MATERIALS AND METHODS

### Data collection

RNALocate v2.0 integrates RNA subcellular localization data from the literature, five databases and 35 RNA-seq datasets. Publications in PubMed (mainly from 2016 to 2021) were screened with the following keyword combinations: (localization name) AND (RNA molecule). ‘Localization name’ represents the subcellular localization name, and ‘RNA molecule’ represents RNA symbols or RNA category names. Finally, we reviewed over 35 000 published studies that included 38 508 RNA subcellular localization entries. The other 174 752 entries were integrated from five other databases, including CSCD, EVmiRNA, exoRBase, PomBase and TAIR (21,33–36). RNA-seq datasets from GEO were screened with the following criteria: species (*Homo sapiens* or *Mus musculus*), sequencing type (bulk RNA-seq or small RNA-seq), condition (delete unknown) and replicate ( $\geq 2$ ) and publication date (after 2016). RNALocate v2.0 adds over 200 new samples from 35 datasets of RNA-seq data with subcellular localization information.

To facilitate elucidating the role of RNA localization at the subcellular level, more annotation information was collected, including RNA subcellular localization conditions, methods and corresponding figures from the literature, RNA homology information from NCBI Gene (37), RNA interactions from RNAInter (38), and RNA-related diseases from MNDR v3.0 (39). Simultaneously, the transcript sequences from Refseq (40) and miRBase (41) were also included. For RNA-seq datasets, GEO accession, localization, sample condition, and PubMed ID were provided. In addition, the RNA expression and Gene Ontology (GO) enrichment results of the top 50 RNAs for each sample were also incorporated (42).

### Data processing

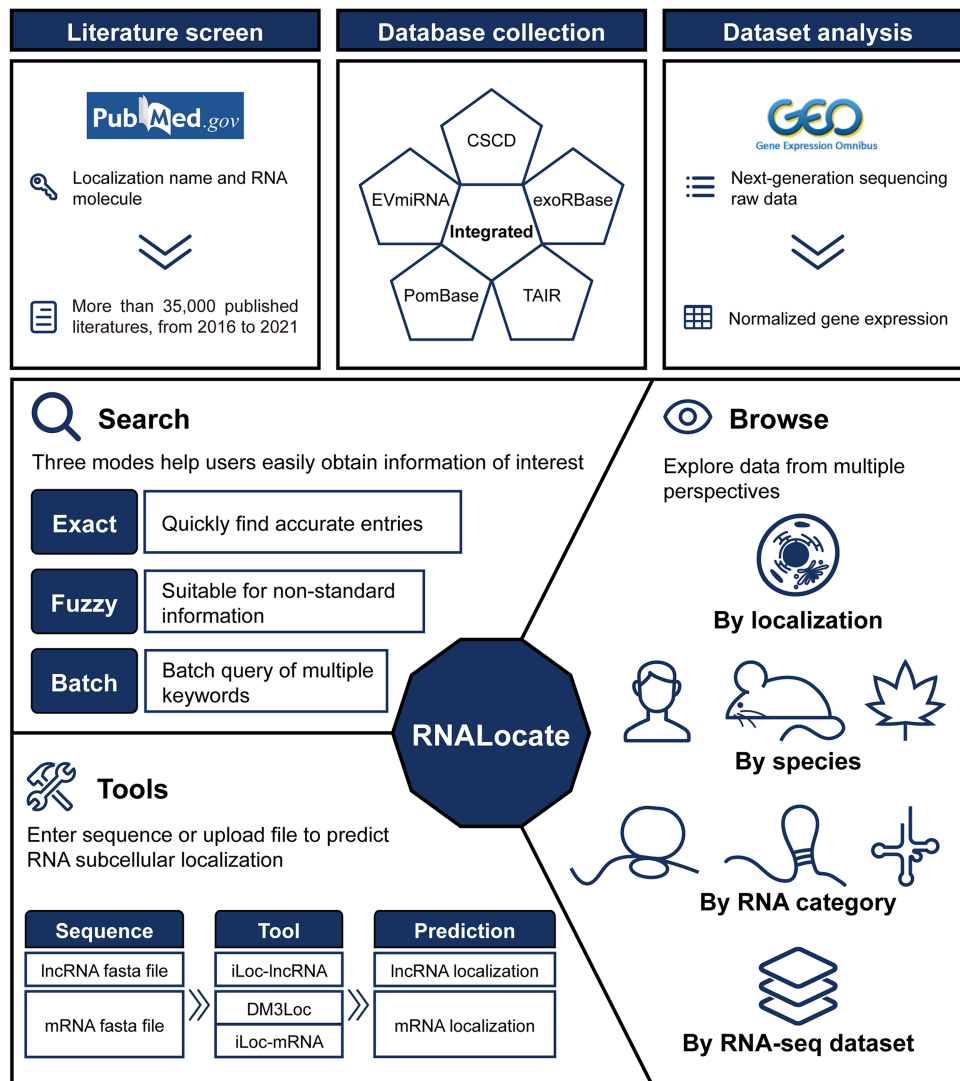
Integrating multisource data requires unifying them into common reference databases to annotate various RNAs. Major types of RNA symbols were used: (i) miRNA symbols from the miRBase database, (ii) messenger RNA (mRNA), small nucleolar RNA (snoRNA), small nuclear RNA (snRNA) and lncRNA symbols from the NCBI Gene database, (iii) ribosomal RNA (rRNA) and piwi interacting RNA (piRNA) symbols from the RNACentral (43) database and (iv) circular RNA (circRNA) symbols from the circBase (44) and exoRBase databases. Then, we reconstructed a hierarchical structure for all of the localizations according to the cellular component annotation curated in Gene Ontology. Additionally, miRBase accession, NCBI Gene ID, Ensembl Gene stable ID, RNACentral identifier, circBase ID, exoRBase ID and their external links were also provided, which can help to efficiently retrieve a substantial amount of RNA-associated information from external resources. For the convenience of users, the RNA-associated information also contains RNA names from the literature, aliases, and sequences, among others.

In particular, we screened and processed 203 samples from 35 RNA-seq datasets that had labels of the subcellular locations in 26 conditions and 13 cell lines. All datasets contained 15 subcellular locations (Figure 2B, Supplementary Table S1). The raw data were downloaded and processed by the NCBI SRA Toolkit v2.10.5 for format conversion, and then adaptor contaminants and low-quality bases were removed using Trimmomatic v0.39 (45). The processed clean reads were aligned to the human and mouse reference genomes (GRCh38 and GRCm38 from GENCODE) with gene annotations (Release 34 and M25 from GENCODE), and the gene expression of each sample was estimated using HISAT2 v2.1.0, SAMtools v1.4 and featureCounts v2.0.1 (46,47). The RNA expression levels were normalized by transcripts per million (TPM). All the data consisted of two independent biological replicates per sample (except for samples from APEX-seq, which have at least two replicates). In order to further analysis, we standardized the genes in each dataset similar to the approach of LncATLAS. Genes with TPM  $> 0$  in all replicates of at least one sample were retained (gene expressed in some replicates but not expressed in others were excluded). And removed genes with a greater than twofold difference between replicates.

## RESULTS

### New data and annotations

To improve the accuracy of our database, we carefully calibrated all of the data in the first release of the database and deleted 6739 entries that represent protein subcellular localizations and unclear localizations. In addition, we merged 2897 entries in which all of the information was the same except for cell lines or tissue types. In summary, RNALocate v2.0 contains 213 260 experimentally validated RNA subcellular localization entries, including 38 508 manually curated entries from the literature and 174 752 entries



**Figure 1.** Overview of the RNALocate v2.0 database.

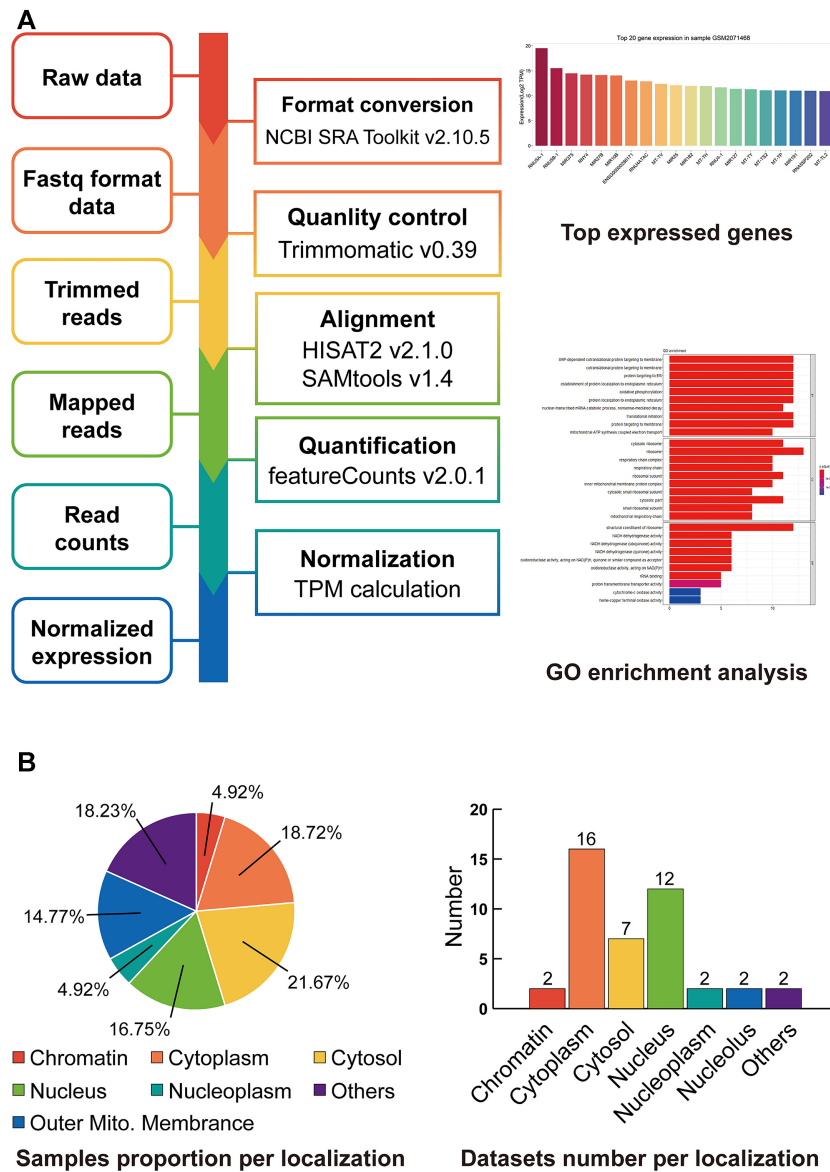
from databases. These entries involved 112,304 nonredundant RNAs and 16 newly added RNA types (such as circRNA, lincRNA, mtRNA, scRNA, scaRNA and Y RNA). The 129 new subcellular locations (such as chromatin, insoluble cytoplasm, mitochondrial cloud, and plasma membrane) were also added.

The distribution of the subcellular localizations among different RNA types is shown in Figure 3A and Supplementary Table S2. The number of species in RNALocate v2.0 increased from 65 to 104 compared with the first version. The species cover seven categories: apicomplexa, euglenozoa, fungi, metazoa, rhodophyta, viridiplantae, and viruses. The top three species are *Homo sapiens*, *Mus musculus*, and *Saccharomyces cerevisiae*, as shown in Figure 3B. Other model species, such as *Drosophila melanogaster*, *Rattus norvegicus* and zebrafish (*Danio rerio*), have also been documented in RNALocate v2.0. Of note, some RNA subcellular localizations that only occur under certain conditions are also recorded in our database.

### Features and utilities of RNALocate v2.0

RNALocate v2.0 provides a user-friendly platform for searching, browsing and profiling RNA subcellular localization data. To improve its search capability, RNALocate v2.0 provides search function for data from literature and RNA-seq dataset, respectively. For search from literature page, it enables an optimized query with a new function of fuzzy and batch search. 'Fuzzy Search' can help users search entries using nonstandardized RNA names and subcellular localization. Meanwhile, 'batch search' supports queries by a list of official symbols/IDs or a file upload to obtain associated entries.

Apart from basic annotations, such as RNA information, localization information, other subcellular localizations and ncRNA disease information, we modified the corresponding homology and interaction data in detail. The 'RNA-RNA interaction' presents only when both RNAs have subcellular localization information. Similarly, 'homology information' shows homologous RNA-associated



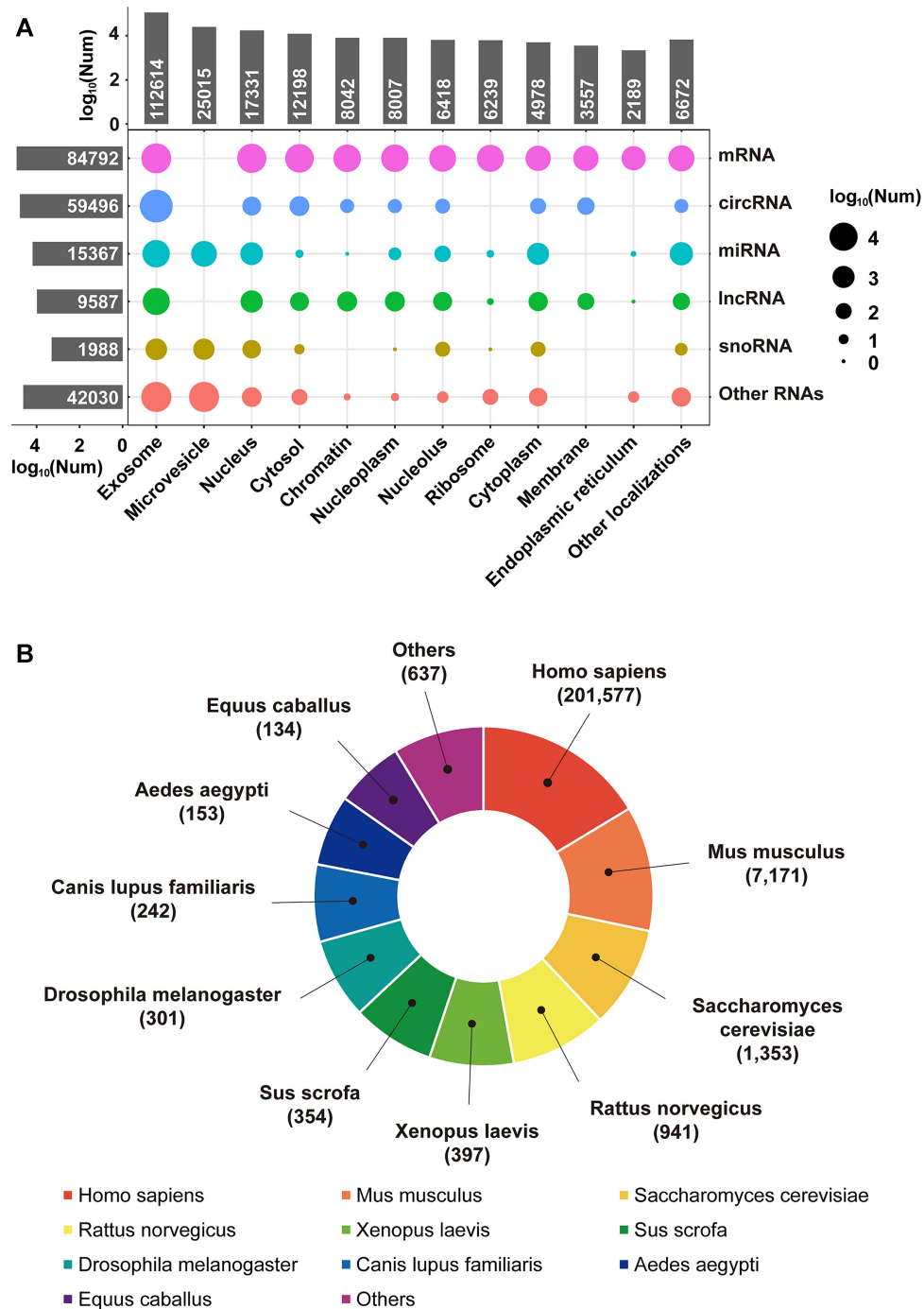
**Figure 2.** Summary of the RNA-seq datasets. (A) The workflow of RNA-seq datasets (left). Top 20 gene expression levels and top 50 gene functional annotations for each sample (right). (B) The proportion of samples and the number of datasets in 15 subcellular localizations.

entries instead of homologous genes. All of the information links to their corresponding databases. In addition, we added the method of RNA subcellular localization from literatures, and also included the corresponding conditions and figures.

To illustrate the different subcellular localizations of each RNA from RNA-seq datasets, the detail page of ‘Search From RNA-seq Dataset’ shows the basic information and subcellular localization in each RNA-seq dataset. Basic information included gene symbol, ensemble gene stable id, genome location and gene type. The latter included: (i) Subcellular localization (chromatin, cytoplasm and nucleoplasm) in different conditions (only in *Mus musculus*); (ii) Different subcellular localizations in individual cell type; (iii) Subcellular localizations revealed by APEX-Seq; (iv) Single subcellular localization in different conditions and

(v) Subcellular localizations of cytoplasm and nucleus in different cell lines (Supplementary Figure S1). The search result page of literature and the detail page of RNA-seq dataset can be switched to each other. ‘Browse By RNA-seq dataset’ shows the sample information, gene expression and gene GO enrichment analysis for each dataset on the ‘Browse’ page. GEO accessions, locations, sampling conditions and other information were included in detail page. And also provided a histogram of the top 20 RNA expressions and the result of the top 50 RNA functional annotations for each sample (Figure 2A). In addition, all of the RNA expression in each dataset can be downloaded.

In response to the diverse needs of users, RNALocate v2.0 incorporates three prediction tools: DM3Loc, iLoc-lncRNA and iLoc-mRNA (all prediction tools were trained on RNALocate v1.0). They are used to predict the



**Figure 3.** Statistics on RNALocate v2.0. (A) The distribution of 25 RNA categories in 171 subcellular localizations. (B) Number of entries in the top 10 species.

subcellular localizations of lncRNAs (iLoc-lncRNA) or mRNAs (DM3Loc and iLoc-mRNA).

### CONCLUSION AND FUTURE PERSPECTIVES

Here, we present a resource of RNA subcellular localization information, RNALocate v2.0, generated by information obtained from the literature, databases and RNA-seq datasets. It contains more than 213 000 RNA subcellular

localization entries, guiding and helping researchers perform further studies. RNALocate v2.0 integrates RNA-seq data with subcellular localization to quantify the expression of RNAs at the subcellular level. In addition, RNALocate v2.0 also incorporates three prediction tools for the various needs of users.

The biological functions of RNAs are usually influenced by their localizations. The fact that RNAs are located at multiple subcellular localizations also increases the

complexity of the cell. The analysis of the protein-protein interaction network at the subcellular level has been confirmed to have a unique effect different from the cellular level, and corresponding methods have also emerged, such as CellWhere and ComPPI (48–50). Because of this, we expect that continuing to expand and improve RNALocate v2.0 can also help explore the RNA-RNA interaction network at the subcellular level in the future. Thus, RNALocate is the most comprehensive map of the subcellular localization of the transcriptome and it can satisfy different requirements.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Key Research and Development Project of China [2017YFA0105001, 2019YFA0801800]; National Natural Science Foundation of China [82070109, 81770104, 62002153]; Guangdong Basic and Applied Basic Research Foundation [2019A1515010784, 2019A1515110701]; China Postdoctoral Science Foundation [2020M682623, 2020M682785]; Paul K. and Diane Shumaker Endowment Fund at University of Missouri. Funding for open access charge: National Key Research and Development Project of China.

*Conflict of interest statement.* None declared.

## REFERENCES

- Buxbaum, A.R., Haimovich, G. and Singer, R.H. (2015) In the right place at the right time: visualizing and understanding mRNA localization. *Nat. Rev. Mol. Cell Biol.*, **16**, 95–109.
- Mofatteh, M. and Bullock, S.L. (2017) SnapShot: subcellular mRNA localization. *Cell*, **169**, 178–178.
- Berkovits, B.D. and Mayr, C. (2015) Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature*, **522**, 363–367.
- Lecuyer, E., Yoshida, H., Parthasarathy, N., Alm, C., Babak, T., Cerovina, T., Hughes, T.R., Tomancak, P. and Krause, H.M. (2007) Global analysis of mRNA localization reveals a prominent role in organizing cellular architecture and function. *Cell*, **131**, 174–187.
- Carlevaro-Fita, J., Rahim, A., Guigo, R., Vardy, L.A. and Johnson, R. (2016) Cytoplasmic long noncoding RNAs are frequently bound to and degraded at ribosomes in human cells. *RNA*, **22**, 867–882.
- Thul, P.J., Akesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund, A., Bjork, L., Breckels, L.M. *et al.* (2017) A subcellular map of the human proteome. *Science*, **356**, eaal3321.
- UniProt, C. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
- Lau, W.Y.V., Hoad, G.R., Jin, V., Winsor, G.L., Madyan, A., Gray, K.L., Laird, M.R., Lo, R. and Brinkman, F.S.L. (2021) PSORTdb 4.0: expanded and redesigned bacterial and archaeal protein subcellular localization database incorporating new secondary localizations. *Nucleic Acids Res.*, **49**, D803–D808.
- Orre, L.M., Vesterlund, M., Pan, Y., Arslan, T., Zhu, Y., Fernandez Woodbridge, A., Frings, O., Fredlund, E. and Lehtio, J. (2019) SubCellBarCode: proteome-wide mapping of protein localization and relocalization. *Mol. Cell*, **73**, 166–182.
- Huang, Z., Ma, L., Wang, Y., Pan, Z., Ren, J., Liu, Z. and Xue, Y. (2015) MiCroKiTS 4.0: a database of midbody, centrosome, kinetochore, telomere and spindle. *Nucleic Acids Res.*, **43**, D328–D334.
- Hooper, C.M., Castleden, I.R., Tanz, S.K., Aryamanesh, N. and Millar, A.H. (2017) SUBA4: the interactive data analysis centre for Arabidopsis subcellular protein locations. *Nucleic Acids Res.*, **45**, D1064–D1074.
- Huang, Y., Wang, J., Zhao, Y., Wang, H., Liu, T., Li, Y., Cui, T., Li, W., Feng, Y., Luo, J. *et al.* (2021) cncRNAdb: a manually curated resource of experimentally supported RNAs with both protein-coding and noncoding function. *Nucleic Acids Res.*, **49**, D65–D70.
- Hirsch, S.M., Sundaramoorthy, S., Davies, T., Zhuravlev, Y., Waters, J.C., Shirasu-Hiza, M., Dumont, J. and Canman, J.C. (2018) FLIRT: fast local infrared thermogenetics for subcellular control of protein function. *Nat. Methods*, **15**, 921–923.
- Salvatore, M., Warholm, P., Shu, N., Basile, W. and Elofsson, A. (2017) SubCons: a new ensemble method for improved human subcellular localization predictions. *Bioinformatics*, **33**, 2464–2470.
- Savojardo, C., Martelli, P.L., Fariselli, P., Profiti, G. and Casadio, R. (2018) BUSCA: an integrative web server to predict subcellular localization of proteins. *Nucleic Acids Res.*, **46**, W459–W466.
- Savojardo, C., Bruciaferri, N., Tartari, G., Martelli, P.L. and Casadio, R. (2020) DeepMito: accurate prediction of protein sub-mitochondrial localization using convolutional neural networks. *Bioinformatics*, **36**, 56–64.
- Marx, V. (2015) Mapping proteins with spatial proteomics. *Nat. Methods*, **12**, 815–819.
- Stadler, C., Rexhepaj, E., Singan, V.R., Murphy, R.F., Pepperkok, R., Uhlen, M., Simpson, J.C. and Lundberg, E. (2013) Immunofluorescence and fluorescent-protein tagging show high correlation for protein localization in mammalian cells. *Nat. Methods*, **10**, 315–323.
- Wen, X., Gao, L., Guo, X., Li, X., Huang, X., Wang, Y., Xu, H., He, R., Jia, C. and Liang, F. (2018) lncSLdb: a resource for long non-coding RNA subcellular localization. *Database*, **2018**, bay085.
- Mas-Ponte, D., Carlevaro-Fita, J., Palumbo, E., HERNANDEZ Pulido, T., Guigo, R. and Johnson, R. (2017) LncATLAS database for subcellular localization of long noncoding RNAs. *RNA*, **23**, 1080–1087.
- Liu, T., Zhang, Q., Zhang, J., Li, C., Miao, Y.R., Lei, Q., Li, Q. and Guo, A.Y. (2019) EVmiRNA: a database of miRNA profiling in extracellular vesicles. *Nucleic Acids Res.*, **47**, D89–D93.
- Wang, D., Zhang, Z., Jiang, Y., Mao, Z., Wang, D., Lin, H. and Xu, D. (2021) DM3Loc: multi-label mRNA subcellular localization prediction and analysis based on multi-head self-attention mechanism. *Nucleic Acids Res.*, **49**, e46.
- Garg, A., Singhal, N., Kumar, R. and Kumar, M. (2020) mRNAloc: a novel machine-learning based in-silico tool to predict mRNA subcellular localization. *Nucleic Acids Res.*, **48**, W239–W243.
- Cao, Z., Pan, X., Yang, Y., Huang, Y. and Shen, H.B. (2018) The lncLocator: a subcellular localization predictor for long non-coding RNAs based on a stacked ensemble classifier. *Bioinformatics*, **34**, 2185–2194.
- Su, Z.D., Huang, Y., Zhang, Z.Y., Zhao, Y.W., Wang, D., Chen, W., Chou, K.C. and Lin, H. (2018) iLoc-lncRNA: predict the subcellular location of lncRNAs by incorporating octamer composition into general PseKNC. *Bioinformatics*, **34**, 4196–4204.
- Zhang, T., Tan, P., Wang, L., Jin, N., Li, Y., Zhang, L., Yang, H., Hu, Z., Zhang, L., Hu, C. *et al.* (2017) RNALocate: a resource for RNA subcellular localizations. *Nucleic Acids Res.*, **45**, D135–D138.
- 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.
- Morf, J., Wingett, S.W., Farabella, I., Cairns, J., Furlan-Magaril, M., Jimenez-Garcia, L.F., Liu, X., Craig, F.F., Walker, S., Segonds-Pichon, A. *et al.* (2019) RNA proximity sequencing reveals the spatial organization of the transcriptome in the nucleus. *Nat. Biotechnol.*, **37**, 793–802.
- Xia, C., Fan, J., Emanuel, G., Hao, J. and Zhuang, X. (2019) Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *PNAS*, **116**, 19490–19499.
- Mayer, A. and Churchman, L.S. (2017) A detailed protocol for subcellular RNA sequencing (subRNA-seq). *Curr. Protoc. Mol. Biol.*, **120**, 4.29.1–4.29.18.
- 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.
- Zhang, Z.Y., Yang, Y.H., Ding, H., Wang, D., Chen, W. and Lin, H. (2021) Design powerful predictor for mRNA subcellular location prediction in Homo sapiens. *Brief. Bioinform.*, **22**, 526–535.

33. Li,S., Li,Y., Chen,B., Zhao,J., Yu,S., Tang,Y., Zheng,Q., Li,Y., Wang,P., He,X. *et al.* (2018) exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic Acids Res.*, **46**, D106–D112.
34. 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.
35. Lock,A., Rutherford,K., Harris,M.A., Hayles,J., Oliver,S.G., Bahler,J. and Wood,V. (2019) PomBase 2018: user-driven reimplementaation of the fission yeast database provides rapid and intuitive access to diverse, interconnected information. *Nucleic Acids Res.*, **47**, D821–D827.
36. Berardini,T.Z., Reiser,L., Li,D., Mezheritsky,Y., Muller,R., Strait,E. and Huala,E. (2015) The Arabidopsis information resource: making and mining the “gold standard” annotated reference plant genome. *Genesis*, **53**, 474–485.
37. Brown,G.R., Hem,V., Katz,K.S., Ovetsky,M., Wallin,C., Ermolaeva,O., Tolstoy,I., Tatusova,T., Pruitt,K.D., Maglott,D.R. *et al.* (2015) Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.*, **43**, D36–D42.
38. 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.
39. 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.
40. Haft,D.H., DiCuccio,M., Badretin,A., Brover,V., Chetvernin,V., O’Neill,K., Li,W., Chitsaz,F., Derbyshire,M.K., Gonzales,N.R. *et al.* (2018) RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.*, **46**, D851–D860.
41. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
42. The Gene Ontology,C. (2019) The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.*, **47**, D330–D338.
43. The,R.C. (2019) RNACentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res.*, **47**, D221–D229.
44. Glazar,P., Papavasileiou,P. and Rajewsky,N. (2014) circBase: a database for circular RNAs. *RNA*, **20**, 1666–1670.
45. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
46. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and Genome Project Data Processing, S. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
47. 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.
48. Ma,W. and Mayr,C. (2018) A membraneless organelle associated with the endoplasmic reticulum enables 3’UTR-mediated protein-protein interactions. *Cell*, **175**, 1492–1506.
49. Zhu,L., Malatras,A., Thorley,M., Aghoghobe,I., Mer,A., Duguez,S., Butler-Browne,G., Voit,T. and Duddy,W. (2015) CellWhere: graphical display of interaction networks organized on subcellular localizations. *Nucleic Acids Res.*, **43**, W571–W575.
50. Veres,D.V., Gyurko,D.M., Thaler,B., Szalay,K.Z., Fazekas,D., Korcsmaros,T. and Csermely,P. (2015) ComPPI: a cellular compartment-specific database for protein-protein interaction network analysis. *Nucleic Acids Res.*, **43**, D485–D493.