Nucleic Acids Research, 2025, **53**, D284–D292
https://doi.org/10.1093/nar/gkae872
Advance access publication date: 15 October 2024
Database issue

RNALocate v3.0: Advancing the Repository of RNA Subcellular Localization with Dynamic Analysis and Prediction

Le Wu[1,†], Luqi Wang[1,†], Shijie Hu[1,2,†], Guangjue Tang[1,†], Jia Chen[1,†], Ying Yi[3], Hailong Xie[1], Jiahao Lin[1], Mei Wang[4], Dong Wang[3,5,*], Bin Yang[3,*] and Yan Huang[6,*]

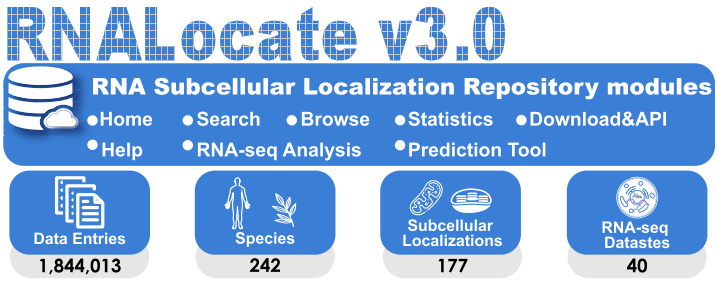
[1]Department of Bioinformatics, School of Basic Medical Sciences, Southern Medical University, No.1023, South Shatai Road, Baiyun District, Guangzhou 510515, China
[2]Department of Pathology, Harbin Medical University, 157th Rd of Baojian, Nangang Distinct, Harbin 150081, China
[3]Dermatology Hospital, Southern Medical University, No.2, Lujing Road, Yuexiu District, Guangzhou 510091, China
[4]State Key Laboratory of Organ Failure Research, Department of Developmental Biology, School of Basic Medical Sciences, Southern Medical University, No.1023, South Shatai Road, Baiyun District, Guangzhou 510515, China
[5]Department of Bioinformatics, Guangdong Province Key Laboratory of Molecular Tumor Pathology, School of Basic Medical Sciences, Southern Medical University, No.1023, South Shatai Road, Baiyun District, Guangzhou 510515, China
[6]Cancer Research Institute, School of Basic Medical Sciences, Southern Medical University, No.1023, South Shatai Road, Baiyun District, Guangzhou 510515, China

[*]To whom correspondence should be addressed. Tel: +86 20 61648279; Fax: +86 20 61648279; Email: huangyan24@smu.edu.cn
Correspondence may also be addressed to Bin Yang. Tel: +86 20 61648279; Fax: 86 20 61648279; Email: yangbin1@smu.edu.cn
Correspondence may also be addressed to Dong Wang. Tel: +86 20 61648279; Fax: 86 20 61648279; Email: wangdong79@smu.edu.cn
[†]The first five authors should be regarded as Joint First Authors.

Abstract

Subcellular localization of RNA is a crucial mechanism for regulating diverse biological processes within cells. Dynamic RNA subcellular localizations are essential for maintaining cellular homeostasis; however, their distribution and changes during development and differentiation remain largely unexplored. To elucidate the dynamic patterns of RNA distribution within cells, we have upgraded RNALocate to version 3.0, a repository for RNA-subcellular localization (http://www.rnalocate.org/ or http://www.rna-society.org/rnalocate/). RNALocate v3.0 incorporates and analyzes RNA subcellular localization sequencing data from over 850 samples, with a specific focus on the dynamic changes in subcellular localizations under various conditions. The species coverage has also been expanded to encompass mammals, non-mammals, plants and microbes. Additionally, we provide an integrated prediction algorithm for the subcellular localization of seven RNA types across eleven subcellular compartments, utilizing convolutional neural networks (CNNs) and transformer models. Overall, RNALocate v3.0 contains a total of 1 844 013 RNA-localization entries covering 26 RNA types, 242 species and 177 subcellular localizations. It serves as a comprehensive and readily accessible data resource for RNA-subcellular localization, facilitating the elucidation of cellular function and disease pathogenesis.

Graphical abstract



Introduction

RNA, a highly complex molecule universally involved in various cellular biological processes, has increasingly garnered attention regarding the relationship between its distribution and function within cells (1). Numerous studies indicate that RNA localization impacts cellular physiological functions at

Received: July 21, 2024. **Revised:** September 18, 2024. **Editorial Decision:** September 20, 2024. **Accepted:** September 24, 2024

subcellular, cellular, tissue and organismal levels (2–6). This widespread biological phenomenon occurs across different cell types and species, observed under conditions of homeostasis, stimulation or cellular stress (2,4,7). Despite the increasing recognition of the close relationship between RNA subcellular localization and function, there remains a significant limitation in the comprehensive integration of dynamic subcellular RNA localization data. Zhou *et al.* mapped the dynamic subcellular localization of RNA during human embryonic stem cells (hESCs) differentiation, revealing various RNA localization patterns that provided new insights into hESCs pluripotency maintenance and differentiation (8). Similarly, Hwang *et al.* demonstrated the crucial role of dynamic RNA subcellular localization in regulating Xenopus oocytes maturation (9). Fonseca *et al.* investigated Arabidopsis thaliana roots and found that nitrate treatments induce dynamic changes in messenger RNA (mRNA) nucleocytoplasmic distribution, highlighting the critical role of RNA localization dynamics in fine-tuning gene expression during the nitrate response and identifying a key adaptive mechanism in plants (10). During early Drosophila embryogenesis, 70% of mRNAs exhibited subcellular localization and dynamic movement, which were crucial for establishing anterior-posterior polarity (11). These findings collectively underscore the critical importance of dynamic RNA subcellular localization in regulating cell differentiation fate and influencing developmental processes. Moreover, a recent study also proposed a comprehensive framework to provide a dynamic overview of RNA and protein subcellular localization (12). To further elucidate the dynamic and specific localization of RNA at subcellular resolution and its association with complex biological processes, it is imperative to integrate, analyze and summarize RNA dynamic subcellular localization data under various conditions and cell stages.

Hence, we have now upgraded RNALocate from version 2.0 to version 3.0 (http://www.rnalocate.org/ or http://www.rna-society.org/rnalocate/). RNALocate v3.0 integrates manual curation of numerous literature, other experimentally validated databases, prediction algorithms and RNA sequencing data from 40 datasets under a unified framework (Figure 1). Meanwhile, leveraging experimental validation data, we have developed a new multi-label RNA localization prediction tool covering seven RNA types and eleven subcellular localizations. Additionally, the website framework has been redesigned to facilitate faster retrieval and browsing of entries through an enhanced, user-friendly interface. These improvements significantly enhance the operation of this system, enabling users to quickly and accurately access RNA-associated subcellular localizations deposited in the database. It provides a reliable and comprehensive data resource, assisting researchers in better understanding transcriptome dynamics at subcellular resolution.

Materials and methods

Data collection and organization

RNALocate v3.0 comprises three types of data: experimentally validated data, computationally predicted data, and RNA sequencing data pertaining to subcellular localizations. We have identified and added 191 803 RNA subcellular localization entries by manually reviewing the published literature, and integrated nine other related experimentally validated databases (Supplementary Table S1) (13–19). Besides, RNALocate utilizes both single-label and multi-label prediction algorithms that have become available in recent years. The focus is particularly on those developed within the last five years, including DeepLncLoc (20), DM3Loc (21), GraphLncLoc (22), iLoc-lncRNA (23), iLoc-LncRNA (2.0) (24), iLoc-mRNA (25), lncLocator (26), Locate-R (27), mRNALoc (28), mRNALocater (29), RNAlight (30), Clarion (31), EL-RMLocNet (32), LncLocFormer (33), miRNALoc (34), DeepLocRNA (35) to predict the subcellular localization of mRNA, long noncoding RNA (lncRNA), microRNA (miRNA) and small nucleolar RNA (snoRNA). Furthermore, RNALocate v3.0 integrates and comprehensively analyzes RNA sequencing data related to subcellular localizations under different conditions from the Gene Expression Omnibus (GEO), Sequence Read Archive (SRA), ArrayExpress and European Nucleotide Archive (ENA) (36–38).

To standardize the data and enhance its reference value, we meticulously linked data from disparate sources to authoritative reference databases, providing detailed annotations in RNALocate v3.0. Major types of RNA symbols were utilized in the study: (i) miRNA symbols from miRBase (39) and NCBI Gene database (40); (ii) circular RNA (circRNA) symbols from circBase (41), circBank (42) and exoRBase (18); (iii) lncRNA symbols from NCBI Gene and LncBook (43); (iv) piwi-interacting RNA (piRNA) symbols from piRBase (44) and RNAcentral (45); and (v) other RNAs from NCBI Gene, Ensembl (46), EnsemblRapid (46) and RNAcentral. Subcellular localization terms were derived from the cellular component annotations curated in the Gene Ontology. Additionally, RNA homology information was obtained from the NCBI Gene, RNA-related diseases from RNADisease v4.0 (47), and RNA interactions from RNAInter v4.0 (48). For the convenience of users, the RNA-associated information also contains RNA names and RNA-related functions extracted from the literature, as well as aliases, sequences and genome positions for miRNAs, circRNAs and lncRNAs from LncBook, among others.

RNA-seq analysis

We screened, processed, and analyzed 858 samples from 40 RNA-seq datasets annotated with the subcellular locations across 44 tissues and cell lines to visualize the dynamic RNA expression patterns within different subcellular compartments under various conditions. All datasets included 32 subcellular locations, and some also encompassed the RNA content of the entire cell (Supplementary Table S2). Raw data were subjected to quality control using FastQC v0.12.1 (https://github.com/s-andrews/FastQC) after download. Adapter contaminants and low-quality bases were subsequently removed using Cutadapt v1.18 (49) and Trimmomatic v0.39 (50). The processed clean reads were aligned to the reference genomes of *Homo sapiens*, *Mus musculus*, *Rattus norvegicus* and *Xenopus laevis* (GENCODE GRCh38.p14 (v45) (51), GENCODE GRCm39 (vM34), Ensembl mRatBN7.2 (release-112) and Xenbase XENLA_10.1 (52)) using HISAT2 v2.2.1 (53). Only uniquely mapped reads were extracted for further analysis. Aligned bam files underwent sorting and indexing using SAMtools v1.5 (54). Gene expression of each sample was estimated using featureCounts v2.0.1 (55).

The RNA expression levels were normalized by transcripts per million (TPM). Each subcellular location in every dataset

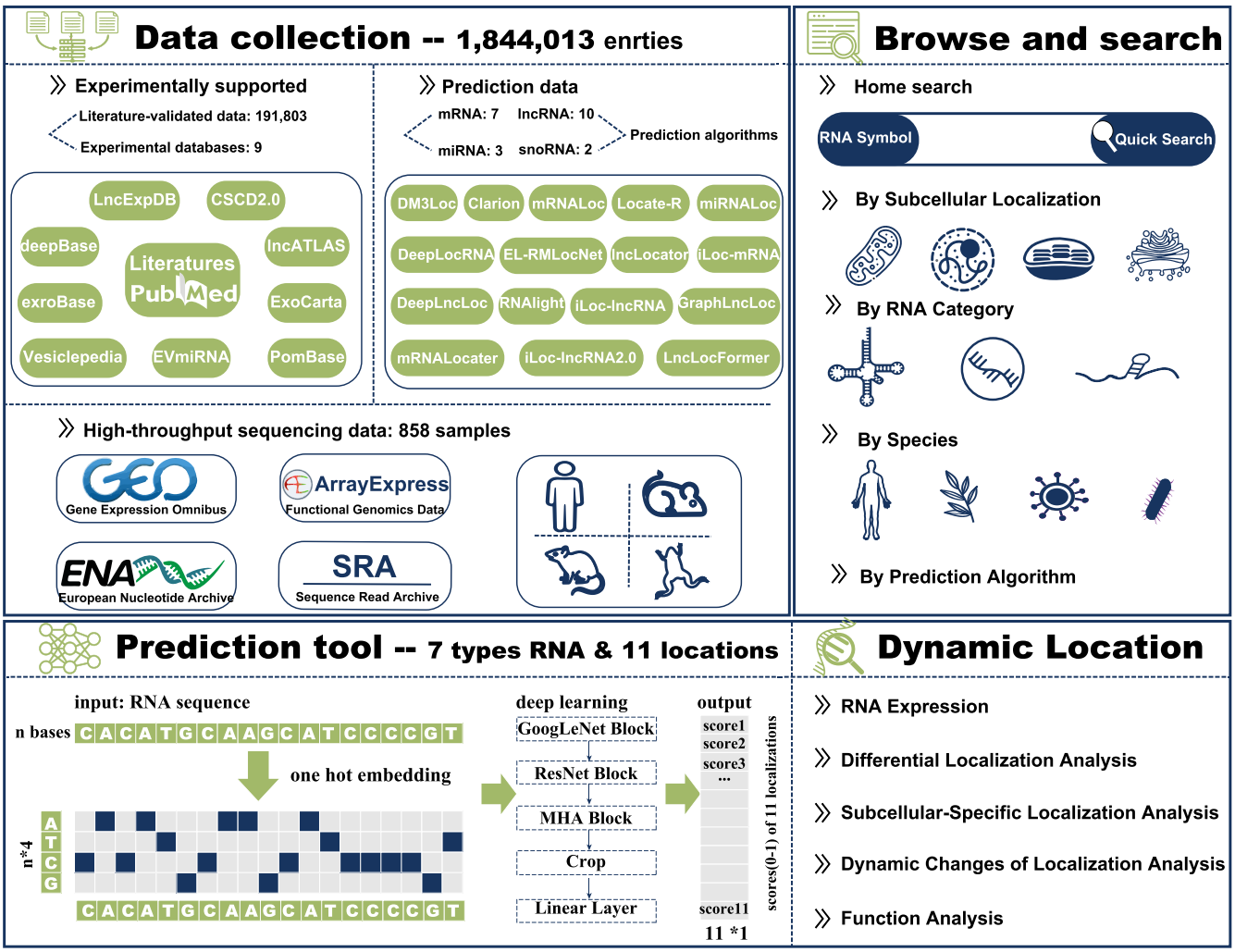


Figure 1. Overview of the RNALocate v3.0 repository.

included a minimum of two independent biological replicates under each condition. Genes with TPM > 1 in at least two samples per dataset were retained for further analysis. For each gene, if at least two biological replicates demonstrated TPM > 0 in a specific subcellular location under a given condition within a dataset, the median or mean TPM was employed as its expression value; otherwise, the expression value was set to 0. Differential expression analysis of various subcellular localizations under a single condition (with at least two independent biological replicates) was conducted using DESeq2 (56) and edgeR (57). Genes exhibiting significant up-regulation with a false discovery rate (FDR) < 0.05 and a fold change (FC) > 1.2 were classified as being enriched in specific subcellular localizations, and the subcellular-specific RNAs represent the intersection of differentially up-regulated RNAs between a given subcellular localization and all other subcellular localizations. We performed comprehensive analyses of RNAs with compartment-specific localization, focusing on their secondary structures, RNA-binding protein (RBP) binding preferences, conservation and functional enrichment, which were conducted on datasets for *Homo Sapiens* and *Mus musculus*. For each RNA, the longest transcript was used to predict secondary structures and minimum free energy (MFE) using RNAfold (58), and then the normalized MFE (NMFE) was computed. Potential RBPs were identified with RBPmap (59) using high stringency settings. PhastCons element annotations (hg38.phastCons100way.bw and mm39.phastCons35way.bw) were downloaded from UCSC Table Browser (60). Functional annotation was performed on differentially expressed mRNAs and the target genes of differentially expressed miRNAs predicted by miRWalk (61), including GO (62) and KEGG (63) analyses. Typically, the top 20 most significantly related pathways or functions are highlighted (Figure 2).

Meanwhile, we provided the relative ratios among different subcellular localizations under the same or multiple conditions, including various tissues and cell lines. Ratio 1 represents the TPM of a specific RNA localization divided by the total TPM in the cell. Ratio 2 compares the TPM of an RNA localization to the average TPM across all localizations in the same condition, highlighting its relative abundance. Ratio 3 compares the TPM of an RNA localization under a specific condition to the overall average TPM across all conditions. Ratio 4 compares the TPM of an RNA localization with that of a reference compartment (e.g., nucleus or cytoplasm), assessing the RNA's enrichment in the specific compartment.

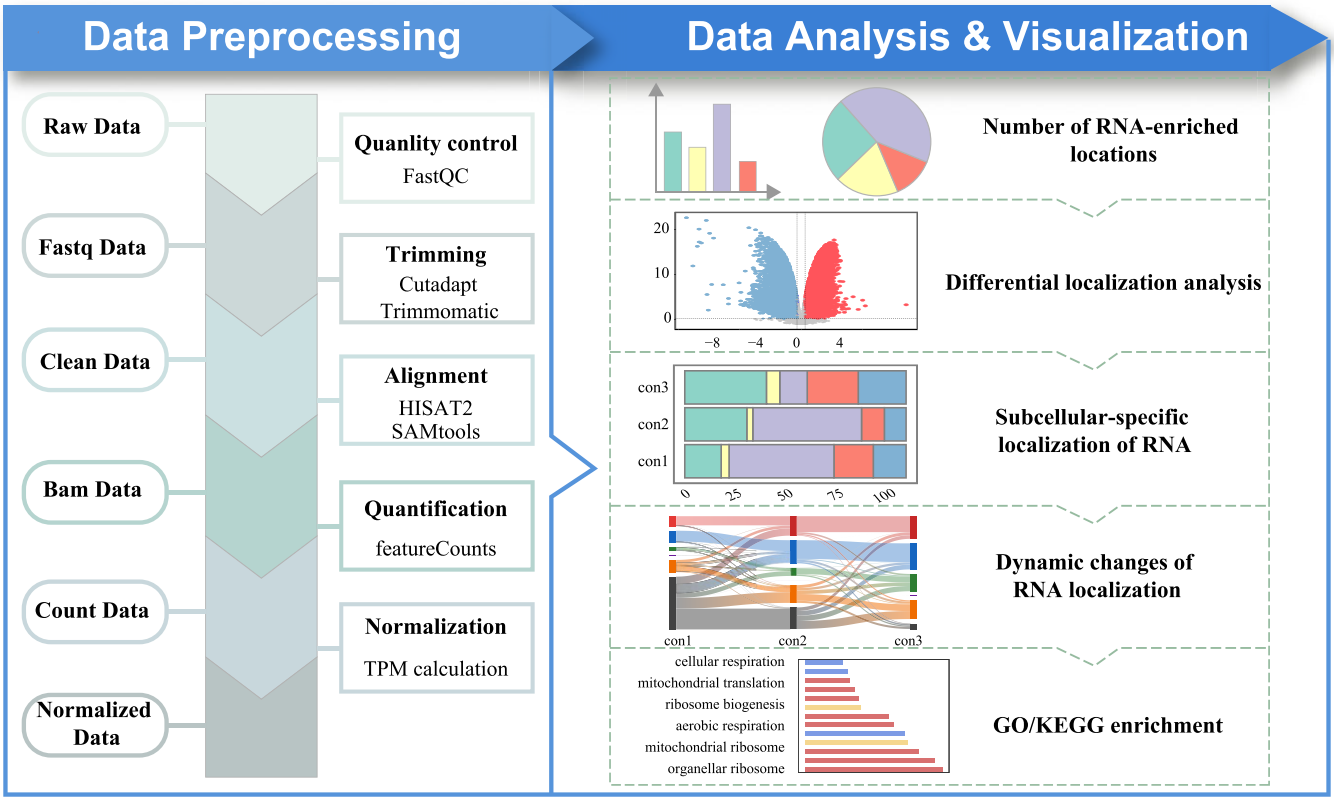


Figure 2. Introduction and usage of the RNA-seq Analysis page.

Prediction tool

RNALocate provides an integrated prediction algorithm for the subcellular localization of seven RNA types across eleven different subcellular compartments. This algorithm employs convolutional neural networks (CNNs) and a multi-head self-attention mechanism, validated through our experimental data. The seven RNA types include mRNA, lncRNA, miRNA, piRNA, snoRNA, small nuclear RNA (snRNA) and circRNA, and the eleven subcellular localizations are extracellular region, nucleoplasm (nucleus), nucleolus (nucleus), chromatin (nucleus), cytosol (cytoplasm), mitochondrion (cytoplasm), ribosome (cytoplasm), endoplasmic reticulum (cytoplasm), membrane, nucleus, and cytoplasm.

The prediction tool comprises four major modules: (i) GoogLeNet Blocks with a pooling layer, (ii) sixteen ResNet Blocks with pooling layers, (iii) a six-layer multi-head self-attention Block, and (iv) cropping and linear layers. The tool applies one-hot encoding for DNA/RNA sequences (A = [1, 0, 0, 0], C = [0, 1, 0, 0], G = [0, 0, 1, 0], T/U = [0, 0, 0, 1], O = [0, 0, 0, 0]) as the model's input feature, with a maximum sequence length of 8192 bp, and ultimately predicts the subcellular localization of RNA. The tool employs a GoogLeNet Block to simulate various sequence K-mer fragments (1-mer, 3-mer, 5-mer, and 7-mer) by configuring convolutional kernels of varying sizes, thereby capturing sequence features. The feature space is expanded to 64 dimensions, and a pooling mechanism condenses the sequence length to 4096 bp. Next, the ResNet Block, based on the ResNet-50 architecture, is integrated with pooling layers to further learn sequence features. This stage expands the feature space dimension to 2048 and aggregates the sequence length into 256 bins, with each bin encapsulating 32 bp of sequence information. The tool then incorporates a multi-head self-attention Block designed to model interactions between these bins. Within this Block, positional information is enriched with Relative Positional Encoding (RoPE), enhancing the model's sensitivity to the sequence's context. Subsequently, a cropping layer is implemented to eliminate potentially unreliable or erroneous information from the extremities of the sequence, which may be less reliable than the central regions. Finally, a multi-layer linear neural network predicts the outputs of RNA localization. Notably, the pooling layer uses attention-based pooling, as opposed to traditional max pooling, for enhanced performance. The training and testing datasets for our predictive tool were derived from the experimental data of RNALocate v3.0, specifically those with a score of 0.3 or higher, with a total of 559 651 entries. The model was trained and validated using a 5-fold cross-validation approach to ensure robustness and reliability. Additional algorithmic details are provided in the Supplementary Data. The framework of our prediction tool is illustrated in Supplementary Figure S1.

Results

RNALocate statistics

RNALocate v3.0 contains a total of 1 258 724 experimentally validated entries, with 191 803 entries manually curated (marked by * in RNALocate) and 1 066 921 entries obtained from other databases, covering 242 species, 26 RNA types and 177 subcellular locations (Figure 3A–C and Supplementary Table S3). Additionally, the database includes 319 625 predicted entries associated with mRNAs, 245 999 predicted entries associated with lncRNAs, 18 129 predicted entries asso-

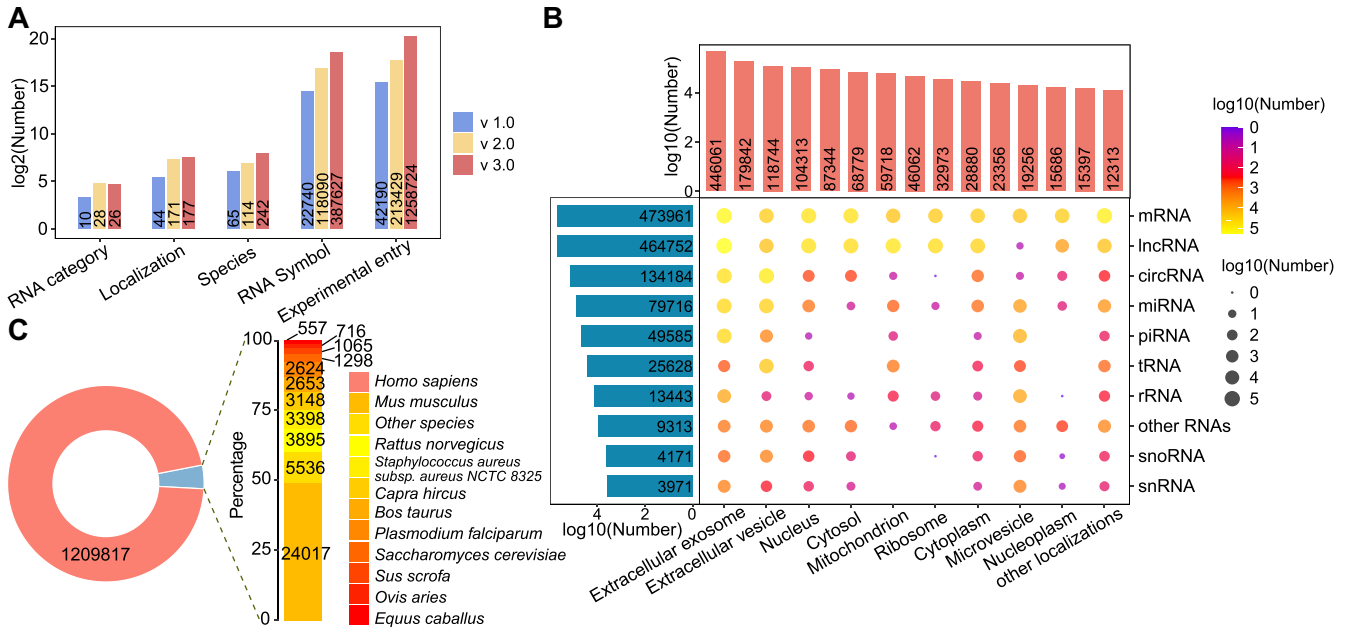


Figure 3. Statistics of RNALocate v3.0. (**A**) Features and development across RNALocate versions. (**B**) The distribution of experimentally validated RNA-localization associations for 26 types of RNA in 177 subcellular localizations. (**C**) Number of entries in the top 12 species with experimentally validated entries exceeding 500.

ciated with miRNAs and 1 536 predicted entries associated with snoRNAs for human.

Database usage

RNALocate v3.0 provides a user-friendly platform tailored to meet diverse research needs. In addition to offering essential information such as RNA information, subcellular localizations, references and official annotations, it also includes experimental validation methods and classifications of strong and weak evidence, which can be accessed on the 'Detail' page. Notably, users can access data in three ways: (i) a quick search on the 'Home' page based on the RNA symbol, RNA ID, subcellular localization or Gene Ontology term; (ii) a search on the 'Search' page utilizing 'Exact Search', 'Fuzzy Search' or 'Batch Search' options; and (iii) browsing data on the 'Browse' page by RNA category, subcellular localization, prediction algorithm or species. The search results page displays basic information for all entries, including RNA symbols, RNA categories, species, tissue/cell lines, localizations and scores. Users can further refine search results by employing filter options (score, species, RNA category, and localization) available on the left side. The experimental validation data, prediction data, and RNA sequence information in the database are accessible on the 'Download & API' page.

To illustrate the RNA sequencing data of different subcellular localizations, the 'RNA-seq Analysis' page provides two methods for searching the analysis results related to RNA expression in different localizations, as well as dynamic changes of RNA localization under various conditions. In addition, RNALocate v3.0 provides a prediction tool on the 'Tool' page, which allows users to input the sequence of a specific type of RNA (in FASTA format) and receive the top three predicted subcellular localizations along with corresponding scores. We comprehensively evaluated our predictive model using a five-fold cross-validation approach, and compared it to six established multi-label algorithms (DeepLocRNA, DM3Loc, miRNALoc, LncLocFormer, EL-RMLocNet and Clarion). Key metrics, including accuracy, precision, F1 score and recall were assessed. While some algorithms outperformed in specific RNA types or subcellular localizations (e.g. EL-RMLocNet for miRNA, and DeepLocRNA for snoRNA), our model consistently achieved an accuracy exceeding 0.75 across all RNA types. It also demonstrated superior performance across all subcellular localizations, excelling in overall predictive reliability and coverage. In instances where true positives were absent, metrics such as F1 score, precision, or recall may have been reported as NA or zero. These findings underscore the robust accuracy and stability of our model across diverse scenarios, as detailed in Supplementary Figure S2 and Supplementary Table S4.

Dynamic subcellular localization analysis

In this section, we have analyzed 40 datasets and provided detailed visualization results, including data from our previous study GSE206328, which investigated the subcellular RNA distribution in hESCs and its changes during hESC differentiation. On the 'RNA-seq Analysis from Dataset' page, the enrichment locations of each RNA type under various conditions are intuitively visualized ('Subcellular Localization Number'). Using the GSE206328 dataset as a case study, we analyzed the localization patterns of RNAs across five subcellular components and identified the number of RNA-enriched locations on day 5 (Figure 4A). Overall, more than half of the RNAs were found in multiple subcellular components. While the majority of RNAs exhibited overlapping distributions among multiple components, some RNAs showed site-specific distributions, suggesting their association with specific locations (Figure 4B). Subsequently, we conducted differential localization analyses of various types of RNAs under the same conditions to determine their subcellular-specific localization,

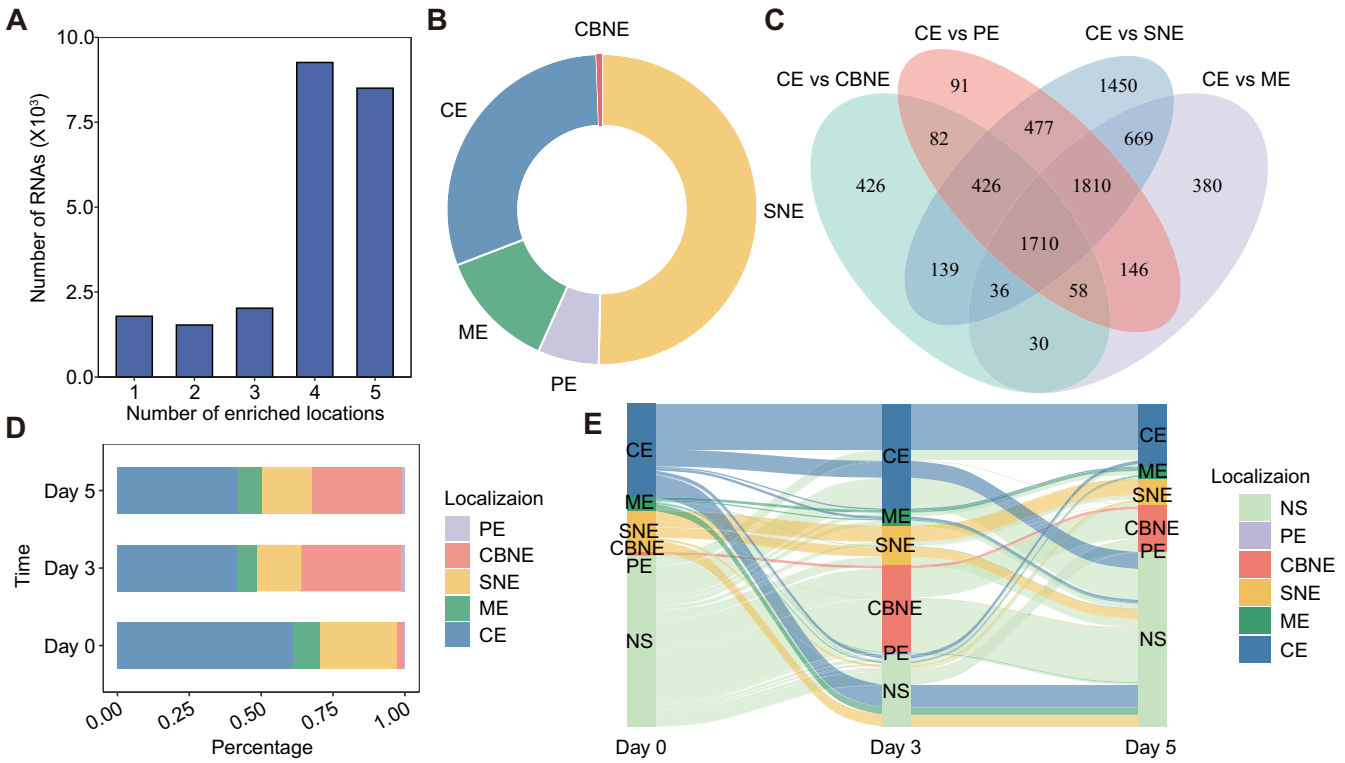


Figure 4. Subcellular localization of RNA during hESC differentiation for GSE206328. (**A**) Number of RNA-enriched locations in hESC on day 5. (**B**) Donut chart showing the distribution of site-specific RNA on day 5. (**C**) Number of CE-specific localization of RNAs on day 5 using DEseq2. (**D**) Bar plot displaying the number of subcellular-specific RNAs among the five subcellular fractions during differentiation. (**E**) Sankey diagram demonstrating the dynamic change of subcellular-specific RNAs during hESC differentiation. NS: non-specific; RNA localization was not specific, i.e. RNA localized to at least two subcellular fractions after differentiation.

revealing the dynamic changes in RNA localization during cellular development. We also provide GO/KEGG enrichment analyses of the differentially expressed RNAs for each subcellular localization. There are 1 710 cytoplasmic extract (CE)-specific localization of RNAs on day 5 using DEseq2, including 1702 mRNAs and 8 lncRNAs, obtained from the intersection of 2907 upregulated RNAs for CE versus chromatin-bound nuclear extract (CBNE), 6717 upregulated RNAs for CE versus soluble nuclear extract (SNE), 4800 upregulated RNAs for CE versus pellet extract (PE) and 4839 upregulated RNAs for CE versus membrane extract (ME) (Figure 4C). These CE-specific RNAs were enriched in various universal biological processes, such as translation and energy metabolism. Furthermore, the top 20 RBPs identified by RBPmap are primarily engaged in RNA processing, splicing, transport and regulation of gene expression, including key RBPs such as SRSF2, HNRNPL and PUF60. Additionally, histograms were constructed to evaluate the NMFE and phastCons conservation scores of these RNAs. The majority of RNAs exhibited low conservation, with NMFE values predominantly centered around −0.3. We further observed the dynamic localization of subcellular-specific RNAs during hESC differentiation. From day 0 to day 3 of differentiation, the number of specific RNAs for all four subcellular components except SNE increased, especially for CBNE; however, on day 5, the number of all specific RNAs decreased (Figure 4D and E). Additionally, we demonstrated changes in the levels of subcellular components for a single RNA during differentiation. On the 'RNA-seq Analysis from RNA' page, the expression of RNAs in subcellular localization across different datasets and the relative expression (ratio) of subcellular-specific RNAs are visualized, allowing users to customize these visualizations by RNA symbol/ID and species.

Service optimization

Due to the vast number of entries, efficient and thorough data searching is crucial. RNALocate v3.0 has been redesigned using the Django Model-Template-View (MTV) framework, resulting in significant improvements in search and browsing speeds while markedly enhancing user experience. Key enhancements include (i) optimizing the browsing module to address previous limitations in displaying localized entries caused by large data volumes and (ii) adding filtering options to allow users to refine results by score, species, RNA category and localization, facilitating easier access to desired entries.

Conclusion and future perspectives

Explaining the diversity and complexity of RNA localization is essential to fully understand cellular architecture. We present a comprehensive RNA subcellular localization resource, RNALocate v3.0, which comprises over 1.8 million entries for RNA-subcellular localization associations, more than eight times that in the previous version. Nearly 1 260 000 entries are derived from experimental data, covering 26 RNA types and 177 subcellular localizations, with species coverage increasing from 104 to 242. Furthermore, RNALocate v3.0 provides comprehensive analyses of RNA subcellular localizations derived from high-throughput sequencing data, in-

corporating over 850 subcellular localization-associated samples among various cell lines and tissues. In addition, it introduces a predictive tool for mRNA, lncRNA, miRNA, piRNA, snoRNA, snRNA and circRNA localization, developed using experimentally validated data from various cell lines and tissues, which has significantly enhanced the model's performance.

We have focused on the field of RNA dynamic subcellular localization and integrated substantial relevant data. Moving forward, we plan to delve deeper into this area of study. Based on a significant amount of experimentally validated RNA subcellular localization data and DNA and protein subcellular data from other databases, we would further explore RNA–DNA, RNA–RNA, and RNA–protein interactions at subcellular resolution, helping researchers better understand biological processes and disease mechanisms. Furthermore, RNALocate currently focuses exclusively on sequencing data analysis for subcellular localization in four species. Next, we plan to expand this dataset to contain subcellular localization sequencing data from a broader spectrum of species, incorporating diverse environmental conditions, tissues and cell lines. This effort aims to uncover the potential roles and mechanisms of RNA localization in influencing biological processes and regulating cell fate. In addition, our prediction tool is based on RNA sequences and does not account for factors such as cell lines, cellular state, gene expression and other variables. To achieve more accurate prediction of RNA subcellular localization, we will incorporate these factors to improve our algorithm. We are committed to continuously maintaining and updating RNALocate, making RNALocate v3.0 the most comprehensive RNA-subcellular localization resource and a robust platform for RNA-subcellular localization analysis, thereby meeting diverse research needs.

Data availability

RNALocate is freely accessible to all users without any login requirement at: http://www.rnalocate.org/ or http://www.rna-society.org/rnalocate/.

Supplementary data

Supplementary Data are available at NAR Online.

Funding

National Key Research and Development Project of China [2020YFA0113300, 2022YFA0806303, 2021YFC2500300]; National Natural Science Foundation of China [82370106, 82070109]; Guangdong Basic and Applied Basic Research Foundation [2024A1515011769, 2022A1515011253]. Funding for open access charge: National Key Research and Development Project of China [2020YFA0113300, 2022YFA0806303, 2021YFC2500300]; National Natural Science Foundation of China [82370106, 82070109]; Guangdong Basic and Applied Basic Research Foundation [2024A1515011769, 2022A1515011253].

Conflict of interest statement

None declared.

References

1. Bridges,M.C., Daulagala,A.C. and Kourtidis,A. (2021) LNCcation: lncRNA localization and function. *J. Cell Biol.*, **220**, e202009045.
2. Dermit,M., Dodel,M., Lee,F.C.Y., Azman,M.S., Schwenzer,H., Jones,J.L., Blagden,S.P., Ule,J. and Mardakheh,F.K. (2020) Subcellular mRNA localization regulates ribosome biogenesis in migrating cells. *Dev. Cell*, **55**, 298–313.
3. Pilaz,L.J., Liu,J., Joshi,K., Tsunekawa,Y., Musso,C.M., D'Arcy,B.R., Suzuki,I.K., Alsina,F.C., Kc,P., Sethi,S., *et al.* (2023) Subcellular mRNA localization and local translation of Arhgap11a in radial glial progenitors regulates cortical development. *Neuron*, **111**, 839–856.
4. Guo,C.J., Ma,X.K., Xing,Y.H., Zheng,C.C., Xu,Y.F., Shan,L., Zhang,J., Wang,S., Wang,Y., Carmichael,G.G., *et al.* (2020) Distinct processing of lncRNAs contributes to non-conserved functions in stem cells. *Cell*, **181**, 621–636.
5. Mikl,M., Eletto,D., Nijim,M., Lee,M., Lafzi,A., Mhamedi,F., David,O., Sain,S.B., Handler,K. and Moor,A.E. (2022) A massively parallel reporter assay reveals focused and broadly encoded RNA localization signals in neurons. *Nucleic Acids Res.*, **50**, 10643–10664.
6. Gasparski,A.N., Moissoglu,K., Pallikkuth,S., Meydan,S., Guydosh,N.R. and Mili,S. (2023) mRNA location and translation rate determine protein targeting to dual destinations. *Mol. Cell*, **83**, 2726–2738.
7. Das,S., Vera,M., Gandin,V., Singer,R.H. and Tutucci,E. (2021) Intracellular mRNA transport and localized translation. *Nat. Rev. Mol. Cell Biol.*, **22**, 483–504.
8. Zhou,F., Tan,P., Liu,S., Chang,L., Yang,J., Sun,M., Guo,Y., Si,Y., Wang,D., Yu,J., *et al.* (2024) Subcellular RNA distribution and its change during human embryonic stem cell differentiation. *Stem Cell Rep.*, **19**, 126–140.
9. Hwang,H., Yun,S., Arcanjo,R.B.,,Divyanshi, Chen,S., Mei,W., Nowak,R.A., Kwon,T. and Yang,J. (2022) Regulation of RNA localization during oocyte maturation by dynamic RNA-ER association and remodeling of the ER. *Cell Rep.*, **41**, 111802.
10. Fonseca,A., Riveras,E., Moyano,T.C., Alvarez,J.M., Rosa,S. and Gutiérrez,R.A. (2024) Dynamic changes in mRNA nucleocytoplasmic localization in the nitrate response of Arabidopsis roots. *Plant Cell Environ.*, **47**, 4227–4245.
11. Bourke,A.M., Schwarz,A. and Schuman,E.M. (2023) De-centralizing the Central Dogma: mRNA translation in space and time. *Mol. Cell*, **83**, 452–468.
12. Villanueva,E., Smith,T., Pizzinga,M., Elzek,M., Queiroz,R.M.L., Harvey,R.F., Breckels,L.M., Crook,O.M., Monti,M., Dezi,V., *et al.* (2024) System-wide analysis of RNA and protein subcellular localization dynamics. *Nat. Methods*, **21**, 60–71.
13. Rutherford,K.M., Lera-Ramírez,M. and Wood,V. (2024) PomBase: a Global Core Biodata Resource-growth, collaboration, and sustainability. *Genetics*, **227**, iyae007.
14. Chitti,S.V., Gummadi,S., Kang,T., Shahi,S., Marzan,A.L., Nedeva,C., Sanwlani,R., Bramich,K., Stewart,S., Petrovska,M., *et al.* (2024) Vesiclepedia 2024: an extracellular vesicles and extracellular particles repository. *Nucleic Acids Res.*, **52**, D1694–D1698.
15. Keerthikumar,S., Chisanga,D., Ariyaratne,D., Al Saffar,H., Anand,S., Zhao,K., Samuel,M., Pathan,M., Jois,M., Chilamkurti,N, *et al.* (2016) ExoCarta: a Web-Based Compendium of Exosomal Cargo. *J. Mol. Biol.*, **428**, 688–692.
16. Feng,J., Chen,W., Dong,X., Wang,J., Mei,X., Deng,J., Yang,S., Zhuo,C., Huang,X., Shao,L., *et al.* (2022) CSCD2: an integrated interactional database of cancer-specific circular RNAs. *Nucleic Acids Res.*, **50**, D1179–D1183.
17. Xie,F., Liu,S., Wang,J., Xuan,J., Zhang,X., Qu,L., Zheng,L. and Yang,J. (2021) deepBase v3.0: expression atlas and interactive analysis of ncRNAs from thousands of deep-sequencing data. *Nucleic Acids Res.*, **49**, D877–D883.
18. Lai,H., Li,Y., Zhang,H., Hu,J., Liao,J., Su,Y., Li,Q., Chen,B., Li,C., Wang,Z., *et al.* (2022) exoRBase 2.0: an atlas of mRNA, lncRNA

and circRNA in extracellular vesicles from human biofluids. *Nucleic Acids Res.*, **50**, D118–D128.
19. Li,Z., Liu,L., Jiang,S., Li,Q., Feng,C., Du,Q., Zou,D., Xiao,J., Zhang,Z. and Ma,L. (2021) LncExpDB: an expression database of human long non-coding RNAs. *Nucleic Acids Res.*, **49**, D962–D968.
20. Zeng,M., Wu,Y., Lu,C., Zhang,F., Wu,F.X. and Li,M. (2022) DeepLncLoc: a deep learning framework for long non-coding RNA subcellular localization prediction based on subsequence embedding. *Brief. Bioinform.*, **23**, bbab360.
21. 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.
22. Li,M., Zhao,B., Yin,R., Lu,C., Guo,F. and Zeng,M. (2023) GraphLncLoc: long non-coding RNA subcellular localization prediction using graph convolutional networks based on sequence to graph transformation. *Brief. Bioinform.*, **24**, bbac565.
23. 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.
24. Zhang,Z.-Y., Sun,Z.-J., Yang,Y.-H. and Lin,H. (2022) Towards a better prediction of subcellular location of long non-coding RNA. *Front. Comput. Sci.*, **16**, 165903.
25. 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.
26. 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.
27. Ahmad,A., Lin,H. and Shatabda,S. (2020) Locate-R: subcellular localization of long non-coding RNAs using nucleotide compositions. *Genomics*, **112**, 2583–2589.
28. 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.
29. Tang,Q., Nie,F., Kang,J. and Chen,W. (2021) mRNALocater: enhance the prediction accuracy of eukaryotic mRNA subcellular localization by using model fusion strategy. *Mol. Ther.*, **29**, 2617–2623.
30. Yuan,G.H., Wang,Y., Wang,G.Z. and Yang,L. (2023) RNAlight: a machine learning model to identify nucleotide features determining RNA subcellular localization. *Brief. Bioinform.*, **24**, bbac509.
31. Bi,Y., Li,F., Guo,X., Wang,Z., Pan,T., Guo,Y., Webb,G.I., Yao,J., Jia,C. and Song,J. (2022) Clarion is a multi-label problem transformation method for identifying mRNA subcellular localizations. *Brief. Bioinform.*, **23**, bbac467.
32. Asim,M.N., Ibrahim,M.A., Malik,M.I., Zehe,C., Cloarec,O., Trygg,J., Dengel,A. and Ahmed,S. (2022) EL-RMLocNet: an explainable LSTM network for RNA-associated multi-compartment localization prediction. *Comput. Struct. Biotechnol. J.*, **20**, 3986–4002.
33. Zeng,M., Wu,Y., Li,Y., Yin,R., Lu,C., Duan,J. and Li,M. (2023) LncLocFormer: a Transformer-based deep learning model for multi-label lncRNA subcellular localization prediction by using localization-specific attention mechanism. *Bioinformatics*, **39**, btad752.
34. Meher,P.K., Satpathy,S. and Rao,A.R. (2020) miRNALoc: predicting miRNA subcellular localizations based on principal component scores of physico-chemical properties and pseudo compositions of di-nucleotides. *Sci. Rep.*, **10**, 14557.
35. Wang,J., Horlacher,M., Cheng,L. and Winther,O. (2024) DeepLocRNA: an interpretable deep learning model for predicting RNA subcellular localization with domain-specific transfer-learning. *Bioinformatics*, **40**, btae065.
36. Sayers,E.W., Bolton,E.E., Brister,J.R., Canese,K., Chan,J., Comeau,D.C., Connor,R., Funk,K., Kelly,C., Kim,S., *et al.* (2022) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **50**, D20–D26.
37. Sarkans,U., Füllgrabe,A., Ali,A., Athar,A., Behrangi,E., Diaz,N., Fexova,S., George,N., Iqbal,H., Kurri,S., *et al.* (2021) From ArrayExpress to BioStudies. *Nucleic Acids Res.*, **49**, D1502–D1506.
38. Yuan,D., Ahamed,A., Burgin,J., Cummins,C., Devraj,R., Gueye,K., Gupta,D., Gupta,V., Haseeb,M., Ihsan,M., *et al.* (2024) The European Nucleotide Archive in 2023. *Nucleic Acids Res.*, **52**, D92–D97.
39. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
40. 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.
41. Glažar,P., Papavasileiou,P. and Rajewsky,N. (2014) circBase: a database for circular RNAs. *RNA*, **20**, 1666–1670.
42. Liu,M., Wang,Q., Shen,J., Yang,B.B. and Ding,X. (2019) Circbank: a comprehensive database for circRNA with standard nomenclature. *RNA Biol.*, **16**, 899–905.
43. Li,Z., Liu,L., Feng,C., Qin,Y., Xiao,J., Zhang,Z. and Ma,L. (2023) LncBook 2.0: integrating human long non-coding RNAs with multi-omics annotations. *Nucleic Acids Res.*, **51**, D186–D191.
44. Wang,J., Shi,Y., Zhou,H., Zhang,P., Song,T., Ying,Z., Yu,H., Li,Y., Zhao,Y., Zeng,X., *et al.* (2022) piRBase: integrating piRNA annotation in all aspects. *Nucleic Acids Res.*, **50**, D265–D272.
45. RNAcentral Consortium (2021) RNAcentral 2021: secondary structure integration, improved sequence search and new member databases. *Nucleic Acids Res.*, **49**, D212–D220.
46. Harrison,P.W., Amode,M.R., Austine-Orimoloye,O., Azov,A.G., Barba,M., Barnes,I., Becker,A., Bennett,R., Berry,A., Bhai,J., *et al.* (2024) Ensembl 2024. *Nucleic Acids Res.*, **52**, D891–D899.
47. Chen,J., Lin,J., Hu,Y., Ye,M., Yao,L., Wu,L., Zhang,W., Wang,M., Deng,T., Guo,F., *et al.* (2023) RNADisease v4.0: an updated resource of RNA-associated diseases, providing RNA-disease analysis, enrichment and prediction. *Nucleic Acids Res.*, **51**, D1397–D1404.
48. Kang,J., Tang,Q., He,J., Li,L., Yang,N., Yu,S., Wang,M., Zhang,Y., Lin,J., Cui,T., *et al.* (2022) RNAInter v4.0: RNA interactome repository with redefined confidence scoring system and improved accessibility. *Nucleic Acids Res.*, **50**, D326–D332.
49. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.
50. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
51. Frankish,A., Carbonell-Sala,S., Diekhans,M., Jungreis,I., Loveland,J.E., Mudge,J.M., Sisu,C., Wright,J.C., Arnan,C., Barnes,I., *et al.* (2023) GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Res.*, **51**, D942–D949.
52. Fisher,M., James-Zorn,C., Ponferrada,V., Bell,A.J., Sundararaj,N., Segerdell,E., Chaturvedi,P., Bayyari,N., Chu,S., Pells,T., *et al.* (2023) Xenbase: key features and resources of the Xenopus model organism knowledgebase. *Genetics*, **224**, iyad018.
53. Kim,D., Langmead,B. and Salzberg,S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.
54. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
55. 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.
56. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.

57. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
58. Lorenz,R., Bernhart,S.H., Höner Zu Siederdissen,C., Tafer,H., Flamm,C., Stadler,P.F. and Hofacker,I.L. (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, **6**, 26.
59. Paz,I., Kosti,I., Ares,M. Jr., Cline,M. and Mandel-Gutfreund,Y. (2014) RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.*, **42**, W361–W367.
60. Siepel,A., Bejerano,G., Pedersen,J.S., Hinrichs,A.S., Hou,M., Rosenbloom,K., Clawson,H., Spieth,J., Hillier,L.W., Richards,S., *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.
61. Sticht,C., De La Torre,C., Parveen,A. and Gretz,N. (2018) miRWalk: an online resource for prediction of microRNA binding sites. *PLoS One*, **13**, e0206239.
62. Gene Ontology Consortium (2015) Gene Ontology Consortium: going forward. *Nucleic Acids Res.*, **43**, D1049–D1056.
63. Kanehisa,M., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.*, **44**, D457–D462.

Received: July 21, 2024. **Revised:** September 18, 2024. **Editorial Decision:** September 20, 2024. **Accepted:** September 24, 2024