

RNAInter v4.0: RNA interactome repository with redefined confidence scoring system and improved accessibility

Juanjuan Kang^{1,2,†}, Qiang Tang^{3,4,†}, Jun He^{5,†}, Le Li⁶, Nianling Yang⁵, Shuiyan Yu⁵, Mengyao Wang⁵, Yuchen Zhang⁵, Jiahao Lin², Tianyu Cui², Yongfei Hu², Puwen Tan², Jun Cheng^{1,2}, Hailong Zheng², Dong Wang^{2,*}, Xi Su^{1,*}, Wei Chen^{2,3,4,*} and Yan Huang^{2,7,8,*}

¹Affiliated Foshan Maternity & Child Healthcare Hospital, Southern Medical University (Foshan Maternity & Child Healthcare Hospital), Foshan 528000, China, ²Department of Bioinformatics, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China, ³School of Basic Medical Sciences, Chengdu University of Traditional Chinese Medicine, Chengdu 611137, China, ⁴Innovative Institute of Chinese Medicine and Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu 611730, China, ⁵Department of Bioinformatics, Fujian Key Laboratory of Medical Bioinformatics, School of Medical Technology and Engineering, Fujian Medical University, Fuzhou 350122, China, ⁶The Sixth Affiliated Hospital of Guangzhou Medical University, Qingyuan People's Hospital, Qingyuan, 511518, China, ⁷Cancer Research Institute, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China and ⁸Shunde Hospital, Southern Medical University (The First People's Hospital of Shunde), Foshan 528308, China

Received September 15, 2021; Revised October 07, 2021; Editorial Decision October 08, 2021; Accepted October 09, 2021

ABSTRACT

Establishing an RNA-associated interaction repository facilitates the system-level understanding of RNA functions. However, as these interactions are distributed throughout various resources, an essential prerequisite for effectively applying these data requires that they are deposited together and annotated with confidence scores. Hence, we have updated the RNA-associated interaction database RNAInter (RNA Interactome Database) to version 4.0, which is freely accessible at <http://www.rnainter.org> or <http://www.rna-society.org/rnainter/>. Compared with previous versions, the current RNAInter not only contains an enlarged data set, but also an updated confidence scoring system. The merits of this 4.0 version can be summarized in the following points: (i) a redefined confidence scoring system as achieved by integrating the trust of experimental evidence, the trust of the scientific community and the types of tissues/cells, (ii) a redesigned fully functional database that enables for a more rapid retrieval and browsing of interactions via an upgraded user-

friendly interface and (iii) an update of entries to >47 million by manually mining the literature and integrating six database resources with evidence from experimental and computational sources. Overall, RNAInter will provide a more comprehensive and readily accessible RNA interactome platform to investigate the regulatory landscape of cellular RNAs.

INTRODUCTION

Benefiting from the advances of small-scale experiments and high-throughput sequencing technology, a growing number of RNA-associated interactions have been revealed over the past decades. These interactions have been implicated in almost all physiological and pathological conditions and involve processes such as cell growth and development, tumorigenesis and tumor invasion. For example, long non-coding RNA (lncRNA) BCRT1 can interact with miR-1303 to promote breast cancer progression (1), and RNA-binding protein activities can regulate mRNA processing of specific gene sets to affect developmental processes (2). With the exception of some established RNA–RNA and RNA–protein interactions, a number of other types of RNA-associated interactions, which also play important biological functions, have been recognized. With

*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 Yan Huang. Tel: +86 20 61648279; Fax: +86 20 61648279; Email: huangyan24@smu.edu.cn
Correspondence may also be addressed to Wei Chen. Tel: +86 28 61800086; Fax: +86 28 61800086; Email: greatchen@ncst.edu.cn
Correspondence may also be addressed to Xi Su. Tel: +86 20 61648279; Fax: +86 20 61648279; Email: 849527809@qq.com

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

use of modular domains to establish chromosome conformations, RNA can interact with DNA and then regulate gene expression (3). Moreover, non-coding RNAs are associated with drug resistance and have provided a new class of targets for drug discovery (4), while histone modifications have been reported to be involved in the transcriptional regulation of RNA through biological processes such as RNA splicing (5). Hence, a prerequisite for system-level understanding of RNA biological functions requires that RNA-associated interactions be integrated under one common framework.

Over the past few years, through integrating experimentally validated and computationally predicted RNA-associated interactions, we have generated three versions of RNA Interactome Databases, RAID (6), RAID v2.0 (7) and RNAInter v3.0 (8). These databases have provided information on a number of interactions, such as RNA–RNA (RRI), RNA–protein (RPI), RNA–DNA (RDI), RNA–Histone modification (RHI) and RNA–Compound (RCI), and have been widely used in the scientific community (9,10). As a result of our efforts to continuously update the previous versions of RNAInter, >40 million interactions have now been included, which has provided a comprehensive RNA interactome platform for researchers. However, these interactions were derived from a variety of resources, such as manually text mining, database integration and computational prediction. Therefore, to more accurately evaluate each interaction, it will be critical to optimize the algorithm of confidence scores. With this goal in mind, we have now updated RNAInter to version 4.0 (<http://www.rnainter.org> or <http://www.rna-society.org/rnainter/>). In this version, the confidence scoring system has been redefined, as achieved by integrating the trust of experimental evidence, the trust of the scientific community and the types of tissues/cells. Moreover, we re-designed the website frame and updated its entries by manually mining the literature and integrating six databases. These improvements will substantially facilitate the operation of this system as users can quickly and accurately retrieve and browse RNA associated interactions deposited in the database.

DATABASE ORGANIZATION

Data collection and integration

RNAInter integrates experimentally validated and computationally predicted RNA interactome data from the literature and databases. Through Pubmed database searches using the same keywords as that in our previous work, over 20 000 new entries describing experimentally validated RNA–RNA and RNA–protein interactions from over 30 000 published reports were screened. The interaction information obtained includes ‘Interactors’, ‘Species’, ‘Tissue or Cell Line’, ‘Target region’ and ‘Supportive evidence’. In addition, RNAInter integrated three experimentally validated databases LncTarD (11), NPInter v4.0 (12) and NoncoRNA (13) as well as three computational predicted databases miRDB v6.0 (14), oRNAmnt (15) and tRFtarget (16), which include RNA–RNA, RNA–protein, RNA–compound and RNA–DNA interactions (Table 1).

Table 1. Overview of interactions as obtained from six databases.

Database resource	Interaction type ^a	Interaction entry	Evidence type ^b	Ref.
LncTarD	RPI/RRI	2823	E	(11)
NPInter v4.0	RPI/RRI/RDI	1 100 659	E/C	(12)
NoncoRNA	RCI	8034	E	(13)
miRDB v6.0	RRI	6 831 595	C	(14)
oRNAmnt	RPI	44 290 806	C	(15)
tRFtarget	RRI	13 58 674	C	(16)

^aRRI is short for RNA–RNA interaction, RPI is for RNA–protein interaction, RDI is for RNA–DNA interaction and RCI is for RNA–compound Interaction.

^bE represents Experimental validation; C represents Computational prediction.

Accordingly, the number of integrated databases in RNAInter has increased from 35 to 39, and consists of more comprehensive RNA interactome data.

All new data were compared with interactions as generated from RNAInter v3.0 to eliminate redundancies. For interactions that were repeated, ‘Tissue or Cell Line’, ‘Target region’, ‘Methods’ and ‘References’ were integrated to the old interactions. Moreover, as data were derived from different sources, we standardized the names of tissues (or cell lines) and methods for all interactions.

Data annotation

For all new interactors, symbols and assigned IDs used by renowned databases were standardized: miRNA from miRBase (17), circRNA from circBase (18), transfer RNA-derived fragments (tRFs) from tRFdb (19), compound from PubChem Compound (20) and other interactors from the NCBI Gene (21). For the convenience of users, aliases, descriptions and other IDs from DrugBank (22), OMIM (23), Ensembl (24), HGNC (25), HPRD (26) and Uniprot (27) were all included in the interactor information.

Like that as performed in previous versions, we also obtained the RNA editing/localization/modification/structure/expression patterns. ‘RNA editing’ consisted of information on editing position, base change and genetic region from DARDED (28), Lncediting (29) and RADAR (30). ‘RNA localization’ included subcellular localization and tissue/cell lines from RNALocate (31). ‘RNA modification’ involved position and type of modification and genomic context from RMBase (32). ‘RNA structure’ displayed the putative RNA secondary structures for each transcript as calculated by RNAs-structure software (33). ‘Expression pattern’ showed RNA expression values in each stage during human (or mouse) spermatogenesis (34,35) and hematopoietic stem cell (HSC) lineage commitment (36,37), and expression correlation coefficients of the two interactors were determined. Some additional functional modules were also provided. For example, disease associations from MNDR v3.0 (38) and tissue specific expression from the GTEx project (39). Homology interactions were pre-calculated as based on the orthology/paralogy gene sets from NCBI Gene (21). An overview of data integration and annotations are shown in Figure 1.

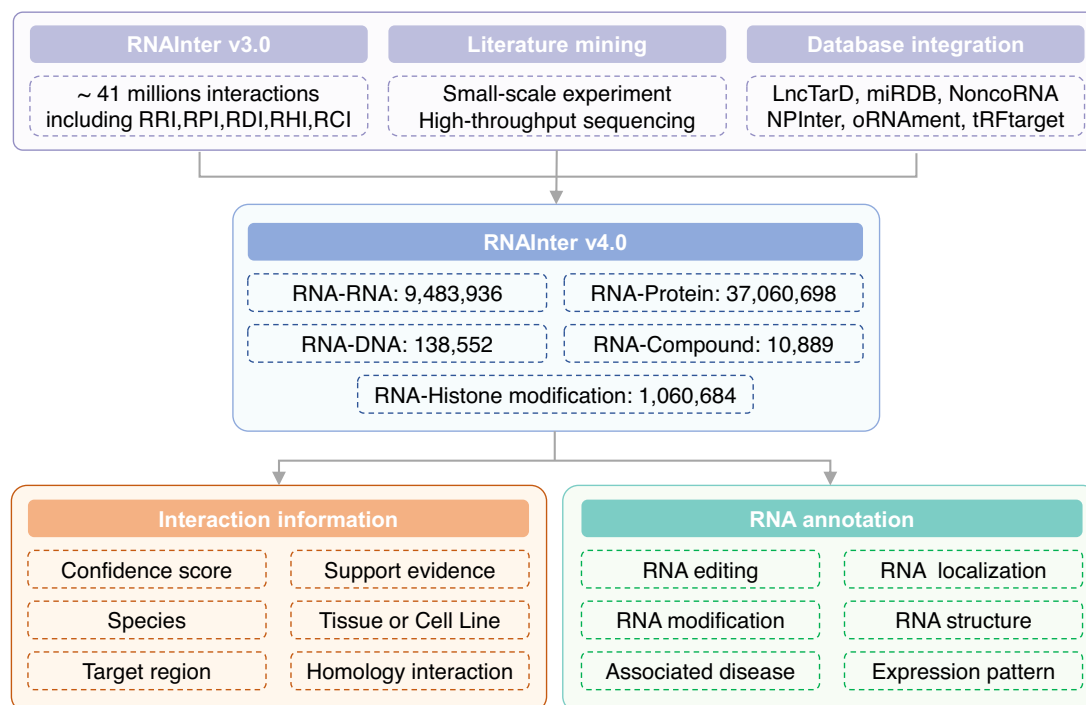


Figure 1. Overview of data integration and annotation.

Confidence scoring system

In RAID v2.0, the supported methods were divided into strong (e.g. RNA immunoprecipitation and luciferase reporter assay) and weak (e.g. ChIP-seq and CLIP-seq) experimentally validated evidence and computationally predicted evidence [e.g. miRanda (40) and TargetScan (41)]. A confidence score was calculated using a sigmoid function by integrating different sources of evidence. In this version, the new confidence scoring system was defined by integrating the trust of experimental evidence (E), trust of the scientific community (S) and types of tissues/cells (T). Such an approach increases the scoring reliability.

With regard to trust of experimental evidence, a small-scale experiment is more reliable than a large-scale screening. Therefore, interactions supported by publications describing few interactions should have higher confidence score than those supported by publications describing many interactions (42). This metric, E, can be calculated as follows:

$$E = \sum_i \frac{\log(10)}{\log(n_i + 1)}$$

where i is the number of publications or prediction tools supporting the interaction and n_i represents the interaction number described or predicted by the i -th publication or prediction tool.

Trust of the scientific community can be reflected by the number of citations and publication years as derived from Google Scholar. The greater the number of citations the higher the confidence score (42). This metric, S, can be cal-

culated as follows:

$$S = \sum_i \log\left(\frac{r_i + 1}{2021 - y_i + 1} + 1\right)$$

where i is the number of publications or prediction tools supporting the interaction, r_i represents the citations of the i -th publication or prediction tool, and y_i is the publication year of the i -th publication or prediction tool.

For the types of tissues/cells, the greater the number of tissues/cells in which an interaction was detected, the higher the confidence score for that interaction. This metric, T, can be calculated as follows:

$$T = \text{type of the tissues/cells}$$

These three metrics were scaled individually in a range from 0 to 1, with the weighted Euclidean distance then calculated as the confidence score.

$$\text{Score} = \sqrt{\alpha E^2 + \beta(S^2 + T^2)}$$

where α and β range from 0–1.

We evaluated the new confidence scoring system based on the three different levels of supporting evidence. The interactions can be divided into three levels according to the different sources of evidence from which they were derived. Interactions with strong evidence were supported by strong experiments, while those with weak evidence by weak experiments and those with predicted evidence were only supported by predicted methods. As interactions with strong evidence and those with predicted evidence are the two more explicit interaction sets. We set the interactions with strong evidence as positive dataset, and those with predicted

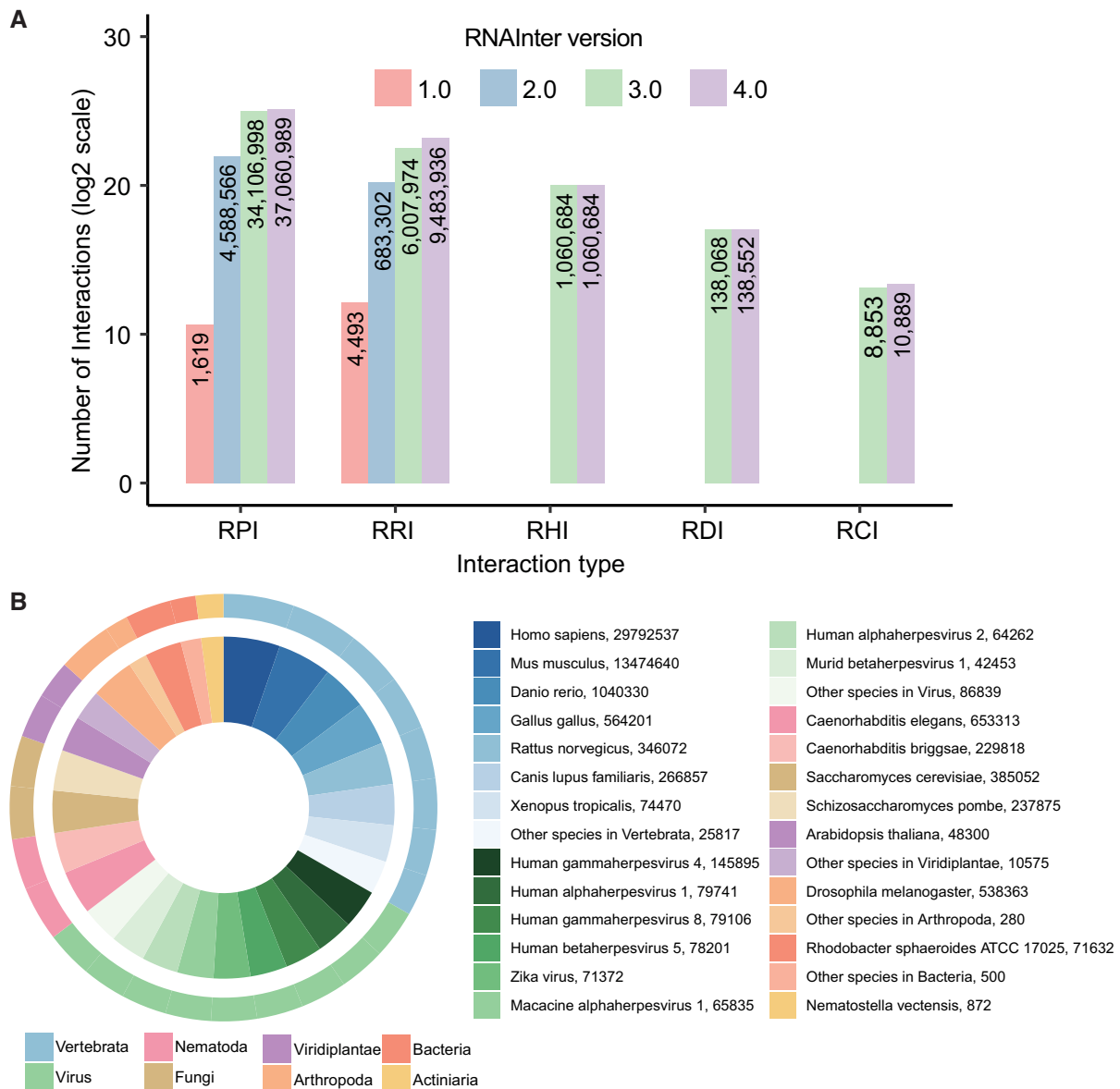


Figure 2. Statistics of RNAInter v4.0. (A) Distributions of five interaction types (RCI/RDI/RHI/RPI/RRI) among RNAInter versions. (B) Number of interactions in species. The outer circle indicates the category of species in the inner circle.

evidence as negative dataset. Then the area under the ROC curve (AUC) were applied to select the optimal weight combination through iterative traversal with a 0.05 step size. Final confidence scores were then log₂-transformed and scaled from 0 to 1. In this way, interactions as reported within highly cited papers and involving detection in greater numbers of tissues/cells would receive a higher confidence score.

RESULTS

RNAInter statistics

In this version, through the scanning of >30 000 published reports and comprehensively evaluating RNA interactome (RNA–RNA/Protein/Compound/DNA) data, >6 million new entries were added, and 2 million entries were up-

dated. From these >8 million interactions, over 600 000 interactions were experimentally validated, while the remaining were computationally predicted. All 156 species, including two new species (*Schizosaccharomyces pombe* and *Rhodobacter sphaeroides* ATCC 17025), represent eight categories (actiniaria, arthropoda, bacteria, fungi, nematode, vertebrata, viridiplantae and virus). Figure 2B contains a summary of the number of interactions in species containing >40 000 interactions. The number of interactions in other species was combined according to species categories. Detailed interaction numbers of each interaction type in all 156 species were listed in Supplementary Table S1. The top three species with the greatest number of interactions were *Homo sapiens*, *Mus musculus* and *Danio rerio*, which all belong to vertebrata. Such findings indicate that the RNA interactome of these three species were more extensively studied than that of other species.

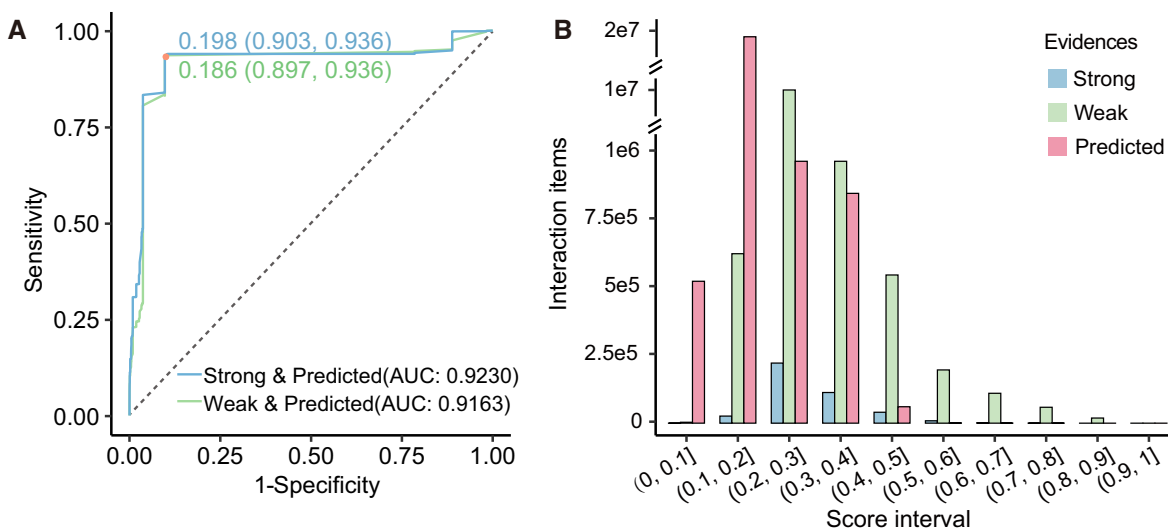


Figure 3. Evaluation of confidence scores. (A) ROC for distinguishing between experimental and predicted interactions. (B) Interaction number for each score interval.

Evaluation of confidence score

The AUCs calculated by different weight combinations were listed in Supplementary Table S2. The AUC of distinguishing interactions with strong and predicted evidences was the highest and up to 0.9230 when the α and β were 0.85 and 0.25. Finally, we chose this weight combination to integrate the three metrics (E, S and T). As shown in Figure 3A, the optimal cutoff was 0.198 with the specificity of 0.903 and the sensitivity of 0.936 for distinguishing interactions with strong and predicted evidences. At the same time, the performance of the scoring system was also satisfactory for distinguishing interactions with weak and predicted evidences. The AUC was also the highest and up to 0.9163, and the optimal cutoff was 0.186 with the specificity of 0.897 and sensitivity of 0.936 (Figure 3A).

In addition, we also evaluated the score distribution of interactions with strong, weak and predicted evidences. As shown in Figure 3B, the greatest enrichment score intervals of interactions with experimental evidence (blue and green bar) were from 0.2 to 0.3, while those of interactions with predicted evidence (red bar) were from 0.1 to 0.2. The mean scores of interactions with strong and weak evidences were 0.2886 and 0.2767, which was obviously >0.1814 , mean of interactions with predicted evidence. These results demonstrate that this new confidence scoring system can effectively estimate the reliability of RNA-associated interactions with more objective metrics and can be used to filter interactions of interest from vast arrays of interactions.

Service optimization

Due to the large number of interactions, it is essential that database searches be performed quickly and thoroughly browsed. The RNAInter database, which was redesigned as based on the Django Model-View-Controller (MVC) framework, achieves this goal by enabling a quick retrieval and browsing of interactions with use of an improved user-friendly interface. As a result, the speed of searching and

browsing with this new version is considerably faster than that of the old version.

We optimized the browse module to display all related entries without being limited by the number of entries, which resolves the problem that the browse module cannot display interactions when the amount of data is too large. A filter option has also been provided to further screen results by interactor, interactor category and species, which allows users to more readily locate interactions of interest. Thus, the result page consisted of query condition, interaction display and filter option.

Increasing evidence has accrued suggesting that RNAs, especially ncRNAs, are involved in the progression of a number of diseases (43). Most of these RNAs are involved with regulating pathways related to the occurrence and development of diseases by interacting with target molecules (44). To promote research on RNA associated interactions in disease processes, we embedded disease associations derived from MNDR v3.0 into RNAInter. We also added human tissue expression data from the GTEx project in the 'Expression pattern' module. In this way, expressions of two interactors in different tissues are simultaneously displayed.

CONCLUSIONS

In RNAInter v4.0, we have increased the number of RNA associated interactions from the literature and other databases and optimized the website to enhance its speed and convenience. Most importantly, by integrating multi-dimensional information, we generated a confidence scoring system which can accurately evaluate each interaction and enable researchers to more readily screen RNA interactions of interest. For example, interactions without experimental support but with high confidence scores could be reliable candidates for RNA functional studies. This confidence scoring system will also contribute to a better understanding of the biological functions associated with these RNA interactions.

With the technological advances in experimental measurements and computational techniques, an explosive amount of literature will be generated with regard to descriptions of RNA interactome. It will be time-consuming and virtually unmanageable to manually review and collect RNA associated interactions from literature. Accordingly, the development of automated text mining tools for full-scale RNA interactome scanning will be fundamental for advances in this field. The RNAInter described in the present work provides one such reliable RNA interactome corpus for the development of a text mining tool, which will facilitate the application of artificial intelligence-based text mining algorithms.

Overall, RNAInter provides an upgraded, comprehensive platform which will substantially contribute to the efficacy of functional RNA interactome research. Finally, it should be noted that this RNAInter database represents the current, but not final, version as it will be continuously updated and improved, once new data on RNA interactome become available.

SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NAR Online.

FUNDING

National Key Research and Development Project of China [2019YFA0801800]; National Natural Science Foundation of China [82070109, 31771471, 81770104, 62002153]; Guangdong Basic and Applied Basic Research Foundation [2019A1515010784]; China Postdoctoral Science Foundation [2020M682623, 2020M682785]; Fujian Medical University Research Foundation of Talented Scholars [XR-CZX2020027]; Foshan Medicine Dengfeng Project (2019-2021). Funding for open access charge: Postdoctoral research funding of Affiliated Foshan Maternity and Child Healthcare Hospital.

Conflict of interest statement. None declared.

REFERENCES

- Liang, Y., Song, X., Li, Y., Chen, B., Zhao, W., Wang, L., Zhang, H., Liu, Y., Han, D., Zhang, N. *et al.* (2020) LncRNA BCRT1 promotes breast cancer progression by targeting miR-1303/PTBP3 axis. *Mol. Cancer*, **19**, 85.
- Brinegar, A.E. and Cooper, T.A. (2016) Roles for RNA-binding proteins in development and disease. *Brain Res.*, **1647**, 1–8.
- Mumbach, M.R., Granja, J.M., Flynn, R.A., Roake, C.M., Satpathy, A.T., Rubin, A.J., Qi, Y., Jiang, Z., Shams, S., Louie, B.H. *et al.* (2019) HiChIRP reveals RNA-associated chromosome conformation. *Nat. Methods*, **16**, 489–492.
- Matsui, M. and Corey, D.R. (2017) Non-coding RNAs as drug targets. *Nat. Rev. Drug Discov.*, **16**, 167–179.
- Rahhal, R. and Seto, E. (2019) Emerging roles of histone modifications and HDACs in RNA splicing. *Nucleic Acids Res.*, **47**, 4911–4926.
- Zhang, X., Wu, D., Chen, L., Li, X., Yang, J., Fan, D., Dong, T., Liu, M., Tan, P., Xu, J. *et al.* (2014) RAID: a comprehensive resource for human RNA-associated (RNA-RNA/RNA-protein) interaction. *RNA*, **20**, 989–993.
- Yi, Y., Zhao, Y., Li, C., Zhang, L., Huang, H., Li, Y., Liu, L., Hou, P., Cui, T., Tan, P. *et al.* (2017) RAID v2.0: an updated resource of RNA-associated interactions across organisms. *Nucleic Acids Res.*, **45**, D115–D118.
- 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.
- Aschenbrenner, A.C., Moukhtaroudi, M., Kramer, B., Oestreich, M., Antonakos, N., Nuesch-Germano, M., Gkizeli, K., Bonaguro, L., Reusch, N., Bassler, K. *et al.* (2021) Disease severity-specific neutrophil signatures in blood transcriptomes stratify COVID-19 patients. *Genome Med.*, **13**, 7.
- Zhang, Y., Jia, C. and Kwok, C.K. (2021) Predicting the interaction biomolecule types for lncRNA: an ensemble deep learning approach. *Brief. Bioinform.*, **22**, bbaa228.
- Zhao, H., Shi, J., Zhang, Y., Xie, A., Yu, L., Zhang, C., Lei, J., Xu, H., Leng, Z., Li, T. *et al.* (2020) LncTarD: a manually-curated database of experimentally-supported functional lncRNA-target regulations in human diseases. *Nucleic Acids Res.*, **48**, D118–D126.
- 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.
- Li, L., Wu, P., Wang, Z., Meng, X., Zha, C., Li, Z., Qi, T., Zhang, Y., Han, B., Li, S. *et al.* (2020) NoncoRNA: a database of experimentally supported non-coding RNAs and drug targets in cancer. *J. Hematol. Oncol.*, **13**, 15.
- Chen, Y. and Wang, X. (2020) miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res.*, **48**, D127–D131.
- Benoit Bouvrette, L.P., Bovaird, S., Blanchette, M. and Lecuyer, E. (2020) oRNAment: a database of putative RNA binding protein target sites in the transcriptomes of model species. *Nucleic Acids Res.*, **48**, D166–D173.
- Li, N., Shan, N., Lu, L. and Wang, Z. (2021) tRFtarget: a database for transfer RNA-derived fragment targets. *Nucleic Acids Res.*, **49**, D254–D260.
- Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
- Glazar, P., Papavasileiou, P. and Rajewsky, N. (2014) circBase: a database for circular RNAs. *RNA*, **20**, 1666–1670.
- Kumar, P., Mudunuri, S.B., Anaya, J. and Dutta, A. (2015) tRFdb: a database for transfer RNA fragments. *Nucleic Acids Res.*, **43**, D141–D145.
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B.A., Thiessen, P.A., Yu, B. *et al.* (2021) PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.*, **49**, D1388–D1395.
- Maglott, D., Ostell, J., Pruitt, K.D. and Tatusova, T. (2011) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **39**, D52–D57.
- Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z. *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.
- Amberger, J.S., Bocchini, C.A., Schiettecatte, F., Scott, A.F. and Hamosh, A. (2015) OMIM.org: Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and genetic disorders. *Nucleic Acids Res.*, **43**, D789–D798.
- 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.
- Tweedie, S., Braschi, B., Gray, K., Jones, T.E.M., Seal, R.L., Yates, B. and Bruford, E.A. (2021) Genenames.org: the HGNC and VGNC resources in 2021. *Nucleic Acids Res.*, **49**, D939–D946.
- Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A. *et al.* (2009) Human Protein Reference Database–2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
- UniProt, C. (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480–D489.
- Kiran, A. and Baranov, P.V. (2010) DARNED: a Database of RNA Editing in humans. *Bioinformatics*, **26**, 1772–1776.
- Gong, J., Liu, C., Liu, W., Xiang, Y., Diao, L., Guo, A.Y. and Han, L. (2017) LNCediting: a database for functional effects of RNA editing in lncRNAs. *Nucleic Acids Res.*, **45**, D79–D84.

30. Ramaswami,G. and Li,J.B. (2014) RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.*, **42**, D109–D113.
31. Cui,T., Dou,Y., Tan,P., Ni,Z., Liu,T., Wang,D., Huang,Y., Cai,K., Zhao,X., Xu,D. *et al.* (2021) RNALocate v2.0: an updated resource for RNA subcellular localization with increased coverage and annotation. *Nucleic Acids Res.*, <https://doi.org/10.1093/nar/gkab825>.
32. 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.
33. Bellaousov,S., Reuter,J.S., Seetin,M.G. and Mathews,D.H. (2013) RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.*, **41**, W471–W474.
34. Wang,M., Liu,X., Chang,G., Chen,Y., An,G., Yan,L., Gao,S., Xu,Y., Cui,Y., Dong,J. *et al.* (2018) Single-cell RNA sequencing analysis reveals sequential cell fate transition during human spermatogenesis. *Cell Stem Cell*, **23**, 599–614.
35. Chen,Y., Zheng,Y., Gao,Y., Lin,Z., Yang,S., Wang,T., Wang,Q., Xie,N., Hua,R., Liu,M. *et al.* (2018) Single-cell RNA-seq uncovers dynamic processes and critical regulators in mouse spermatogenesis. *Cell Res.*, **28**, 879–896.
36. Velten,L., Haas,S.F., Raffel,S., Blaszkiewicz,S., Islam,S., Hennig,B.P., Hirche,C., Lutz,C., Buss,E.C., Nowak,D. *et al.* (2017) Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.*, **19**, 271–281.
37. Zhou,F., Li,X., Wang,W., Zhu,P., Zhou,J., He,W., Ding,M., Xiong,F., Zheng,X., Li,Z. *et al.* (2016) Tracing haematopoietic stem cell formation at single-cell resolution. *Nature*, **533**, 487–492.
38. 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.
39. GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
40. John,B., Enright,A.J., Aravin,A., Tuschl,T., Sander,C. and Marks,D.S. (2004) Human MicroRNA targets. *PLoS Biol.*, **2**, e363.
41. Agarwal,V., Bell,G.W., Nam,J.W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, e05005.
42. Li,T., Wernersson,R., Hansen,R.B., Horn,H., Mercer,J., Slodkowitz,G., Workman,C.T., Rigina,O., Rapacki,K., Staerfeldt,H.H. *et al.* (2017) A scored human protein-protein interaction network to catalyze genomic interpretation. *Nat. Methods*, **14**, 61–64.
43. Esteller,M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, **12**, 861–874.
44. Coffre,M. and Koralov,S.B. (2017) miRNAs in B cell development and lymphomagenesis. *Trends Mol. Med.*, **23**, 721–736.