Published online 17 September 2019 *Nucleic Acids Research, 2020, Vol. 48, Database issue* ***D189–D197***
doi: 10.1093/nar/gkz804

RNAInter in 2020: RNA interactome repository with increased coverage and annotation

Yunqing Lin[1,2,†], Tianyuan Liu[1,†], Tianyu Cui[1,†], Zhao Wang[2], Yuncong Zhang[2], Puwen Tan[1], Yan Huang[3], Jia Yu[4] and Dong Wang[1,3,5,6,*]

[1]Department of Bioinformatics, School of Basic Medical Sciences, Southern Medical University, Guangzhou 510515, China, [2]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China, [3]Shunde Hospital, Southern Medical University (The First People's Hospital of Shunde), Foshan 528308, China, [4]State Key Laboratory of Medical Molecular Biology, Department of Biochemistry & Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences (CAMS) & Peking Union Medical College (PUMC), Beijing 100730, China, [5]Dermatology Hospital, Southern Medical University, Guangzhou 510091, China and [6]Center for Informational Biology, University of Electronic Science and Technology of China, Chengdu 611731, China

Received August 08, 2019; Revised September 03, 2019; Editorial Decision September 09, 2019; Accepted September 10, 2019

ABSTRACT

Research on RNA-associated interactions has exploded in recent years, and increasing numbers of studies are not limited to RNA–RNA and RNA–protein interactions but also include RNA–DNA/compound interactions. To facilitate the development of the interactome and promote understanding of the biological functions and molecular mechanisms of RNA, we updated RAID v2.0 to RNAInter (RNA Interactome Database), a repository for RNA-associated interactions that is freely accessible at http://www.rna-society.org/rnainter/ or http://www.rna-society.org/raid/. Compared to RAID v2.0, new features in RNAInter include (i) 8-fold more interaction data and 94 additional species; (ii) more definite annotations organized, including RNA editing/localization/modification/structure and homology interaction; (iii) advanced functions including fuzzy/batch search, interaction network and RNA dynamic expression and (iv) four embedded RNA interactome tools: RIscoper, IntaRNA, PRIdictor and DeepBind. Consequently, RNAInter contains >41 million RNA-associated interaction entries, involving more than 450 thousand unique molecules, including RNA, protein, DNA and compound. Overall, RNAInter provides a comprehensive RNA interactome resource for researchers and paves the way to investigate the regulatory landscape of cellular RNAs.

INTRODUCTION

RNA-associated interactions involve many physiological and pathological processes, such as cell growth and development, cell differentiation and inflammation (1–4). With the rapid development of biotechnology techniques, new RNA-associated interactions are being discovering continuously. These new techniques include Degradome-seq (5), LIGR-seq (6), MARIO (7) and PARIS (8) for the detection of RNA–RNA interactions (RRIs); dCLIP (9), PAR-CLIP (10), RIP-seq (11) and uvCLAP (12) for the detection of RNA–protein interactions (RPIs) and ChIRP-seq (13), ChOP-seq (14), diMARGI (15) and GRO-seq (16) for the detection of RNA–DNA interactions (RDIs) (see description in Supplementary Table S1). Recently, the regulatory roles of drug-associated miRNAs and lncRNAs in drug resistance have been a research focus (17–19). Transcription factors (TFs) and histone modifications contribute to the transcriptional regulation of RNA, which participates in various biological processes (20,21). The integration of these is therefore a prerequisite for RNA-related biomarker or mechanistic studies. However, many databases have manually collected and identified RNA-associated interactions through experimental validation and computational prediction from the literature and high-throughput sequencing. The majority of these resources focus on certain types of interactions with insufficient molecular information. Thus, numbers of annotations about RNA and other interactors, such as target sites, RNA editing and RNA modification, should be included. Currently, a global view of the RNA interactome with comprehensive annotations is not available across most species.

Here, we updated RAID v2.0 (22) to RNAInter (RNA Interactome Database, http://www.rna-society.org/

*To whom correspondence should be addressed. Tel: +86 20 61648279; Fax: +86 20 61648279; Email: wangdong@ems.hrbmu.edu.cn or wangdong79@smu.edu.cn
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

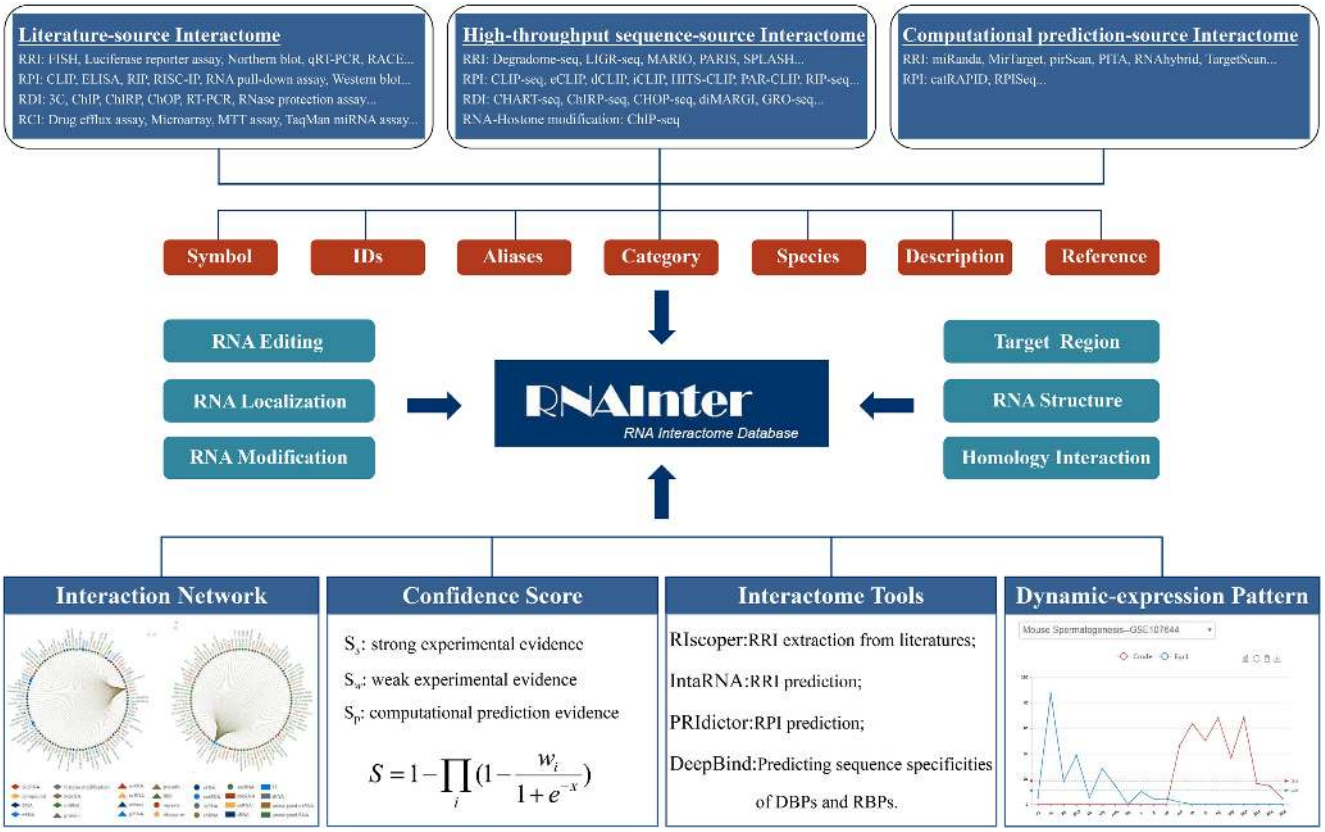

Figure 1. Overview of the RNAInter database.

rnainter/ or http://www.rna-society.org/raid/) to address these challenges. RNAInter establishes a repository of integrated experimentally validated and computationally predicted RNA-associated interactions through manual curation of the literature, along with another 35 resources under one common framework (Figure 1, Table 1). It also supports interaction network, RNA dynamic expression and four RNA interactome tools: RIscoper (23), IntaRNA (24), PRIdictor (25) and DeepBind (26) (Figures 1 and 4). In total, RNAInter integrated >41 million RNA-associated interactions across 154 species. It will provide a valuable resource for better understanding the RNA interactome.

DATA ORGANIZATION

Data collection

RNAInter integrated experimentally validated and computationally predicted RNA interactome data from the literature and another 35 resources (Table 1). Literature within PubMed (mainly from 2016 to 2019) was screened with the following keyword combinations: (RNA molecule) AND (other molecule) AND (interaction). The keyword in brackets represents (i) RNA molecule: RNA symbols or RNA category names and (ii) other molecule: RNA symbols or RNA category names, protein symbols or 'transcription factor' or 'RNA-binding protein' or 'protein', gene symbols or 'chromosome', compound symbols or 'compound' or 'drug', or histone modification symbols or 'histone modification'; and (iii) interaction: 'bind' or 'interact' or 'regular' or 'target'. Finally, we reviewed over 31 000 published studies that included 419 522 RNA-associated interactions. Diverse RNA-associated interactions were also integrated from 24 experimentally validated databases and 14 computationally predicted databases (22,27–60) (see details in Table 1).

To facilitate elucidating the role of RNA in molecular interactions, more annotation information for the interactors was collected, including RNA modification sites from RMBase v2.0 (61), RNA subcellular localization from RNALocate (62), and RNA editing sites from RADAR (63), DARNED (64) and Lnceditting (65). Simultaneously, the transcript and protein sequences from Refseq (66) and miRBase (67) were included to visualize the structure of RNA and represent target sites by miRanda, RIsearch (68) (tools for predicting RRIs), or PRIdictor (tool for predicting RPIs). The experimentally verified RNA-binding sites in proteins documented in the RBPDB (69), RsiteDB (70) and PDB (71) databases were also incorporated. Furthermore, we integrated the orthology/paralogy gene sets from miRBase and NCBI Gene (72) to reveal the conservation of homologous RNA-associated interactions across species.

Data procession

Integrating multisource data requires unifying them into common reference databases to annotate various interactors. Four major types of interactor symbols were used: (i) miRNA symbols from the miRBase database, (ii) DNA,

Table 1. Overview of curated interaction data from 35 resources

Evidence type	Interaction type	Interaction entry	Database resource	Reference
Experimental validation	RCI	4525	SM2miR	(27)
		4113	ncDR	(28)
		822	EmDL	(29)
	RDI	138 062	LnChrom	(30)
	RPI	1 530 693	POSTAR2	(31)
		199 835	TransmiR v2.0	(32)
	RRI	258 818	RISE	(33)
		155 622	LncRNA2Target v2.0	(34)
		7904	VIRmiRNA	(35)
		3028	LncACTdb 2.0	(36)
		2680	NPInter v3.0	(37)
		1846	OncomiRDB	(38)
		1213	ncRDeathDB	(39)
		559	miR2Disease	(40)
		405	sRNATarBase 3.0	(41)
		81	MNDR v2.0	(42)
		60	LncRNADisease 2.0	(43)
	RHI/RPI	9 515 123	ChIPBase v2.0	(44)
	RPI/RRI	1 246 631	starBase v2.0	(45)
		737 835	miRTarBase	(46)
Computational prediction	RPI	23 304 537	RNAct	(47)
	RRI	1 956 709	miRDB	(48)
		1 557 635	miRanda	(49)
		547 003	piRTarBase	(50)
		247 731	RepTar	(51)
		191 123	TargetScan	(52)
		149 817	EIMMo	(53)
		106 471	DroID	(54)
		74 884	ZIKV - CDB	(55)
		243	HumanViCe	(56)
		14	miRcode	(57)
Experimental validation/ Computational prediction	RRI	538 529	VmiReg	(58)
	RPI/RRI	5 272 396	RAID v2.0	(22)
		327 123	RAIN	(59)
		110 293	ViRBase	(60)

Table 2. The features and developments of RNAInter

Feature	RAID v1.0	RAID v2.0	RNAInter
Interaction entry*	6112 (6112)	5 272 396 (2 426 181)	41 322 577 (13,653,108)
RNA symbol	2070	118 878	381 319
Species coverage	1	60	154
Interaction type	RNA-protein/RNA-RNA	RNA-Protein/RNA-RNA	RNA-Protein/RNA-RNA/RNA-Compound/RNA-DNA/RNA-Histone modification
RNA category	lncRNA/miRNA/mRNA/rRNA/snoRNA	circRNA/lncRNA/miRNA/miscRNA/mRNA/pseudogene/rRNA/scRNA/snRNA/snoRNA/snRNA/tRNA	circRNA/lncRNA/miRNA/miscRNA/mRNA/pseudogene/rRNA/scRNA/snRNA/snoRNA/snRNA/sRNA/tRNA/eRNA/ncRNA/piRNA/repeats/ribozyme/scaRNA/shRNA/sRNA
Detailed information	Basic annotations/Evidence support/Reference/Tissue or cell line	Basic annotations/Evidence support/Interactor homolog/Integrated confidence score/Reference/RNA-binding sites	Basic annotations/Evidence support/Interactor homolog/Integrated confidence score/Reference/RNA-binding sites/Homology interaction/RNA editing/RNA localization/RNA modification/Target region
Data visualization	Predicted binding sites/ Interaction network	Predicted binding sites	Predicted binding sites/Interaction network/RNA dynamic expression/RNA structure
Web application	-	Advanced filter search	Exact search/Batch search/Fuzzy search/Four interactome tools: RIscoper, IntaRNA, PRIdictor, DeepBind

*The number in brackets counts interactions entries verified by experimental methods.

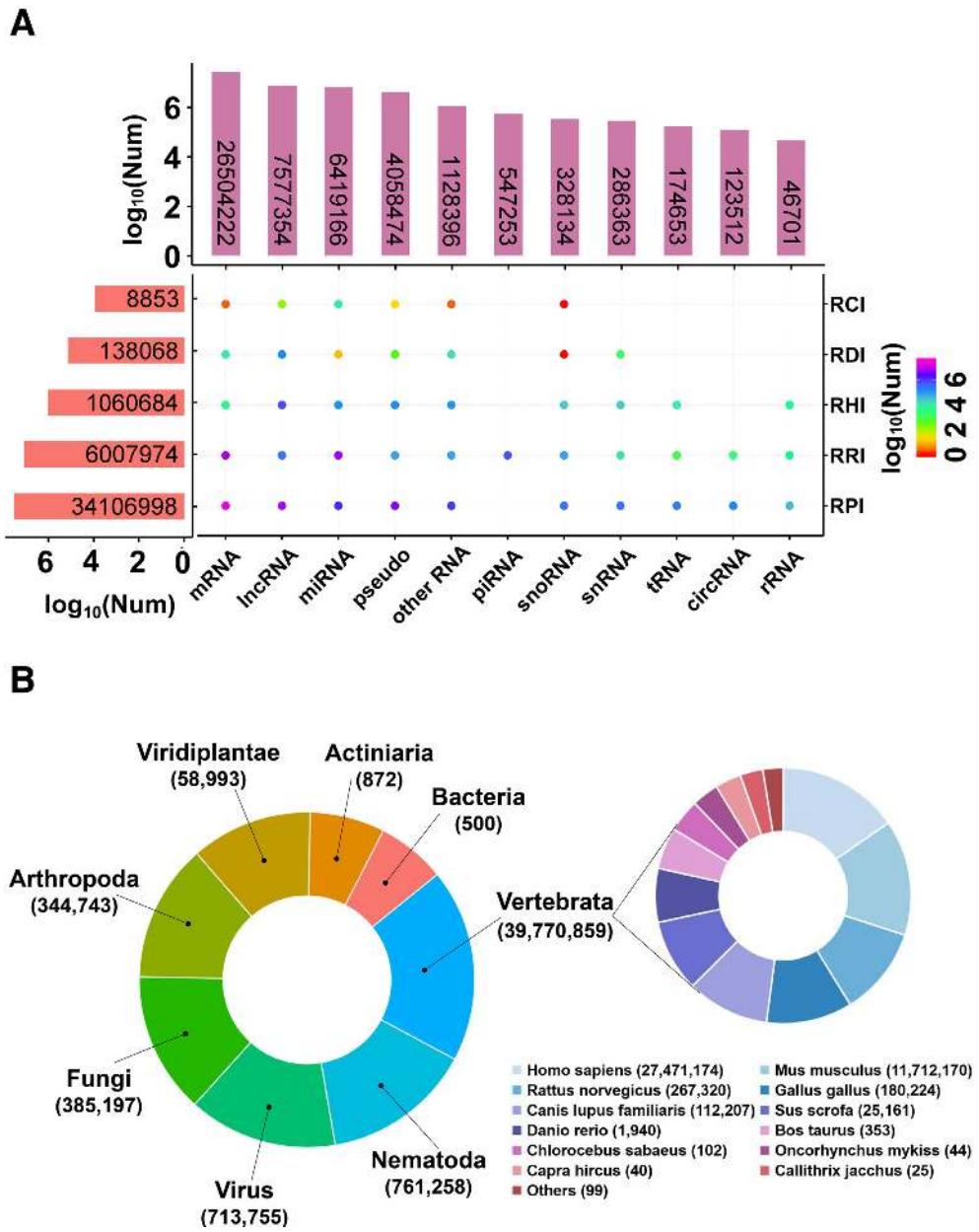

Figure 2. Statistics on RNAInter. (**A**) The distribution of five interaction types (RCI/RDI/RHI/RPI/RRI) in 22 RNA categories. The category 'other RNA' includes eRNA, ncRNA, others, repeats, ribozyme, scaRNA, scRNA, shRNA, sncRNA sRNA, unassigned RNA and unknown. (**B**) Number of interactions in vertebrata, nematoda, virus, arthropoda, fungi, viridiplantae, actiniaria, bacteria (left) and 28 species belonging to vertebrata (right).

RNA and protein symbols from the NCBI Gene or Ensembl (73) database, (iii) compound symbols from the PubChem Compound (74) database and (iv) histone modification symbols from the ChIPBase v2.0 database. Notably, each histone undergoes various modifications, and we separated RNA–histone modification interactions (RHIs) from RPIs to specify the relationship between RNA and histone modification. Additionally, Entrez ID, Ensembl Gene ID, miRBase accession, PubChem Compound CID and their external links are also provided, which can efficiently retrieve a substantial amount of genome-associated information from external resources. For the convenience of users, interactor information also included NCBI Aliases, DrugBank Aliases, OMIM ID, HGNC ID, HPRD ID, UniprotKB protein accession, among others. The software 'RNAstructure' (75) was used to predict RNA secondary structure.

In particular, we collected and processed four single-cell RNA-seq (scRNA-seq) data sets from the Gene Expression Omnibus (GEO) (76) to visualize the RNA molecular dynamic expression pattern during diverse stages of human (or mouse) spermatogenesis and HSC lineage commitment (77,78). Firstly, scRNA-seq reads were downloaded and processed to remove adaptor contaminants and low-quality bases using trimmomatics v0.36 (79). The processed clean reads were aligned to the human and mouse reference genome (hg38 and mm10 from GENCODE) using

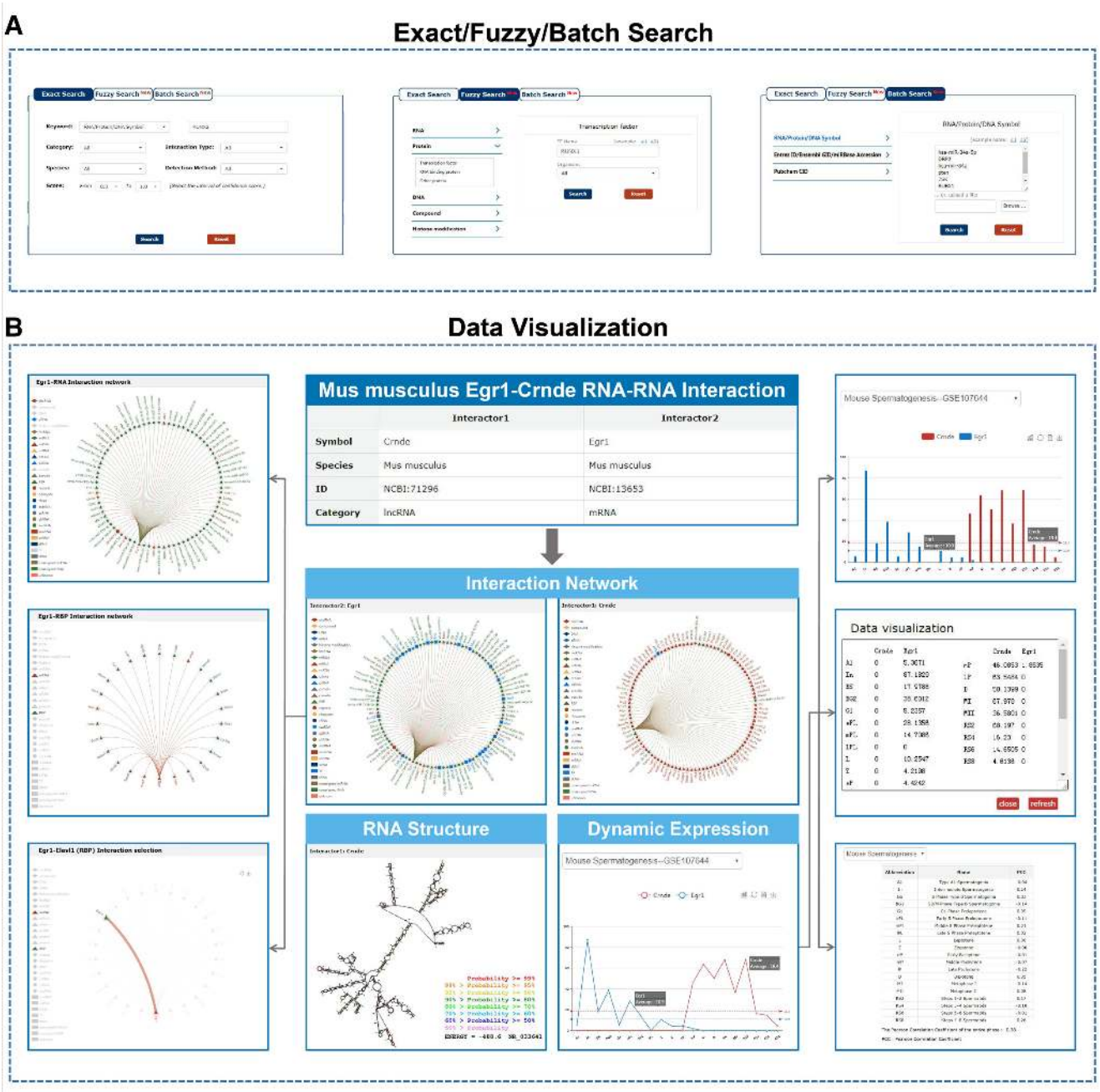

Figure 3. New search function and data visualization of the RNA interactome. (**A**) Presentation of exact, fuzzy and batch search described in the search options. (**B**) Visualization of the interaction network, RNA structure and RNA dynamic expression.

TopHat v2.0.12 (80). The HTSeq v0.11.0 (81) was used to estimate the gene expression of each single-cell. The transcript copy number, counted by distinct unique molecular identifiers (UMIs), was obtained by removing duplicated transcripts according to the UMI information. For a given cell, the number of UMIs represents the transcript number of each gene. Secondly, we filtered out cells with fewer than 2000 genes and 10 000 transcripts to retain high-quality cells. In total, we obtained 2414 human bone marrow cells (GSE75478), 99 mouse precursor-haematopoietic stem cells (GSE67120), 2,435 human testicular cells (GSE106487) and 1136 mouse spermatogenic cells (GSE107644). The RNA expression levels were normalized by transcripts per million (TPM). Finally, we evaluated the correlation between two RNAs with the Pearson correlation coefficient (PCC) during human (or mouse) spermatogenesis and HSC lineage commitment.

RESULTS

RNAInter statistics

In summary, RNAInter contains 41 322 577 RNA-associated interactions, including 34 106 998 RPIs, 6 007

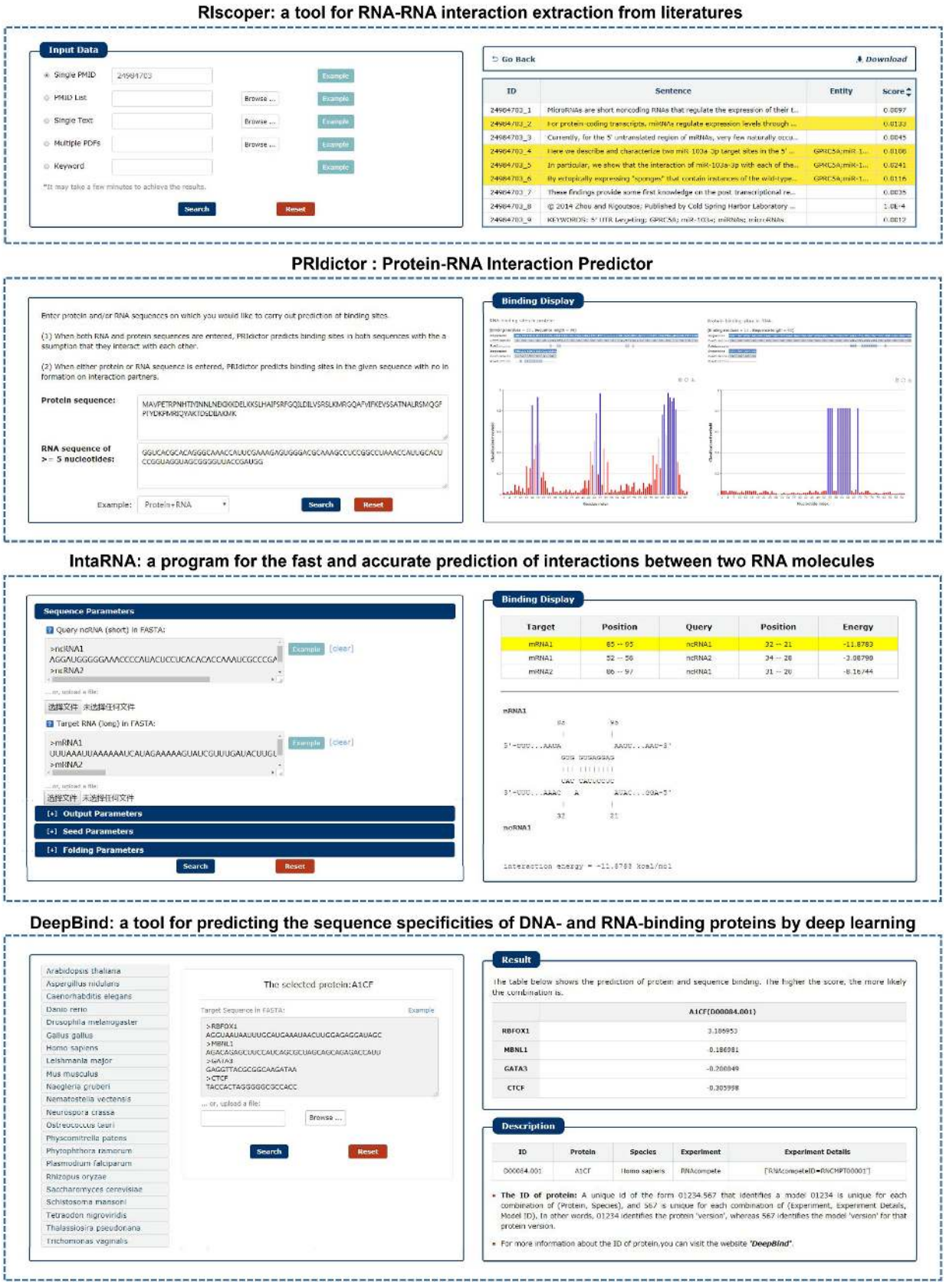

Figure 4. Snapshot of four RNA interactome tools in RNAInter: RIscoper, PRIdictor, IntaRNA and DeepBind (left: input option, right: result presentation).

974 RRIs, 1 060 684 RHIs, 138 068 RDIs and 8853 RNA-compound interactions (RCIs) (Figure 2A, Table 1). These interactions involve 381 319 nonredundant RNAs and 42 215 nonredundant proteins, 33 970 newly added nonredundant DNAs, 425 nonredundant compounds and 61 nonredundant histone modifications. RNAInter involved 22 RNA types, eight of which added for the first time, including enhancer RNA (eRNA), Piwi-interacting RNA (piRNA), repeats, ribozyme, short hairpin RNA (shRNA), small Cajal body-specific RNA (scaRNA), small RNA (sRNA) and noncoding RNA (indefinite classified ncRNA) (Table 2). The distribution of the five types of interactions among different RNAs is shown in Figure 2A. The number of organisms in RNAInter increased from 60 to 154 compared with that in RAID v2.0 (Table 2). All the species covered nine categories (actiniaria, arthropoda, bacteria, fungi, mycetozoa, nematode, vertebrata, viridiplantae, virus). *Homo sapiens* and *Mus musculus* interactions took up the main part of the vertebrata (Figure 2B). Other model organisms, such as *Drosophila melanogaster, Rattus norvegicus, Saccharomyces cerevisiae* and zebrafish (*Danio rerio*), have also been documented in RNAInter.

Data feature and utility

Then, we expanded the RNA-compound, RNA–DNA and RNA–histone modification interactions in RNAInter. Apart from basic annotation, support evidence, RNA-binding sites and references, we focused on the multifaceted supplementation of the details of RNA editing/localization/modification/structure/dynamic expression, the interaction network, the target region and the homology interaction in detail. 'RNA editing' provides editing position, editing type and genetic region. 'RNA localization' includes subcellular localization and the tissue or cell line. 'RNA modification' involves the modification position, modification type and genetic region. Moreover, 'Homology interaction' shows the conservative interactions across organisms documented in RNAInter. 'Target region' shows the target locus in RHI/RPI/RRI and data accession from the literature or high-throughput sequencing with their sample resources. All this information links to their corresponding databases.

RNAInter provides a user-friendly platform for searching, browsing, visualizing and profiling RNA interactome data. To improve the search capability, RNAInter enables an optimized query with a new function of fuzzy and batch search. Fuzzy Search can help users to search interactions using unstandardized or uncertained interactor name under selected molecular category, then the result of interactions will be presented by selecting interactors in candidate list. Meanwhile, Batch Search supports for inputting a list of official symbols/IDs or uploading a file with text format to obtain multiple molecular categories associated interactions. Thus, users can select 'Exact Search' to filter the search results, or 'Fuzzy Search' to further focus on interactors of interest, or 'Batch Search' to customize their query content in batch (Figure 3A). Taking the load time into account, RNAInter offers the download option for over 2 million entries on the 'Browse' page. 'RNA structure' represents the putative RNA secondary structure for each transcript. In addition, 'Interaction network' is offered to picture the top 100 interactions ranked by integrative confidence score in RNAInter. Users can also select specific categories of RNA-associated interactions by clicking the different icons of interactor to conceal uninterested interactions for superior view. Click any edge of the network can jump to a detailed page of the corresponding entry (Figure 3B). To illustrate the RNA molecular dynamic expression pattern, 'Dynamic expression' shows the line chart of RNA expression values in each stage during human (or mouse) spermatogenesis and HSC lineage commitment and their expression correlation in each stage and entire phage with PCC (Figure 3B). The images of the interaction network and dynamic expression pattern can be downloaded.

Extended toolkit

In response to the diverse needs of users, RNAInter embeds four interactome tools: RIscoper, IntaRNA, PRIdictor and DeepBind. RIscoper is a tool for RNA–RNA interaction extraction from the literature. IntaRNA is a program for the fast and accurate prediction of interactions between two RNA molecules. PRIdictor is a protein–RNA interaction predictor. DeepBind predicted the sequence specificities of DNA- and RNA-binding proteins by deep learning (Figure 4).

CONCLUSIONS AND PERSPECTIVES

RNAInter is an update of RAID v2.0, a comprehensive resource for RNA interactome data obtained from the literature and other databases, containing over 41 million RNA-associated interactions of RCI, RDI, RHI, RPI and RRI. With detailed interactome information, visualized interaction network and RNA dynamic expression, enhanced search functions, and embedded RNA interactome tools, RNAInter depicts a system-level RNA interactome landscape with guides and help researchers to perform further studies. We expect RNAInter to update the manual curation of RNA interactome data and expand the available information about RNAs and other molecules in the future. Continuously integrating high-throughput data, including scRNA-seq, to provide more precise depiction of the dynamic expression pattern of RNAs illuminates the role of RNA across organisms. We may optimize the confidence score strategy with the emergence of new mass sequencing technologies, experimental methods and prediction algorithms. At the same time, more RNA-associated applications are docking with our database. Eventually, RNAInter will present the most comprehensive map of the RNA interactome to satisfy different requirements.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Natural Science Foundation of China [81770104]. Funding for open access charge: National Natural Science Foundation of China[81770104].

Conflict of interest statement. None declared.

REFERENCES

1. Lal,A., Navarro,F., Maher,C.A., Maliszewski,L.E., Yan,N., O'Day,E., Chowdhury,D., Dykxhoorn,D.M., Tsai,P., Hofmann,O. *et al.* (2009) miR-24 Inhibits cell proliferation by targeting E2F2, MYC, and other cell-cycle genes via binding to "seedless" 3′UTR microRNA recognition elements. *Mol. Cell*, **35**, 610–625.
2. Nowakowski,T.J., Rani,N., Golkaram,M., Zhou,H.R., Alvarado,B., Huch,K., West,J.A., Leyrat,A., Pollen,A.A., Kriegstein,A.R. *et al.* (2018) Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. *Nat. Neurosci.*, **21**, 1784–1792.
3. Baumjohann,D. and Ansel,K.M. (2013) MicroRNA-mediated regulation of T helper cell differentiation and plasticity. *Nat. Rev. Immunol.*, **13**, 666–678.
4. Yoshikawa,T., Wu,J., Otsuka,M., Kishikawa,T., Suzuki,N., Takata,A., Ohno,M., Ishibashi,R., Yamagami,M., Nakagawa,R. *et al.* (2017) Repression of MicroRNA function mediates Inflammation-associated colon tumorigenesis. *Gastroenterology*, **152**, 631–643.
5. Shi,M., Hu,X., Wei,Y., Hou,X., Yuan,X., Liu,J. and Liu,Y. (2017) Genome-Wide profiling of small RNAs and degradome revealed conserved regulations of miRNAs on Auxin-Responsive genes during fruit enlargement in peaches. *Int. J. Mol. Sci.*, **18**, E2599.
6. Sharma,E., Sterne-Weiler,T., O'Hanlon,D. and Blencowe,B.J. (2016) Global mapping of human RNA–RNA interactions. *Mol. Cell*, **62**, 618–626.
7. Nguyen,T.C., Cao,X., Yu,P., Xiao,S., Lu,J., Biase,F.H., Sridhar,B., Huang,N., Zhang,K. and Zhong,S. (2016) Mapping RNA–RNA interactome and RNA structure in vivo by MARIO. *Nat. Commun.*, **7**, 12023.
8. Lu,Z., Zhang,Q.C., Lee,B., Flynn,R.A., Smith,M.A., Robinson,J.T., Davidovich,C., Gooding,A.R., Goodrich,K.J., Mattick,J.S. *et al.* (2016) RNA duplex map in living cells reveals Higher-Order transcriptome structure. *Cell*, **165**, 1267–1279.
9. Rosenberg,M., Blum,R., Kesner,B., Maier,V.K., Szanto,A. and Lee,J.T. (2017) Denaturing CLIP, dCLIP, pipeline identifies discrete RNA footprints on Chromatin-Associated proteins and reveals that CBX7 targets 3′ UTRs to regulate mRNA expression. *Cell Syst.*, **5**, 368–385.
10. Hafner,M., Landthaler,M., Burger,L., Khorshid,M., Hausser,J., Berninger,P., Rothballer,A., Ascano,M. Jr, Jungkamp,A.C., Munschauer,M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
11. Zhao,J., Ohsumi,T.K., Kung,J.T., Ogawa,Y., Grau,D.J., Sarma,K., Song,J.J., Kingston,R.E., Borowsky,M. and Lee,J.T. (2010) Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell*, **40**, 939–953.
12. Maticzka,D., Ilik,I.A., Aktas,T., Backofen,R. and Akhtar,A. (2018) uvCLAP is a fast and non-radioactive method to identify in vivo targets of RNA-binding proteins. *Nat. Commun.*, **9**, 1142.
13. Chu,C., Qu,K., Zhong,F.L., Artandi,S.E. and Chang,H.Y. (2011) Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. *Mol. Cell*, **44**, 667–678.
14. Akhade,V.S., Arun,G., Donakonda,S. and Rao,M.R. (2014) Genome wide chromatin occupancy of mrhl RNA and its role in gene regulation in mouse spermatogonial cells. *RNA Biol.*, **11**, 1262–1279.
15. Sridhar,B., Rivas-Astroza,M., Nguyen,T.C., Chen,W., Yan,Z., Cao,X., Hebert,L. and Zhong,S. (2017) Systematic mapping of RNA-Chromatin interactions in vivo. *Curr. Biol.: CB*, **27**, 602–609.
16. Yang,L., Lin,C., Jin,C., Yang,J.C., Tanasa,B., Li,W., Merkurjev,D., Ohgi,K.A., Meng,D., Zhang,J. *et al.* (2013) lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature*, **500**, 598–602.
17. Lin,A., Hu,Q., Li,C., Xing,Z., Ma,G., Wang,C., Li,J., Ye,Y., Yao,J., Liang,K. *et al.* (2017) The LINK-A lncRNA interacts with PtdIns(3,4,5)P3 to hyperactivate AKT and confer resistance to AKT inhibitors. *Nat. Cell Biol.*, **19**, 238–251.
18. Malek,E., Jagannathan,S. and Driscoll,J.J. (2014) Correlation of long non-coding RNA expression with metastasis, drug resistance and clinical outcome in cancer. *Oncotarget*, **5**, 8027–8038.
19. Rodriguez-Barrueco,R., Nekritz,E.A., Bertucci,F., Yu,J., Sanchez-Garcia,F., Zeleke,T.Z., Gorbatenko,A., Birnbaum,D., Ezhkova,E., Cordon-Cardo,C. *et al.* (2017) miR-424(322)/503 is a breast cancer tumor suppressor whose loss promotes resistance to chemotherapy. *Genes Dev.*, **31**, 553–566.
20. Zhang,H.M., Kuang,S., Xiong,X., Gao,T., Liu,C. and Guo,A.Y. (2015) Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. *Brief. Bioinform.*, **16**, 45–58.
21. Elsheikh,S.E., Green,A.R., Rakha,E.A., Powe,D.G., Ahmed,R.A., Collins,H.M., Soria,D., Garibaldi,J.M., Paish,C.E., Ammar,A.A. *et al.* (2009) Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. *Cancer Res.*, **69**, 3802–3809.
22. 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.
23. Zhang,Y., Liu,T., Chen,L., Yang,J., Yin,J., Zhang,Y., Yun,Z., Xu,H., Ning,L., Guo,F. *et al.* (2019) RIscoper: a tool for RNA–RNA interaction extraction from the literature. *Bioinformatics*, **35**, 3199–3202.
24. Mann,M., Wright,P.R. and Backofen,R. (2017) IntaRNA 2.0: enhanced and customizable prediction of RNA–RNA interactions. *Nucleic Acids Res.*, **45**, W435–W439.
25. Tuvshinjargal,N., Lee,W., Park,B. and Han,K. (2016) PRIdictor: protein-RNA Interaction predictor. *Biosystems*, **139**, 17–22.
26. Alipanahi,B., Delong,A., Weirauch,M.T. and Frey,B.J. (2015) Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.*, **33**, 831–838.
27. Liu,X., Wang,S., Meng,F., Wang,J., Zhang,Y., Dai,E., Yu,X., Li,X. and Jiang,W. (2013) SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics*, **29**, 409–411.
28. Dai,E., Yang,F., Wang,J., Zhou,X., Song,Q., An,W., Wang,L. and Jiang,W. (2017) ncDR: a comprehensive resource of non-coding RNAs involved in drug resistance. *Bioinformatics*, **33**, 4010–4011.
29. Xie,W., Yan,H. and Zhao,X.M. (2017) EmDL: extracting miRNA-drug interactions from literature. *IEEE/ACM Trans. Comput. Biol. Bioinf.*, doi:10.1109/TCBB.2017.2723394.
30. Yu,F., Zhang,G., Shi,A., Hu,J., Li,F., Zhang,X., Zhang,Y., Huang,J., Xiao,Y., Li,X. *et al.* (2018) LnChrom: a resource of experimentally validated lncRNA-chromatin interactions in human and mouse. *Database*, **2018**, bay039.
31. Zhu,Y., Xu,G., Yang,Y.T., Xu,Z., Chen,X., Shi,B., Xie,D., Lu,Z.J. and Wang,P. (2019) POSTAR2: deciphering the post-transcriptional regulatory logics. *Nucleic Acids Res.*, **47**, D203–D211.
32. Tong,Z., Cui,Q., Wang,J. and Zhou,Y. (2019) TransmiR v2.0: an updated transcription factor-microRNA regulation database. *Nucleic Acids Res.*, **47**, D253–D258.
33. Gong,J., Shao,D., Xu,K., Lu,Z., Lu,Z.J., Yang,Y.T. and Zhang,Q.C. (2018) RISE: a database of RNA interactome from sequencing experiments. *Nucleic Acids Res.*, **46**, D194–D201.
34. Cheng,L., Wang,P., Tian,R., Wang,S., Guo,Q., Luo,M., Zhou,W., Liu,G., Jiang,H. and Jiang,Q. (2019) LncRNA2Target v2.0: a comprehensive database for target genes of lncRNAs in human and mouse. *Nucleic Acids Res.*, **47**, D140–D144.
35. Qureshi,A., Thakur,N., Monga,I., Thakur,A. and Kumar,M. (2014) VIRmiRNA: a comprehensive resource for experimentally validated viral miRNAs and their targets. *Database*, **2014**, bau103.
36. Wang,P., Li,X., Gao,Y., Guo,Q., Wang,Y., Fang,Y., Ma,X., Zhi,H., Zhou,D., Shen,W. *et al.* (2019) LncACTdb 2.0: an updated database of experimentally supported ceRNA interactions curated from low- and high-throughput experiments. *Nucleic Acids Res.*, **47**, D121–D127.
37. Hao,Y., Wu,W., Li,H., Yuan,J., Luo,J., Zhao,Y. and Chen,R. (2016) NPInter v3.0: an upgraded database of noncoding RNA-associated interactions. *Database*, **2016**, baw057.
38. Wang,D., Gu,J., Wang,T. and Ding,Z. (2014) OncomiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs. *Bioinformatics*, **30**, 2237–2238.
39. Wu,D., Huang,Y., Kang,J., Li,K., Bi,X., Zhang,T., Jin,N., Hu,Y., Tan,P., Zhang,L. *et al.* (2015) ncRDeathDB: A comprehensive

bioinformatics resource for deciphering network organization of the ncRNA-mediated cell death system. *Autophagy*, **11**, 1917–1926.
40. Jiang,Q., Wang,Y., Hao,Y., Juan,L., Teng,M., Zhang,X., Li,M., Wang,G. and Liu,Y. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–D104.
41. Wang,J., Liu,T., Zhao,B., Lu,Q., Wang,Z., Cao,Y. and Li,W. (2016) sRNATarBase 3.0: an updated database for sRNA-target interactions in bacteria. *Nucleic Acids Res.*, **44**, D248–D253.
42. Cui,T., Zhang,L., Huang,Y., Yi,Y., Tan,P., Zhao,Y., Hu,Y., Xu,L., Li,E. and Wang,D. (2018) MNDR v2.0: an updated resource of ncRNA-disease associations in mammals. *Nucleic Acids Res.*, **46**, D371–D374.
43. Bao,Z., Yang,Z., Huang,Z., Zhou,Y., Cui,Q. and Dong,D. (2019) LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.*, **47**, D1034–D1037.
44. Zhou,K.R., Liu,S., Sun,W.J., Zheng,L.L., Zhou,H., Yang,J.H. and Qu,L.H. (2017) ChIPBase v2.0: decoding transcriptional regulatory networks of non-coding RNAs and protein-coding genes from ChIP-seq data. *Nucleic Acids Res.*, **45**, D43–D50.
45. Li,J.H., Liu,S., Zhou,H., Qu,L.H. and Yang,J.H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.*, **42**, D92–D97.
46. Chou,C.H., Shrestha,S., Yang,C.D., Chang,N.W., Lin,Y.L., Liao,K.W., Huang,W.C., Sun,T.H., Tu,S.J., Lee,W.H. *et al.* (2018) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.*, **46**, D296–D302.
47. Lang,B., Armaos,A. and Tartaglia,G.G. (2019) RNAct: Protein-RNA interaction predictions for model organisms with supporting experimental data. *Nucleic Acids Res.*, **47**, D601–D606.
48. Wong,N. and Wang,X. (2015) miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.*, **43**, D146–D152.
49. Betel,D., Koppal,A., Agius,P., Sander,C. and Leslie,C. (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.*, **11**, R90.
50. Wu,W.S., Brown,J.S., Chen,T.T., Chu,Y.H., Huang,W.C., Tu,S. and Lee,H.C. (2019) piRTarBase: a database of piRNA targeting sites and their roles in gene regulation. *Nucleic Acids Res.*, **47**, D181–D187.
51. Elefant,N., Berger,A., Shein,H., Hofree,M., Margalit,H. and Altuvia,Y. (2011) RepTar: a database of predicted cellular targets of host and viral miRNAs. *Nucleic Acids Res.*, **39**, D188–D194.
52. Agarwal,V., Bell,G.W., Nam,J.W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, **4**, e05005.
53. Gaidatzis,D., van Nimwegen,E., Hausser,J. and Zavolan,M. (2007) Inference of miRNA targets using evolutionary conservation and pathway analysis. *BMC Bioinformatics*, **8**, 69.
54. Murali,T., Pacifico,S., Yu,J., Guest,S., Roberts,G.G. 3rd and Finley,R.L. Jr (2011) DroID 2011: a comprehensive, integrated resource for protein, transcription factor, RNA and gene interactions for Drosophila. *Nucleic Acids Res.*, **39**, D736–D743.
55. Pylro,V.S., Oliveira,F.S., Morais,D.K., Cuadros-Orellana,S., Pais,F.S., Medeiros,J.D., Geraldo,J.A., Gilbert,J., Volpini,A.C. and Fernandes,G.R. (2016) ZIKV - CDB: A collaborative database to guide research linking SncRNAs and ZIKA virus disease symptoms. *PLoS Negl. Trop. Dis.*, **10**, e0004817.
56. Ghosal,S., Das,S., Sen,R. and Chakrabarti,J. (2014) HumanViCe: host ceRNA network in virus infected cells in human. *Front. Genet.*, **5**, 249.
57. Jeggari,A., Marks,D.S. and Larsson,E. (2012) miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. *Bioinformatics*, **28**, 2062–2063.
58. Shao,T., Zhao,Z., Wu,A., Bai,J., Li,Y., Chen,H., Jiang,C., Wang,Y., Li,S., Wang,L. *et al.* (2015) Functional dissection of virus-human crosstalk mediated by miRNAs based on the VmiReg database. *Mol. Biosyst.*, **11**, 1319–1328.
59. Junge,A., Refsgaard,J.C., Garde,C., Pan,X., Santos,A., Alkan,F., Anthon,C., von Mering,C., Workman,C.T., Jensen,L.J. *et al.* (2017) RAIN: RNA–protein association and interaction networks. *Database*, **2017**, baw167.
60. Li,Y., Wang,C., Miao,Z., Bi,X., Wu,D., Jin,N., Wang,L., Wu,H., Qian,K., Li,C. *et al.* (2015) ViRBase: a resource for virus-host ncRNA-associated interactions. *Nucleic Acids Res.*, **43**, D578–D582.
61. 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.
62. 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.
63. Ramaswami,G. and Li,J.B. (2014) RADAR: a rigorously annotated database of A-to-I RNA editing. *Nucleic Acids Res.*, **42**, D109–D113.
64. Kiran,A. and Baranov,P.V. (2010) DARNED: a DAtabase of RNa EDiting in humans. *Bioinformatics*, **26**, 1772–1776.
65. 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.
66. Haft,D.H., DiCuccio,M., Badretdin,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.
67. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
68. Wenzel,A., Akbasli,E. and Gorodkin,J. (2012) RIsearch: fast RNA–RNA interaction search using a simplified nearest-neighbor energy model. *Bioinformatics*, **28**, 2738–2746.
69. Cook,K.B., Kazan,H., Zuberi,K., Morris,Q. and Hughes,T.R. (2011) RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.*, **39**, D301–D308.
70. Shulman-Peleg,A., Nussinov,R. and Wolfson,H.J. (2009) RsiteDB: a database of protein binding pockets that interact with RNA nucleotide bases. *Nucleic Acids Res.*, **37**, D369–D373.
71. Burley,S.K., Berman,H.M., Bhikadiya,C., Bi,C., Chen,L., Di Costanzo,L., Christie,C., Dalenberg,K., Duarte,J.M., Dutta,S. *et al.* (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.*, **47**, D464–D474.
72. 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.
73. Cunningham,F., Achuthan,P., Akanni,W., Allen,J., Amode,M.R., Armean,I.M., Bennett,R., Bhai,J., Billis,K., Boddu,S. *et al.* (2019) Ensembl 2019. *Nucleic Acids Res.*, **47**, D745–D751.
74. Kim,S., Thiessen,P.A., Bolton,E.E., Chen,J., Fu,G., Gindulyte,A., Han,L., He,J., He,S., Shoemaker,B.A. *et al.* (2016) PubChem substance and compound databases. *Nucleic Acids Res.*, **44**, D1202–D1213.
75. 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.
76. 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.
77. 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.
78. 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.
79. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
80. Trapnell,C., Pachter,L. and Salzberg,S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.
81. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.