

NPInter v5.0: ncRNA interaction database in a new era

Yu Zheng^{1,2,†}, Huaxia Luo^{1,†}, Xueyi Teng^{1,†}, Xinpei Hao^{1,3}, Xiaoyu Yan^{1,3}, Yiheng Tang¹, Wanyu Zhang¹, Yuanxin Wang^{1,3}, Peng Zhang¹, Yanyan Li¹, Yi Zhao⁴, Runsheng Chen^{1,*} and Shunmin He^{1,3,*}

¹Key Laboratory of RNA Biology, Center for Big Data Research in Health, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China, ²College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China, ³University of Chinese Academy of Sciences, Beijing 100049, China and ⁴Bioinformatics Research Group, Key Laboratory of Intelligent Information Processing, Advanced Computing Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Received September 15, 2022; Revised October 16, 2022; Editorial Decision October 17, 2022; Accepted October 21, 2022

ABSTRACT

Noncoding RNAs (ncRNAs) play key regulatory roles in biological processes by interacting with other biomolecules. With the development of high-throughput sequencing and experimental technologies, extensive ncRNA interactions have been accumulated. Therefore, we updated the NPInter database to a fifth version to document these interactions. ncRNA interaction entries were doubled from 1 100 618 to 2 596 695 by manual literature mining and high-throughput data processing. We integrated global RNA–DNA interactions from iMARGI, ChAR-seq and GRID-seq, greatly expanding the number of RNA–DNA interactions (from 888 915 to 8 329 382). In addition, we collected different types of RNA interaction between SARS-CoV-2 virus and its host from recently published studies. Long noncoding RNA (lncRNA) expression specificity in different cell types from tumor single cell RNA-seq (scRNA-seq) data were also integrated to provide a cell-type level view of interactions. A new module named RBP was built to display the interactions of RNA-binding proteins with annotations of localization, binding domains and functions. In conclusion, NPInter v5.0 (<http://bigdata.ibp.ac.cn/npinter5/>) provides informative and valuable ncRNA interactions for biological researchers.

INTRODUCTION

The ENCODE project has revealed that up to 80% of the human genome is transcribed but <2% of the genome encodes proteins, with the vast majority of RNA transcripts being ncRNAs (1). ncRNAs can regulate cell physiology

and shape cellular functions in the whole biological regulatory network through interactions with other biomolecules, such as proteins, mRNAs and the genome (2). For example, microRNAs (miRNAs), well-known small single-stranded ncRNAs, usually bind to the 3'-untranslated region (UTR) of target mRNAs and drive post-transcriptional gene regulation (3). The lncRNA, *XIST*, can regulate dosage compensation by interacting with X chromosome in the female embryo (4). The study of ncRNA interaction is important in ncRNA function research.

Given the importance of ncRNA interactions in biological regulation, we constructed and released the first version of NPInter in 2006 to gather and arrange ncRNA interactions from experimental validation and high-throughput data (5). In the subsequent 15 years, we have updated and improved the database in three versions (6–8). In recent years, many experimental and high-throughput technologies have been developed and widely applied to identify RNA interactions, such as PARIS (9), SPLASH (10), LIGR-seq (11), CLIP-seq (12), ChIRP-seq (13) and MARGI (14). ncRNA interactome data has expanded extraordinarily and covers numerous organisms and tissues. We therefore updated NPInter to a fifth version to integrate these expanding resources and to provide comprehensive functional annotations to help ncRNA researchers look into the whole transcriptomic regulatory networks. In addition to a doubling of entries compared with the last version, we have also added global RNA–DNA interactions detected by iMARGI (15), ChAR-seq (16) and GRID-seq (17), which greatly expand the number of interactions between ncRNAs and chromosomes. To contribute to COVID-19 research, we also collected different types of RNA interaction between SARS-CoV-2 virus and its host from recently published studies. In addition, lncRNA expression specificity in different cell types from tumor scRNA-seq data were integrated into this version to

*To whom correspondence should be addressed. Tel: +86 10 64887032; Fax: +86 10 64887032; Email: heshunmin@ibp.ac.cn
Correspondence may also be addressed to Runsheng Chen. Tel: +86 10 64888543; Fax: +86 10 64871293 Email: rschen@ibp.ac.cn

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

provide a cell-type level view of interactions. Based on user feedback, we built the RBP module, which deposits RNA-binding proteins (RBPs) with annotations of localization, binding domains and functions. All interaction data can be freely downloaded from the download page.

DATA COLLECTION AND ANNOTATION

In NPInter v5.0, new RNA interactions from manual literature mining and high-throughput datasets were integrated with NPInter v4.0 interactions (8). Interactions from different sources were added to the database with redundancy removed. Biomolecules were then allocated with standard IDs from reference databases according to their categories. The overall workflow of data integration is shown in Figure 1.

Interactions curated from literature mining

To collect experimentally validated interactions between ncRNAs and biomolecules (proteins, RNAs and DNAs), we manually curated interaction pairs from public literature documented in PubMed. Over ten thousand related papers published between May 2019 and March 2022 were collected. After manual and stringent curation, we obtained 8587 interactions with experimental evidence from 5143 articles. The keywords for the PubMed literature search are listed in the Supplementary Material. In addition, we specifically collected SARS-CoV-2 related RNA interactions from literature published since 2020.

RNA–DNA interactions from ChIRP-seq and CHART-seq data

To collect RNA–DNA interactions from high-throughput datasets, we downloaded and processed 156 raw SRA ChIRP-seq datasets (13) and 117 CHART-seq datasets (18). CHART-seq data was newly added in this version. Raw sequence reads were mapped to the reference genome (hg19, mm9) using bwa (19), and then MACS2 (20) was employed to call peaks. Genome binding sites of each ncRNA were annotated with GENCODE annotation (such as intron, exon and UTR) (21), and sites in intergenic regions were annotated with the nearest gene. To display the genome-wide distribution of RNA–DNA interactions, we calculated binding peak numbers for each ncRNA across 1 Mb window and visualized it using BioCircos.js (22).

Global RNA–DNA interactions

For global RNA–DNA interactions, we collected data from techniques including iMARGI, ChAR-seq and GRID-seq, which globally localize chromatin-associated ncRNAs, much more efficiently than the single RNA-targeting methods of ChIRP-seq and CHART-seq. ncRNA molecules were extracted and annotated with disease information from raw datasets of these three technologies. Information of experimental technologies for each interaction was retained. We compared the RNA–DNA binding sites with ncRNA annotations deposited in NONCODE v6 database (23) and allocated NONCODE IDs to these ncRNAs. Annotation and distribution of genome-wide binding sites on

chromosomes were processed as described in the previous section.

RNA–protein interactions from CLIP-seq datasets

To gather RNA–protein interactions, we preprocessed 130 GSE datasets with 308 protein targets from the Gene Expression Omnibus (GEO) (24) database, covering human and mouse interactome data. STAR (25) was used to map raw sequences to reference genomes (hg19 for human, mm9 for mouse), and then piranha software (26) was applied to call peaks interacting with specific RNA-binding proteins. To ensure the quality of calling peaks, *P* value threshold was set to 0.001. RNA–protein binding sites were also compared with ncRNA annotations in NONCODE v6 database and ncRNAs were allocated NONCODE IDs. Evolutionary conservation score of binding site was calculated for each interaction using PhastCons (27).

miRNA targets from Ago-CLIP datasets

We processed Argonaute targeted CLIP-seq (AGO-CLIP) datasets based on the method described in NPInter v4.0 (8). The *P* value threshold was also set to 0.001 in software piranha for good quality of peaks. BEDOPS (28) was used to extract overlaps of miRNA binding sites deposited in miRanda (29) and TargetScan (30) and miRNA targeting peaks derived from Ago-CLIP datasets. RNAs with such overlapping binding sites are predicted to be credible miRNA interaction partners. Moreover, we also assigned standard molecule IDs to these molecules.

Re-annotation, removal of redundancy and integration

We re-annotated biomolecules in NPInter with standard IDs from reference databases. Specifically, lncRNAs, miRNAs, circRNAs, mRNAs and proteins were annotated with NONCODE IDs (23), miRBase IDs (31), circBase IDs (32), Ensembl IDs (33) and UniProt IDs (34) respectively. Additionally, an Ensembl ID and a RefSeq ID (35) for each molecule were provided if available, which facilitates biomolecule search. We integrated tissues, experiments, data sources and references for a large number of interaction pairs collected from different sources and allocated new interaction IDs for these interactions after eliminating repetitive entries.

lncRNA annotation from scRNA-seq data

In NPInter v5.0, we newly integrated lncRNA information with cell type and cell state from cancer scRNA-seq datasets. Based on cancer types in TCGA (<https://portal.gdc.cancer.gov/>), we collected information of over 50 tumor scRNA-seq datasets from the literature and downloaded related data sources from GEO. After screening, 20 scRNA-seq datasets representing 20 cancer types were selected with complete cell type annotation and available expression count matrix. The Seurat (36) package was used to accomplish a standard pre-processing workflow, which included normalization, clustering and non-linear dimensional reduction for raw expressing count matrix. Then, we

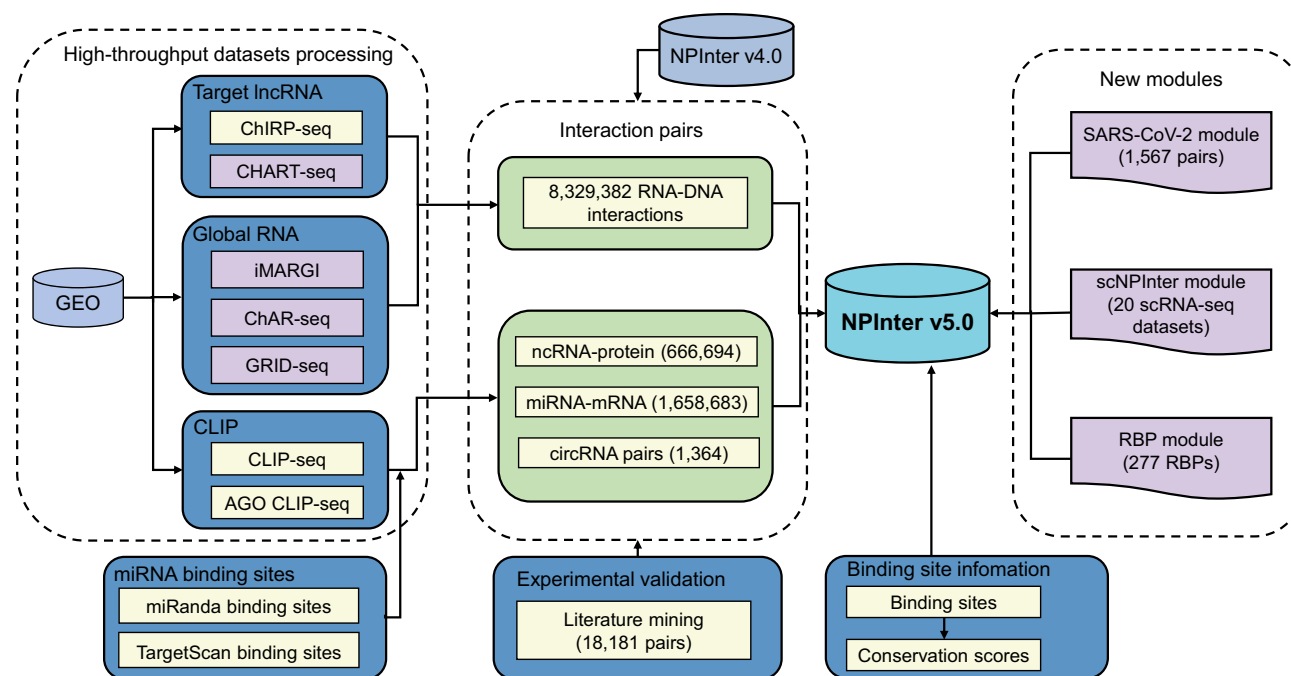


Figure 1. The overall workflow of NPInter v5.0 data integration. ncRNA interactions were obtained by literature mining and high-throughput datasets processing. ChIRP-seq, CLIP-seq and AGO CLIP-seq datasets were processed to obtain RNA–DNA, lncRNA–protein and miRNA–mRNA interactions, respectively. CHART-seq, iMARGI, ChAR-seq and GRID-seq data were processed to obtain RNA–DNA interactions, which are newly added in this version of NPInter. Purple frames (SARS-CoV-2, scNPInter, RBP) indicate new modules integrated into NPInter v5.0.

manually assigned cell type to each cluster according to marker genes extracted from the original paper and identified specifically expressed lncRNAs in each cell type. Interactions related to these cell type-specific lncRNAs were collected and organized for display in the scNPInter module on the website.

DATABASE CONTENT AND STRUCTURE

Interactions and annotated information

The volume of data is greatly expanded in NPInter v5.0. We doubled the number of interactions (without counting RNA–DNA bindings) from 1 100 618 in NPInter v4.0 to 2 596 695 interactions in NPInter v5.0 covering 60 species by literature mining and high-throughput data processing (Table 1). Interactions were attributed to different categories according to the specific types of biomolecules involved (such as ‘lncRNA–protein’, ‘miRNA–mRNA’, etc) (Table 2). In particular, we collected 8,586 experimentally validated interactions via literature mining from 5143 articles published after 2019, a larger number than in NPInter v4.0 (1221 papers). The integration of interactions from global RNA–DNA interactome technologies caused the number of RNA–DNA interaction pairs to surge from 888 915 to 8 329 382 (Table 3).

Apart from basic information such as species, cell line and experimental methods, biomolecules and interactions were provided with more detailed annotation, including interaction level, interaction class, tags of interactions and Phast-Cons sequence conservation score of the binding site. Levels were defined as ‘RNA–protein’, ‘RNA–RNA’, ‘RNA–DNA’ and ‘RNA–TF’ based on the types of interacting

Table 1. Statistics of interactions from different data sources

Data sources	Interactions
Literature mining	18 181
CLIP-seq data	661 938
miRNA–mRNA interactions from Ago CLIP-seq data	1 787 679
Other database ^a	128 897

^aRNA interactions from RISE database.

Table 2. Statistics of different categories of interactions

Data sources	Interactions
lncRNA–protein	662 956
lncRNA–miRNA	155 140
lncRNA–mRNA	17 451
miRNA–protein	3656
miRNA–mRNA	1 658 683
miRNA–circRNA	1267
circRNA–protein	82
circRNA–mRNA	15
Others ^a	97 445

^aRNA interactions of other categories (such as ‘snoRNA–pseudogene’, ‘snRNA–pseudogene’, etc).

Table 3. Statistics of RNA–DNA interactions

Technologies	Interactions
ChIRP-seq data	2 450 577
CHART-seq data	915 101
iMARGI, ChAR-seq and GRID-seq	4 963 704

molecules. According to the interacting classes, interacting pairs were classified as different types (such as ‘binding’, ‘regulatory’, etc). Tags including ‘miRNA target interaction’, ‘ncRNA–protein binding’ and others were applied to briefly describe the mechanism of each interaction with visualization displayed on the website. For each binding site detected in CLIP-seq, we provide an average PhastCons score to present evolutionary conservation. All the detail information of interactions can be viewed on the Interaction Profile page (<http://bigdata.ibp.ac.cn/npinter5/browse/>). For each interacting molecule, such as lncRNA, circRNA and miRNA, we re-annotated with NONCODE IDs, circBase IDs and miRBase IDs respectively. We arranged the type, description, alias, interaction tables and related disease annotation into the Molecule Profile page (e.g. <http://bigdata.ibp.ac.cn/npinter5/molecule/22789/>). Users can easily browse and view these data on our website.

We added ncRNA interactions from several new sources to NPInter v5.0, which greatly expanded our data volume. This is detailed as follows.

Global RNA–DNA interactions. Several studies have shown that ncRNA–chromatin interactions have impact on the transcription of genes. For example, *Drosophila* ncRNAs *roX1* and *roX2* are involved in dosage compensation (37) and the lncRNA, *HOTAIR*, recruits proteins by interacting with chromatin (38). We included ncRNA–DNA interactions derived from ChIRP-seq datasets in the previous version, and we newly integrated 915 101 pairs of RNA–DNA interactions from CHART-seq in NPInter v5.0 because this method is widely applied. However, both ChIRP and CHART technologies are limited to one RNA at a time, and lack a global view of all potential RNA–chromatin interactions (17). Recently, iMARGI, ChAR-seq and GRID-seq technologies have been developed and greatly increased the number of RNA–DNA interactions by their ability to globally localize chromatin-associated ncRNAs. This enables the depiction of a more comprehensive RNA–DNA network. We therefore integrated interactions from these three techniques into NPInter v5.0. A total of 8 329 382 RNA–DNA interaction pairs involving 16 806 ncRNAs (15 in NPInter v4.0) were collected, greatly expanding the data volume.

SARS-CoV-2 related RNA interactions. Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) has caused large-scale outbreaks of coronavirus disease 2019 (COVID-19) all over the world. It is an RNA virus whose success as a pathogen relies on its abilities to repurpose host RBPs and to evade antiviral RBPs (39). Numerous interactions between the virus and host have been described, which may play essential roles in its evasion. For example, host cellular La-related protein1 (LARP1) binds genomic and subgenomic SARS-CoV-2 RNAs to repress SARS-CoV-2 replication in infected human cells (40), and the viral RBP, ORF9b, may contribute to SARS-CoV-2 infection by interacting with virus RNA (41). Also, the SARS-CoV-2 genome engages in different interactions with cellular host RNA (42). Multiple studies have captured thousands of RNA in-

teractions related to SARS-CoV-2 (39–44). However, these data have not been integrated. To provide a comprehensive reference for SARS-CoV-2 research to find antiviral drug targets, we produced a dedicated SARS-CoV-2 RNA interaction collection including five panels for different types of interactions between virus and host. We manually extracted 1567 interactions from studies that employed RNA interactome capture technologies such as cRIC (45), RAP-MS (46) and ChIRP-MS (47). This is the first RNA interaction resource for COVID-19 and we will keep updating the newest findings in this area.

Interactions of lncRNAs with cell type-specific expression from scRNA-seq data. The tumor microenvironment (TME) is a highly heterogeneous and constantly evolving milieu composed of various cell types ranging from normal to malignant cells (48). These cell types perform different functions in the tumor. Cell type-specific gene expression helps to confer cellular functions and ncRNAs such as lncRNAs have distinct cellular functions in the TME (49). For example, pancreatic ductal adenocarcinoma (PDAC) cancer cells showed heterogeneity and were divided into four major subclusters, one of which marked by lncRNA *MEG3* was mostly contributed to metastatic PDAC (50). scRNA-seq was considered as an unprecedented technology to explore cell type-specific expression compared with bulk RNA-seq in which various cell types contribute to an averaged signal. Cell type-specific expression of lncRNAs such as *DLX6-AS1*, *LOC646329* and *H19* has been identified in scRNA-seq datasets (51,52). Here, we preferentially acquired cell type-specific lncRNAs from tumor scRNA-seq data and annotated lncRNAs with expression heterogeneity. After screening, 20 scRNA-seq datasets (including 883 330 cells) representing 20 cancer types were collected and processed to identify specifically expressed lncRNAs in different cell types, and their interaction pairs existing in our database were presented. Cell type distribution of each scRNA-seq dataset was displayed and 1422 lncRNAs from 248 cell types were identified. We extracted marker genes and functional pathways from corresponding studies to depict cell states for each cell type. Information of cell type-specific lncRNAs was provided in detail, and interaction pairs could be accessed in the Molecular Profile page. We believe that these interactions with cell specific molecules will contribute to the study of cancer.

An example of database usage

Users can retrieve various types of interactions for a given molecule in NPInter. Here we take lncRNA *HOTAIR* as an example, which is one of the most extensively studied lncRNAs. Users could search *HOTAIR* at webpages of ‘Home’ and ‘Browse’. (Figure 2A). You can also search it at ‘RBP’ webpage but obtain no results (Figure 2A). Then click the human record on the search result and jump to the molecule page. This page contains molecule basic information, traditional interactions, DNA binding information and disease annotation for a specific molecule, which can all be downloaded (Figure 2B). 93 interaction records of *HOTAIR* were retrieved in the traditional interaction panel (Figure 2B). The interaction partners consisted of miRNAs and pro-

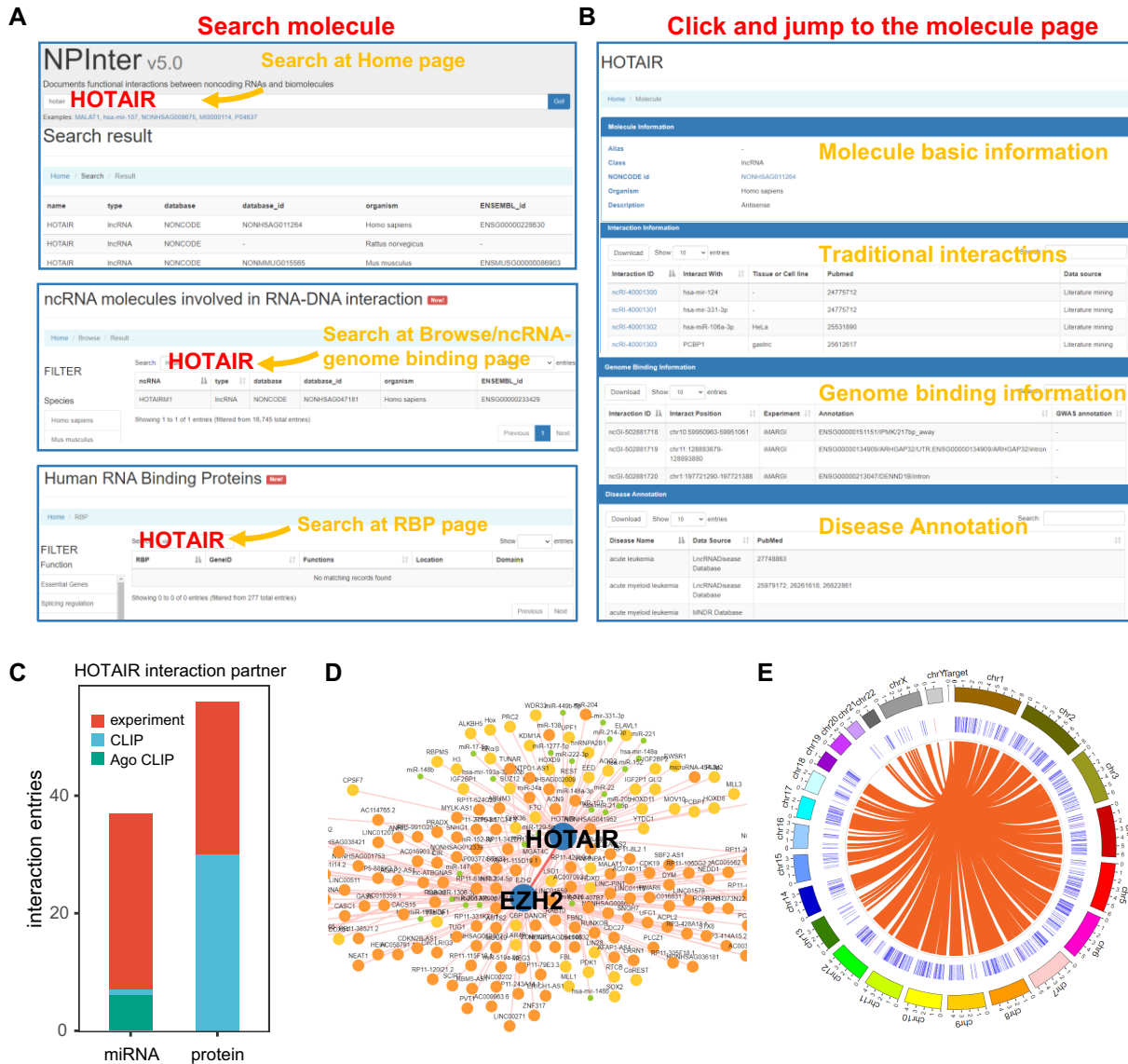


Figure 2. An example in NPInter v5.0. (A) Search lncRNA HOTAIR at Home page (top), Browse/ncRNA-genome bindings page (middle) and RBP page (bottom), respectively. (B) Click and jump to the molecule page of HOTAIR which contains the annotations of molecule basic information, traditional interactions, genome binding information and disease annotation. (C) Number of interaction entries of HOTAIR from different types and data sources. Experiment refers to low-throughput experiments such as dual-luciferase reporter gene assay, RNA pull-down assay, RNA immunoprecipitation assay, etc. (D) The lncRNA HOTAIR and protein EZH2 interaction network. (E) HOTAIR genome binding diagram in the molecule page.

teins, and a large portion of the entries was validated experimentally (Figure 2C). *HOTAIR* has been proved to directly bind polycomb repressive complex 2 (PRC2) to coordinate chromatin occupancy (53). The components of PRC2, including EZH2, SUZ12 and EED, are all collected as *HOTAIR* interaction partners in our database. There is an interaction profile page for every interaction record, such as interaction between *HOTAIR* and EZH2 (ncRI-40001312), in which the interaction network is displayed (Figure 2D). The DNA target sites of *HOTAIR* are listed in the genome binding information panel of the molecule page (Figure 2B). These target sites are visualized in the same page (Figure 2E). In summary, NPInter provides comprehensive annotations of interactions.

Comparison with other databases

Currently, there are several databases for RNA interactions such as starBase (54), LncTarD (55), miRTarBase (56) and RNAInter (57). starBase dedicated to identify RNA-RNA and protein-RNA interactions from high-throughput sequencing data of CLIP-seq. LncTarD documented experimentally-supported functional lncRNA-target regulations in human diseases, and miRTarBase deposited interactions between miRNAs and their target genes from manually curated articles and CLIP-seq data. RNAInter integrated all interaction data from NPInter v4.0. NPInter aims to arrange and provide comprehensive interactions for ncRNAs (such as lncRNAs, miRNAs and circRNAs, etc). Compared with these similar databases,

Table 4. Comparison with other databases

	starBase	LncTarD	miRTarBase	RNAInter	NPInter
Types					
RNA–protein	✓	✓		✓	✓
RNA–RNA	✓	✓	✓	✓	✓
RNA–DNA				✓	✓
COVID-19					✓
Methods					
Literature mining		✓	✓		✓
CLIP-seq	✓			✓	✓
Ago CLIP-seq	✓		✓	✓	✓
ChIRP-seq,				✓	✓
CHART-seq					✓
iMARGI,					✓
ChAR-seq,					
GRID-seq					

NPInter provides more multidimensional interactions for ncRNAs (RNA–protein, RNA–RNA, RNA–DNA) (see Table 4). Also, our database contains both experimental validated RNA interactions by literature mining and data generated by high-throughput sequencing experiments with disease and functional annotation. Moreover, NPInter covers global RNA–DNA interactions and integrated SARS-CoV-2 RNA interactions, which other databases have not documented.

We integrated abundant RNA interactions with experimental evidence from published literature (18 181 interactions from 7311 articles) and applied strict *P* value thresholds in reliable softwares (such as macs2 and piranha) when processing high-throughput data. In summary, NPInter is dedicated to providing multidimensional interactions for each ncRNA while maintaining high data quality.

Service update

We updated existing modules and rearranged our web interface to incorporate the new modules mentioned above for better user interaction and experience. The Browse module was updated to be categorized as traditional interactions and ncRNA-genome bindings. The former contains interactions of RNA–protein, RNA–RNA and RNA–TF derived from the existing Browse module, while the latter deposits chromatin-associated ncRNAs. The SARS-CoV-2 module was built to exhibit SARS-CoV-2 related interactions and annotations of each biomolecule. Also, the scNPInter module was developed for data visualization and display of cell type-specific lncRNAs and their interaction pairs from scRNA-seq datasets. Additionally, we integrated a new module named RBP to provide the interactions of RNA-binding proteins with corresponding annotations.

RBP module. RBPs are proteins that bind to double or single stranded RNA to play essential roles in RNA-mediated gene regulation (58). The regulatory functions of RBPs are critical for normal human physiology because defects in RBP function are associated with diverse genetic and somatic disorders, such as neurodegeneration, autoimmune defects, and cancer (59). Many researchers are interested in the interactions between RBPs and their targets, as can be seen from the user feedback of NPInter v4.0. For NPInter v5.0, we built a new RBP module for browse. To

work towards exhibiting more detailed binding and function of the human RBPs we collected, we annotated 277 RBPs with localization, RNA binding domains, and function from ENCODE project phase III (60). Interactions between a certain RBP and its targets can be easily browsed and downloaded on our website. Users can filter RBPs with different properties. We hope that this module will provide a useful and convenient resource portal for the study of the RNA-RBP network.

CONCLUSION

Unlike protein-protein interactions, which have been extensively investigated, ncRNA interactions have received less attention and remain largely unknown. To well document the ncRNA interactions and make a comprehensive complement to the biological regulatory network, we constructed the NPInter database and have updated it to a fifth version. In NPInter v5.0, we doubled the number of ncRNA interactions primarily by literature mining and high-throughput data processing. The volume of experimentally validated ncRNA interactions has been greatly expanded from the large accumulation of recently published literature. Each interaction pair was annotated with detailed information and conservation scores, and each molecule was annotated with disease information. In addition to ChIRP-seq and newly added CHART-seq which can only target one chromatin-associated ncRNA, we have integrated global RNA–DNA interactions derived from iMARGI, ChAR-seq and GRID-seq technologies. This has greatly increased the number of RNA–DNA interactions and chromatin-associated ncRNAs. To provide reference for the study of COVID-19, we also collected different types of SARS-CoV-2–RNA interactions between virus and host. Specificity of lncRNA expression in different cell types from tumor scRNA-seq data were also integrated in this version to provide a cell-type level view of interactions. Based on user feedback, a new RBP module was presented to provide RBP-based interactions classified according to annotation of localization, binding domains and functions. Compared with other RNA interaction databases, NPInter v5.0 covers global RNA–DNA interactions and integrated SARS-CoV-2–RNA interactions with functional annotations.

The study of RNA interaction has expanded into various areas, such as tumor biology and COVID-19 study. Moreover, new technologies for identifying interactions are constantly being developed. We will continuously update and maintain the NPInter database to document such interactions. There are several techniques that investigate RNA interaction directly in a small number of cells or at single cell level, such as LACE-seq (61). Detecting RNA interactions at the cellular level is becoming more and more feasible, although we can only identify cell type-specific lncRNAs and show related interactions due to current limitation of technology. We will keep updating our scNPInter section when these single cell related techniques are developed and widely applied. Furthermore, NPInter is a member of our systematic platform for noncoding RNAs, which provides databases and webtools for analysis of ncRNAs, including NONCODE (23), piRBase (62) and smProt (63). Consequently, we hope that NPInter will provide useful informa-

tion of RNA interactions and be a valuable web service for the scientific community.

DATA AVAILABILITY

NPInter v5.0 is free to access, browse, search and download at <http://bigdata.ibp.ac.cn/npinter5>.

SUPPLEMENTARY DATA

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

ACKNOWLEDGEMENTS

Data analysis and computing resources were supported by the Center for Big Data Research in Health (<http://bigdata.ibp.ac.cn>), Institute of Biophysics, Chinese Academy of Sciences.

FUNDING

National Key R&D Program of China [2021YFF0703701, 2021YFF0704500]; Informatization Plan of Chinese Academy of Sciences [CAS-WX2021SF-0203]; Strategic Priority Research Program of the Chinese Academy of Sciences [XDB38040300]; National Science Foundation of China [91940306, 31871294, 31970647, 81902519, 32200478]; China Postdoctoral Science Foundation [2022M713311]; Special investigation on science and technology basic resources of the MOST, China [2019FY100102]; National Genomics Data Center, China. Funding for open access charge: National Key R&D Program of China [2021YFF0703701].

Conflict of interest statement. None declared.

REFERENCES

- Djebali,S., Davis,C.A., Merkel,A., Dobin,A., Lassmann,T., Mortazavi,A., Tanzer,A., Lagarde,J., Lin,W., Schlesinger,F. *et al.* (2012) Landscape of transcription in human cells. *Nature*, **489**, 101–108.
- Panni,S., Lovering,R.C., Porras,P. and Orchard,S. (2020) Non-coding RNA regulatory networks. *Biochim. Biophys. Acta, Gene Regul. Mech.*, **1863**, 194417.
- Bartel,D.P. (2009) MicroRNAs: Target recognition and regulatory functions. *Cell*, **136**, 215–233.
- Penny,G.D., Kay,G.F., Sheardown,S.A., Rastan,S. and Brockdorff,N. (1996) Requirement for Xist in X chromosome inactivation. *Nature*, **379**, 131–137.
- Wu,T., Wang,J., Liu,C., Zhang,Y., Shi,B., Zhu,X., Zhang,Z., Skogerboe,G., Chen,L., Lu,H. *et al.* (2006) NPInter: the noncoding RNAs and protein related biomacromolecules interaction database. *Nucleic Acids Res.*, **34**, D150–D152.
- Yuan,J., Wu,W., Xie,C., Zhao,G., Zhao,Y. and Chen,R. (2014) NPInter v2.0: an updated database of ncRNA interactions. *Nucleic Acids Res.*, **42**, D104–D108.
- 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 J. Biol. Databases Curation*, **2016**, baw057.
- 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.
- 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.
- Aw,J.G.A., Shen,Y., Wilm,A., Sun,M., Lim,X.N., Boon,K.-L., Tapsin,S., Chan,Y.-S., Tan,C.-P., Sim,A.Y.L. *et al.* (2016) In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell*, **62**, 603–617.
- 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.
- Licatalosi,D.D., Mele,A., Fak,J.J., Ule,J., Kaykci,M., Chi,S.W., Clark,T.A., Schweitzer,A.C., Blume,J.E., Wang,X. *et al.* (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, **456**, 464–469.
- 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.
- 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.*, **27**, 602–609.
- Wu,W., Yan,Z., Nguyen,T.C., Bouman Chen,Z., Chien,S. and Zhong,S. (2019) Mapping RNA–chromatin interactions by sequencing with iMARGI. *Nat. Protoc.*, **14**, 3243–3272.
- Bell,J.C., Jukam,D., Teran,N.A., Risca,V.I., Smith,O.K., Johnson,W.L., Skotheim,J.M., Greenleaf,W.J. and Straight,A.F. (2018) Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *Elife*, **7**, e27024.
- Li,X., Zhou,B., Chen,L., Gou,L.-T., Li,H. and Fu,X.-D. (2017) GRID-seq reveals the global RNA–chromatin interactome. *Nat. Biotechnol.*, **35**, 940–950.
- Simon,M.D., Wang,C.I., Kharchenko,P.V., West,J.A., Chapman,B.A., Alekseyenko,A.A., Borowsky,M.L., Kuroda,M.I. and Kingston,R.E. (2011) The genomic binding sites of a noncoding RNA. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 20497–20502.
- Li,H. and Durbin,R. (2010) Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinform. Oxf. Engl.*, **26**, 589–595.
- Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of chip-Seq (MACS). *Genome Biol.*, **9**, R137.
- Frankish,A., Diekhans,M., Ferreira,A.-M., Johnson,R., Jungreis,I., Loveland,J., Mudge,J.M., Sisu,C., Wright,J., Armstrong,J. *et al.* (2019) GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.*, **47**, D766–D773.
- Cui,Y., Chen,X., Luo,H., Fan,Z., Luo,J., He,S., Yue,H., Zhang,P. and Chen,R. (2016) BioCircos.js: an interactive circos javascript library for biological data visualization on web applications. *Bioinform. Oxf. Engl.*, **32**, 1740–1742.
- Zhao,L., Wang,J., Li,Y., Song,T., Wu,Y., Fang,S., Bu,D., Li,H., Sun,L., Pei,D. *et al.* (2021) NONCODEV6: an updated database dedicated to long non-coding RNA annotation in both animals and plants. *Nucleic Acids Res.*, **49**, D165–D171.
- 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.
- Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
- Uren,P.J., Bahrami-Samani,E., Burns,S.C., Qiao,M., Karginov,F.V., Hodges,E., Hannon,G.J., Sanford,J.R., Penalva,L.O.F. and Smith,A.D. (2012) Site identification in high-throughput RNA–protein interaction data. *Bioinform. Oxf. Engl.*, **28**, 3013–3020.
- 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.
- Neph,S., Kuehn,M.S., Reynolds,A.P., Haugen,E., Thurman,R.E., Johnson,A.K., Rynes,E., Maurano,M.T., Vierstra,J., Thomas,S. *et al.* (2012) BEDOPS: high-performance genomic feature operations. *Bioinform. Oxf. Engl.*, **28**, 1919–1920.
- 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.
- Agarwal,V., Bell,G.W., Nam,J.-W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, e05005.

31. Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
32. Glažar, P., Papavasileiou, P. and Rajewsky, N. (2014) circBase: a database for circular RNAs. *RNA*, **20**, 1666–1670.
33. 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.
34. UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480–D489.
35. O’Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D. *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733–D745.
36. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M. *et al.* (2021) Integrated analysis of multimodal single-cell data. *Cell*, **184**, 3573–3587.
37. Franke, A. and Baker, B. S. (1999) The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in drosophila. *Mol. Cell*, **4**, 117–122.
38. Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Bruggmann, S. A., Goodnough, H., Helms, J. A., Farnham, P. J., Segal, E. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by non-coding RNAs. *Cell*, **129**, 1311–1323.
39. Lee, S., Lee, Y.-S., Choi, Y., Son, A., Park, Y., Lee, K.-M., Kim, J., Kim, J.-S. and Kim, V. N. (2021) The SARS-CoV-2 RNA interactome. *Mol. Cell*, **81**, 2838–2850.
40. Schmidt, N., Lareau, C. A., Keshishian, H., Ganskih, S., Schneider, C., Hennig, T., Melanson, R., Werner, S., Wei, Y., Zimmer, M. *et al.* (2021) The SARS-CoV-2 RNA–protein interactome in infected human cells. *Nat. Microbiol.*, **6**, 339–353.
41. Kamel, W., Noerenberg, M., Cerikan, B., Chen, H., Järvelin, A. I., Kammoun, M., Lee, J. Y., Shuai, N., Garcia-Moreno, M., Andrejeva, A. *et al.* (2021) Global analysis of protein–RNA interactions in SARS-CoV-2-infected cells reveals key regulators of infection. *Mol. Cell*, **81**, 2851–2867.
42. Ziv, O., Price, J., Shalamova, L., Kamenova, T., Goodfellow, I., Weber, F. and Miska, E. A. (2020) The Short- and Long-Range RNA–RNA interactome of SARS-CoV-2. *Mol. Cell*, **80**, 1067–1077.
43. Flynn, R. A., Belk, J. A., Qi, Y., Yasumoto, Y., Wei, J., Alfajaro, M. M., Shi, Q., Mumbach, M. R., Limaye, A., DeWeirdt, P. C. *et al.* (2021) Discovery and functional interrogation of SARS-CoV-2 RNA–host protein interactions. *Cell*, **184**, 2394–2411.
44. Zhang, S., Huang, W., Ren, L., Ju, X., Gong, M., Rao, J., Sun, L., Li, P., Ding, Q., Wang, J. *et al.* (2022) Comparison of viral RNA–host protein interactomes across pathogenic RNA viruses informs rapid antiviral drug discovery for SARS-CoV-2. *Cell Res.*, **32**, 9–23.
45. Garcia-Moreno, M., Noerenberg, M., Ni, S., Järvelin, A. I., González-Almela, E., Lenz, C. E., Bach-Pages, M., Cox, V., Avolio, R., Davis, T. *et al.* (2019) System-wide profiling of RNA-binding proteins uncovers key regulators of virus infection. *Mol. Cell*, **74**, 196–211.
46. McHugh, C. A. and Guttman, M. (2018) RAP-MS: a method to identify proteins that interact directly with a specific RNA molecule in cells. *Methods Mol. Biol. Clifton NJ*, **1649**, 473–488.
47. Chu, C., Zhang, Q. C., da Rocha, S. T., Flynn, R. A., Bharadwaj, M., Calabrese, J. M., Magnuson, T., Heard, E. and Chang, H. Y. (2015) Systematic discovery of Xist RNA binding proteins. *Cell*, **161**, 404–416.
48. Labani-Motlagh, A., Ashja-Mahdavi, M. and Loskog, A. (2020) The tumor microenvironment: a milieu hindering and obstructing antitumor immune responses. *Front. Immunol.*, **11**, 940.
49. Park, E.-G., Pyo, S.-J., Cui, Y., Yoon, S.-H. and Nam, J.-W. (2022) Tumor immune microenvironment lncRNAs. *Brief. Bioinform.*, **23**, bbab504.
50. Pan, H., Diao, H., Zhong, W., Wang, T., Wen, P. and Wu, C. (2021) A cancer cell cluster marked by lincRNA MEG3 leads pancreatic ductal adenocarcinoma metastasis. *Front. Oncol.*, **11**, 656564.
51. Liu, S. J., Nowakowski, T. J., Pollen, A. A., Lui, J. H., Horlbeck, M. A., Attenello, F. J., He, D., Weissman, J. S., Kriegstein, A. R., Diaz, A. A. *et al.* (2016) Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol.*, **17**, 67.
52. Zhou, J., Xu, J., Zhang, L., Liu, S., Ma, Y., Wen, X., Hao, J., Li, Z., Ni, Y., Li, X. *et al.* (2019) Combined single-cell profiling of lncRNAs and functional screening reveals that H19 is pivotal for embryonic hematopoietic stem cell development. *Cell Stem Cell*, **24**, 285–298.
53. Tsai, M.-C., Manor, O., Wan, Y., Mosammamaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E. and Chang, H. Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, **329**, 689–693.
54. 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.
55. 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.
56. Huang, H.-Y., Lin, Y.-C.-D., Li, J., Huang, K.-Y., Shrestha, S., Hong, H.-C., Tang, Y., Chen, Y.-G., Jin, C.-N., Yu, Y. *et al.* (2020) miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Res.*, **48**, D148–D154.
57. Kang, J., Tang, Q., He, J., Li, L., Yang, N., Yu, S., Wang, M., Zhang, Y., Lin, J., Cui, T. *et al.* (2021) RNAInter v4.0: RNA interactome repository with redefined confidence scoring system and improved accessibility. *Nucleic Acids Res.*, **50**, D326–D332.
58. Kwon, S. C., Yi, H., Eichelbaum, K., Föhr, S., Fischer, B., You, K. T., Castello, A., Krijgsvelde, J., Hentze, M. W. and Kim, V. N. (2013) The RNA-binding protein repertoire of embryonic stem cells. *Nat. Struct. Mol. Biol.*, **20**, 1122–1130.
59. Van Nostrand, E. L., Pratt, G. A., Shishkin, A. A., Gelboin-Burkhart, C., Fang, M. Y., Sundararaman, B., Blue, S. M., Nguyen, T. B., Surka, C., Elkins, K. *et al.* (2016) Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP). *Nat. Methods*, **13**, 508–514.
60. Van Nostrand, E. L., Freese, P., Pratt, G. A., Wang, X., Wei, X., Xiao, R., Blue, S. M., Chen, J.-Y., Cody, N. A. L., Dominguez, D. *et al.* (2020) A large-scale binding and functional map of human RNA-binding proteins. *Nature*, **583**, 711–719.
61. Su, R., Fan, L.-H., Cao, C., Wang, L., Du, Z., Cai, Z., Ouyang, Y.-C., Wang, Y., Zhou, Q., Wu, L. *et al.* (2021) Global profiling of RNA-binding protein target sites by LACE-seq. *Nat. Cell Biol.*, **23**, 664–675.
62. 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.
63. Li, Y., Zhou, H., Chen, X., Zheng, Y., Kang, Q., Hao, D., Zhang, L., Song, T., Luo, H., Hao, Y. *et al.* (2021) SmProt: a reliable repository with comprehensive annotation of small proteins identified from ribosome profiling. *Genomics Proteomics Bioinformatics*, **19**, 602–610.