

NPInter v4.0: an integrated database of ncRNA interactions

Xueyi Teng^{1,2,†}, Xiaomin Chen^{1,2,†}, Hua Xue^{1,2,†}, Yiheng Tang^{1,2}, Peng Zhang^{1,2},
Quan Kang¹, Yajing Hao^{1,2}, Runsheng Chen^{1,2,3,*}, Yi Zhao^{4,*} and Shunmin He^{1,2,*}

¹Key Laboratory of RNA Biology, Center for Big Data Research in Health, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China, ²University of Chinese Academy of Sciences, Beijing 100049, China, ³Guangdong Geneway Decoding Bio-Tech Co. Ltd, Foshan 528316, 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 12, 2019; Revised October 08, 2019; Editorial Decision October 09, 2019; Accepted October 11, 2019

ABSTRACT

Noncoding RNAs (ncRNAs) play crucial regulatory roles in a variety of biological circuits. To document regulatory interactions between ncRNAs and biomolecules, we previously created the NPInter database (<http://bigdata.ibp.ac.cn/npinter>). Since the last version of NPInter was issued, a rapidly growing number of studies have reported novel interactions and accumulated numerous high-throughput interactome data. We have therefore updated NPInter to its fourth edition in which are integrated 600 000 new experimentally identified ncRNA interactions. ncRNA–DNA interactions derived from ChIRP-seq data and circular RNA interactions have been included in the database. Additionally, disease associations were annotated to the interacting molecules. The database website has also been redesigned with a more user-friendly interface and several additional functional modules. Overall, NPInter v4.0 now provides more comprehensive data and services for researchers working on ncRNAs and their interactions with other biomolecules.

INTRODUCTION

RNAs are not just intermediate molecules between DNA and protein. Over the recent decades, large numbers of non-coding RNAs (ncRNAs) have been found that do not encode for proteins and instead play regulatory roles by interacting with biomolecules. For example, the lncRNA Xist mediates X chromosome silencing through an interaction with chromatin DNA (1). MicroRNAs interact with the 3'

UTR region of their target mRNAs and regulate their post-transcriptional repression (2). Studying the ncRNA interactions is thus important for understanding the regulatory network among biomolecules.

With the advancement of high-throughput sequencing technology, a number of new methods have been developed to investigate interactions pertaining to RNAs. Approaches such as CLIP-seq (3), PARIS (4), CLASH (5), ChIRP-seq (6) and GRID-seq (7) have the ability to globally find the interacting partners of specific target ncRNAs. We initially established the NPInter database (8) in 2006 in order to organize and classify such ncRNA interactions and have later upgraded and expanded the database for the past 10 or so years (9,10). In the former version NPInter v3.0 (2016), we already added interactions detected by newly developed methodology, and included functional modules such as binding prediction and network viewing to facilitate its usage. However, since the release of NPInter v3.0, quite a number of articles related to ncRNA interactions have been published, accompanied by a large amount of high-throughput sequencing data. This paper thus describes the update of NPInter to the fourth edition, which includes the integration of newly identified ncRNA interactions and even more user-friendly web services. Specially, we first included circular RNA (circRNA) interactions and ncRNA–DNA interactions detected by ChIRP-seq data. Information on diseases associated with each biomolecule and interaction was collected to improve their function annotation. We also redesigned the entire website to provide a more friendly user interface. In order to assist users in coping with the dramatically increased data size, the new browse module now contains several convenient ways for users to search for target interactions. All interaction data can be freely downloaded from the download page.

*To whom correspondence should be addressed. Tel: +86 10 6488 7032; Fax: +86 10 6488 7842; Email: heshunmin@ibp.ac.cn
Correspondence may also be addressed to Yi Zhao. Tel: +86 10 6260 1010; Fax: +86 10 6260 1356; Email: biozy@ict.ac.cn
Correspondence may also be addressed to Runsheng Chen. Tel: +86 10 6488 8543; Fax: +86 10 6487 1293 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.

DATA COLLECTION AND ANNOTATION

For the NPInter v4.0, we collected interaction data primarily through manual literature mining and processing of high-throughput sequencing data. Data from different sources were subsequently integrated with redundant entries removed. For the convenience of searching, we annotated the involved biomolecules with commonly used molecule IDs. An overview of data integration workflow is shown in Figure 1.

Interactions recorded in literature and databases

To search for experimentally validated ncRNA interactions, we conducted mining of literature published between April 2015 and April 2019 and found 1221 papers. The keywords used in searching the PubMed database are listed in the Supplementary Material. We only included interactions with experimental evidence. Besides, interaction data from the RISE database (11) were integrated into NPInter v4.0.

RNA–protein interactions from CLIP-seq processing

We searched the Gene Expression Omnibus (GEO) (12) for CLIP-seq datasets released in the past 4 years and downloaded raw sequencing data of 69 datasets. We also downloaded 338 ENCODE eCLIP datasets (13). With these CLIP-seq data, we mapped the reads to the reference genome (hg19 for human, mm9 for mouse) using GSNAP (14) and called peaks with the Piranha software (15). We compared the RNA–protein binding sites with ncRNA annotations from the NONCODE v5 database (16) and assigned a NONCODE ID to each site overlapping an ncRNA. The binding score for each ncRNA–protein interaction was then calculated using LncPro (17). We also computed the mean PhastCons sequence conservation score (18) of the binding sites for each ncRNA.

miRNA target extraction from Argonaute CLIP-seq

For Argonaute CLIP-seq datasets, raw data were processed as described earlier. We then used BEDOPS (19) to identify the TargetScan (20) and miRanda (21) miRNA binding sites that overlapped with Argonaute binding peaks. RNAs with predicted target sites that also located to Argonaute CLIP-seq peaks were believed to interact with the miRNAs for which the targets were predicted.

RNA–DNA interactions from ChIRP-seq processing

In NPInter v4.0, we have included RNA–DNA interactions derived from all ChIRP-seq datasets published on the GEO database (12). The HISAT2 software (22) was employed to map ChIRP-seq DNA fragments to the reference genome and peak calling was then done by MACS2 (23). Genome binding sites were annotated with GENCODE annotations (e.g. UTR, intron, and exon) (24), while sites in intergenic regions were annotated with the nearest gene. To visualize the genome-wide distribution of the interactions, we counted the number of interactions inside each 1 Mb window across the genome and visualized it by a heatmap using BioCircos.js (25).

Annotation, redundancy removal and integration

For all biomolecules involved in interactions, we assigned the IDs used by renowned databases. ncRNAs were annotated with NONCODE IDs (16), miRBase IDs (26) or circBase IDs (27), while proteins were annotated with UniProt IDs (28). Ensembl IDs (29), UniGene IDs (30) and RefSeq IDs (31) were also provided where available. Since the data were derived from different sources, we standardized the names of tissues (or cell lines) and experiments. Then, we eliminated redundancies and integrated the repetitive interactions. We combined experiments, tissues, references, sources and binding sites of redundant entries and allocated new interaction IDs for these interactions.

Disease annotation

Disease annotations were collected from several databases, including LncRNADisease (32), MNDR (33), eDGAR (34) and circRNADisease (35). Biomolecules and their interactions were labeled with the relevant disease information. For RNA–DNA interactions, we also downloaded risk GWAS sites from the GWAS Catalog database (36) and marked the corresponding genome regions with this information.

DATABASE CONTENT AND SERVICES

Interactions and associated information

In the NPInter v4.0, we added to the preexisting 491 416 entries in NPInter v3.0, a total of 609 242 new interactions (not including 888 915 ncRNA–genome binding interactions) obtained from different data sources (see Table 1) and including 35 organisms. These interactions cover most kinds of ncRNAs, including lncRNA, miRNA, circRNA, snoRNA, snRNA, etc. (see Table 2). Using these data, we tried to apply some function analysis. By allocating the interactions to cell types, we found many interactions occurring in over 200 cancer cell lines belonging to 50 kinds of cancers. For some of common cancers, more than 100 tissue-specific interactions were found (Figure 2A). We also found numbers of common interactions across different cancer types (Figure 2B). By using our new Function search module to search interactions with multiple gene lists with certain functions, we also acquired numerous putative interactions related to these functions (Figure 2C).

We furnished each interaction entry with annotations of both molecules as well as detailed annotations of the interactions, including the interaction level, the interaction class, tags, organism, tissue or cell type, experimental description, the interaction description, the data source and binding sites. The interaction level is defined according to the molecular types of the interacting molecules (such as ‘RNA–RNA’, ‘RNA–protein’, etc.). For each binding site, we calculated the average PhastCons (18) score across the nucleotides to represent sequence conservation in different organisms. To access the binding potential of ncRNA–protein interactions captured by the CLIP-seq datasets, we applied the LncPro software (17) to calculate a structure-based binding score. On the Interaction Profile page, users can view all the details described earlier as well as the reference information documenting the interactions.

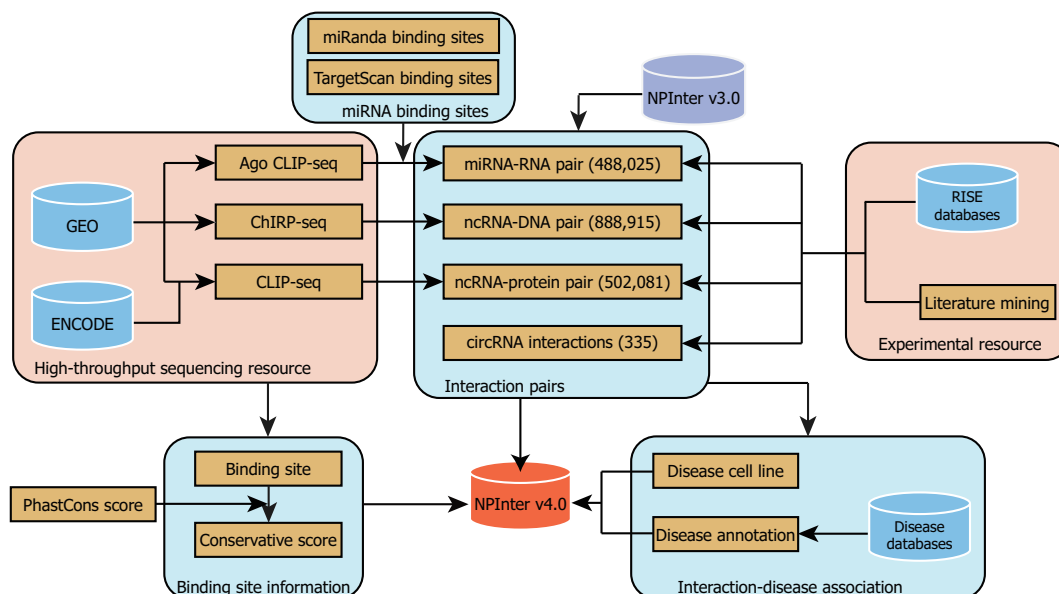


Figure 1. An overview of the NPInter v4.0 data integration.

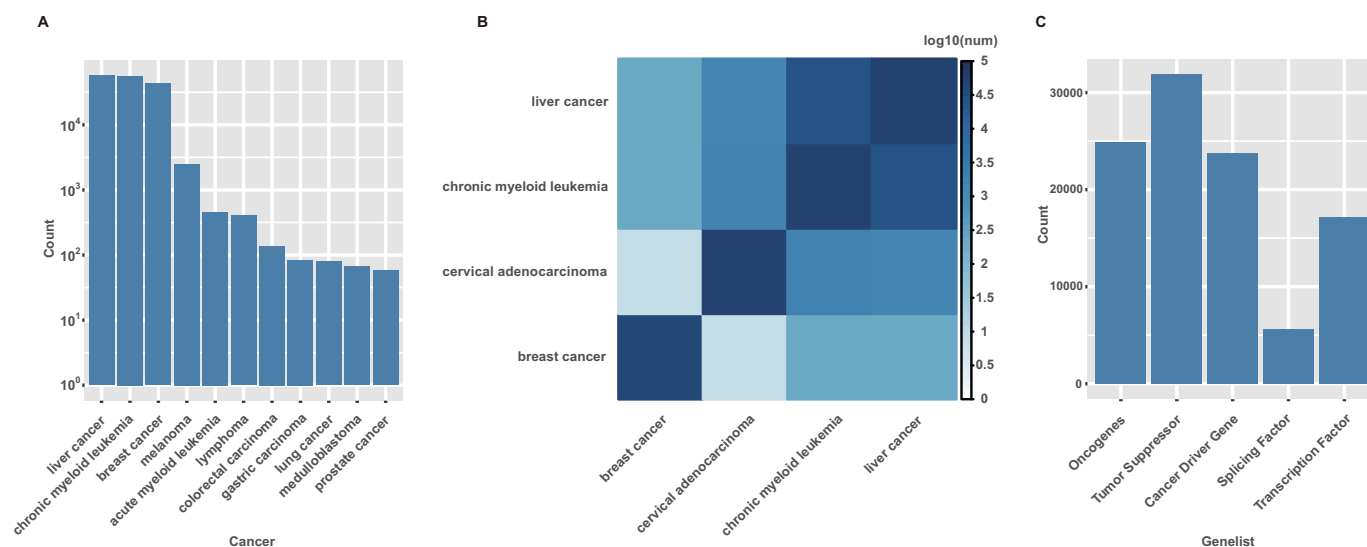


Figure 2. Function analysis using NPInter v4 interactions. (A) Counts of cancer cell line-specific interactions. (B) Heatmap for overlapped interactions among several types of cancers. Interaction number has been calculated by log 10. The values on the diagonal are the numbers of cancer-specific interactions. (C) Counts of interactions searched by gene sets with specific functions.

For each participating molecule, we organized its IDs, molecular type, organism, description, aliases and related diseases into the Molecular Profile page. Users can search for individual molecules and view their details and their interaction partners on this page.

In addition, we integrated some new data sources and annotations, which greatly expanded the coverage of our database. We will describe them in detail in the following paragraphs.

ncRNA–DNA interactions. Previous studies have reported that ncRNAs have the ability to interact with genomic DNA and affect transcription. The best known example is the lncRNA Xist, which functions in dosage compensation by

Table 1. Statistics of interactions from different data sources

Data sources	Interactions
Literature mining	9595
High-throughput data (exclude Ago CLIP-seq and ChIRP-seq)	498 083
Predicted miRNA binding with Ago	464 043
CLIP-seq data	
ChIRP-seq data	888 915
Other database	129 585

interacting with the X chromosome (1). With the accumulation of newly released ChIRP-seq datasets, we included 888 915 RNA–DNA interactions in NPInter v4.0. For each

Table 2. Statistics of different types of ncRNA interactions

Interaction type	Interactions
lncRNA interactions	658 171
miRNA interactions	488 025
snoRNA interactions	61 700
snRNA interactions	12 789
circRNA interactions	335

ChIRP targeted ncRNA, we organized the corresponding genome binding sites into a table on the Molecular Profile page.

circRNA-associated interactions. Recent evidence suggests that circRNAs play critical roles in diverse biological processes via interactions with other biomolecules (37). The human circRNA CDR1as has 74 miR-7 seed matches that allow it to function as an miR-7 sponge (38). Exon–intron circRNAs can interact with snRNAs and enhance the expression of their parental genes (39). To collect such interactions, we searched the literature for circRNA-associated interactions and integrated them into NPInter v4.0.

Disease annotation. Previous research has validated that ncRNAs are associated with a large number of diseases (40,41). In most cases, ncRNAs exert regulatory roles through interactions with molecules involved in diseases. For example, BACE1-AS can form a duplex with BACE1 mRNA and upregulate BACE1 translation in Alzheimer's disease-related cells (42). These interactions highlight the importance of adding known ncRNA–disease associations to our interaction data. To facilitate ncRNA–disease research, we have therefore included disease annotation for molecules involved in NPInter interactions. Users can find them on both the Interaction Profile page and the Molecular Profile page. Besides, we have provided users with an application to search for associated molecules and interactions using disease names.

Service update

We redesigned the entire web interface for the NPInter v4.0. The new NPInter UI, which was constructed using Django and Bootstrap, is even more user-friendly and convenient. We also optimized original NPInter modules such as BLAST and Cytoscape. The BLAST module now provides a detailed Results page with links to related ncRNA molecules. Besides, we have substantially updated the Browse module and have added a Biocircos.js module.

New browse module. Based on users' feedbacks, we have learned that many researchers want to search interactions by tissue, interaction class, interaction level and data source. The old NPInter browse module is not convenient for searches on such interactions. In NPInter v4.0, we paid much attention to improving the Browse module. Users can now filter the interactions using multiple parameters such as tissues, organisms, data sources, interaction classes or interaction levels. They can also select multiple types in one field and search for keywords in the filtered results. We thus believe that the new module will allow users to browse the database in a convenient and time-saving manner.

Biocircos.js module. In the NPInter v4.0, we have added a large amount of ncRNA–DNA interactions derived from high-throughput data. For each ChIRP targeted ncRNA, we can call thousands of peaks across the whole genome. This makes it difficult for users to obtain an overview of all interactions. To achieve such an overview, we applied our previously developed module Biocircos.js (25) to plot interactions across the entire genome. We linked all hotspots interacting with the molecule and depicted the interaction density among the genome.

Function search module. To help user find interactions that function in diseases or important biological processes, we collected some gene lists with certain functions, including oncogenes from ONGene (43), tumor suppressors from TSGene (44), cancer driver genes from 20/20 rule (45) and CGC (46), transcription factors from HOCOMOCO (47) and JASPAR (48), and alternative splicing factors from MisasDB (49). We also processed GDC TCGA cancer expression profiles (50). We filtered differentially expressed lncRNAs with a standard of fold change ≥ 2 and rank sum test P -value ≤ 0.01 . Besides, tissue expression profiles were downloaded from GTEx project (51). We divided the tissue Transcripts Per Million (TPM) by the average TPM of other tissues to calculate a tissue-specific fold change (TSFC). We selected lncRNAs whose TSFC ≥ 8 and TPM ≥ 1 as tissue-specific expressed lncRNAs. We provided searching service with these gene/lncRNA lists in the Function page. We believe the Function search module will facilitate the identification of valuable interactions in various biological processes.

CONCLUSION

Overall, NPInter v4.0 has significantly increased the data size obtained by adding all recently identified ncRNA interactions reported in the literature and submitted to data collections. We have organized the interaction entries along with detailed annotations and prediction scores. Each associated molecule has been annotated with relevant types of IDs and can be searched with nucleotide sequences by the BLAST module. We have further integrated circRNA interactions and ncRNA–DNA interactions captured by ChIRP-seq data. Numerous ncRNA binding regions on the genome presented by newly added Biocircos.js module extended the coverage of ncRNA regulatory network in NPInter. To highlight the links between ncRNA interactions and diseases, we annotated disease association for participating molecules. The new website interface also provides much more convenient services. Compared to other similar databases such as starBase (52) and RAID (53), we have been more focused on providing detailed annotations for interactions, not just for molecules. Visualization modules and predictive scores are also integrated in order to add confidence to the interactions.

In recent years, research on ncRNAs has been a hotspot in the scientific community. Newly invented high-throughput methods will keep on providing large numbers of interactions from various organisms and cell types. We will regularly update and maintain the database. Together with our online ncRNA research platform, which contains

NONCODE (16), CNCI (54) and ncFANs (55), we hope to provide a comprehensive and informative data source on ncRNA interaction network and a series of web services for RNA research spanning from identification to function.

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. We thank Geir Skogerboe, Jianjun Luo and Yiwei Niu for carefully reading of our article.

FUNDING

National Natural Science Foundation of China [31701117, 31970647]; National Key R&D Program of China [2016YFC0901702, 2017YFC0907503]; 13th Five-Year Informatization Plan of Chinese Academy of Sciences [XXH13505-05]. Funding for open access charge: National Key R&D Program of China [2016YFC0901702, 2017YFC0907503].

Conflict of interest statement. None declared.

REFERENCES

- 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.
- Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
- Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.-D. and Gage, F.H. (2009) An RNA code for the FOX2 splicing regulator revealed by mapping RNA–protein interactions in stem cells. *Nat. Struct. Mol. Biol.*, **16**, 130–137.
- 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.
- Helwak, A., Kudla, G., Dudnakova, T. and Tollervey, D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, **153**, 654–665.
- 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.
- 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.
- 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. (2013) 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*, **2016**, baw057.
- Gong, J., Shao, D., Xu, K., Lu, Z., Lu, Z.J., Yang, Y.T. and Zhang, Q.C. (2017) RISE: a database of RNA interactome from sequencing experiments. *Nucleic Acids Res.*, **46**, D194–D201.
- 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.* (2012) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.
- Encode Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
- Wu, T.D. and Nacu, S. (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.
- 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. and Smith, A.D. (2012) Site identification in high-throughput RNA–protein interaction data. *Bioinformatics*, **28**, 3013–3020.
- Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., Zhao, L., Li, X., Teng, X., Sun, X. *et al.* (2017) NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.*, **46**, D308–D314.
- Lu, Q., Ren, S., Lu, M., Zhang, Y., Zhu, D., Zhang, X. and Li, T. (2013) Computational prediction of associations between long non-coding RNAs and proteins. *BMC Genomics*, **14**, 651.
- 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. *Bioinformatics*, **28**, 1919–1920.
- Agarwal, V., Bell, G.W., Nam, J.-W. and Bartel, D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, **4**, e05005.
- 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.
- Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*, **12**, 357–360.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, 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.* (2018) 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. *Bioinformatics*, **32**, 1740–1742.
- Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2018) 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.
- UniProt Consortium (2018) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
- Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M.R., Armean, I., Bennett, R., Bhai, J., Billis, K., Boddu, S. *et al.* (2018) Ensembl 2019. *Nucleic Acids Res.*, **47**, D745–D751.
- NCBI Resource Coordinators (2018) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **46**, D8–D19.
- 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.* (2015) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.*, **44**, D733–D745.
- Bao, Z., Yang, Z., Huang, Z., Zhou, Y., Cui, Q. and Dong, D. (2018) LncRNADisease 2.0: an updated database of long non-coding RNA-associated diseases. *Nucleic Acids Res.*, **47**, D1034–D1037.
- Cui, T., Zhang, L., Huang, Y., Yi, Y., Tan, P., Zhao, Y., Hu, Y., Xu, L., Li, E. and Wang, D. (2017) MNDR v2. 0: an updated resource of ncRNA–disease associations in mammals. *Nucleic Acids Res.*, **46**, D371–D374.
- Babbi, G., Martelli, P.L., Profiti, G., Bovo, S., Savojardo, C. and Casadio, R. (2017) eDGAR: a database of disease–gene associations with annotated relationships among genes. *BMC Genomics*, **18**, 554.
- Zhao, Z., Wang, K., Wu, F., Wang, W., Zhang, K., Hu, H., Liu, Y. and Jiang, T. (2018) circRNA disease: a manually curated database of

- experimentally supported circRNA–disease associations. *Cell Death Dis.*, **9**, 475.
36. Buniello,A., MacArthur,J.A.L., Cerezo,M., Harris,L.W., Hayhurst,J., Malangone,C., McMahon,A., Morales,J., Mountjoy,E., Sollis,E. *et al.* (2018) The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
 37. Qu,S., Yang,X., Li,X., Wang,J., Gao,Y., Shang,R., Sun,W., Dou,K. and Li,H. (2015) Circular RNA: a new star of noncoding RNAs. *Cancer Lett.*, **365**, 141–148.
 38. Memczak,S., Jens,M., Elefsinioti,A., Torti,F., Krueger,J., Rybak,A., Maier,L., Mackowiak,S.D., Gregersen,L.H., Munschauer,M. *et al.* (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495**, 333–338.
 39. Li,Z., Huang,C., Bao,C., Chen,L., Lin,M., Wang,X., Zhong,G., Yu,B., Hu,W., Dai,L. *et al.* (2015) Exon–intron circular RNAs regulate transcription in the nucleus. *Nat. Struct. Mol. Biol.*, **22**, 256–264.
 40. Lalevée,S. and Feil,R. (2015) Long noncoding RNAs in human disease: emerging mechanisms and therapeutic strategies. *Epigenomics*, **7**, 877–879.
 41. Lu,M., Zhang,Q., Deng,M., Miao,J., Guo,Y., Gao,W. and Cui,Q. (2008) An analysis of human microRNA and disease associations. *PLoS One*, **3**, e3420.
 42. Wapinski,O. and Chang,H.Y. (2011) Long noncoding RNAs and human disease. *Trends Cell Biol.*, **21**, 354–361.
 43. Liu,Y., Sun,J. and Zhao,M. (2017) ONGene: a literature-based database for human oncogenes. *J. Genet. Genomics*, **44**, 119–121.
 44. Zhao,M., Kim,P., Mitra,R., Zhao,J. and Zhao,Z. (2015) TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res.*, **44**, D1023–D1031.
 45. Vogelstein,B., Papadopoulos,N., Velculescu,V.E., Zhou,S., Diaz,L.A. and Kinzler,K.W. (2013) Cancer genome landscapes. *Science*, **339**, 1546–1558.
 46. Sondka,Z., Bamford,S., Cole,C.G., Ward,S.A., Dunham,I. and Forbes,S.A. (2018) The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer*, **18**, 696–705.
 47. Kulakovskiy,I.V., Vorontsov,I.E., Yevshin,I.S., Sharipov,R.N., Fedorova,A.D., Rumynskiy,E.I., Medvedeva,Y.A., Magana-Mora,A., Bajic,V.B., Papatsenko,D.A. *et al.* (2017) HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.*, **46**, D252–D259.
 48. Khan,A., Fornes,O., Stigliani,A., Gheorghe,M., Castro-Mondragon,J.A., van der Lee,R., Bessy,A., Cheneby,J., Kulkarni,S.R., Tan,G. *et al.* (2017) JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.*, **46**, D260–D266.
 49. Xing,Y., Zhao,X., Yu,T., Liang,D., Li,J., Wei,G., Liu,G., Cui,X., Zhao,H. and Cai,L. (2016) MiasDB: a database of molecular interactions associated with alternative splicing of human pre-mRNAs. *PLoS One*, **11**, e0155443.
 50. Grossman,R.L., Heath,A.P., Ferretti,V., Varmus,H.E., Lowy,D.R., Kibbe,W.A. and Staudt,L.M. (2016) Toward a shared vision for cancer genomic data. *N. Eng. J. Med.*, **375**, 1109–1112.
 51. Lonsdale,J., Thomas,J., Salvatore,M., Phillips,R., Lo,E., Shad,S., Hasz,R., Walters,G., Garcia,F., Young,N. *et al.* (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
 52. Li,J.-H., Liu,S., Zhou,H., Qu,L.-H. and Yang,J.-H. (2013) 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.
 53. Yi,Y., Zhao,Y., Li,C., Zhang,L., Huang,H., Li,Y., Liu,L., Hou,P., Cui,T., Tan,P. *et al.* (2016) RAID v2.0: an updated resource of RNA-associated interactions across organisms. *Nucleic Acids Res.*, **45**, D115–D118.
 54. Sun,L., Luo,H., Bu,D., Zhao,G., Yu,K., Zhang,C., Liu,Y., Chen,R. and Zhao,Y. (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.*, **41**, e166.
 55. Liao,Q., Xiao,H., Bu,D., Xie,C., Miao,R., Luo,H., Zhao,G., Yu,K., Zhao,H., Skogerbø,G. *et al.* (2011) ncFANs: a web server for functional annotation of long non-coding RNAs. *Nucleic Acids Res.*, **39**, W118–W124.