Nucleic Acids Research, 2024, **52**, D1365–D1369
https://doi.org/10.1093/nar/gkad828
Advance access publication date: 11 October 2023
Database issue

LncRNADisease v3.0: an updated database of long non-coding RNA-associated diseases

Xiao Lin[1,†], Yingyu Lu[1,†], Chenhao Zhang[1], Qinghua Cui[1,2,*], Yi-Da Tang[3,*], Xiangwen Ji[3,*] and Chunmei Cui[1,*]

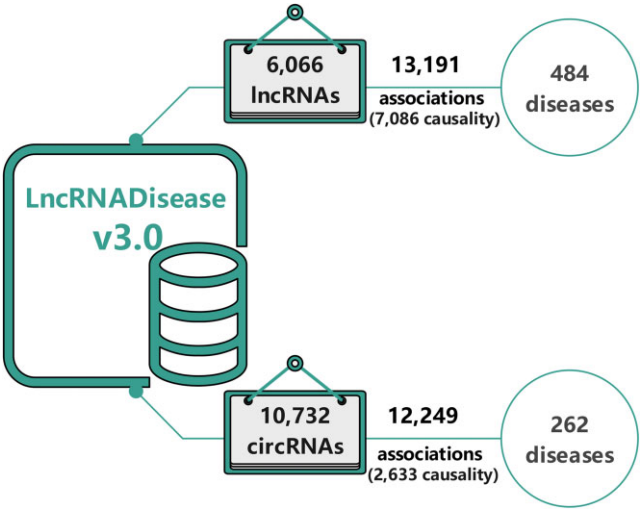
[1]Department of Biomedical Informatics, Center for Noncoding RNA Medicine, State Key Laboratory of Vascular Homeostasis and Remodeling, School of Basic Medical Sciences, Peking University, 38 Xueyuan Rd, Beijing 100191, China
[2]School of Sports Medicine, Wuhan Institute of Physical Education, No.461 Luoyu Rd. Wuchang District, WuHan 430079, Hubei Province, China
[3]Department of Cardiology and Institute of Vascular Medicine, State Key Laboratory of Vascular Homeostasis and Remodeling, Peking University Third Hospital, 49 Huayuanbei Road, Beijing 100191, China

[*]To whom the correspondence should be addressed. Tel: +86 10 82801001; Fax: +86 10 82801001; Email: cuiqinghua@bjmu.edu.cn
Correspondence may also be addressed to Yi-Da Tang. Email: tangyida@bjmu.edu.cn
Correspondence may also be addressed to Xiangwen Ji. Email: jxw01@pku.edu.cn
Correspondence may also be addressed to Chunmei Cui. Email: ccm328@bjmu.edu.cn
[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Abstract

Systematic integration of lncRNA-disease associations is of great importance for further understanding their underlying molecular mechanisms and exploring lncRNA-based biomarkers and therapeutics. The database of long non-coding RNA-associated diseases (LncRNADisease) is designed for the above purpose. Here, an updated version (LncRNADisease v3.0) has curated comprehensive lncRNA (including circRNA) and disease associations from the burgeoning literatures. LncRNADisease v3.0 exhibits an over 2-fold increase in experimentally supported associations, with a total of 25440 entries, compared to the last version. Besides, each lncRNA-disease pair is assigned a confidence score based on experimental evidence. Moreover, all associations between lncRNAs/circRNAs and diseases are classified into general associations and causal associations, representing whether lncRNAs or circRNAs can directly lead to the development or progression of corresponding diseases, with 15721 and 9719 entries, respectively. In a case study, we used the data of LncRNADisease v3.0 to calculate the phenotypic similarity between human and mouse lncRNAs. This database will continue to serve as a valuable resource for potential clinical applications related to lncRNAs and circRNAs. LncRNADisease v3.0 is freely available at http://www.rnanut.net/lncrnadisease.

Graphical abstract



Received: July 27, 2023. **Revised:** September 4, 2023. **Editorial Decision:** September 19, 2023. **Accepted:** September 21, 2023

Introduction

Non-coding RNA was initially regarded as transcriptional noise, which has been overturned with the advancement of multiple technologies. As a class of non-coding RNAs, long non-coding RNAs (lncRNAs) with more than 200nts at length, are involved in diverse biological processes such as chromatin remodeling (1), modulation of DNA and RNA modification by interacting with epigenetic regulatory complex (2), regulation at transcriptional and posttranscriptional level (3). Correspondingly, the dysregulation of lncRNAs is broadly associated with diverse diseases (4), including cancers (5,6), cardiovascular diseases (7), and endocrine system diseases (8), and more. Systematic arranging this knowledge can greatly facilitate the development of lncRNAs as biomarkers or therapeutic targets in diseases. For this purpose, LncRNADisease was established in June 2012 (9), and a subsequent version was released in June 2018 (10), which manually curated >10000 experimentally supported lncRNA and disease associations and provided a user-friendly web server to obtain the annotation of lncRNAs and their related diseases.

The evidence on associations between lncRNAs and diseases is still rapidly increasing benefiting from the advancements of RNA sequencing technologies. For instance, a myriad of lncRNAs including MALAT1, NEAT1 and DANCR engaging in immune responses or inflammatory processes during COVID-19 infection have been reported (11–13). Notably, there has been an exponential growth in the number of studies investigating circular RNAs (circRNAs), characterized by the closed loop structure, and their associations with diseases in recent years (14,15). Furthermore, myriad researches focusing on the causal associations of ncRNA and disease are available, which holds the potential to accelerate our understanding of the role of lncRNAs in pathogenesis (16,17). Therefore, it becomes necessary to integrate new lncRNA and disease associations to progress lncRNA/circRNA-based therapeutics. Here, we introduce LncRNADisease v3.0 with several improvements compared to the previous version. First, LncRNADisease v3.0 has collected 25440 experimentally supported lncRNA/circRNA-disease association entries covering 6066 lncRNAs, 10732 circRNAs, and 566 diseases, which increases >2-fold entries compared to the last version. Then, all associations have been summarized into two classes, i.e. general associations and causal associations. In addition, the associations with potential clinical applications are provided. LncRNADisease v3.0 can be freely available at http://www.rnanut.net/lncrnadisease.

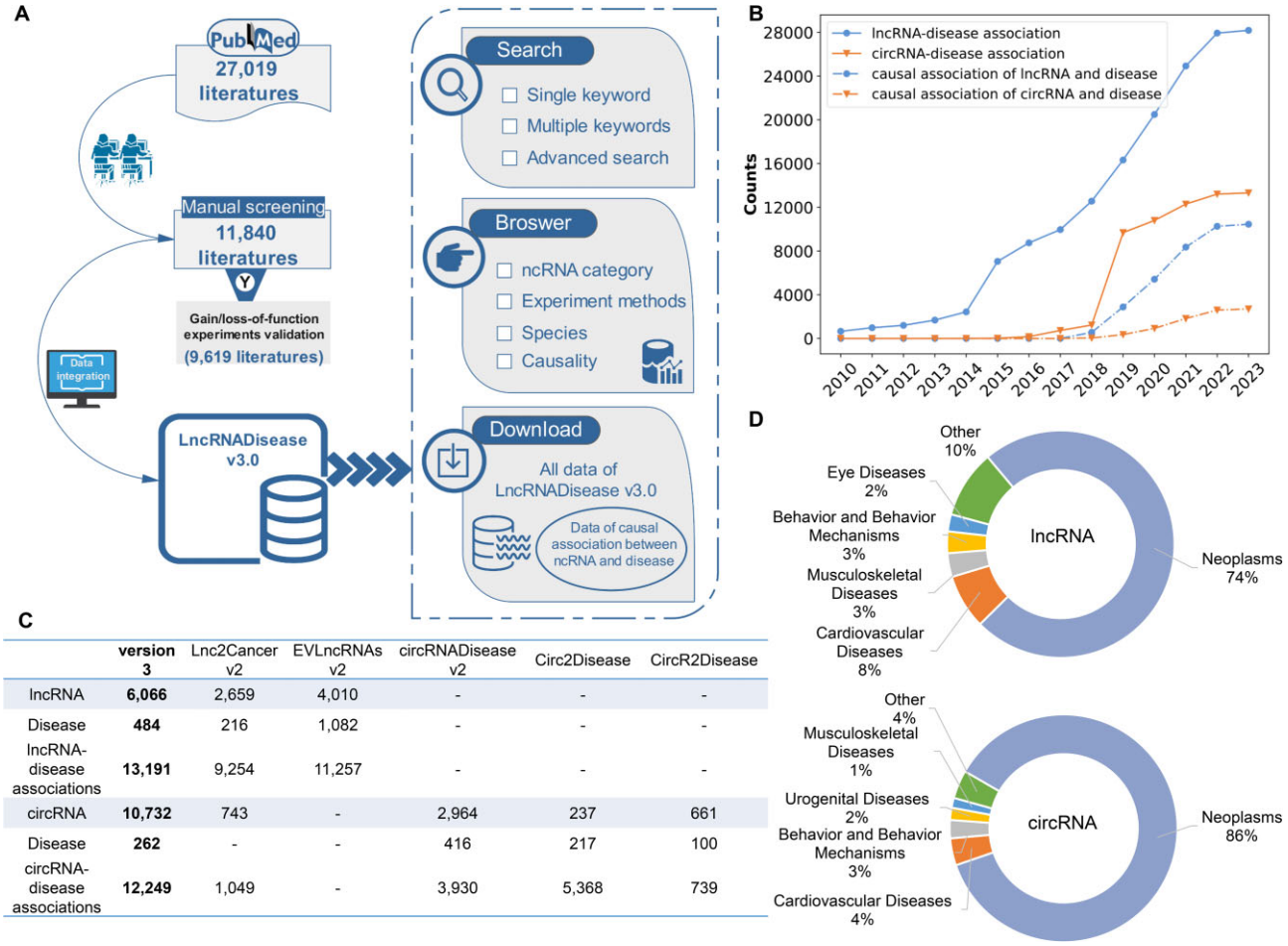
Data collection and overview

The workflow of LncRNADisease v3.0 is shown in Figure 1A. To incorporate new lncRNA and disease associations, we firstly employed the keywords with the combination of 'disease or cancer' and 'lncRNA or lincRNA or long non-coding RNA', as well as 'circRNA or circular RNA', to retrieve relevant literatures in PubMed ranging from June 2018 to January 2023. As a result, 27019 publications met the inclusion criteria. Then we manually extracted the experimentally supported associations between lncRNAs/circRNAs and diseases, in which either strong (e.g. qRT-PCR and ChIP) or weak experimental evidence (high-throughput sequencing) are both included, and detailed experimental methods are also recorded for calculating the confidence score of associations. The process of calculating the confidence score is similar to LncRNADisease v2.0, with weights of 0.75 and 1.0 for weak and strong evidence, respectively. And 11840 papers contain the lncRNA/circRNA and disease associations. Meanwhile, given that causal associations can offer more powerful clues for revealing the roles of lncRNAs/circRNAs in diseases, all associations were further annotated to determine whether they represent causal relationships. This mainly relies on whether gain/loss-of-function experiments were conducted (17). Besides, the molecules or biological pathways that lncRNAs/circRNAs interact with or modulate and potential clinical applications are also integrated into LncRNADisease v3.0. We annotated the lncRNA names with standard symbols from NONCODE (18), RefSeq (19) and Ensembl (20). The disease names were also standardized using the Disease Ontology (DO, http://www.disease-ontology.org) (21) and the Medical Subject Headings (MeSH, https://www.ncbi.nlm.nih.gov/mesh).

As a result, LncRNADisease v3.0 integrates 25440 experimentally supported association entries covering 6066 lncRNAs, 10732 circRNAs and 566 diseases, which has been expanded by more than two times in data size compared to the last version. For lncRNA-disease associations, among 13191 entries, 280, 109 and 18 are from mice, rats, and rabbits comprising 154, 86 and 18 lncRNAs, respectively. Similarly, there are 155 and 27 circRNA-disease associations for mice and rats, respectively, covering 150 and 27 circRNAs. There are 7086 causal association entries out of 13191 lncRNA-disease entries and 2633 causal association entries out of 12249 circRNA-disease entries. The distribution of lncRNA-disease association entries per year is depicted in Figure 1B. We observed a rapid elevation in the number of both ncRNA-disease entries and the causal associations, particularly within the last five years. This accumulation of causal associations serves as a valuable resource for uncovering the pathogenesis or developing new ncRNA-based therapies. In addition, there is a remarkably 10-fold increase in the number of circRNAs and disease associations compared to the LncRNADisease v2.0, reflecting more efforts have been made to explore the functions of circRNAs in diseases in recent years. As depicted in Figure 1C, the circRNA and disease associations in LncRNADisease v3.0 represent a substantial supplement to the other similar databases such as circRNADisease (22), Circ2Disease (23) and CircR2Disease (24). Likewise, the number of lncRNA-disease association in LncRNADisease v3.0 also demonstrate a subtle advantage. We categorized diseases based on the disease hierarchy of the MeSH database and observed that cancers have the most association entries for both lncRNAs and circRNAs. Additionally, cardiovascular diseases account for 8% of lncRNA-disease associations and 4% of circRNA-disease entries (Figure 1D).

Database construction and utility

The LncRNADisease v3.0 website is freely available at http://www.rnanut.net/lncrnadisease, which is built with the PHP framework and utilizes MySQL as the database management system to store all association data. The user interface is implemented using HTML5 and JavaScript to enable interactive functionality. Users can easily assess all the data by navigating to the 'Browse', 'Search' or 'Download' page. On the 'Browse' page, the association entries are grouped based on multiple classified criteria, including the type of ncRNAs



C

	version 3	Lnc2Cancer v2	EVLncRNAs v2	circRNADisease v2	Circ2Disease	CircR2Disease
lncRNA	**6,066**	2,659	4,010	-	-	-
Disease	**484**	216	1,082	-	-	-
lncRNA-disease associations	**13,191**	9,254	11,257	-	-	-
circRNA	**10,732**	743	-	2,964	237	661
Disease	**262**	-	-	416	217	100
circRNA-disease associations	**12,249**	1,049	-	3,930	5,368	739

Figure 1. Overview of the LncRNADisease v3.0 database. (**A**) The schematic framework of LncRNADisease v3.0. (**B**) Number of lncRNA/circRNA and disease association entries in each year, as well as the number of causal associations. (**C**) The comparison between LncRNADisease v3.0, and other similar databases. (**D**) The distribution of disease categories among lncRNA-disease entries and circRNA-disease entries, where the disease category was determined by the disease hierarchy of MeSH database.

(lncRNA or circRNA), experimental methods validating the association (qPT-PCR, northern blot, microarray and RNA-seq), species (human, mouse, rat, and rabbit), and causality, where users can select one criterion to obtain their interested ncRNA and disease association entries. After submission, all relevant associations are shown, which can be sorted by the confidence score, then the details of corresponding lncRNA, disease, publication, regulation pattern, and clinical application can be obtained by clicking the 'detail' link. To be specific, for the detail of each association, lncRNA/circRNA has been linked to Ensembl and NONCODE, and all associated diseases are provided in a graphical format. Similarly, users can find the disease details linking to DO and MeSH, and it's associated ncRNAs. As for the 'Search' page, users can directly input lncRNAs, circRNAs, or disease names to retrieve association entries. Here, two searching modes, single keyword and semicolon-separated multiple keywords, are both allowed. With an advanced search, the types of experimental methods, species, and the confidence score all become the alternatives according to the users' preferences. And the search result is downloadable. At the same time, users can obtain all data of LncRNADisease v3.0 through the 'Download' page. The detailed guidance of LncRNADisease v3.0 usage is described on the 'Help' page for non-specialist users.

Example of data utility: analysis of phenotypic similarity of orthologous lncRNAs between human and mouse

It is well-established that lncRNAs are less conserved than protein-coding RNAs (25,26), which causes the divergence of phenotypes or functions linked to lncRNAs across species and further poses challenges in the application of animal models to explore the mechanism of lncRNAs. LncRNADisease v3.0 not only collects diseases associated with human lncRNAs but includes those lncRNAs from mouse, rat, and rabbit. To investigate the phenotypic similarity of orthologous lncRNAs, we picked out the orthologs between human and mouse based on the homologous information provided by Ensembl and calculated the Jaccard index using their all associated diseases. As a result, 32 human and mouse orthologous lncRNAs were found, and their phenotypic similarity scores were shown in Table 1. It is observed that phenotypic similarity scores of human and mouse orthologous lncRNAs are relatively small, indicating the poor conservation of lncRNAs across multiple species. NEAT1 and Neat1, which are implicated in organizing nuclear structure and display relatively conserved functions (27), are with most shared diseases, such as Alzheimer's disease, Parkinson's disease, fatty liver disease, and breast cancer. As well as GAS5/Gas5, they are conserved in genomic

Table 1. Phenotypic similarity between human and mouse lncRNAs in LncRNADisease v3.0

Human lncRNA	No.	Mouse lncRNA	No.	No. (common diseases)	Jaccard index
POSTN	1	Postn	1	1	1.0000
NEAT1	92	Neat1	16	8	0.0800
HOTAIR	54	Hotair	8	4	0.0690
SNHG20	17	Snhg20	2	1	0.0556
SNHG14	19	Snhg14	3	1	0.0476
MALAT1	97	Malat1	14	5	0.0472
SNHG1	41	Snhg1	4	2	0.0465
GAS5	64	Gas5	10	3	0.0423
SNHG12	25	Snhg12	1	1	0.0400
CRNDE	29	Crnde	3	1	0.0323
MIAT	33	Miat	5	1	0.0270
ZFAS1	42	Zfas1	4	1	0.0222
XIST	48	Xist	3	1	0.0200
MEG3	57	Meg3	6	1	0.0161
AIRN	2	Airn	1	1	0
FIRRE	4	Firre	1	0	0
SNHG15	20	Snhg15	2	0	0
PEG10	2	Peg10	1	0	0
RPPH1	5	Rpph1	2	0	0
NORAD	25	Norad	1	0	0
HIPK3	6	Hipk3	2	0	0
NRON	5	Nron	1	0	0
RMRP	17	Rmrp	2	0	0
PVT1	64	Pvt1	1	0	0
HOTAIRM1	20	Hotairm1	1	0	0
MT-CO2	1	Ptgs2	2	0	0
MIR22HG	9	Mir22hg	1	0	0
SNHG3	22	Snhg3	2	0	0
TUG1	53	Tug1	9	0	0
KCNQ1OT1	32	Kcnq1ot1	3	0	0
SNHG6	28	Snhg6	1	0	0
FTX	12	Ftx	1	0	0

No., number of diseases associated with human or mouse lncRNAs.

sequence and RNA secondary structure (25), and are both associated with myocardial infarction, coronary artery disease, and idiopathic thrombocytopenic purpura. The function or phenotypic differences of lncRNAs across species are not fully determined by their sequence identity (28). We noticed that TUG1 is involved in 53 diseases, and its ortholog Tug1 is associated with 9 diseases (Table 1). However, there was no overlap in the diseases associated with TUG1/Tug1, although they have a high level of exonic nucleotide conservation, reaching 77% (29). The analysis of phenotypic similarity with more comprehensive lncRNA and disease association data would contribute to describing the selection process of lncRNAs to some degree.

Conclusion

Considering the continuous growth of lncRNA-related studies, especially the significant increase in circRNAs over the past 5 years, and the giant potential of lncRNA in serving as biomarkers or therapeutic targets for diseases, it is of great importance to update LncRNADisease with the latest research findings. LncRNADisease v3.0 integrated more comprehensive experimentally supported lncRNA/circRNA-disease associations in multiple species and scored each pair of lncRNA/circRNA and disease based on experimental methods. LncRNADisease v3.0 also provided the disease causative lncRNA/circRNA information, helping quickly and precisely identify critical ncRNAs for diseases. With a case study, the data of LncRNADisease v3.0 are able to characterize the phenotypic similarity of lncRNAs across species. The experimentally supported lncRNA and disease association we updated represents a valuable resource for future researches, such as uncovering other unknown lncRNA–disease associations with diverse computational algorithms, and screening potential lncRNA targets for diseases. We will continue to update the LncRNADisease database with high-quality associations and incorporate convenient modules to facilitate the studies on the regulatory mechanism of lncRNA in diseases.

Data availability

LncRNADisease v3.0 is a database with online and open access, available at http://www.rnanut.net/lncrnadisease.

Funding

National Natural Science Foundation of China [62025102, 81921001]. Scientific and Technological Research Project of Xinjiang Production and Construction Corps (2021AB028). Funding for open access charge: Scientific and Technological Research Project of Xinjiang Production and Construction Corps (2021AB028).

Conflict of interest statement

None declared.

References

1. Han,P., Li,W., Lin,C.-H., Yang,J., Shang,C., Nurnberg,S.T., Jin,K.K., Xu,W., Lin,C.-Y., Lin,C.-J., *et al.* (2014) A long noncoding RNA protects the heart from pathological hypertrophy. *Nature*, **514**, 102–106.
2. Lee,J.T. (2012) Epigenetic regulation by long noncoding RNAs. *Science*, **338**, 1435–1439.
3. Kopp,F. and Mendell,J.T. (2018) Functional classification and experimental dissection of long noncoding RNAs. *Cell*, **172**, 393–407.
4. Zhang,X.L., Hong,R.Y., Chen,W.Q., Xu,M.W. and Wang,L.F. (2019) The role of long noncoding RNA in major human disease. *Bioorg. Chem.*, **92**, 103214.
5. Huarte,M. (2015) The emerging role of lncRNAs in cancer. *Nat. Med.*, **21**, 1253–1261.
6. Nandwani,A., Rathore,S. and Datta,M. (2021) LncRNAs in cancer: regulatory and therapeutic implications. *Cancer Lett.*, **501**, 162–171.
7. Li,Z.N., Gao,J.L., Sun,D., Jiao,Q., Ma,J., Cui,W.L., Lou,Y.Q., Xu,F., Li,S.S. and Li,H.X. (2022) LncRNA MEG3: potential stock for precision treatment of cardiovascular diseases. *Front. Pharmacol.*, **13**, 1045501.
8. Bi,Y., Wang,Y. and Sun,X.L. (2022) Recent Advances of LncRNA H19 in Diabetes LncRNA H19 in Diabetes. *Horm. Metab. Res.*, **54**, 212–219.
9. Chen,G., Wang,Z., Wang,D., Qiu,C., Liu,M., Chen,X., Zhang,Q., Yan,G. and Cui,Q. (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.*, **41**, D983–D986.
10. 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.
11. Shaath,H., Vishnubalaji,R., Elkord,E. and Alajez,N.M. (2020) Single-cell transcriptome analysis highlights a role for neutrophils

and inflammatory macrophages in the pathogenesis of severe COVID-19. *Cells*, **9**, 2374.
12. Meydan,C., Madrer,N. and Soreq,H. (2020) The neat dance of COVID-19: NEAT1, DANCR, and co-modulated cholinergic RNAs link to inflammation. *Front. Immunol.*, **11**, 590870.
13. Moazzam-Jazi,M., Lanjanian,H., Maleknia,S., Hedayati,M. and Daneshpour,M.S. (2021) Interplay between SARS-CoV-2 and human long non-coding RNAs. *J. Cell. Mol. Med.*, **25**, 5823–5827.
14. Liu,C.-X. and Chen,L.-L. (2022) Circular RNAs: characterization, cellular roles, and applications. *Cell*, **185**, 2016–2034.
15. van Zonneveld,A.J., Kölling,M., Bijkerk,R. and Lorenzen,J.M. (2021) Circular RNAs in kidney disease and cancer. *Nat. Rev. Nephrol.*, **17**, 814–826.
16. Delás,M.J. and Hannon,G.J. (2017) lncRNAs in development and disease: from functions to mechanisms. *Open Biol.*, **7**, 170121.
17. Jia,K., Gao,Y., Shi,J., Zhou,Y., Zhou,Y. and Cui,Q. (2020) Annotation and curation of the causality information in LncRNADisease. *Database*, **2020**, baz150.
18. Fang,S., Zhang,L., Guo,J., Niu,Y., Wu,Y., Li,H., Zhao,L., Li,X., Teng,X., Sun,X., *et al.* (2018) NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Res.*, **46**, D308–D314.
19. O'Leary,N.A., Wright,M.W., Brister,J.R., Ciufo,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.
20. Cunningham,F., Allen,J.E., Allen,J., Alvarez-Jarreta,J., Amode,M.R., Armean,I.M., Austine-Orimoloye,O., Azov,A.G., Barnes,I., Bennett,R., *et al.* (2022) Ensembl 2022. *Nucleic Acids Res.*, **50**, D988–D995.
21. Schriml,L.M., Munro,J.B., Schor,M., Olley,D., McCracken,C., Felix,V., Baron,J.A., Jackson,R., Bello,S.M., Bearer,C., *et al.* (2022) The Human Disease Ontology 2022 update. *Nucleic Acids Res.*, **50**, D1255–D1261.
22. 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.
23. Yao,D., Zhang,L., Zheng,M., Sun,X., Lu,Y. and Liu,P. (2018) Circ2Disease: a manually curated database of experimentally validated circRNAs in human disease. *Sci. Rep.*, **8**, 11018.
24. Fan,C., Lei,X., Fang,Z., Jiang,Q. and Wu,F.-X. (2018) CircR2Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases. *Database*, **2018**, bay044.
25. Johnsson,P., Lipovich,L., Grandér,D. and Morris,K.V. (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim. Biophys. Acta (BBA) Gen. Subj.*, **1840**, 1063–1071.
26. Hezroni,H., Koppstein,D., Schwartz,M.G., Avrutin,A., Bartel,D.P. and Ulitsky,I. (2015) Principles of long noncoding RNA evolution derived from direct comparison of transcriptomes in 17 species. *Cell Rep.*, **11**, 1110–1122.
27. Lin,Y., Schmidt,B.F., Bruchez,M.P. and McManus,C.J. (2018) Structural analyses of NEAT1 lncRNAs suggest long-range RNA interactions that may contribute to paraspeckle architecture. *Nucleic Acids Res.*, **46**, 3742–3752.
28. Diederichs,S. (2014) The four dimensions of noncoding RNA conservation. *Trends Genet.*, **30**, 121–123.
29. Jordan,P.L., Gabrijela,D., Audrey,R.W., Taeyoung,H., Emily,J.-P., Nydia,C., Christian,M., Kyle,T., Christopher,K., Nimrod,D.R., *et al.* (2020) The *Tug1* locus is essential for male fertility. *Genome Biol*, **21**, 237.

Received: July 27, 2023. **Revised:** September 4, 2023. **Editorial Decision:** September 19, 2023. **Accepted:** September 21, 2023