EVLncRNAs: a manually curated database for long non-coding RNAs validated by low-throughput experiments

Bailing Zhou[1,2,†], **Huiying Zhao**[3,†], **Jiafeng Yu**[1,2], **Chengang Guo**[1,2], **Xianghua Dou**[1,2], **Feng Song**[1,2], **Guodong Hu**[1,2], **Zanxia Cao**[1,2], **Yuanxu Qu**[4], **Yuedong Yang**[5], **Yaoqi Zhou**[1,5,*] **and Jihua Wang**[1,2,*]

[1]Shandong Provincial Key Laboratory of Biophysics, Institute of Biophysics, Dezhou University, Dezhou 253023, China, [2]College of Physics and Electronic Information, Dezhou University, Dezhou 253023, China, [3]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD 4059, Australia, [4]Department of Surgery, Beijing Tiantan Hospital, Capital Medical University, Beijing 100069, China and [5]Institute for Glycomics and School of Information and Communication Technology, Griffith University, Gold Coast, QLD 4222, Australia

ABSTRACT

Long non-coding RNAs (lncRNAs) play important functional roles in various biological processes. Early databases were utilized to deposit all lncRNA candidates produced by high-throughput experimental and/or computational techniques to facilitate classification, assessment and validation. As more lncRNAs are validated by low-throughput experiments, several databases were established for experimentally validated lncRNAs. However, these databases are small in scale (with a few hundreds of lncRNAs only) and specific in their focuses (plants, diseases or interactions). Thus, it is highly desirable to have a comprehensive dataset for experimentally validated lncRNAs as a central repository for all of their structures, functions and phenotypes. Here, we established EVLncRNAs by curating lncRNAs validated by low-throughput experiments (up to 1 May 2016) and integrating specific databases (lncRNAdb, LncRANDisease, Lnc2Cancer and PLNlncRBase) with additional functional and disease-specific information not covered previously. The current version of EVLncRNAs contains 1543 lncRNAs from 77 species that is 2.9 times larger than the current largest database for experimentally validated lncRNAs. Seventy-four percent lncRNA entries are partially or completely new, comparing to all existing experimentally validated databases. The established database allows users to browse, search and download as well as to submit experimentally validated lncRNAs. The database is available at http://biophy.dzu.edu.cn/EVLncRNAs.

INTRODUCTION

The ENCODE project found that three-quarters of the human genome is capable of being transcribed (1) but <3% of which code for proteins (2). The vast majority of the genome transcribe into non-coding RNAs (lncRNAs), many of which are shown to have important biological functions (3). Small non-coding RNAs (<200 nt), such as microRNAs, snoRNAs, siRNAs and piRNAs, have been extensively studied (4,5). The discoveries of two lncRNAs, H19 and Xist (6,7), inspired the search for additional lncRNAs and determination of their biological functions (8). LncRNAs have been shown to be involved in regulating gene expression and other cellular processes and implicated in disease etiology (8–10) including cancers (9,11–13). As more and more lncRNAs are being discovered, a number of databases (14–16) have been established for different purposes.

Early databases were utilized to deposit all lncRNA candidates from high-throughput techniques to facilitate classification, assessment and validation. Examples include LNCipedia for human lncRNA (17), PNRD (Plant noncoding RNA database) (18) and PlantNATsDB for plants (19), and NONCODE for 16 different species (20). Some databases such as NPInter 3.0 (21) contain both manually curated as well as high-throughput experiments. Recognizing potential false positives in high-throughput experiments, five recent databases were built to contain lncRNA

validated by low-throughput experiments only. They are lncRNAdb, a manually curated database with 287 functional lncRNAs from eukaryotic organisms (22), LncRNADisease with 1102 entries of experimentally validated lncRNA–disease associations (23), Lnc2Cancer with manually curated 1057 associations between 531 lncRNAs and 86 human cancers (24), lncRInter with 922 lncRNA interaction pairs between 276 lncRNAs and 597 partners in 15 organisms (25), and PLNlncRbase with manually collected 420 plant lncRNAs in 43 plant species after excluding those from high-throughput experiments (26). All of these datasets were well designed for their respective specific purposes [eukaryotic only without legible disease associations (lncRNAdb), disease-focused (Lnc2Cancer and LncRNADisease), interaction-focused (lncRInter) and plant-focused (PLNlncRbase)]. Thus, it is highly desirable to have a central repository for all lncRNAs validated by low-throughput experiments along with their sequence, structure, function and related disease information.

Here, we manually curated all experimentally validated lncRNAs (EVlncRNAs) published prior to 1 May 2016. We compared these lncRNAs with those collected in several specific databases (lncRNAdb, Lnc2Cancer, LncRNADisease and PLNlncRbase) and curated additional functional and disease-specific information not covered previously. The final dataset contains 1543 lncRNAs from 77 species, 2324 associations between 886 lncRNAs and 338 diseases, 793 entries of functional descriptions of 664 lncRNAs, and 1163 interactions between 445 lncRNAs and their interacting partners. More than 70% entries are partially or completely new, comparing to existing specific databases of experimentally validated lncRNAs.

DATABASE CONSTRUCTION

For this database, we only kept lncRNAs confirmed by low-throughput experiments such as quantitative reverse transcription-polymerase chain reaction (qRT-PCR), knock down, western blot, northern blot and luciferase reporter assays. Western blotting validates the interaction between lncRNAs and proteins often after knocking down (27,28). qRT-PCR and northern blot confirm the expression level of lncRNAs (29,30). Luciferase reporter assays examine the promoter activity of lncRNA or their interacting partners and confirm the interaction between lncRNAs and their partners (31,32). We excluded lncRNAs derived from RNA-seq only because these RNAs were transcribed but not necessarily functional. Moreover, RNA-seq data suffers from potential problems of partial transcripts, sequencing biases and low sequencing depth for lowly expressed transcripts, in particular (33). Furthermore, it is still challenging to separate peptide-coding from non-coding RNAs. Given the uncertainty, we have excluded lncRNAs if RNA-seq data is the only evidence. For similar reasons, lncRNAs with only microarray evidence are also excluded.

All EVLncRNAs entries were curated as follows. We first searched the PubMed with keywords matching 'lncRNA', 'long non coding RNA', 'long ncRNA' or 'long non-code RNA' along with their plural forms. We then separated those articles covered in previous databases (lncRNAdb, LncRNADisease, Lnc2Cancer and PLNlncRbase) from those completely new ones, which led to 1267 new publications. From these literatures, we manually extracted information on newly validated lncRNAs, their related diseases and interactions, and found 155 lncRNAs not listed in any existing databases for experimentally validated lncRNAs.

Next, the articles referenced as well as not referenced by previous databases were investigated for locating additional information required by our database. We manually examined all entries in lncRNAdb, LncRNADisease, Lnc2Cancer and PLNlncRbase. We excluded 767 lncRNAs collected by PLNlncRbase that was obtained from high-throughput microarray or RNA-seq. For lncRNAs collected by above databases but lacking function annotation, disease association or other information, we manually curated information from literature. In short, we added disease associated to entries in lncRNAdb, function annotations to entries in LncRNADisease, other interaction and disease information to entries in Lnc2Cancer, and interaction information to entries in PLNlncRbase. Only 399 lncRNAs indexed in previous databases were incorporated into our database without finding new information, compared to the fact that 989 entries incorporated were modified with new information. This modification along with 155 novel lncRNAs leads to 1144 (74%, 1144/1543) lncRNAs with new information not covered by existing databases of experimentally validated lncRNAs.

DATABASE CONTENT

Our database EVLncRNAs is intended to be a comprehensive database that includes all species and covers functional and disease-specific roles for all lncRNAs validated by one or more low-throughput experiments. In EVLncRNAs, lncRNAs were named following the HUGO Gene nomenclature Committee (HGNC) (34) with alias included. Sequence and position information of lncRNAs were annotated according to NCBI with links pointing to accessions in the NCBI (35) and Ensembl (36) provided if known. All data in EVLncRNAs were organized using MySQL. The web interface was designed by PHP. The web services were built on Apache Tomcat. The EVLncRNAs database is available at http://biophy.dzu.edu.cn/EVLncRNAs/.

The database contains a total of 1543 lncRNAs from 77 species along with their annotated functions, interaction partners and relevant diseases, if known. Table 1 provides statistics of experimentally validated datasets for comparison. EVlncRNAs has nearly three times more lncRNAs than the largest existing dataset and covers 6 new species, 89 new diseases and 405 new interactions. Each entry contains general information, such as lncRNA official name, alias, species, chromosome, start site, end site, chain, exon number, assembly version, lncRNA class (such as lincRNA, antisense, exonic, intronic etc.) and accession numbers to NCBI and Ensembl. Moreover, we have included 14 peptide-coding lncRNAs.

EVLncRNAs contains 2324 associations between 886 lncRNAs and 338 diseases along with information on experimental methods, experimental samples, expression pattern, dysfunction types, detailed description and original articles (PubMed ID and web link). Among the entries, 89 diseases and 346 lncRNAs–disease associations were manually

Table 1. Comparison of EVLncRNAs with other experimentally validated lncRNA databases

Database	#LncRNAs	#Species	#Diseases	#Disease associations	#Functions (Excl. Interactions)	#Interactions
lncRNAdb	287	71	N/A	N/A	287	307
LncRNADisease	321	14	221	1102	N/A	475
Lnc2Cancer	531	1	86	1057	N/A	N/A
PLNlncRbase	420*	43	N/A	N/A	420*	N/A
lncRInter	276	15	N/A	N/A	N/A	922
EVLncRNAs	1543	77	338	2324	793	1163

*After excluding 767 lncRNA from high-throughput microarray and RNA-seq.

Figure 1. Browse page of EVLncRNAs.

curated from recent literatures that never appeared in other databases. The 346 lncRNAs–disease associations comprise of 76 associations between previously known lncRNAs and diseases, and 270 associations on new lncRNAs or diseases. Experimentally validated functions of lncRNAs were classified according to expression, mutation, interaction, locus and epigenetics (23).

For those lncRNAs that are not associated with diseases, we collected their functional information such as gene regulation, cell differentiation or embryonic development (interaction information is collected separately, see below). There are 166 novel entries in 793 entries from recent literature.

The database also contains 1163 entries of 445 lncRNAs interacting with other molecules, where we included the name of interaction target, level of interaction, type of interaction, description of interaction and original publication (PubMed ID and linked). Here, 'level of interaction' represents levels of interaction partner including DNA, RNA

and protein, and 'types of interactions' were classified into binding, regulation and co-expression; 'description of interaction' means descriptions of lncRNA interacting process from the literature. Among these entries, 500 entries were curated from recent literatures. In other words, 43% (500/1163) of interactions entries are new.

USER INTERFACE

The interface of EVLncRNAs allows users to browse, search and download. As shown Figure 1, there are four pull-down menus: Species, Disease, Interaction and Peptide-related in the 'Browse' menu (located in labeled position 1). Each menu contains sub-menus. Users can browse relevant entries in EVLncRNAs by clicking any pull-down menus or their sub-menus. Alternatively, users can click navigation bar (located in position 2) on the left to view the corresponding page or entry. Using 'HOXA11-AS' as an exam-

Welcome to EVLncRNAs

Home Browse Search Tools Download Submit Help

CLOSE

LncRNA

ID	Name	Alias	Species	NCBI accession	Ensembl
EL0569	HIF1A-AS1	5'aHIF-1A; 5'aHIF1alpha	Homo sapiens	NR_047116	ENSG00000258777

DISEASES

ID	lncRNA Name	Disease	Method	Sample	Expression pattern	Function type	Description	PMID	Source
EL0569	HIF1A-AS1	kidney cancer	N/A	N/A	N/A	expression	Diagnostic and discriminative marker	24373479	LncRNADisease
EL0569	HIF1A-AS1	Cardiovascular disease	N/A	N/A	N/A	N/A	Clopidogrel reduces apoptosis and promotes proliferation of human vascular endothelial cells induced by palmitic acid via suppression of the long non-coding RNA HIF1A-AS1 in vitro.	25761653	LncRNADisease
EL0569	HIF1A-AS1	kidney cancer	qPCR etc.	renal cancer tissue	up-regulated	expression	The antitumor DNA topoisomerase I (Top1) inhibitor camptothecin (CPT) increases the cellular levels of two antisense lncRNAs at the 5' (5'aHIF-1 alpha) and 3' (3'aHIF-1 alpha) ends of the human HIF1A-AS1 gene.	21897117	LncRNADisease Lnc2Cancer
EL0569	HIF1A-AS1	non-small cell lung cancer	qPCR etc.	blood (serum), NSCLC tissue	up-regulated	expression	The levels of XIST (P < 0.05) and HIF1A-AS1 (P < 0.05) were significantly increased in tumor tissues or serum from NSCLC patients as compared to those of control group. Moreover, serum levels of XIST and HIF1A-AS1 were significantly decreased after surgical treatment as compared to pre-operative.	26339353	Lnc2Cancer
EL0569	HIF1A-AS1	aneurysm of the thoracoabdominal aorta	RT-PCR	VSMCs, serum of TAA patients	up-regulated	expression	HIF1a-AS1 was overexpressed in the thoracoabdominal aorta aneurysm and the interaction between HIF1a-AS1 and apoptotic proteins plays a key role in the proliferation and apoptosis of VSMCs in vitro, which may contribute to the pathogenesis of thoracoabdominal aorta aneurysm.	25550800	
EL1459	XIST	non-small cell lung cancer	qPCR etc.	blood (serum), NSCLC tissue	up-regulated	expression	The levels of XIST (P < 0.05) and HIF1A-AS1 (P < 0.05) were significantly increased in tumor tissues or serum from NSCLC patients as compared to those of control group. Moreover, serum levels of XIST and HIF1A-AS1 were significantly decreased after surgical treatment as compared to pre-operative.	26339353	Lnc2Cancer

Function (not disease relevant)

ID	lncRNA Name	Method	Sample	Expression pattern	Function type	Description	PMID	Source
EL0569	HIF1A-AS1	Real-time polymerase chain reaction (PCR), western blot analysis, lncRNA PCR arrays and chromatin immunoprecipitation	human bone marrow stromal cells	down-regulated	interaction	HIF1α-AS1 is an essential mediator of osteoblast differentiation.	26460121	

INTERACTION

ID	lncRNA Name	Interaction target	Level of interaction	Type of interaction	Description	PMID	Source
EL0569	HIF1A-AS1	HOXD10	RNA-DNA	regulation	Low expression of HIF1α-AS1 was sufficient to block the expression of HOXD10. Downregulation of HOXD10 by HIF1α-AS1 interfered with acetylation, and subsequently resulted in the inhibition of osteoblast differentiation.	26460121	

Contact blzhou@126.com for inquiries and feedback.
Copyright@Dezhou University

Figure 2. Search result page of EVLncRNAs.

ple (position 3), the page of this entry displays the lncRNA information (position 4), its related disease (position 5), function (position 6) and interaction information (position 7). If an lncRNA involves the function of interacting with other molecules, its interaction function will be described in more details in a following table. In the 'Search' page, EVLncRNAs enables users to search by any keywords, such as lncRNA name, alias, disease name, experimental methods, associated components and level of interaction. EVLncRNAs offers 'fuzzy' searching capabilities and returns all matching records. Figure 2 shows the search result of 'HIF1A-AS1' with specific link to the lncRNA found (labeled in position 1) and original articles (with PMID, labeled in 2) along with its disease association (position 3), function (position 4) and interaction information (position 5) as labeled. EVLncRNAs supports users to download the database data in the 'Download' page. In addition, EVLncRNAs provides links to the tools for lncRNA prediction (37,38) in the 'Tools' page. EVLncRNAs also permits users to submit novel experimentally validated lncRNAs and related diseases or associated components in the 'Submit' page. The submitted record will be included in the next release after manually checked by our curators. A detailed user tutorial is available in the 'Help' page.

FUTURE EXTENSIONS

It is expected that the number of experimentally validated lncRNAs will continue to increase. EVLncRNAs will be manually curated and updated in a regular basis. We will also gradually add new tools for analysis and prediction of lncRNA, when they are available.

DISCUSSION AND CONCLUSION

LncRNAs have been found to play an essential role in gene regulation and diseases (3,11,12). However, the majority of databases for lncRNAs relied on computational prediction and/or high-throughput experiments with uncertain accuracy. A few experimentally validated lncRNA databases (22–26) have limited information available. The purpose of this work is to build a comprehensive dataset for all lncRNAs that were validated by low-throughput experiments.

The current version of EVLncRNAs (up to 1 May 2016) contains the largest collection of experimentally validated lncRNAs across 77 species. This is close to three times larger than any other experimentally validated datasets. There are 869 lncRNAs for human, 220 for animal, 428 for plant and 26 for microbe (Supplementary Figure S1). The top 20 most studied species on lncRNA were shown in Figure 3. The human and mouse used as a human disease model are top two species with the most identified lncRNAs. The next species is *Arabidopsis thaliana*, which is a model organism for plant research. About 34 lncRNAs have orthologs in human, mouse and/or rat, which mean the same functional lncRNAs in different species, such as XIST of human and Xist of mouse, and 23 lncRNAs have orthologs in plant
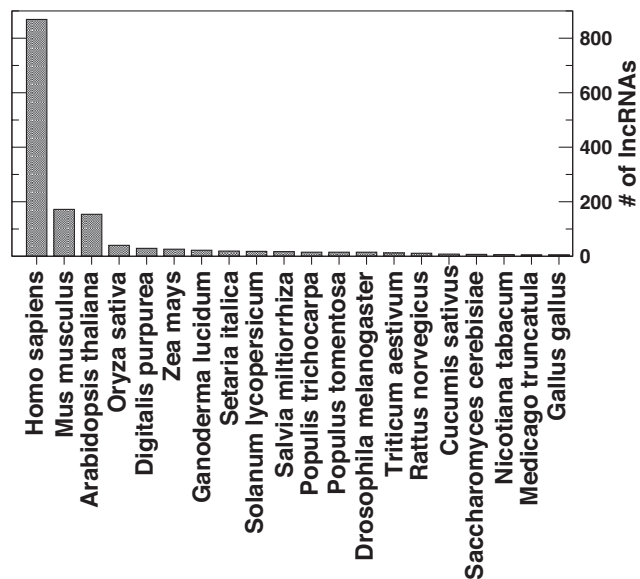
Figure 3. Top 20 species for the number of experimentally validated lncR-NAs.

species. This should provide important information for evolution studies.

Disease association is an important component of EVL-ncRNAs. Diseases were classified into cancers (147 types) and others. Top 15 cancers and top 15 other diseases associated with the highest number of lncRNAs are shown in Supplementary Figures S2 and 3, respectively. That the number of cancer-related lncRNAs is greater than those related with other diseases is likely due to the fact that more researchers are working on cancers because even in LncRNADisease (specifically designed to include lncRNAs relevant to human disease) collections, the majority (627/1070 = 59%) are related to neoplastic conditions. That is, the dominance of cancer-related lncRNAs is not caused by the contribution from the entries from Lnc2Cancer collecting lncRNAs relevant to cancers. Currently, the disease associated with the most number of lncRNAs (>100) is hepatocelluar carcinoma. Many lncRNAs are associated with more than one disease. Those lncRNAs associated with more than 10 diseases are shown in Supplementary Figure S4. H19, for example, is associated with 40 cancers and 24 other diseases.

In addition, EVLncRNAs contains 1163 entries of interaction information between 445 lncRNAs and their associated partners. As shown in Supplementary Figure S5, the highest number of lncRNAs is involved in regulation. This is followed by binding and co-expression.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. 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.
2. Consortium,E.P., Dunham,I. and Kundaje,A. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
3. Ponting,C.P., Oliver,P.L. and Reik,W. (2009) Evolution and functions of long noncoding RNAs. *Cell*, **136**, 629–641.
4. Stefani,G. and Slack,F.J. (2008) Small non-coding RNAs in animal development. *Nat. Rev. Mol. Cell Biol.*, **9**, 219–230.
5. Carthew,R.W. and Sontheimer,E.J. (2009) Origins and mechanisms of miRNAs and siRNAs. *Cell*, **136**, 642–655.
6. Brannan,C.I., Dees,E.C., Ingram,R.S. and Tilghman,S.M. (1990) The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.*, **10**, 28–36.
7. Brockdorff,N., Ashworth,A., Kay,G.F., McCabe,V.M., Norris,D.P., Cooper,P.J., Swift,S. and Rastan,S. (1992) The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell*, **71**, 515–526.
8. Mercer,T.R., Dinger,M.E. and Mattick,J.S. (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.*, **10**, 155–159.
9. Hon,C.-C., Ramilowski,J.A., Harshbarger,J., Bertin,N., Rackham,O.J., Gough,J., Denisenko,E., Schmeier,S., Poulsen,T.M. and Severin,J. (2017) An atlas of human long non-coding RNAs with accurate 5′ ends. *Nature*, **543**, 199–204.
10. Toiber,D., Leprivier,G. and Rotblat,B. (2017) Long noncoding RNA: noncoding and not coded. *Cell Death Discov.*, **3**, 16104.
11. Gibb,E.A., Brown,C.J. and Lam,W.L. (2011) The functional role of long non-coding RNA in human carcinomas. *Mol. Cancer*, **10**, 38.
12. Van Roosbroeck,K., Pollet,J. and Calin,G.A. (2013) miRNAs and long noncoding RNAs as biomarkers in human diseases. *Exp. Rev. Mol. Diagn.*, **13**, 183–204.
13. Yang,Y., Chen,L., Gu,J., Zhang,H., Yuan,J., Lian,Q., Lv,G., Wang,S., Wu,Y., Yang,Y.C.T. *et al.* (2017) Recurrently deregulated lncRNAs in hepatocellular carcinoma. *Nat. Commun.*, **8**, 14411.
14. Yotsukura,S., duVerle,D., Hancock,T., Natsume-Kitatani,Y. and Mamitsuka,H. (2017) Computational recognition for long non-coding RNA (lncRNA): software and databases. *Brief. Bioinform.*, **18**, 9–27.
15. Jalali,S., Kapoor,S., Sivadas,A., Bhartiya,D. and Scaria,V. (2015) Computational approaches towards understanding human long non-coding RNA biology. *Bioinformatics*, **31**, 2241–2251.
16. Fritah,S., Niclou,S.P. and Azuaje,F. (2014) Databases for lncRNAs: a comparative evaluation of emerging tools. *RNA*, **20**, 1655–1665.
17. Volders,P.-J., Verheggen,K., Menschaert,G., Vandepoele,K., Martens,L., Vandesompele,J. and Mestdagh,P. (2015) An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res.*, **43**, D174–D180.
18. Yi,X., Zhang,Z., Ling,Y., Xu,W. and Su,Z. (2015) PNRD: a plant non-coding RNA database. *Nucleic Acids Res.*, **43**, D982–D989.
19. Chen,D., Yuan,C., Zhang,J., Zhang,Z., Bai,L., Meng,Y., Chen,L.L. and Chen,M. (2012) PlantNATsDB: a comprehensive database of

plant natural antisense transcripts. *Nucleic Acids Res.*, **40**, D1187–D1193.

20. Zhao,Y., Li,H., Fang,S., Kang,Y., wu,W., Hao,Y., Li,Z., Bu,D., Sun,N., Zhang,M.Q. *et al.* (2016) NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.*, **44**, D203–D208.

21. 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.

22. Quek,X.C., Thomson,D.W., Maag,J.L.V., Bartonicek,N., Signal,B., Clark,M.B., Gloss,B.S. and Dinger,M.E. (2015) lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.*, **43**, D168–D173.

23. 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.

24. Ning,S., Zhang,J., Wang,P., Zhi,H., Wang,J., Liu,Y., Gao,Y., Guo,M., Yue,M. and Wang,L. (2016) Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic Acids Res.*, **44**, D980–D985.

25. Liu,C.-J., Gao,C., Ma,Z., Cong,R., Zhang,Q. and Guo,A.-Y. lncRInter: a database of experimentally validated long non-coding RNA interaction. *J. Genet. Genomics*, **44**, 265–268.

26. Xuan,H., Zhang,L., Liu,X., Han,G., Li,J., Li,X., Liu,A., Liao,M. and Zhang,S. (2015) PLNlncRbase: a resource for experimentally identified lncRNAs in plants. *Gene*, **573**, 328–332.

27. Yang,F., Huo,X.-s., Yuan,S.-x., Zhang,L., Zhou,W.-p., Wang,F. and Sun,S.-h. (2013) Repression of the long noncoding RNA-LET by histone deacetylase 3 contributes to hypoxia-mediated metastasis. *Mol. Cell*, **49**, 1083–1096.

28. Yang,F., Zhang,L., Huo,X.s., Yuan,J.H., Xu,D., Yuan,S.X., Zhu,N., Zhou,W.P., Yang,G.S. and Wang,Y.Z. (2011) Long noncoding RNA high expression in hepatocellular carcinoma facilitates tumor growth through enhancer of zeste homolog 2 in humans. *Hepatology*, **54**, 1679–1689.

29. Zhang,B., Arun,G., Mao,Yuntao S., Lazar,Z., Hung,G., Bhattacharjee,G., Xiao,X., Booth,Carmen J., Wu,J., Zhang,C. *et al.* (2012) The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep.*, **2**, 111–123.

30. Xiang,J.-F., Yin,Q.-F., Chen,T., Zhang,Y., Zhang,X.-O., Wu,Z., Zhang,S., Wang,H.-B., Ge,J., Lu,X. *et al.* (2014) Human colorectal cancer-specific CCAT1-L lncRNA regulates long-range chromatin interactions at the MYC locus. *Cell Res.*, **24**, 513–531.

31. Yang,F., Xue,X., Bi,J., Zheng,L., Zhi,K., Gu,Y. and Fang,G. (2013) Long noncoding RNA CCAT1, which could be activated by c-Myc, promotes the progression of gastric carcinoma. *J. Cancer Res. Clin. Oncol.*, **139**, 437–445.

32. Chakravarty,D., Sboner,A., Nair,S.S., Giannopoulou,E., Li,R., Hennig,S., Mosquera,J.M., Pauwels,J., Park,K., Kossai,M. *et al.* (2014) The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer. *Nat. Commun.*, **5**, 5383.

33. Conesa,A., Madrigal,P., Tarazona,S., Gomez-Cabrero,D., Cervera,A., McPherson,A., Szczesniak,M.W., Gaffney,D.J., Elo,L.L., Zhang,X.G. *et al.* (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol.*, **17**, 13.

34. Yates,B., Braschi,B., Gray,K.A., Seal,R.L., Tweedie,S. and Bruford,E.A. (2017) Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Res.*, **45**, D619–D625.

35. NCBI Resource Coordinators (2017) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **45**, D12–D17.

36. Aken,B.L., Achuthan,P., Akanni,W., Amode,M.R., Bernsdorff,F., Bhai,J., Billis,K., Carvalho-Silva,D., Cummins,C., Clapham,P. *et al.* (2017) Ensembl 2017. *Nucleic Acids Res.*, **45**, D635–D642.

37. Hu,L., Xu,Z., Hu,B. and Lu,Z.J. (2017) COME: a robust coding potential calculation tool for lncRNA identification and characterization based on multiple features. *Nucleic Acids Res.*, **45**, e2.

38. Kong,L., Zhang,Y., Ye,Z.-Q., Liu,X.-Q., Zhao,S.-Q., Wei,L. and Gao,G. (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.*, **35**, W345–W349.