

# NSDNA: a manually curated database of experimentally supported ncRNAs associated with nervous system diseases

Jianjian Wang<sup>1,†</sup>, Yuze Cao<sup>1,†</sup>, Huixue Zhang<sup>1,†</sup>, Tianfeng Wang<sup>1</sup>, Qinghua Tian<sup>1</sup>, Xiaoyu Lu<sup>1</sup>, Xiaoyan Lu<sup>1</sup>, Xiaotong Kong<sup>1</sup>, Zhaojun Liu<sup>1</sup>, Ning Wang<sup>1</sup>, Shuai Zhang<sup>1</sup>, Heping Ma<sup>2,\*</sup>, Shangwei Ning<sup>3,\*</sup> and Lihua Wang<sup>1,\*</sup>

<sup>1</sup>Department of Neurology, The Second Affiliated Hospital of Harbin Medical University, Harbin 150081, China,

<sup>2</sup>Department of Physiology, Emory University School of Medicine, Atlanta, GA 30322, USA and <sup>3</sup>College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China

Received August 09, 2016; Revised October 17, 2016; Editorial Decision October 18, 2016; Accepted October 20, 2016

## ABSTRACT

**The Nervous System Disease NcRNAome Atlas (NSDNA) (<http://www.bio-bigdata.net/nsdna/>) is a manually curated database that provides comprehensive experimentally supported associations about nervous system diseases (NSDs) and noncoding RNAs (ncRNAs). NSDs represent a common group of disorders, some of which are characterized by high morbidity and disabilities. The pathogenesis of NSDs at the molecular level remains poorly understood. ncRNAs are a large family of functionally important RNA molecules. Increasing evidence shows that diverse ncRNAs play a critical role in various NSDs. Mining and summarizing NSD–ncRNA association data can help researchers discover useful information. Hence, we developed an NSDNA database that documents 24 713 associations between 142 NSDs and 8593 ncRNAs in 11 species, curated from more than 1300 articles. This database provides a user-friendly interface for browsing and searching and allows for data downloading flexibility. In addition, NSDNA offers a submission page for researchers to submit novel NSD–ncRNA associations. It represents an extremely useful and valuable resource for researchers who seek to understand the functions and molecular mechanisms of ncRNA involved in NSDs.**

## INTRODUCTION

Nervous system diseases (NSDs) represent a common class of diseases, some of which are characterized by high mor-

bidity and disabilities. For example, stroke causes serious long-term disabilities and has been the third leading cause of death in most developed countries for many decades (1). Other NSDs, such as Alzheimer's and Parkinson's diseases, cause progressive neuron damage in the brain, severely affecting quality of life and putting great burdens on families and society (2). NSDs are remarkably complex, and the underlying causes involve changes at multiple levels, including genome, transcriptome, epigenetic mechanisms and environmental factors (3). Therefore, one of the primary challenges in NSD research is to uncover the precise molecular mechanisms underlying the diseases.

Non-coding RNAs (ncRNAs) are a family of functional RNA molecules that include microRNAs (miRNAs), long noncoding RNAs (lncRNA), small interfering RNA (siRNA), small nucleolar RNAs (snoRNA), piwi-interacting RNA (piRNA), among others (4). There is now overwhelming evidence that changes in ncRNA expression levels have been associated with NSDs (5). For example, miRNAs (~22 nt small ncRNA molecules that repress mRNA target expression) are the hallmark of various NSDs, such as epilepsy (6), multiple sclerosis (7) and Huntington disease (8). lncRNAs (>200 nt RNA molecules) have been found in recent years to play a role in nervous system development, plasticity, evolution and disease (9). For example, the lncRNA HOTAIR acts as an endogenous 'sponge' of miR-141, resulting in significant reduction of glioma growth (10). To date, many ncRNA-related databases have been developed, such as HMDD (11), lncRNA Disease (12) and lnc2Cancer (13). However, a high quality and well-curated public resource specifically for NSDs–ncRNAs association data remains unavailable.

\*To whom correspondence should be addressed. Tel: +86 451 86605788; Fax: +86 451 86605788; Email: wanglh211@163.com

Correspondence may also be addressed to Shangwei Ning. Email: ningsw@ems.hrbmu.edu.cn

Correspondence may also be addressed to Heping Ma. Email: heping.ma@emory.edu

†These authors contributed equally to the work as the first authors.

**Table 1.** Statistics for the NSD-ncRNA entries in the NSDNA database

Species	miRNA	lncRNA	siRNA	snoRNA	piRNA	Total
Homo sapiens	12 361	4223	4	6	-	16 594
Mus musculus	5284	159	5	10	2	5460
Rattus norvegicus	2103	120	4	-	105	2332
Drosophila	75	-	-	-	-	75
Macaca mulatta	71	-	-	-	-	71
Pan troglodytes	56	-	-	-	-	56
Sus scrofa	48	-	-	-	-	48
Caenorhabditis elegans	35	-	-	-	-	35
Marmoset	25	-	-	-	-	25
Gallus gallus	16	-	-	-	-	16
Aedes albopictus cells	1	-	-	-	-	1
Total	20 075	4502	13	16	107	24 713

Currently, a huge amount of ncRNAs and NSDs data have accumulated in a short period of time with the rapid development of this field. However, the data are distributed among many massive articles that makes it inconvenient for researchers who want to further explore the relationship between NSDs and ncRNAs. It is worth noting that researchers have made great efforts to develop databases to explore NSD-related information across several fields. For example, the MethylomeDB database collects DNA methylation profiles of the brain (14), the Allen Brain Atlas integrates gene expression data and neuroanatomical information (15) and the EpilepsyGene and MDPD databases present gene mutation information for epilepsy and Parkinson's disease, respectively (16,17). However, no resource is currently focused on collecting experimentally supported ncRNAs–NSDs associations across various species.

To bridge this gap, we developed NSDNA, a manually curated database of experimentally supported NSD and ncRNA associations, which seeks to provide a high quality, comprehensive and specialist resource of NSD-related ncRNA deregulation across various species. The current version of NSDNA documents 24 713 entries of associations between 142 NSDs and 8593 ncRNAs in 11 species, curated from more than 1300 articles. We expect NSDNA to serve as a useful resource for researchers who wish to further investigate the relationship between ncRNAs and NSDs.

## DATABASE CONSTRUCTION

To ensure a high quality database, we referred to the steps used to assemble the databases HMDD, Lnc2Cancer and ViRBase (11,13,18). We mainly applied the following three steps in the data collection process: collecting disease and ncRNA information, searching for relevant articles and extracting useful information from the selected articles. First, NSD terminologies were organized based upon the controlled vocabulary of the Medical Subject Heading disease categories that contains ~600 NSDs (19). The ncRNA information was collected and integrated from various resources: miRNA symbols were taken from miRBase (20), lncRNA symbols were taken from NONCODE (21) and lncRNAdb (22), siRNA symbols were taken from siRecords (23), snoRNA symbols were taken from snoRNA-LBMEdb (24) and piRNA symbols were taken from piRNABank (25). In total, we collected ~600 NSDs and more than 600 000 ncRNA terms in the first step. We also applied the

ncRNA category names for those and other ncRNAs, such as ribosomal RNA and transfer RNA. Second, we used a simple script to select all abstracts in the PubMed database using the following keyword combinations: 'each nervous system disease' and/or 'each ncRNA symbol' or 'ncRNA category name' (18). Then, we downloaded all published literature and available supplementary files describing the associations between NSDs and ncRNAs. Third, we manually extracted experimentally supported ncRNA–NSD associations from selected articles by at least two researchers.

Finally, all NSDNA data were stored and managed using MySQL. The web interface was built using JSP. The scripts for the data processing programs were written in Java. The web service is running on an Apache Tomcat web server. The NSDNA database is freely available at <http://www.bio-bigdata.net/nsdna/> and <http://www.bio-bigdata.com/nsdna/>.

## DATABASE CONTENT

After systematically reviewing more than 1300 published articles, a total of 24 713 entries representing 142 NSDs and 8593 ncRNAs in 11 species were manually collected. The number of ncRNA-associated entries for each species is listed in Table 1. Each entry contains detailed information, including disease and ncRNA name, species, the ncRNA expression pattern (e.g. upregulated, downregulated), experimental methods (e.g. microarray, Real-time PCR), experimentally verified targets of ncRNA, detected tissue (e.g. brain, blood), associated treatment (e.g. levodopa, Interferon-Beta), detailed description and corresponding literature (PubMed ID and publication year). To verify altered ncRNA expression in NSDs, the researchers used various experimental methods, including high-throughput and low-throughput experiments. Therefore, we divided the data into high-throughput and low-throughput experiments, as previously described (26). High-throughput experiments include microarray and next-generation sequencing, whereas low-throughput experiments include real-time PCR, Northern blot, Western blot and luciferase reporter assays, and so on.

## USER INTERFACE

NSDNA provides a user-friendly web interface for an easy database query (Figure 1). In the 'Browse' page, users can

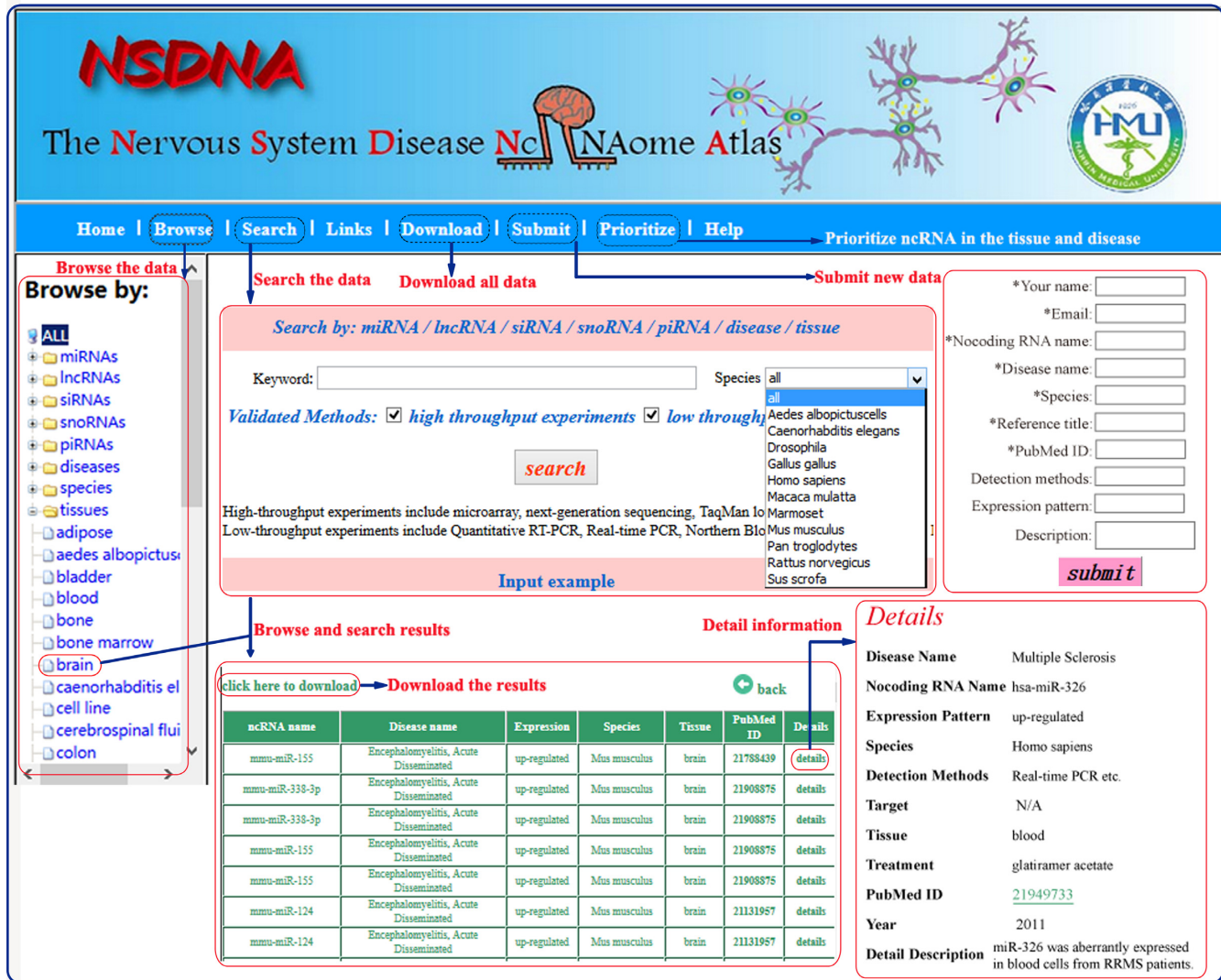
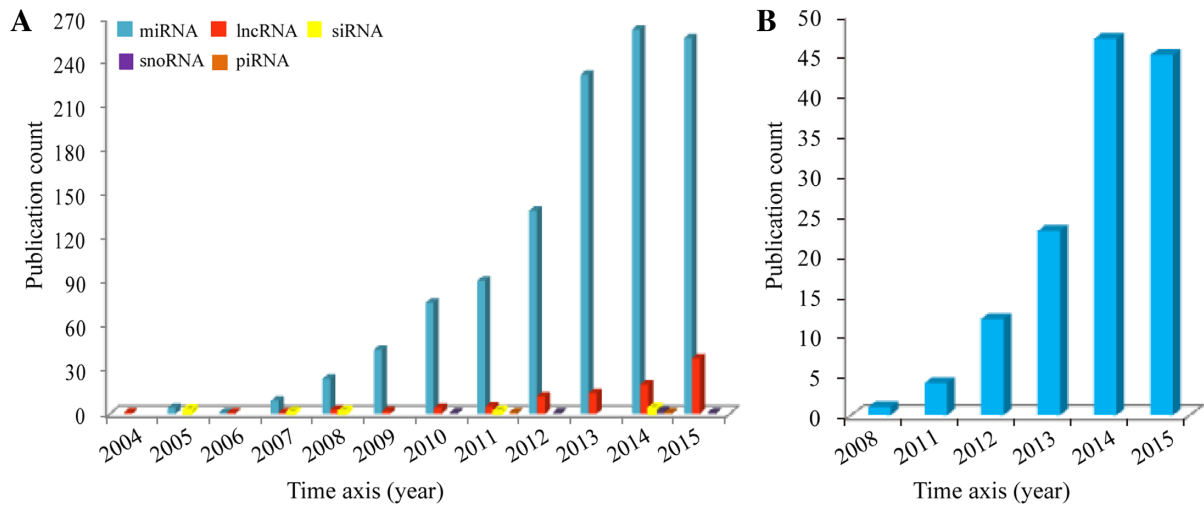


Figure 1. A schematic workflow of the NSDNA database.

browse by miRNAs, lncRNAs, siRNAs, snoRNAs, piRNAs, diseases, species and tissue name. By clicking on a particular node, corresponding results are displayed. In the 'Search' page, NSDNA allows users to search by miRNAs, lncRNAs, siRNAs, snoRNAs, piRNAs, disease and tissue name. NSDNA enables users to select all or particular species from the species pull-down menu. In addition, NSDNA provides an option that allows users to select low-throughput (Real-time PCR, Northern Blot, Luciferase Reporter Assay, etc.) or high-throughput (microarray, next-generation sequencing, etc.) experimental data or both. The search results are organized as tables in the web interface, and users can obtain more extensive information by clicking on the 'Details' hyperlink. NSDNA supports 'fuzzy' searching capabilities, where all possible results matching the keywords will be displayed. NSDNA allows users to download all the obtained data in the 'Download' page. In addition, NSDNA invites users to submit novel NSD-ncRNA associations for updating the database in the 'Submit' page. Once approved by the submission review committee, the submit-

ted record is included in the database and made available to the public in the following release. We provide a detailed database use tutorial in the 'Help' page.

Moreover, NSDNA provides users a prioritization page that enables researchers to prioritize ncRNAs in a specific tissue and disease. For miRNAs, we used low-throughput experiments confirmed miRNA as the gold standard and prioritize the set of miRNAs supported by high-throughput experiments based on semantic similarities of functional genomic data as previously described (27). In brief, this prioritized method ranked candidate miRNAs according to their semantic similarity and distance proximity to the gold standard. For lncRNAs, we used low-throughput experiments confirmed lncRNA as the gold standard and prioritize the set of lncRNAs supported by high-throughput experiments based on functional similarity as previously described (28). Finally, we integrated the tool and the prioritized results into the NSDNA database.



**Figure 2.** (A) Annual publication counts in NSDNA. (B) Annual publication counts of circulating miRNAs in NSDNA.

## DISCUSSION

NSDs are a common group of disorders, encompassing a broad range of clinical features, such as seizures, dementia and paralysis, among others. Researchers and neurologists have been trying to decipher the precise molecular and cellular regulatory mechanisms involved in these disorders. In recent years, emerging work has focused on exploring potential mechanisms involving ncRNAs in NSDs (29). Consequently, NSD–ncRNA association data have been accumulating rapidly in this decade. However, most data are dispersed in many independent studies. It has become increasingly clear that an integrated collection of NSD-associated ncRNA data are critical for a complete understanding of NSD processes. Thus, we developed NSDNA, an NSD-specific database that provides a comprehensive resource on diverse ncRNA–NSD associations across various species.

With the advance of disease-related ncRNA research, several ncRNAs (mainly miRNAs and lncRNAs) associated databases were established, such as miR2Disease (30), HMDD (11) and LncRNADisease (12). These databases are crucial for studying ncRNA functions in human diseases. However, to the best of our knowledge, none of these resources were developed to specifically collect experimentally supported NSD–ncRNA association data. For example, miR2Disease only documents 450 NSD-related entries of associations between 179 miRNAs and 26 NSDs, while LncRNADisease only includes 196 NSD-related entries of associations between 74 lncRNAs and 37 NSDs, and only human data are available in these databases. NSDNA contains general NSDs, nearly 6 and 4 times the number of NSDs than the miR2Disease and LncRNADisease databases, respectively. Compared with miR2Disease, NSDNA also includes data regarding mutations in ncRNA and their association with NSDs. For example, in the NSDNA database, we also found that a SNP (rs11014002) located in miR-603 precursor exhibits a protective effect toward Alzheimer disease risk (31). Moreover, miR2Disease has not been updated since 2009. In addition to recording more NSD–ncRNA associations, NSDNA has several advantages compared to previous studies (Supple-

mentary Table S1). First, NSDNA includes many ncRNAs categories (miRNAs, lncRNAs, siRNAs, snoRNAs and piRNAs), while the above-mentioned databases only include one ncRNA category. Second, NSDNA includes data for many species, and users can retrieve and download all species or particular species data. Third, data on NSD-associated circulating miRNA, treatment-associated miRNA and miRNA target were also added to the NSDNA. Thus, NSDNA is a specialized database that provides a comprehensive resource on ncRNA and NSD associations. We believe that NSDNA will be beneficial to researchers who wish to further dissect mechanisms involved in NSDs.

We also counted the number of published articles each year in NSDNA. We found that the number of NSD-associated ncRNA articles have generally increased, suggesting that the research on ncRNAs may be one of the hottest topics in the NSD field in this decade (Figure 2A). Particularly, the number of miRNA- and lncRNA-related articles have been steadily increasing, and hundreds of miRNAs and lncRNAs have been identified due to the development of high-throughput technologies. In recent years, new evidence has shown that circulating miRNA levels appear stable in fluids and can be used as biomarkers for diagnosis and treatment of NSDs (32). Therefore, we focused on circulating miRNA studies and found that the number of circulating miRNAs publications has increased dramatically from 2013 to 2015, suggesting that circulating miRNAs have also become a hot topic in NSD research (Figure 2B).

Because more than half of the entries were human miRNA-associated data, we constructed a human miRNA–NSD bipartite network, where nodes represent miRNAs or NSDs and the lines denote experimentally supported relationships between miRNAs and NSDs (Supplementary Figure S1A). We listed the number of miRNAs associated with the NSDs, the number of NSDs associated with the miRNAs, and the five most highly connected nodes (Supplementary Figure S1B and S1C). For instance, the disease with the highest connectivity is glioblastoma that is

associated with 839 miRNAs. These results indicate that glioblastoma has received wide attention in miRNA-related research and also has a complex regulatory mechanism mediated by miRNA. In addition, hsa-miR-146a is the most highly connected miRNA node, which is associated with 34 NSDs and might be widely associated with NSDs.

## CONCLUSION AND FUTURE EXTENSIONS

In conclusion, NSDs are common complex disorders that researchers have been trying to understand for years. Recent studies have shed light on ncRNAs' role in NSDs and the mechanisms involved. NSDNA not only provides a comprehensive NSD-specialized database with experimental support but also offers a more global perspective on ncRNA functions in NSDs. In the future, we will continue to manually collect newly validated ncRNA–NSD associations and update the database every two months. In addition, we plan to integrate more sources such as functional annotations and provide additional tools such as an ncRNA–NSD association prediction tool. NSDNA will serve as a valuable resource for deciphering ncRNA mechanisms and improving the diagnosis and treatment of NSDs.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

National Natural Science Foundation of China [81371324, 81571166 and 31401090]; Specialized Research Fund for the Doctoral Program of Higher Education of China [20132307110008]; Foundation of Heilongjiang Medical Science Institute [201406]; Health and Family Planning Commission of Heilongjiang Province [2016-052 and 2016-072]. Funding for open access charge: National Natural Science Foundation of China [81371324, 81571166 and 31401090].

*Conflict of interest statement.* None declared.

## REFERENCES

- Brott,T. and Bogousslavsky,J. (2000) Treatment of acute ischemic stroke. *N. Engl. J. Med.*, **343**, 710–722.
- Silberberg,D., Anand,N.P., Michels,K. and Kalaria,R.N. (2015) Brain and other nervous system disorders across the lifespan - global challenges and opportunities. *Nature*, **527**, S151–S154.
- Silbereis,J.C., Pochareddy,S., Zhu,Y., Li,M. and Sestan,N. (2016) The cellular and molecular landscapes of the developing human central nervous system. *Neuron*, **89**, 248–268.
- Guttman,M. and Rinn,J.L. (2012) Modular regulatory principles of large non-coding RNAs. *Nature*, **482**, 339–346.
- Esteller,M. (2011) Non-coding RNAs in human disease. *Nat. Rev. Genet.*, **12**, 861–874.
- Tan,C.L., Plotkin,J.L., Veno,M.T., von Schimmelmann,M., Feinberg,P., Mann,S., Handler,A., Kjems,J., Surmeier,D.J., O'Carroll,D. *et al.* (2013) MicroRNA-128 governs neuronal excitability and motor behavior in mice. *Science*, **342**, 1254–1258.
- Zhu,S., Pan,W., Song,X., Liu,Y., Shao,X., Tang,Y., Liang,D., He,D., Wang,H., Liu,W. *et al.* (2012) The microRNA miR-23b suppresses IL-17-associated autoimmune inflammation by targeting TAB2, TAB3 and IKK-alpha. *Nat. Med.*, **18**, 1077–1086.
- Marti,E., Pantano,L., Banez-Coronel,M., Llorens,F., Minones-Moyano,E., Porta,S., Sumoy,L., Ferrer,I. and Estivill,X. (2010) A myriad of miRNA variants in control and Huntington's disease brain regions detected by massively parallel sequencing. *Nucleic Acids Res.*, **38**, 7219–7235.
- Briggs,J.A., Wolvetang,E.J., Mattick,J.S., Rinn,J.L. and Barry,G. (2015) Mechanisms of long non-coding RNAs in mammalian nervous system development, plasticity, disease, and evolution. *Neuron*, **88**, 861–877.
- Bian,E.B., Ma,C.C., He,X.J., Wang,C., Zong,G., Wang,H.L. and Zhao,B. (2016) Epigenetic modification of miR-141 regulates SKA2 by an endogenous 'sponge' HOTAIR in glioma. *Oncotarget*, **7**, 30610–30625.
- Li,Y., Qiu,C., Tu,J., Geng,B., Yang,J., Jiang,T. and Cui,Q. (2014) HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.*, **42**, D1070–D1074.
- 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.
- Ning,S., Zhang,J., Wang,P., Zhi,H., Wang,J., Liu,Y., Gao,Y., Guo,M., Yue,M., Wang,L. *et al.* (2016) Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic Acids Res.*, **44**, D980–D985.
- Xin,Y., Chanrion,B., O'Donnell,A.H., Milekic,M., Costa,R., Ge,Y. and Haghighi,F.G. (2012) MethyloDB: a database of DNA methylation profiles of the brain. *Nucleic Acids Res.*, **40**, D1245–D1249.
- Sunkin,S.M., Ng,L., Lau,C., Dolbeare,T., Gilbert,T.L., Thompson,C.L., Hawrylycz,M. and Dang,C. (2013) Allen brain atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res.*, **41**, D996–D1008.
- Ran,X., Li,J., Shao,Q., Chen,H., Lin,Z., Sun,Z.S. and Wu,J. (2015) EpilepsyGene: a genetic resource for genes and mutations related to epilepsy. *Nucleic Acids Res.*, **43**, D893–D899.
- Tang,S., Zhang,Z., Kavitha,G., Tan,E.K. and Ng,S.K. (2009) MDPD: an integrated genetic information resource for Parkinson's disease. *Nucleic Acids Res.*, **37**, D858–D862.
- Li,Y., Wang,C., Miao,Z., Bi,X., Wu,D., Jin,N., Wang,L., Wu,H., Qian,K., Li,C. *et al.* (2015) ViRBase: a resource for virus-host ncRNA-associated interactions. *Nucleic Acids Res.*, **43**, D578–D582.
- Nelson,S.J., Schopen,M., Savage,A.G., Schulman,J.L. and Arluk,N. (2004) The MeSH translation maintenance system: structure, interface design, and implementation. *Stud. Health Technol. Inform.*, **107**, 67–69.
- Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
- 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.
- Quek,X.C., Thomson,D.W., Maag,J.L., 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.
- Ren,Y., Gong,W., Zhou,H., Wang,Y., Xiao,F. and Li,T. (2009) siRecords: a database of mammalian RNAi experiments and efficacies. *Nucleic Acids Res.*, **37**, D146–D149.
- Lestrade,L. and Weber,M.J. (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.*, **34**, D158–D162.
- Sai Lakshmi,S. and Agrawal,S. (2008) piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.*, **36**, D173–D177.
- Liu,X., Wang,S., Meng,F., Wang,J., Zhang,Y., Dai,E., Yu,X., Li,X. and Jiang,W. (2013) SM2miR: a database of the experimentally validated small molecules' effects on microRNA expression. *Bioinformatics*, **29**, 409–411.
- Wang,P., Ning,S., Wang,Q., Li,R., Ye,J., Zhao,Z., Li,Y., Huang,T. and Li,X. (2013) mirTarPri: improved prioritization of microRNA targets through incorporation of functional genomics data. *PLoS One*, **8**, e53685.
- Wang,P., Ning,S., Zhang,Y., Li,R., Ye,J., Zhao,Z., Zhi,H., Wang,T., Guo,Z. and Li,X. (2015) Identification of lncRNA-associated

- competing triplets reveals global patterns and prognostic markers for cancer. *Nucleic Acids Res.*, **43**, 3478–3489.
29. Barry, G. (2014) Integrating the roles of long and small non-coding RNA in brain function and disease. *Mol. Psychiatry*, **19**, 410–416.
  30. Jiang, Q., Wang, Y., Hao, Y., Juan, L., Teng, M., Zhang, X., Li, M., Wang, G. and Liu, Y. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–D104.
  31. Zhang, C., Lu, J., Liu, B., Cui, Q. and Wang, Y. (2016) Primate-specific miR-603 is implicated in the risk and pathogenesis of Alzheimer's disease. *Aging (Albany NY)*, **8**, 272–290.
  32. Westphal, M. and Lamszus, K. (2015) Circulating biomarkers for gliomas. *Nat. Rev. Neurol.*, **11**, 556–566.