



Original article

CircR2Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases

Chunyan Fan¹, Xiujuan Lei^{1,*}, Zengqiang Fang¹, Qinghua Jiang² and Fang-Xiang Wu^{3,*}

¹School of Computer Science, Shaanxi Normal University, Xi'an 710119, China, ²Center for Bioinformatics, School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China and, ³Department of Mechanical Engineering and Division of Biomedical Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

*Corresponding author: Tel: +86 29 85310164; Fax: +86 29 85310661; Email: xjlei@snnu.edu.cn

Correspondence may also be addressed to Fang-Xiang Wu. Tel: +(306) 966 5280; Fax: +(306) 966 5427; Email: faw341@mail.usask.ca

Citation details: Fan,C., Lei,X., Fang,Z. *et al.* CircR2Disease: a manually curated database for experimentally supported circular RNAs associated with various diseases. *Database* (2018) Vol. 2018: article ID bay044; doi:10.1093/database/bay044

Received 24 February 2018; Revised 12 April 2018; Accepted 16 April 2018

Abstract

CircR2Disease is a manually curated database, which provides a comprehensive resource for circRNA deregulation in various diseases. Increasing evidences have shown that circRNAs play critical roles in transcriptional, post-transcriptional and translational regulation. Therefore, the aberrant expression of circRNAs has been associated with a group of diseases. It is significant to develop a high-quality database to deposit the deregulated circRNAs in diseases. The current version of CircR2Disease contains 725 associations between 661 circRNAs and 100 diseases by reviewing existing literatures. Each entry in the CircR2Disease contains detailed information for the circRNA–disease relationship, including circRNA name, coordinates and gene symbol, disease name, expression patterns of circRNA, experimental techniques, a brief description of the circRNA–disease relationship, year of publication and the PubMed ID. CircR2Disease provides a user-friendly interface to browse, search and download as well as to submit novel disease-related circRNAs. CircR2Disease could be very beneficial for researches to investigate the mechanism of disease-related circRNAs and explore the appropriate algorithms for predicting novel associations.

Database URL: <http://bioinfo.snnu.edu.cn/CircR2Disease/>

Introduction

Circular RNAs (circRNAs) are a class of recently re-discovered endogenous non-coding RNAs (ncRNAs) and have been found in various tissues and cell lines across most lives including archaea, plants and animals (1–4). CircRNAs are mostly generated from back-splicing events, a process in which the downstream 5' splice site and the 3' splice site are covalently linked to form covalently closed loops (5, 6). Normally, circRNAs regulate gene expression at transcriptional or post-transcriptional levels by titrating microRNAs (miRNAs), regulating transcription and interfering with splicing (2, 7, 8). In recent years, with the applications of microarray, RNA-seq and other techniques, the deregulated circRNAs are widely detected in a broad spectrum of diseases, including gliomas (9, 10), esophageal cancer (11), hepatoma carcinoma (12) and so on. In addition, circRNAs have the characteristics of universality, tissue/cell specific specificity, conservatism, stability (13–16), as well as easy to detect in human blood (17) or saliva (18). Thus, circRNAs are becoming the ideal class of molecular biomarkers for disease diagnose and treatment.

Recent studies have constructed several databases for circRNAs. For example, CircBase merged and unified several circRNAs datasets into a standardized database, including circRNA IDs, genomic coordinates and best transcripts and so on. (19). The CircNet provided novel circRNAs, expression profiles, circRNA isoforms and circRNA-miRNA-mRNA regulatory networks (20). The Tissue-Specific CircRNA Database (TSCD) identified the tissue-specific circRNAs and characterized the features and functions of circRNAs (13). CircInteractome allowed researchers to search the potential interactions of circRNAs with RNA-binding proteins (RBPs) and miRNAs, as well as designed specific circRNA divergent primers and circRNA-directed siRNAs (21). The starBase systematically identified the RNA-RNA and protein-RNA interaction networks from CLIP-seq datasets (22). SomamiR 2.0 contained the cancer somatic mutations in miRNA and competing endogenous RNAs (ceRNAs) including mRNAs, long non-coding RNAs (lncRNAs) and circRNAs (23). PlantCircNet provided visualized plant circRNA-miRNA-mRNA networks specific for plants (24). Most of these databases focused on the identification, expression, evolution or function of circRNAs by high-throughput sequencing technologies. Furthermore, Circ2Traits linked circRNAs and diseases by combining the miRNA-disease associations, disease associated SNPs and Argonaute interaction sites (25). A cancer-specific circRNA database (CSCD) contained the potential cellular localization, miRNA response element sites and RNA binding protein sites, open reading frames and alternative splicing events of parent genes through predicting methods (26).

Although Circ2Traits and CSCD have laid a significant foundation for the studies of disease-related circRNAs, these associations are confirmed using computational methods. Therefore, the studies of disease-related circRNAs are still limited and no database focused on the experimentally supported associations between circRNAs and diseases.

To bridge the gap, it is highly desirable to develop a high-quality circRNA-disease association database to study the roles of circRNAs in diseases. Here, we manually curated experimentally validated circRNA-disease associations in circR2Disease from existing literatures prior to 31 March 2018. The final database contains 725 experimentally supported associations between 661 circRNAs and 100 diseases. Furthermore, we summarized the usage of data sets in circR2Disease, which helps to mine the mechanism of the relationship between circRNAs and diseases as well as to predict the novel associations.

Data collection and database content

The experimentally validated associations between deregulated circRNAs and the occurrence of human diseases were collected through several steps as previously described (27–31). First, we searched the PubMed database with keywords matching 'circRNA', 'circular RNA', 'circRNA cancer', 'circRNA disease', 'circRNA tumor' and 'circRNA neoplasm'. Then, we retrieved the entries that describe the associations between circRNAs and diseases manually from these publications. In circR2Disease database, we collected 739 entries that include 725 circRNA-disease associations, 661 circRNAs and 100 diseases from the published papers prior to 31 March 2018. The curated information includes circRNA name, coordinates and gene symbol, disease name, expression patterns of circRNAs (upregulated or downregulated), experimental techniques (qRT-PCR, RNAi, northern blot, western blot, northern blot, luciferase reporter assays and so on), a brief description of circRNAs from literatures, year of publication. At the same time, the hyperlinks to the circBase ID for circRNA, the MalaCards database for disease and the NCBI PubMed ID for the reference are provided. Furthermore, the links of other useful databases are available in the homepage of circR2Disease database, including CircBase, CSCD, TSCD, Circ2Traits and CircInteractome.

Finally, all data in CircR2Disease are stored and managed using SQL Server (version 2008 r2), which is a mid-lighted database management system. The website is developed based on the .Net, a C# web framework (version 4.5). The web service is built using IIS (version 7.0), a Microsoft web service. The CircR2Disease database is freely available at <http://bioinfo.snnu.edu.cn/CircR2Disease/>.

User interface

The CircR2Disease database provides a user-friendly interface for users to browse, search, download and submit associations between circRNAs and diseases (Figure 1). First, users can browse relevant entries by selecting ‘CircRNA’ or ‘Disease’ on the left to view the corresponding entry. Taking ‘hsa_circ_0005986’ as an example, the page of this entry displays that hsa_circ_0005986 was downregulated in hepatocellular carcinoma. In the search page, users can obtain detailed information on each circRNA-disease associations by inputting the corresponding items such as circRNA name and disease name. CircR2Disease also offers a fuzzy search, whose results will list all potential entries with the full or partial names of keywords. In addition, all data in the database, including circRNA-disease associations, circRNA names and disease names, can be downloaded. The database also provides a submission page, in which users can submit novel identified circRNA-disease associations. Once approved by the review committee, the new associations will be included in the coming release database. Moreover, a detailed tutorial for the usage of the database is available in the ‘Help’ page.

Discussion and conclusions

Accumulating evidences have revealed that circRNAs are closely correlated with different types of diseases such as

atherosclerosis (32), lung adenocarcinoma (33), Alzheimer’s disease (34). Numerous investigations have been carried out to explore the specifically dysregulated circRNAs in diseases, which are considered as promising biomarkers for diagnosis, therapeutic and prognosis. Here, we have developed a database named as CircR2Disease, which integrates the experimentally supported circRNA and disease associations as well as their relationship descriptions. CircR2Disease provides a comprehensive resource for users to browse, search, download and submit the circRNA and disease associations with an easy-to-use web interface.

By analysing the entries from CircR2Disease database, we can find the publications about the aberrant expressed circRNA related with diseases are increasing dramatically (Figure 2), which indicates the investigation of the associations between circRNAs and diseases is becoming one of the hot topics. We construct a circRNA-disease bipartite network to describe the top 10 disease-related circRNAs (Figure 3). Based on the network knowledge, we infer that the node with more links was more important, and the loss of this node would have a great influence on the network. With the network, the circRNA hsa_circ_0000284 has the highest connectivity with diseases, and the gastric cancer has the highest connectivity with circRNAs. In addition, the concept of disease spectrum width (DSW) was introduced by a previous study, and we applied it to calculate

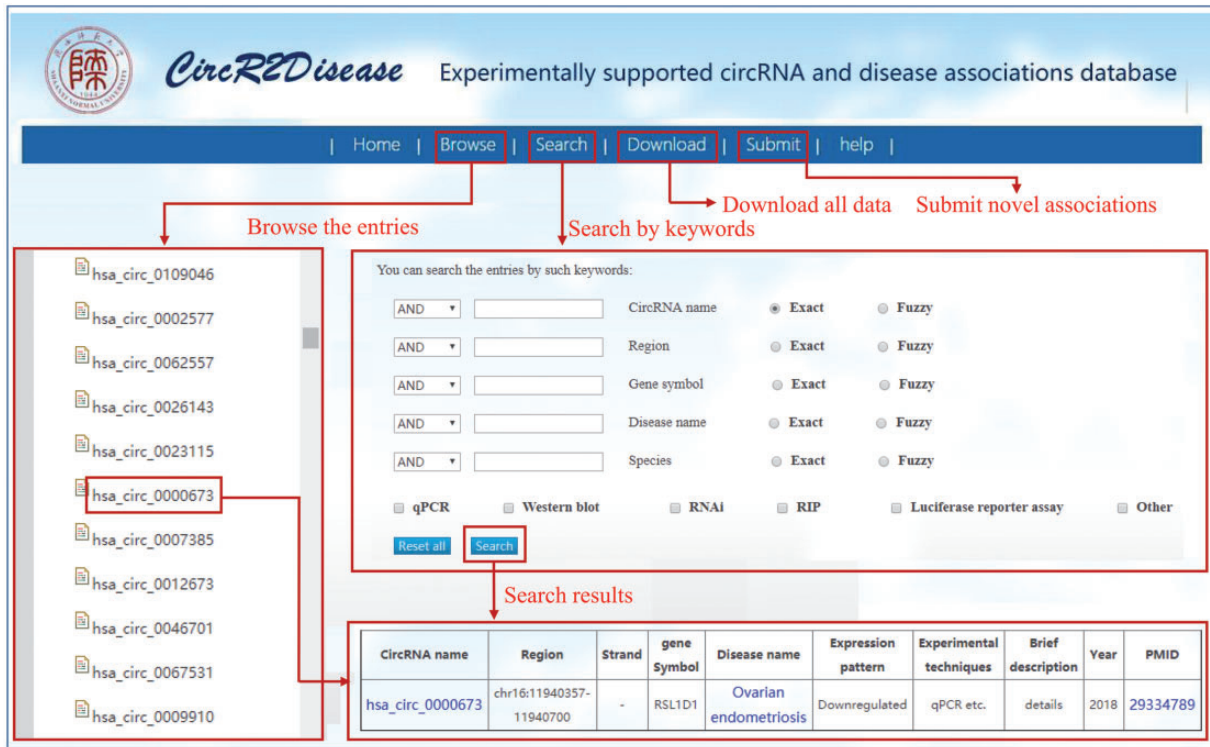


Figure 1. A schematic workflow of CircR2Disease.

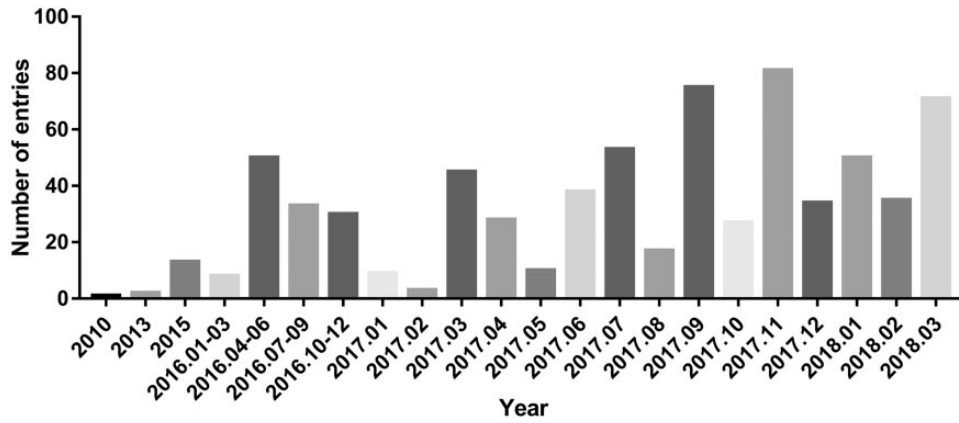


Figure 2. The number of entries in recent years.

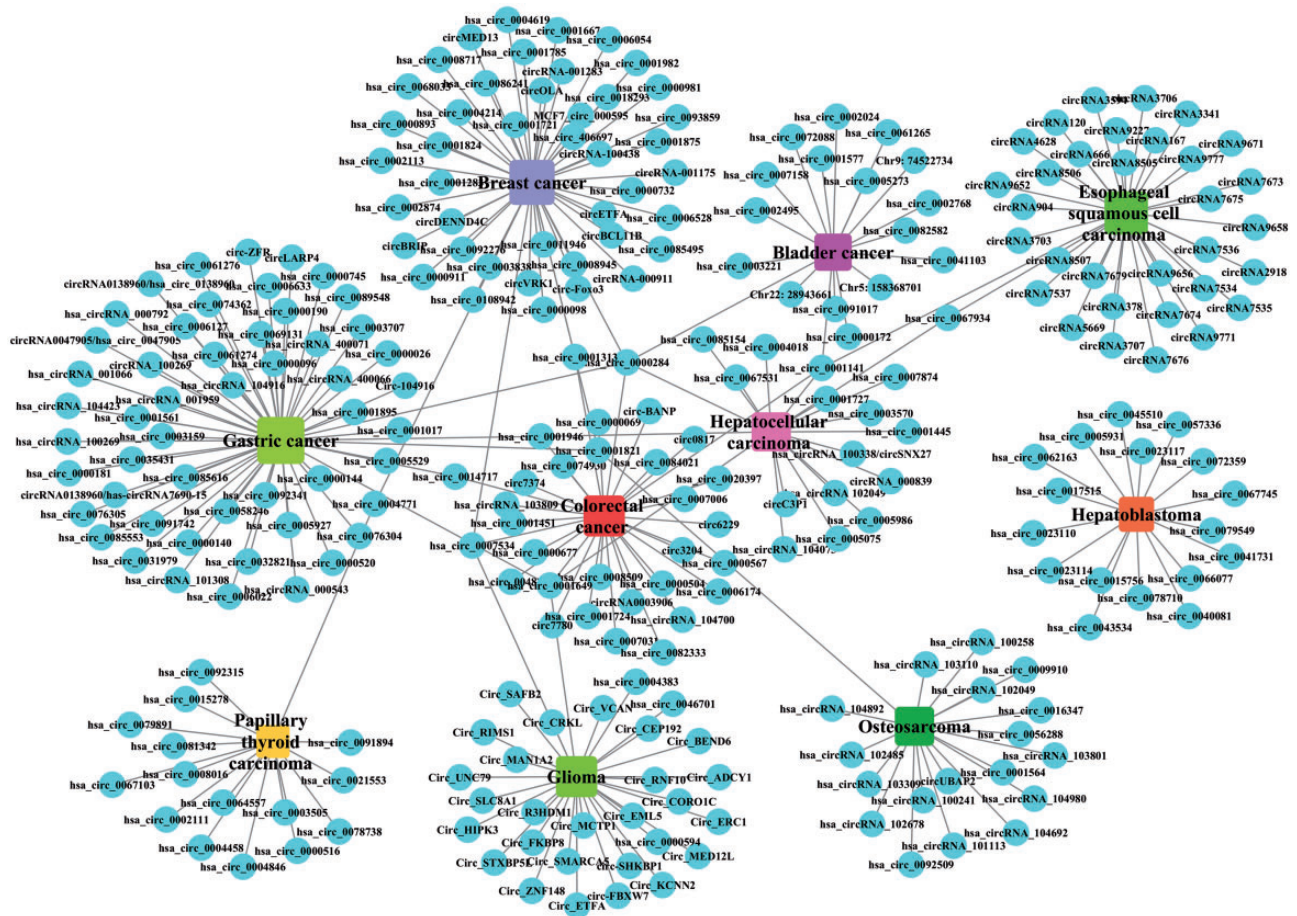


Figure 3. A bipartite network composed of top 10 disease-related circRNAs. Rectangles and circles corresponded to disease names and circRNA names, respectively. An edge corresponds to the experimentally circRNA-disease associations. The size of node corresponds to the degree of these nodes.

the DSW of circRNAs (35). For one circRNA i , $DSW = n(i)/N$ where, $n(i)$ represents the number of disease related with circRNA i , N represents the total number of diseases related with circRNAs. Here, we used DSW of a circRNA to be a metric to evaluate the importance in diseases. As a result, the top 10 circRNAs with the largest DSWs are

shown in Figure 4A. Similarly, the circRNA spectrum width (CSW) of a disease was introduced as a novel metric for one disease and the top 10 diseases with biggest CSWs are listed in Figure 4B. The result shows that the values of DSW and CSW were low, which is mainly because the relationships between circRNAs and diseases are still limited.

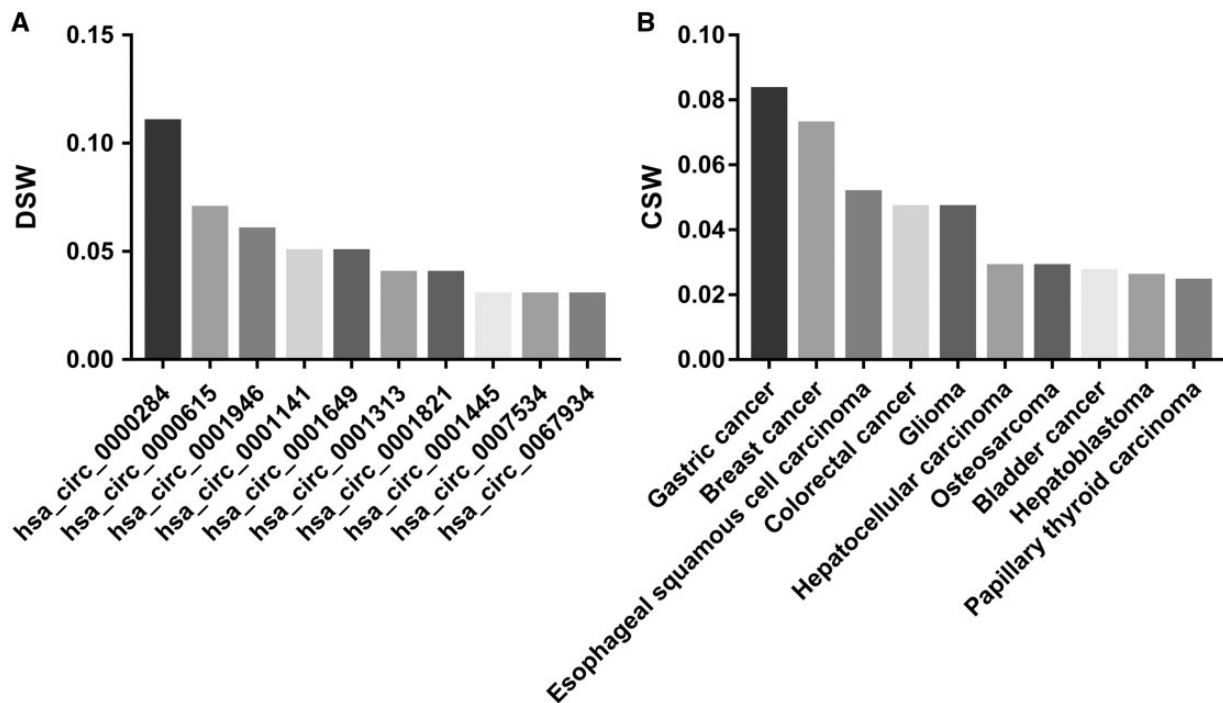


Figure 4. The top 10 circRNAs that with the largest DSWs (A) and the top 10 diseases that with the largest CSWs (B).

In addition, the values of DSW and CSW could be acted as two useful metrics with more associations confirmed.

The CircR2Disease database can be widely used to perform specific researches. Based on the hypothesis that circRNAs with similar functions tend to be associated with similar disease phenotypes and vice versa. Researches can predict the potential circRNA-disease associations through bioinformatics methods, such as network-based methods and machine learning methods. In addition, reports have shown that circRNA can perform their biological functions by binding with miRNA (7) and proteins (36). Therefore, miRNA-disease associations and protein-protein interactions are also useful for predicting circRNA-disease relationships. Furthermore, circRNA is a significant class of molecular ceRNAs, the circRNA function may be indirectly inferred by the lncRNA, mRNAs and pseudogenes. Overall, CircR2Disease will be a useful resource for further research of human disease.

Future extensions

The CircR2Disease provides a high-quality resource for studying the associations between circRNAs and diseases, and further extensions will be developed. It is expected that the number of experimentally validated disease-related circRNAs will continue to increase. The newly validated circRNA-disease associations will be manually curated and CircR2Disease database will be updated every 2

months. Additionally, the experimentally validated circRNA partners including miRNA sponges, proteins or other biological molecules will be integrated. Meanwhile, new tools and algorithms for analysing circRNA-disease associations will be developed and will be integrated into the CircR2Disease database in the future.

Funding

National Natural Science Foundation of China (grant nos. 61672334, 61502290 and 61401263). Funding for open access charge: National Natural Science Foundation of China (61672334).

Conflict of interest. None declared.

References

- Danan, M., Schwartz, S., Edelheit, S. *et al.* (2012) Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Res.*, **40**, 3131–3142.
- Memczak, S., Jens, M., Elefsinioti, A. *et al.* (2013) Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, **495**, 333–338.
- Chen, L., Huang, C., Wang, X. *et al.* (2015) Circular RNAs in eukaryotic cells. *Curr Genomics*, **16**, 312–318.
- Chu, Q., Zhang, X., Zhu, X. *et al.* (2017) PlantcircBase: a database for plant circular RNAs. *Mol. Plant*, **10**, 1126–1128.
- Barrett, S.P., Wang, P.L. and Salzman, J. (2015) Circular RNA biogenesis can proceed through an exon-containing lariat precursor. *Elife*, **4**, e07540.
- Starke, S., Jost, I., Rossbach, O. *et al.* (2015) Exon circularization requires canonical splice signals. *Cell Rep.*, **10**, 103–111.

7. Hansen, T.B., Jensen, T.I., Clausen, B.H. *et al.* (2013) Natural RNA circles function as efficient microRNA sponges. *Nature*, **495**, 384–388.
8. Qu, S., Yang, X., Li, X. *et al.* (2015) Circular RNA: a new star of noncoding RNAs. *Cancer Lett.*, **365**, 141–148.
9. Song, X., Zhang, N., Han, P. *et al.* (2016) Circular RNA profile in gliomas revealed by identification tool UROBORUS. *Nucleic Acids Res.*, **44**, e87.
10. Yang, Y., Gao, X., Zhang, M. *et al.* (2018) Novel Role of FBXW7 Circular RNA in Repressing Glioma Tumorigenesis. *J. Natl. Cancer Inst.*, **110**,
11. Xia, W., Qiu, M., Chen, R. *et al.* (2016) Circular RNA has_circ_0067934 is upregulated in esophageal squamous cell carcinoma and promoted proliferation. *Sci. Rep.*, **6**, 35576.
12. Shang, X., Li, G., Liu, H. *et al.* (2016) Comprehensive circular RNA profiling reveals that hsa_circ_0005075, a new circular RNA biomarker, is involved in hepatocellular carcinoma development. *Medicine*, **95**, e3811.
13. Xia, S., Feng, J., Lei, L. *et al.* (2017) Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief Bioinformatics*, **18**, 984–992.
14. Salzman, J., Chen, R.E., Olsen, M.N. *et al.* (2013) Cell-type specific features of circular RNA expression. *PLoS Genet.*, **9**, e1003777.
15. Zheng, Q., Bao, C., Guo, W. *et al.* (2016) Circular RNA profiling reveals an abundant circHIPK3 that regulates cell growth by sponging multiple miRNAs. *Nat. Commun.*, **7**, 11215.
16. Suzuki, H. and Tsukahara, T. (2014) A view of pre-mRNA splicing from RNase R resistant RNAs. *Int. J. Mol. Sci.*, **15**, 9331–9342.
17. Memczak, S., Papavasiliou, P., Peters, O. *et al.* (2015) Identification and characterization of circular RNAs as a new class of putative biomarkers in human blood. *PLoS One*, **10**, e0141214.
18. Bahn, J.H., Zhang, Q., Li, F. *et al.* (2015) The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. *Clin. Chem.*, **61**, 221–230.
19. Glazar, P., Papavasiliou, P. and Rajewsky, N. (2014) circBase: a database for circular RNAs. *RNA*, **20**, 1666–1670.
20. Liu, Y.C., Li, J.R., Sun, C.H. *et al.* (2016) CircNet: a database of circular RNAs derived from transcriptome sequencing data. *Nucleic Acids Res.*, **44**, D209–D215.
21. Dudekula, D.B., Panda, A.C., Grammatikakis, I. *et al.* (2016) CircInteractome: a web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol.*, **13**, 34–42.
22. Li, J.H., Liu, S., Zhou, H. *et al.* (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.*, **42**, D92–D97.
23. Bhattacharya, A. and Cui, Y. (2016) SomamiR 2.0: a database of cancer somatic mutations altering microRNA-ceRNA interactions. *Nucleic Acids Res.*, **44**, D1005–D1010.
24. Zhang, P., Meng, X., Chen, H. *et al.* (2017) PlantCircNet: a database for plant circRNA-miRNA-mRNA regulatory networks. *Database*, **2017**, 1–8.
25. Ghosal, S., Das, S., Sen, R. *et al.* (2013) Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits. *Front. Genet.*, **4**, 283.
26. Xia, S., Feng, J., Chen, K. *et al.* (2018) CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res.*, **46**, D925–d929.
27. Jiang, Q., Wang, Y., Hao, Y. *et al.* (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–104.
28. Li, Y., Qiu, C., Tu, J. *et al.* (2014) HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.*, **42**, D1070–D1074.
29. Chen, G., Wang, Z., Wang, D. *et al.* (2013) LncRNADisease: a database for long-non-coding RNA-associated diseases. *Nucleic Acids Res.*, **41**, D983–D986.
30. Ning, S., Zhang, J., Wang, P. *et al.* (2016) Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. *Nucleic Acids Res.*, **44**, D980–D985.
31. Wang, P., Zhi, H., Zhang, Y. *et al.* (2015) miRSponge: a manually curated database for experimentally supported miRNA sponges and ceRNAs. *Database*, **2015**, 1–7.
32. Burd, C.E., Jeck, W.R., Liu, Y. *et al.* (2010) Expression of linear and novel circular forms of an INK4/ARF-associated non-coding RNA correlates with atherosclerosis risk. *PLoS Genet.*, **6**, e1001233.
33. Zhao, J., Li, L., Wang, Q. *et al.* (2017) CircRNA expression profile in early-stage lung adenocarcinoma patients. *Cell Physiol. Biochem.*, **44**, 2138–2146.
34. Lukiw, W.J. (2013) Circular RNA (circRNA) in Alzheimer's disease (AD). *Front. Genet.*, **4**, 307.
35. Qiu, C., Chen, G. and Cui, Q. (2012) Towards the understanding of microRNA and environmental factor interactions and their relationships to human diseases. *Sci. Rep.*, **2**, 318.
36. Du, W.W., Yang, W., Liu, E. *et al.* (2016) Foxo3 circular RNA retards cell cycle progression via forming ternary complexes with p21 and CDK2. *Nucleic Acids Res.*, **44**, 2846–2858.