


SCIENTIFIC REPORTS



OPEN

Circ2Disease: a manually curated database of experimentally validated circRNAs in human disease

Dongxia Yao¹, Lei Zhang², Mengyue Zheng², Xiwei Sun¹, Yan Lu¹  & Pengyuan Liu¹

Circular RNAs (circRNAs), a new class of regulatory noncoding RNAs, play important roles in human diseases. While a growing number of circRNAs have been characterized with biological functions, it is necessary to integrate all the information to facilitate studies on circRNA functions and regulatory networks in human diseases. Circ2Disease database contains 273 manually curated associations between 237 circRNAs and 54 human diseases with strong experimental evidence from 120 studies. Each association includes circRNA name, disease name, expression pattern, experimental method, a brief functional description of the circRNA-disease relationship, and other detailed information. The experimentally validated miRNAs that may be 'sponged up' by these circRNAs and their validated targets were also integrated to form a comprehensive regulatory network. Circ2Disease provides a user-friendly interface to browse, search, analyze regulatory network and download data. With the rapidly increasing interest in circRNAs, Circ2Disease will significantly improve our understanding of circRNA deregulation in diseases and is a useful resource for studying posttranscriptional regulatory roles of circRNAs in human diseases.

Circular RNA (circRNA) is a type of RNA molecule which forms a covalently closed continuous loop by back-splicing or lariat¹. Although circRNA was discovered decades ago, it was initially perceived as RNA splicing errors². With the advent of next-generation sequencing (NGS), circRNAs were found to be abundant, stable and evolutionarily conserved among eukaryotes³⁻⁵. So far, tens of thousands of circRNAs have been predicted by bioinformatics methods. These circRNAs are largely classified into three categories, including exonic circRNAs³, circular intronic RNAs (ciRNAs)⁶, and retained-intron circRNAs or exon-intron circRNAs (EIciRNAs)⁷.

CircRNAs have become of great interest in the field of transcriptional regulation^{8,9}. For instance, they can act as miRNA sponges¹⁰, or bind to RNA-associated proteins to form RNA-protein complexes that regulate gene transcription¹¹. Some ciRNAs⁶ and EIciRNAs¹¹ can interfere with pre-mRNA splicing that is a critical step in the posttranscriptional regulation of gene expression. Recent studies have identified circRNAs as novel regulators in many disorders such as Alzheimer's disease¹², diabetes mellitus^{13,14} and cancer progression⁶. Furthermore, due to lack of 3' termini, circRNAs are more resistant to degradation by exonuclease RNase R and possess greater stability than linear RNAs^{3,5,6}. circRNAs are abundant in blood samples¹⁵, saliva¹⁶, and exosomes¹⁷, making them promising diagnostic biomarkers for complex diseases such as cancer.

To facilitate studies on circRNAs, several circRNA databases have been developed, such as circBase¹⁸, Circ2Traits¹⁹, circRNADb²⁰, CircNet²¹, deepBase²², CircInteractome²³, TSCD²⁴, CSCD²⁵ and starBase²⁶. Most of these databases integrated circRNAs information from high throughput sequencing studies and various datasets. For example, circBase combined data from several large-scale circRNA studies¹⁸. starBase displayed regulatory networks from thousands of circRNAs and CLIP-seq datasets²⁶. circRNADb integrated datasets of circRNAs from literatures and their own datasets and contained 32,914 exonic circRNAs from various sources²⁰. TSCD is a tissue-specific circRNA database from RNA sequencing (RNA-seq) datasets and characterized the features

¹Sir Run Run Shaw Hospital and Institute of Translational Medicine, Zhejiang University School of Medicine, Hangzhou, Zhejiang, 310016, China. ²Center for Uterine Cancer Diagnosis & Therapy Research of Zhejiang Province, Women's Hospital and Institute of Translational Medicine, Zhejiang University School of Medicine, Hangzhou, Zhejiang, 310029, China. Dongxia Yao and Lei Zhang contributed equally to this work. Correspondence and requests for materials should be addressed to Y.L. (email: yanlu76@zju.edu.cn) or P.L. (email: pyliu@zju.edu.cn)

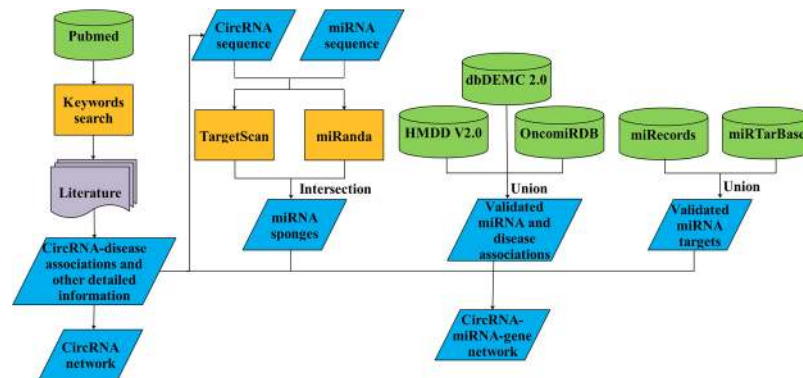


Figure 1. A flowchart of constructing Circ2Disease.

of circRNAs in human and mouse²⁴. CSCD was focused on cancer-specific circular RNAs and identified more than 27 thousands of cancer-specific circRNAs²⁵. Besides these, circNet also utilized transcriptome sequencing datasets to identify the expression of circRNAs²¹. Another database Circ2Traits studied the potential association of circRNAs with human traits and mapped trait related SNPs on circRNA loci¹⁹.

In addition to these large-scale studies, individual or specific circRNA deregulation in various diseases have been reported rapidly. However, these experimentally validated circRNA-disease associations are scattered among various studies. For this reason, we developed a high-quality circRNA-disease association database, named Circ2Disease. The current version of Circ2Disease contains 273 manually curated associations between 237 circRNAs and 54 human diseases with strong experimental evidence from 120 studies. The experimentally validated miRNAs that may be ‘sponged up’ by these circRNAs and their validated targets were also integrated to form a comprehensive regulatory network. With a user-friendly interface, users can easily browse, search, analyze regulatory network and download detailed information for each entry. We hope that this database can help in pushing forward the studies of circRNA-disease associations in the research community.

Methods

Collection of experimentally validated circRNAs. Figure 1 depicts the flowchart of constructing Circ2Disease database. We manually collected all the entries about the associations between circRNA and human disease. Firstly, we searched the PubMed with a list of keywords, such as ‘circular RNA disease’, ‘circRNA disease’, ‘circular RNA cancer’, ‘circRNA cancer’, etc. All the abstracts were viewed to determine whether a study is associated with circRNA deregulation and human disease. Secondly, we carefully reviewed all related literature published prior to 1 November 2017 and curated the detailed information of circRNAs with experimental evidence. To provide high-quality associations with strong experimental evidence, circRNAs only identified by microarray or RNA-seq were not included. If any miRNAs that are sponged by circRNAs were validated experimentally, their miRNA targets, RNA binding protein (RBP) and other up- or down-stream regulatory genes were also retrieved from literature. Thirdly, circRNA and disease names were normalized. In literature, the names of circRNAs and diseases varied from one to another study. To be more readable, we normalized their names in a uniform way. For circRNA names, we unified the names on the basis of circBase, Human circRNA microarray (Arraystar, USA), their original names in particular papers, etc. The names representing the same circRNA were grouped as a collection. Each collection was assigned a name according to the order: original name, circBase id, Arraystar probe id, and others. All these groups are listed on: <http://bioinformatics.zju.edu.cn/Circ2Disease/circRNAgroup.html>. The disease names were also normalized based on the standardized classification vocabulary of Disease Ontology (<http://www.disease-ontology.org/>).

CircRNA-miRNA-gene network construction. As circRNAs may act as miRNA sponges, we predicted potential miRNA response elements on these circRNAs using software TargetScan²⁷ and miRanda²⁸. We separately ran TargetScan with default parameters and miRanda with “strict” parameters. To increase the specificity, 6mer matches were excluded from the TargetScan result. The intersection of the two results was considered as the predicted miRNA sponge. To determine the interactions between these miRNAs and diseases, we integrated three public databases, i.e. HMDD V2.0²⁹, dbDEMOC 2.0³⁰ and OncomiRDB³¹. Experimentally validated miRNA targets from miRecords³² and miRTarBase³³ were also included in the circRNA-miRNA-gene network. For data from miRTarBase, only strong experimental evidence of miRNA-target interactions (MTIs) were retained for further network analysis. The RNA-binding proteins (RBPs) potentially interacted with circRNAs were retrieved from CircInteractome database²³. All these interactions formed a comprehensive circRNA-miRNA-gene regulatory network, which will help users elucidate the pathogenesis of diseases and offer beneficial evidence for diagnosis and therapy.

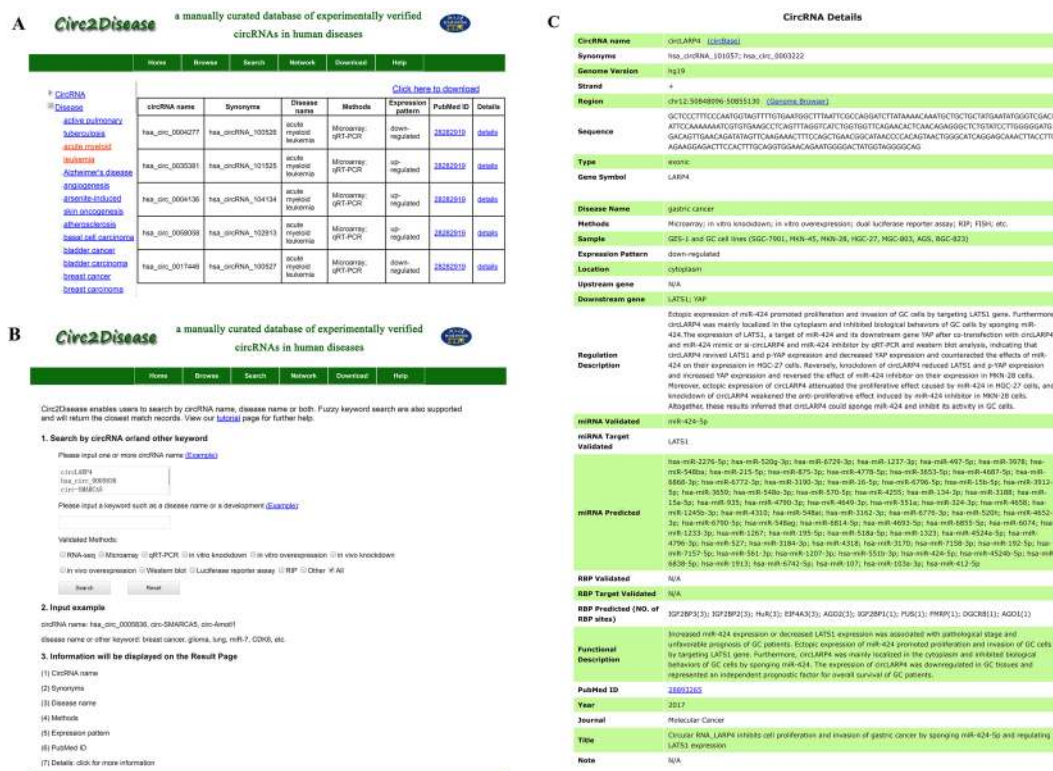


Figure 2. Detail information on the Circ2Disease webpage. **(A)** Browse function. **(B)** Search function. **(C)** Detailed information of circRNA.

Results

Browse and Search. A user-friendly interface was provided in Circ2Disease to facilitate users to browse, search, analyze regulatory network and download all the detailed information on circRNA-disease associations. In the 'Browse' page, a specific circRNA or disease name, and a list of matched entries can be browsed by clicking and a list of matched entries will be displayed on the screen (Fig. 2A). In the 'Search' page, Circ2Disease allows users to query detailed information on each circRNA-disease association by circRNA name and alias name, and/or disease name (Fig. 2B). Circ2Disease also offers fuzzy keyword search function, allowing users to retrieve the closest matching records. In addition, users can filter associations by certain experimental methods. Each item in the result provides a link that takes the user to the detailed circRNA page (Fig. 2C).

Network analysis. Two pages were provided for network analysis in Circ2Disease. They can be accessed by hovering the mouse on the 'Network' menu. By clicking the 'CircRNA' page, a network of all the circRNAs and their up- and down-stream genes that were manually retrieved from literature will be displayed. 'CircRNA-miRNA-Gene' page allows users to search and tune all regulatory networks across circRNA studies (Fig. 3). Users can input circRNA name, disease name, miRNA or gene to search the network. After submitting the query, all the matched entries will be organized as a visualizing network. There is a toolbar above the network. Using the toolbar, several functions can be executed:

- (1) Search: circRNA, miRNA, disease and gene can be input to search the network and it is case-insensitive. The matched node will be turned yellow and displayed in the center of the page.
- (2) Zoom: Scroll the mouse wheel to zoom in and out of the network.
- (3) Display: Grab and drag nodes; tap a node or edge to select; tap blank to unselect.
- (4) Hide/Show: The selected nodes and edges will be hidden or shown.
- (5) Neighbor: Click the "Neighbor" button to select nodes and edges connected to the previously selected nodes.
- (6) Export: Export the displayed network as JPG or PNG file. It can be saved on a local computer.

All these graphical networks were developed based on the cytoscape.js³⁴. We used different colors and shapes to represent the nodes and edges: red ellipse nodes indicate circRNAs, cyan roundrectangle nodes indicate disease names, purple vee nodes indicate miRNAs and green triangle nodes indicate genes.

Download and Help. All the entries can be downloaded in the 'Download' page, and a detailed tutorial of Circ2Disease is available on the 'Help' drop-down menu. A more detailed user guide can be found on the tutorial page.

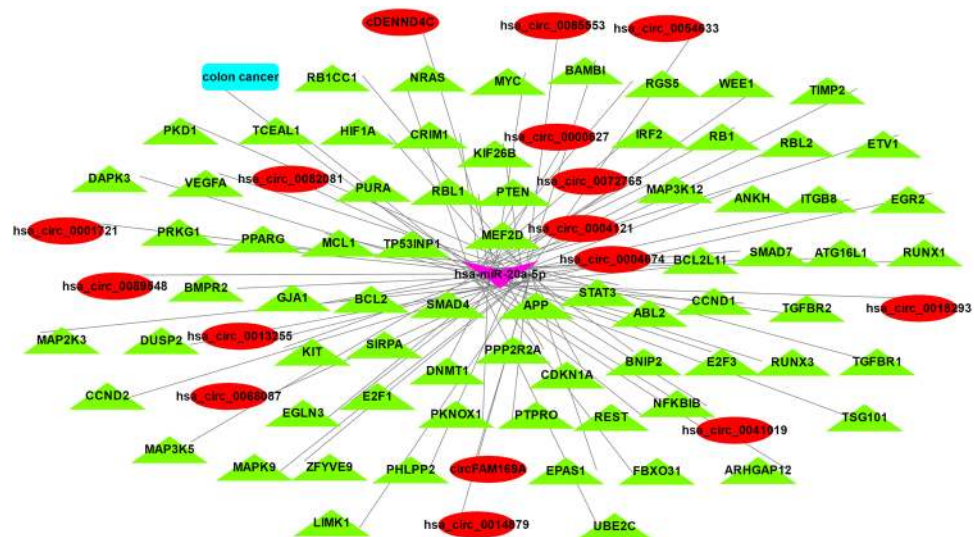


Figure 3. hsa-miR-20a-5p search results in CircRNA-miRNA-Genes network.

Case study 1: circRNA dysregulated in breast carcinoma. To identify circRNAs which are potentially involved in breast carcinoma, we searched a keyword “breast carcinoma” in Circ2Disease. As a result, 29 circRNAs were found in Circ2Disease. Among them, seven circRNAs had been previously assigned a function by direct experimentation. They are circ-ABCB10, cDENND4C, circ-Amot1, circ-Amot1, hsa_circ_0006528, circ-Foxo3 and circGFRA1. The functions of the other 22 circRNAs remain unknown. In the present study, our analysis was focused on these novel downregulated circRNAs, including hsa_circ_0000911, hsa_circ_0018293, hsa_circ_0001283, hsa_circ_0000893, hsa_circ_0000981, hsa_circ_0006054, hsa_circ_0004619 and hsa_circRNA_406697. In addition, Circ2Disease provides RBP-circRNA interaction results, which predicted that hsa_circ_0001283 can bind to AGO2. This indicates that hsa_circ_0001283 may act as a miRNA sponge.

BRCA1 is a tumor suppressor gene in human³⁵; its protein, also called breast cancer type 1 susceptibility protein, is responsible for repairing DNA in breast carcinoma³⁶. According to miRTarBase, BRCA1 is an experimentally validated target of 27 miRNAs such as hsa-miR-146a-5p and hsa-miR-15a-5p. Interestingly, hsa-miR-146a-5p is one of miRNAs that were predicted to target hsa_circ_0001283 in Circ2Disease. Taken together, these results suggest that hsa_circ_0001283 can bind to hsa-miR-146a-5p to regulate the expression of BRCA1 in breast carcinoma. By combining and mining from the validated or predicted circRNA, RBP, and other information such as miRNA-mRNA target interactions, we can prioritize the circRNA-disease association and generate testable hypothesis for further investigation.

Case study 2: circRNA-miRNA-gene network. The miRNA set analysis using TAM tool³⁷ demonstrated that hsa-miR-20a-5p was significantly enriched in immune system, HIV latency, cell proliferation, angiogenesis and apoptosis. This makes it as an attractive candidate potentially involved in cell signaling and disease¹. Thus, we searched hsa-miR-20a-5p in circRNA-miRNA-gene network in Circ2Disease. The network analysis found 71 hsa-miR-20a-5p validated targets such as CDKN1A, VEGFA, and MYC, and 16 circRNAs that were predicted to bind to hsa-miR-20a-5p (Fig. 3). Then, we discovered that these 16 circRNAs were dysregulated in multiple diseases (e.g., coronary artery disease, type 2 diabetes mellitus, osteoarthritis, gastric cancer and so on) by keyword search in “Search” page. Together, these results indicated that these 16 circRNAs may act as sponges to hsa-miR-20a-5p to regulate gene expression of the 71 targets in different diseases. These hypotheses generated from Circ2Disease are worth being investigated in the future.

Discussion

Circ2Disease is the first database for experimentally validated disease-related circRNAs. These data provide the missing cornerstone of functional studies of circRNAs in human diseases. Circ2Disease also constructed a comprehensive circRNA-miRNA-gene network and gives deeper insights into the posttranscriptional regulatory roles of circRNAs in diseases. As the number of circRNA-disease association increases in the future, we will continue to integrate those data into Circ2Disease and will update the database regularly. More importantly, as only a small number of circRNAs have been characterized to be associated with diseases, it is necessary to incorporate more data sources and bioinformatics methods to improve the utility of this database. We believe that Circ2Disease would be a valuable resource for studying circRNAs and will significantly improve our understanding on the relationships between circRNAs and diseases.

References

- Motieghader, H., Kouhsar, M., Najafi, A., Sadeghi, B. & Masoudi-Nejad, A. mRNA-miRNA bipartite network reconstruction to predict prognostic module biomarkers in colorectal cancer stage differentiation. *Mol Biosyst* **13**, 2168–2180, <https://doi.org/10.1039/c7mb00400a> (2017).
- Cocquerelle, C., Mascrez, B., Hetuin, D. & Bailleul, B. Mis-splicing yields circular RNA molecules. *FASEB J* **7**, 155–160 (1993).

3. Jeck, W. R. *et al.* Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* **19**, 141–157, <https://doi.org/10.1261/rna.035667.112> (2013).
4. Salzman, J., Gawad, C., Wang, P. L., Lacayo, N. & Brown, P. O. Circular RNAs are the predominant transcript isoform from hundreds of human genes in diverse cell types. *PLoS One* **7**, e30733, <https://doi.org/10.1371/journal.pone.0030733> (2012).
5. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338, <https://doi.org/10.1038/nature11928> (2013).
6. Zhang, Y. *et al.* Circular intronic long noncoding RNAs. *Mol Cell* **51**, 792–806, <https://doi.org/10.1016/j.molcel.2013.08.017> (2013).
7. Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L. & Brown, P. O. Cell-type specific features of circular RNA expression. *PLoS genetics* **9**, e1003777, <https://doi.org/10.1371/journal.pgen.1003777> (2013).
8. Huang, S. *et al.* The emerging role of circular RNAs in transcriptome regulation. *Genomics* **109**, 401–407, <https://doi.org/10.1016/j.ygeno.2017.06.005> (2017).
9. Chen, L. L. The biogenesis and emerging roles of circular RNAs. *Nat Rev Mol Cell Biol* **17**, 205–211, <https://doi.org/10.1038/nrm.2015.32> (2016).
10. Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384–388, <https://doi.org/10.1038/nature11993> (2013).
11. Li, Z. *et al.* Exon-intron circular RNAs regulate transcription in the nucleus. *Nat Struct Mol Biol* **22**, 256–264, <https://doi.org/10.1038/nsm.2959> (2015).
12. Goswami, C. P. & Nakshatri, H. PROGeneV2: enhancements on the existing database. *BMC cancer* **14**, 970, <https://doi.org/10.1186/1471-2407-14-970> (2014).
13. Stoll, L. *et al.* Circular RNAs as novel regulators of beta-cell functions in normal and disease conditions. *Molecular metabolism* **9**, 69–83, <https://doi.org/10.1016/j.molmet.2018.01.010> (2018).
14. Shan, K. *et al.* Circular Noncoding RNA HIPK3 Mediates Retinal Vascular Dysfunction in Diabetes Mellitus. *Circulation* **136**, 1629–1642, <https://doi.org/10.1161/CIRCULATIONAHA.117.029004> (2017).
15. Memczak, S., Papavasileiou, P., Peters, O. & Rajewsky, N. Identification and Characterization of Circular RNAs As a New Class of Putative Biomarkers in Human Blood. *PLoS One* **10**, e0141214, <https://doi.org/10.1371/journal.pone.0141214> (2015).
16. Bahn, J. H. *et al.* The landscape of microRNA, Piwi-interacting RNA, and circular RNA in human saliva. *Clin Chem* **61**, 221–230, <https://doi.org/10.1373/clinchem.2014.230433> (2015).
17. Li, Y. *et al.* Circular RNA is enriched and stable in exosomes: a promising biomarker for cancer diagnosis. *Cell research* **25**, 981 (2015).
18. Glazar, P., Papavasileiou, P. & Rajewsky, N. circBase: a database for circular RNAs. *RNA* **20**, 1666–1670, <https://doi.org/10.1261/rna.043687.113> (2014).
19. Ghosal, S., Das, S., Sen, R., Basak, P. & Chakrabarti, J. Circ2Traits: a comprehensive database for circular RNA potentially associated with disease and traits. *Front Genet* **4**, 283, <https://doi.org/10.3389/fgene.2013.00283> (2013).
20. Chen, X. *et al.* circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Sci Rep* **6**, 34985, <https://doi.org/10.1038/srep34985> (2016).
21. Liu, Y. C. *et al.* CircNet: a database of circular RNAs derived from transcriptome sequencing data. *Nucleic Acids Res* **44**, D209–215, <https://doi.org/10.1093/nar/gkv940> (2016).
22. Zheng, L. L. *et al.* deepBasev2.0: identification, expression, evolution and function of small RNAs, lncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Res* **44**, D196–202, <https://doi.org/10.1093/nar/gkv1273> (2016).
23. Dudekula, D. B. *et al.* CircInteractome: A web tool for exploring circular RNAs and their interacting proteins and microRNAs. *RNA Biol* **13**, 34–42, <https://doi.org/10.1080/15476286.2015.1128065> (2016).
24. Xia, S. *et al.* Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. *Brief Bioinform*, <https://doi.org/10.1093/bib/bbw081> (2016).
25. Xia, S. *et al.* CSCD: a database for cancer-specific circular RNAs. *Nucleic Acids Res*, <https://doi.org/10.1093/nar/gkx863> (2017).
26. Li, J. H., Liu, S., Zhou, H., Qu, L. H. & Yang, J. H. starBasev2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* **42**, D92–97, <https://doi.org/10.1093/nar/gkt1248> (2014).
27. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, <https://doi.org/10.7554/eLife.05005> (2015).
28. Betel, D., Koppal, A., Agius, P., Sander, C. & Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol* **11**, R90, <https://doi.org/10.1186/gb-2010-11-8-r90> (2010).
29. Li, Y. *et al.* HMDDv2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res* **42**, D1070–1074, <https://doi.org/10.1093/nar/gkt1023> (2014).
30. Yang, Z. *et al.* dbDEM2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res* **45**, D812–D818, <https://doi.org/10.1093/nar/gkw1079> (2017).
31. Wang, D., Gu, J., Wang, T. & Ding, Z. OncomiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs. *Bioinformatics* **30**, 2237–2238, <https://doi.org/10.1093/bioinformatics/btu155> (2014).
32. Xiao, F. *et al.* miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res* **37**, D105–110, <https://doi.org/10.1093/nar/gkn851> (2009).
33. Chou, C. H. *et al.* miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res*, <https://doi.org/10.1093/nar/gkx1067> (2017).
34. Franz, M. *et al.* Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* **32**, 309–311, <https://doi.org/10.1093/bioinformatics/btv557> (2016).
35. Duncan, J. A., Reeves, J. R. & Cooke, T. G. BRCA1 and BRCA2 proteins: roles in health and disease. *Mol Pathol* **51**, 237–247 (1998).
36. Check, W. BRCA: What we know now. *College of American Pathologists* (2006).
37. Lu, M., Shi, B., Wang, J., Cao, Q. & Cui, Q. TAM: a method for enrichment and depletion analysis of a microRNA category in a list of microRNAs. *BMC Bioinformatics* **11**, 419, <https://doi.org/10.1186/1471-2105-11-419> (2010).

Acknowledgements

We would like to give special thanks to Liyuan Zhou for very helpful comments on this paper. This work has been supported in part by National Key R&D program (2016YFC0902700 and 2016YFA0501800), National Natural Science Foundation of China (No. 81772766, 81472420, and 81572256), the Fundamental Research Funds for the Central Universities, and the 1000 talent Plan of China.

Author Contributions

Y.L. and P.L. designed the study. D.Y. designed the database platform and visual interface. D.Y., L.Z., M.Z. and X.S. collected literatures and extracted circRNA information. D.Y. drafted the manuscript. P.L. and Y.L. revised the manuscript. All authors reviewed the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018